Comparación de técnicas de clasificación de múltiples anotadores para la valoración Automática de la calidad de voz

Tomas Echeverri Valencia
Julián Gil González

Universidad Tecnológica de Pereira

Facultad de Ingenierías

Pereira
2014

Comparación de técnicas de clasificación de múltiples anotadores para la valoración Automática de la calidad de voz

Tomas Echeverri Valencia
Julián Gil González

Trabajo de grado para optar al título de
Ingeniero Electrónico

PhD. Mauricio A. Álvarez

Universidad Tecnológica de Pereira

Facultad de Ingenierías

Pereira
2014

Nota de aceptación:

_____

_____

_____

_____

_____

_____

_____

_____

Director de Trabajo de Grado

_____

Jurado

_____

Jurado

Tabla de Contenido

Lista de Figuras

pág.

Lista de Tablas

pág.

Resumen

Actualmente se han hecho más comunes los problemas que afectan la voz. La medicina ha desarrollado técnicas que evalúan la calidad de voz, con el propósito de detectar patologías asociadas al aparato fonador, específicamente aquellas que afectan las cuerdas vocales. Entre las técnicas desarrolladas se identifican principalmente dos enfoques: el análisis acústico y el análisis perceptivo. Estas técnicas presentan algunos inconvenientes: para el análisis acústico se debe contar con las etiquetas verdaderas para definir los patrones de comparación, por otro lado el análisis perceptivo presenta subjetividad en las valoraciones. Estos problemas pueden ser minimizados usando técnicas de aprendizaje supervisado con múltiples anotaciones. En este sentido, se expone el desarrollo de un sistema de valoración automática de la calidad de voz bajo el protocolo GRBAS y basado en técnicas de aprendizaje de máquina para múltiples anotadores. En la etapa de aprendizaje automático para múltiples anotadores, se comparan dos tipos de técnicas, una de ellas basada en Procesos Gaussianos [1], la otra se basa en un modelo de Regresión Logística Multiclase que tiene en cuenta la sensibilidad y especificidad de cada anotador [2]. Las señales de voz se caracterizan usando los coeficientes cepstrales en la escala de frecuencias Mel. La comparación de las técnicas de clasificación nombradas se efectúa en términos de precisión y de las curvas ROC. Los resultados muestran que el clasificador con mejor desempeño para tareas de valoración de la calidad de voz es aquel basado en Procesos Gaussianos, el cual obtuvo un AUC promedio de 0,59 mientras que el clasificador basado en regresión logística multiclase alcanzó un AUC promedio de 0,55. Además los resultados de los experimentos indican que el clasificador de múltiples anotadores basado en Procesos Gaussianos obtuvo mejor rendimiento que los clasificadores típicos que usan “majority voting” para calcular la etiqueta verdadera a partir de las anotaciones.

Agradecimientos

Queremos agradecer al director del proyecto de grado, el Ph.D. Mauricio Álvarez por su apoyo y orientación en el desarrollo de este trabajo.
Gracias a los profesores Julián David Arias y Juan Ignacio Godino por facilitarnos la base de datos de voz usada en este trabajo.

Introducción

En la actualidad, los problemas que afectan la voz se han hecho más comunes, debido principalmente a hábitos poco saludables (consumo de tabaco y alcohol), factores ambientales y el uso excesivo de la voz. Estas afecciones deben ser diagnosticadas y tratadas en sus fases iniciales, con la finalidad de evitar posibles complicaciones, en especial en el caso de cáncer de laringe [3].

La medicina ha desarrollado algunas técnicas que permiten evaluar la calidad de voz, con el objetivo de detectar patologías asociadas al aparato fonador, específicamente aquellas que afectan los pliegues vocales. Este grupo incluye patologías como pólipos, nódulos, quistes, carcinomas, etc. (ver Apéndice A). Entre las técnicas que se han desarrollado, se reconocen principalmente dos enfoques, el análisis perceptivo y el análisis acústico [4].

El análisis perceptivo es una técnica que se basa en la interpretación que tenga un especialista (otorrinolaringólogo) sobre la calidad de voz, a través de sus funciones perceptivas y psico-acústicas. Entre los protocolos empleados para este tipo de análisis se encuentran *the Buffalo Voice Profile Analysis* (BVP), *the Hammarberg scheme*, *the Vocal Profile Analysis scheme* (VPA) y *the GRBAS scale* [4]. El protocolo GRBAS es el más usado para este tipo de análisis. Comprende la especificación de cinco características: Grado de Disfonía (G, Grade of Disphony), Aspereza (R, Roughness), Respiración Dificultosa (B, Breathiness), Astenia o fatiga vocal (A, Asthenicity), y Voz forzada (S, Strainess) [5]. A cada una de estas características se le asigna un valor en el rango $[0, 3]$, donde 0 corresponde a una voz saludable, 1 alteración leve, 2 alteración moderada y 3 alteración severa. Aunque el análisis perceptivo ha sido el más usado en los procedimientos para la valoración de la voz, también ha sido ampliamente criticado debido a

la subjetividad de sus valoraciones, puesto que dichas valoraciones dependen de factores como la experiencia de los especialistas, el protocolo de valoración usado, la fatiga mental y física del especialista, etc. [6].

Por otro lado se encuentra el análisis acústico, el cual es una técnica no invasiva que se basa en el procesamiento digital de la señal del habla. A partir de este procesamiento se extraen de la señal de voz un conjunto de características espectrales y temporales que se supone están relacionadas con su calidad. Dichas características se comparan con patrones definidos y de allí se etiqueta como patológica o normal. El análisis acústico presenta grandes ventajas sobre el análisis perceptivo, debido principalmente a que se elimina en gran proporción el problema de la subjetividad [3]. Sin embargo, este enfoque también posee problemas: para definir los patrones de comparación se debe tener acceso a la etiqueta verdadera (voz normal o voz patológica). En la práctica esto se convierte en un problema, puesto que la única manera de obtener estas etiquetas es a través de métodos invasivos (Biopsias), que son potencialmente peligrosos. Para construir un sistema basado en análisis acústico, en lugar de las etiquetas verdaderas, lo que suele suceder es que se cuenta con un número de anotaciones subjetivas (posiblemente erróneas) provistas por diferentes especialistas con distintos niveles de experiencia, lo cual produce que estas anotaciones tengan un gran nivel de desacuerdo.

Los problemas asociados a las técnicas de valoración de la calidad de voz descritas, pueden ser minimizados a través del uso de sistemas automáticos que usan técnicas del *Aprendizaje de Máquina Supervisado* [6], específicamente aquellas que consideran el caso de múltiples anotadores.

El área de Aprendizaje Automático en el caso de múltiples anotadores es relativamente nueva, esta tiene como propósito, realizar tareas de clasificación o regresión sin tener acceso a las etiquetas verdaderas (Gold Standard). Entre los enfoques que aparecen en la literatura, las referencias [7] y [8], proponen un modelo paramétrico, en el cual usan el esquema de Máxima Esperanza (Expectation-Maximization), para estimar el Gold Standard a partir de múltiples anotaciones ruidosas y luego a partir de este estimado aprender el clasificador; por el contrario en [2] se propone un modelo paramétrico

el cual conjuntamente estima el gold standard y aprende el clasificador, esta solución es más general puesto que puede ser fácilmente extendida para tareas de clasificación multiclase e incluso regresión. Además en [1] se propone un modelo no paramétrico basado en *Procesos Gaussainos*, el cual realiza de manera simultánea la estimación de las etiquetas verdaderas y el entrenamiento del modelo de regresión. Otros enfoques usados para acercarse a este problema incluyen [9], [10], [11].

El aprendizaje automático que incluye múltiples anotaciones es un problema de interés reciente en aprendizaje de máquina. En este sentido, este trabajo tiene como finalidad la comparación de técnicas de clasificación multiclase con múltiples anotadores para la valoración automática de la calidad de voz bajo el protocolo GRBAS. Para este fin, se usan los coeficientes cepstrales para las frecuencias Mel (MFCC) en la etapa de caracterización de la señal del habla, y para la etapa de clasificación se comparan dos algoritmos, uno de ellos basado en Procesos Gaussianos para regresión con múltiples anotadores [1], mientras que el otro se basa en un modelo de regresión logística multiclase, que tiene en cuenta la sensibilidad y especificidad de cada anotador [2]. Los resultados obtenidos se comparan en términos de la precisión y de las curvas ROC.

Objetivos

Objetivo General

- Comparar técnicas de clasificación multiclase con múltiples anotadores para la valoración de la calidad de voz bajo el protocolo GRBAS.

Objetivos Específicos

- Implementar un algoritmo de clasificación para múltiples anotadores, usando regresión logística multiclase con sensibilidades y especificidades por cada anotador [2].
- Implementar un algoritmo de regresión de múltiples anotadores usando Procesos Gaussianos para regresión [1].
- Comparar el desempeño de los algoritmos implementados sobre una base de datos de voz etiquetada por múltiples anotadores bajo el protocolo GRBAS.

Capítulo 1

Técnicas de Aprendizaje Automático Supervisado para Múltiples Anotadores

En el presente capítulo se exponen los fundamentos teóricos de las técnicas de aprendizaje de máquina supervisado para el caso de múltiples anotadores. Se empieza con el análisis del clasificador para múltiples anotadores basado en regresión logística multiclase. Se finaliza con el estudio de un modelo de regresión para múltiples anotadores basado en Procesos Gaussianos.

Típicamente un problema de aprendizaje de máquina supervisado consiste de un conjunto de entrenamiento $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, que contiene N muestras, donde $\mathbf{x}_i \in \mathbf{X}$ es una muestra (regularmente corresponde a un vector de características d dimensional) y $y_i \in Y$ es la correspondiente etiqueta verdadera. Sin embargo para muchos problemas de la vida real, es muy difícil o muy costoso adquirir las etiquetas verdaderas y_i. En lugar de estas etiquetas verdaderas, se disponen de múltiples etiquetas (posiblemente erróneas) $y_i^1, y_i^2, \ldots, y_i^R$ provistas por R expertos o anotadores. En la práctica existe un gran nivel de desacuerdo entre las anotaciones hechas por los expertos. Lo que motiva a modificar las técnicas convencionales de aprendizaje supervisado con el fin de adaptarlas a este caso con múltiples anotaciones [2].

1.1. Regresión Logística Multiclase considerando sensibilidades y especificidades por cada anotador

Se sigue el modelo de clasificación propuesto por [2]. Para facilitar la exposición del modelo, se empieza con el análisis del problema de clasificación binaria. Para finalizar se extiende el modelo para problemas de clasificación multiclase.

1.1.1. Clasificación Binaria

Sea $y_i^j \in \{0, 1\}$ la etiqueta asignada a la muestra $\mathbf{x}_i$ por el j-ésimo anotador. Sea y_i la verdadera etiqueta (desconocida) correspondiente a esta muestra. Cada anotador provee una versión de esta etiqueta verdadera basado en el modelo de las dos monedas sesgadas. Si la etiqueta verdadera es uno, se lanza la moneda con sesgo α^j (sensibilidad). Si la etiqueta verdadera es cero con sesgo β^j (especificidad). En cada caso, si se consigue cara, se mantiene la etiqueta original, de otra manera, se cambia la etiqueta.

La sensibilidad α^j, se define como la probabilidad de que el anotador j-ésimo etiquete una muestra como uno, dado que la muestra verdadera es uno

$$\alpha^j = p(y_i^j = 1 | y_i = 1).$$

Por su parte, la especificidad β^j, es definida como la probabilidad de que el anotador j-ésimo etiquete una muestra como cero, dado que la muestra verdadera es cero

$$\beta^j = p(y_i^j = 0 | y_i = 0).$$

Este modelo puede ser aplicado a cualquier tipo de clasificador, en este caso se elige un clasificador basado en Regresión Logística. Así la probabilidad para la clase positiva está dada por

$$p(y_i = 1 | \mathbf{x}_i, \mathbf{w}) = \boldsymbol{\sigma}(\mathbf{w}^T \mathbf{x}_i),$$

donde la función $\boldsymbol{\sigma}(z)$, se conoce como Logistic Sigmoid y se define como

$$\boldsymbol{\sigma}(z) = \frac{1}{1+e^{-z}}. \tag{1.1}$$

Ahora dado un conjunto de entrenamiento $\mathcal{D}$ de N ejemplos con anotaciones provenientes de R expertos, esto es, $\mathcal{D} = \{\mathbf{x}_i, y_i^1, \ldots, y_i^R\}_{i=1}^N$, la tarea consiste en estimar el vector de ponderación $\mathbf{w}$ además de la sensibilidad $\boldsymbol{\alpha} = [\alpha^1, \alpha^2, \ldots, \alpha^R]$ y la especificidad $\boldsymbol{\beta} = [\beta^1, \beta^2, \ldots, \beta^R]$. Se define la variable $\boldsymbol{\theta}$, como el vector de parámetros que se desean estimar

$$\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{w}\}.$$

Luego la función de verosimilitud de los parámetros $\boldsymbol{\theta}$ dado las muestras $\mathcal{D}$ se puede escribir como

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} p(y_i^1, y_i^2, \ldots, y_i^R)|\mathbf{x}_i, \boldsymbol{\theta}).$$

Dados $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, y_i. y asumiendo que $y_i^1, y_i^2, \ldots, y_i^R$ son independientes, esto es, que los anotadores toman sus decisiones independientemente, la verosimilitud puede ser escrita así

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N} (a_i p_i + b_i(1-p_i)),$$

donde p_i, a_i y b_i, se definen como

$$p_i = \boldsymbol{\sigma}(\mathbf{w}^T \mathbf{x}_i).$$

$$a_i = \prod_{j=1}^{R} (\alpha^j)^{y_i^j} (1-\alpha^j)^{1-y_i^j}.$$

$$b_i = \prod_{j=1}^{R} (\beta^j)^{1-y_i^j} (1-\beta^j)^{y_i^j}.$$

Luego la estimación de los parámetros $\boldsymbol{\theta}$, se efectúa a través, de la maximización del logaritmo de la verosimilitud, esto es

$$\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{w}}\} = \text{argmax}_{\boldsymbol{\theta}}\{\ln(p(\mathcal{D}|\boldsymbol{\theta}))\}$$

Algoritmo EM

El problema de maximización, se lleva a cabo a partir del uso del algoritmo de Máxima-Esperanza o EM por sus siglas en inglés (Expectation-Maximization), el cual es un algoritmo iterativo que calcula la solución de la máxima verosimilitud en presencia de datos ocultos o faltantes [12]. Para este caso se usa como datos ocultos las etiquetas verdaderas y_i. Asumiendo que se conocen las etiquetas verdaderas, la verosimilitud está dada por

$$p(\mathcal{D}, Y|\boldsymbol{\theta}) = \prod_{i=1}^{N}(a_i p_i)^{y_i}(b_i(1-p_i))^{1-y_i}.$$

Igualmente, el logaritmo de la verosimilitud se escribe como

$$\ln[p(\mathcal{D}, Y|\boldsymbol{\theta})] = \sum_{i=1}^{N}[y_i \ln(p_i a_i) + (1-y_i)\ln(b_i(1-p_i))].$$

El algoritmo EM, se compone de dos etapas en cada una de sus iteraciones. Una etapa de Esperanza (etapa E) y una etapa de Maximización (etapa M). La etapa M, involucra la maximización de una cota inferior del logaritmo de la verosimilitud, que es refinada en cada iteración en la etapa E.

- **Etapa E**

  Dado el conjunto de entrenamiento $\mathcal{D}$ y el estimado actual de los parámetros $\boldsymbol{\theta}$, la esperanza condicional (la cual es una cota inferior de la verosimilitud) se calcula como

$$\mathbb{E}[\ln(p(\mathcal{D}, Y|\boldsymbol{\theta}))] = \sum_{i=1}^{N}\mathbb{E}[y_i]\ln(a_i p_i) + (1-\mathbb{E}[y_i])\ln(b_i(1-p_i)), \tag{1.2}$$

donde $\mathbb{E}[y_i] = \mu_i = p(y_i = 1|y_i^1, y_i^2, \ldots, y_i^R, \mathbf{x}_i, \boldsymbol{\theta})$. Usando el teorema de Bayes, se tiene

$$\begin{aligned}\mu_i &= \frac{p(y_i^1, y_i^2, \ldots, y_i^R|y_i = 1)p(y_i = 1|\mathbf{x}_i, \boldsymbol{\theta})}{p(y_i^1, y_i^2, \ldots, y_i^R)} \\ &= \frac{a_i p_i}{a_i p_i + b_i(1 - p_i)}.\end{aligned}$$

- **Etapa M** Basado en el estimado actual de μ_i y el conjunto de entrenamiento $\mathcal{D}$, los parámetros $\boldsymbol{\theta}$ son estimados a partir de la maximización de la esperanza condicional. Al igualar el gradiente de (1.2) a cero (ver Apéndice C), se obtienen estimados para la sensibilidad (α^j) y la especificidad (β^j) mostrados en las ecuaciones (1.3) y (1.4) respectivamente

$$\alpha^j = \frac{\sum_{i=1}^N \mu_i y_i^j}{\sum_{i=1}^N \mu_i}. \tag{1.3}$$

$$\beta^j = \frac{\sum_{i=1}^N (1 - \mu_i)(1 - y_i^j)}{\sum_{i=1}^N (1 - \mu_i)}. \tag{1.4}$$

Por último se debe estimar el vector de ponderaciones $\mathbf{w}$. Debido a la no linealidad de la función Sigmoid (1.1), no existe una solución exacta para $\mathbf{w}$, por lo que se hace necesario el uso de gradiente ascendente basado en métodos de optimización. En este caso se usa el método de Newton-Raphson dado por

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \mathbf{H}^{-1}\mathbf{g},$$

donde $\mathbf{g}$ es el vector gradiente, $\mathbf{H}$ es la matriz Hessiana y η es la longitud de paso. El gradiente $\mathbf{g}$ está dado por

$$\mathbf{g}(w) = \sum_{i=1}^N \left[\mu_i - \boldsymbol{\sigma}(\mathbf{w}^T \mathbf{x}_i)\right] \mathbf{x}_i.$$

La matriz Hessiana se calcula

$$\mathbf{H} = -\sum_{i=1}^{N} \left[\boldsymbol{\sigma}(\mathbf{w}^T\mathbf{x}_i)\right]\left[1-\boldsymbol{\sigma}(\mathbf{w}^T\mathbf{x}_i)\right]\mathbf{x}_i\mathbf{x}_i^T.$$

1.1.2. Clasificación para Múltiples Clases

Se puede extender el modelo para clasificación binaria introduciendo un vector de parámetros multinomiales α_{ck}^j, el cual se define como la probabilidad de que el anotador j le asigne la clase k a una observación, dado que la clase verdadera es c

$$\alpha_{ck}^j = p(y_i^j = k|y_i = c), \quad \sum_{k=1}^{K}\alpha_{ck}^j = 1.$$

Al igual que para el modelo de clasificación binaria, se requiere la estimación de parámetros $\boldsymbol{\theta} = \{\alpha_{ck}^j, \mathbf{w}\}$. Ahora la función de verosimilitud para los parámetros $\boldsymbol{\theta}$, dado el conjunto de entrenamiento se da por

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N}\sum_{c=1}^{K}\left\{\prod_{j=1}^{R}\prod_{k=1}^{K}(\alpha_{ck}^j)^{Z_k^{i,j}}\right\}p_{ic},$$

donde $p_{ic} = p(y_i = c|\mathbf{x}_i, \boldsymbol{\theta})$ y $Z_k^{i,j} = \delta(y_i^j, k)$. Ahora se supone un matriz T de dimensiones $N \times K$, donde cada vector fila representa la etiqueta verdadera desconocida y codificada en notación 1 de K, de la observación i. La función de verosimilitud para los datos completos (D, T) es

$$\begin{aligned} p(D, T|\boldsymbol{\theta}) &= \prod_{i=1}^{N}\prod_{c=1}^{K}\left\{\prod_{j=1}^{R}\prod_{k=1}^{K}(\alpha_{ck}^j)^{Z_k^{i,j}}\right\}^{t_{ic}} p_{ic}^{t_{ic}} \\ &= \prod_{i=1}^{N}\prod_{c=1}^{K}(a_{ic}p_{ic})^{t_{ic}}, \end{aligned}$$

donde

$$a_{ic} = \prod_{j=1}^{R}\prod_{k=1}^{K}(\alpha_{ck}^j)^{Z_k^{i,j}}.$$

El logaritmo de la función de verosimilitud toma la forma

$$\begin{aligned}\ln[p(\mathcal{D}, T|\boldsymbol{\theta})] &= \sum_{i=1}^{N}\sum_{c=1}^{K} t_{ic}\ln(a_{ic}p_{ic}) \\ &= \sum_{i=1}^{N}\sum_{c=1}^{K}[t_{ic}\ln(a_{ic}) + t_{ic}\ln(p_{ic})]. \end{aligned} \tag{1.5}$$

Algoritmo EM

Al igual que en el modelo para clasificación binaria, se usa el algoritmo EM para solucionar el problema de la maximización del logaritmo de la verosimilitud ante datos ocultos (las etiquetas verdaderas t_{ic}). Cabe recordar que dicha maximización se realiza con el fin de obtener una estimación para los parámetros $\boldsymbol{\theta}$.

- **Etapa E**

  Dados los datos del conjunto de entrenamiento $\mathcal{D}$ y las etiquetas verdaderas T, la esperanza condicional toma la forma

$$\mathbb{E}[\ln(p(\mathcal{D}, T|\boldsymbol{\theta}))] = \sum_{i=1}^{N}\sum_{c=1}^{K}\mathbb{E}[t_{ic}]\ln(a_{ic}p_{ic}), \tag{1.6}$$

  donde

$$\mathbb{E}[t_{ic}] = \mu_{ic} = \frac{p(y_i^1, y_i^2, \ldots, y_i^R|y_i = c)p(y_i = c|\mathbf{x}_i, \boldsymbol{\theta})}{\sum_{\forall c} p(y_i^1, y_i^2, \ldots, y_i^R|y_i = c)p(y_i = c|\mathbf{x}_i, \boldsymbol{\theta})}$$

$$\mu_{ic} = \frac{a_{ic}p_{ic}}{\sum_{k=1}^{K} a_{ik}p_{ik}} \tag{1.7}$$

- **Etapa M**

  Ahora tomando como base la actual estimación de μ_{ic} y el conjunto de entrenamiento $\mathcal{D}$, se estiman los parámetros $\boldsymbol{\theta}$ maximizando la esperanza condicional. Igualando a cero el gradiente de (1.6) (ver Apéndice C), se obtiene el estimado para el vector de parámetros multinomiales α_{ck}^j mostrado en la ecuación (1.8)

$$\alpha_{ck}^{j} = \frac{\sum_{\forall i} \mu_{ic} Z_k^{i,j}}{\sum_{\forall i} \mu_{ic}} \tag{1.8}$$

Ahora bien, falta determinar qué es, o mejor cómo se calcula p_{ic}. Por definición (Ver [13], regresión logística multiclase)

$$p_{ic} = \frac{\exp(b_{ic})}{\sum_{m=1}^{K} \exp(b_{im})}, \tag{1.9}$$

donde $b_{ic} = w_c^T \mathbf{x}_i$.

El logaritmo negativo de la verosimilitud $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old})$ con los términos que sólo dependen de w_c es la siguiente.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = \sum_{i=1}^{N} \sum_{c=1}^{K} \mu_{ic} \ln(p_{ic}).$$

Al igual que en el modelo para clasificación binaria, no existe una solución exacta para $\mathbf{w}$ debido a la no-linealidad de la función softmax (1.9), por lo que se usa el método de Newton-Raphson. La actualización para $\mathbf{w}$ está dada por

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \mathbf{H}^{-1} \mathbf{g}, \tag{1.10}$$

donde $\mathbf{g}$ es el gradiente dado por.

$$\mathbf{g}(w) = \sum_{i=1}^{N} \left[\mu_i - \boldsymbol{\sigma}(\mathbf{w}^T \mathbf{x}_i) \right] \mathbf{x}_i, \tag{1.11}$$

$\mathbf{H}$ es la matriz Hessiana, la cual compromete la evaluación de bloques de tamaño $M \times M$ (M, es la dimensión del espacio de entrada $\mathbf{X}$). Cada bloque m, c está dado por

$$\mathbf{H} = -\sum_{i=1}^{N} \left[\boldsymbol{\sigma}(\mathbf{w}_c^T \mathbf{x}_i) \right] \left[\mathbf{I}_{cm} - \boldsymbol{\sigma}(\mathbf{w}_m^T \mathbf{x}_i) \right] \mathbf{x}_i \mathbf{x}_i^T, \tag{1.12}$$

donde $\mathbf{I}_{cm}$ son los elementos de la matriz identidad.

1.2. Regresión con Múltiples Anotadores Usando Procesos Gaussianos

Se siguió el modelo de regresión propuesto por [1]. Hay R anotadores y cada uno de ellos ha anotado N_j observaciones para formar una conjunto de datos $\mathcal{D}_j = \{(\mathbf{x}_i^j, y_i^j)\}_{i=1}^{N_j}$. Por simplicidad, el conjunto de datos se denota como $\mathcal{D}_j = (\mathbf{X}_j, \mathbf{y}_j)$. El modelo asumido para las etiquetas y_i^j dadas por cada anotador es, $y_i^j = y_i + \epsilon^j$, donde y_i corresponde a la etiqueta verdadera, la cual se desconoce y $\epsilon^j \sim \mathcal{N}(0, \sigma_j^2)$. Sea I el número de entradas únicas. Asumiendo la independencia entre los anotadores y el hecho de que cada anotador etiqueta cada observación $\mathbf{x}_i$ independientemente, la probabilidad está dada por:

$$p(\mathbf{y}|Y) = \prod_j \prod_{i \sim j} \mathcal{N}(y_i^j | y_i, \sigma_j^2),$$

donde $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_R\}$, y $i \sim j$ se refiere a "la observación i etiquetada por el anotador j". Asumiendo un Proceso Gaussiano previo para Y dado por $p(Y) = \mathcal{N}(Y|\mathbf{0}, \mathbf{K})$, con un kernel $\mathbf{K}$ calculado usando una función particular de kernel $k(\mathbf{x}, \mathbf{x}')$, puede ser mostrado que la posterior sobre una nueva observación $f(\mathbf{x}_*)$ se calcula como [1] $p(f(\mathbf{x}_*)|\mathbf{y}) = \mathcal{N}(f(\mathbf{x}_*)|\bar{f}(\mathbf{x}_*), k(\mathbf{x}_*, \mathbf{x}'_*))$, donde

$$\begin{aligned} \bar{f}(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{X})(\mathbf{K} + \hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}, \\ k(\mathbf{x}_*, \mathbf{x}'_*) &= k(\mathbf{x}_*, \mathbf{x}'_*) - k(\mathbf{x}_*, \mathbf{X})(\mathbf{K} + \hat{\mathbf{\Sigma}})^{-1}k(\mathbf{X}, \mathbf{x}'_*). \end{aligned} \tag{1.13}$$

Para las anteriores expresiones, $\mathbf{X}$ corresponde a las observaciones de entrada únicas $\mathbf{x}_i$ para la etapa de entrenamiento; la matriz diagonal $\hat{\mathbf{\Sigma}}$ tiene elementos $\hat{\sigma}_i^2$; y el vector $\hat{\mathbf{y}}$ tiene entradas $\hat{y}_i$. Los elementos $\hat{\sigma}_i^2$ y $\hat{y}_i$ están definidos como

$$\frac{1}{\hat{\sigma}_i^2} = \sum_{j \sim i} \frac{1}{\sigma_j^2}, \qquad \hat{y}_i = \hat{\sigma}_i^2 \sum_{j \sim i} \frac{y_i^j}{\sigma_j^2}, \qquad \hat{\mathbf{\Sigma}} = \text{diag}(\hat{y}_i). \tag{1.14}$$

La notación $j \sim i$ se refiere a "tomar en cuenta sólo los anotadores j que anotaron la observación i".

De acuerdo con [1] el logaritmo negativo de la evidencia está dado como:

$$
\begin{aligned}
-\log p(\mathbf{y}) = & \frac{1}{2}\log|\mathbf{K}+\hat{\mathbf{\Sigma}}| + \frac{1}{2}\hat{\mathbf{y}}^{\top}(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}} - \frac{1}{2}\log|\hat{\mathbf{\Sigma}}| \\
& + \frac{1}{2}\sum_{i}\sum_{j\sim i}\frac{(y_i^j)^2}{\sigma_j^2} - \frac{1}{2}\sum_{i}\frac{\hat{y}_i}{\hat{\sigma}_i^2} - \sum_{j}\sum_{i\sim j}\log\frac{1}{\sigma_j} \\
& + \frac{N}{2}\log 2\pi, \qquad (1.15)
\end{aligned}
$$

donde $N = \sum_j N_j$ es el número total de anotaciones.

Se define el vector de parámetros como $\boldsymbol{\theta}$. El vector de parámetros incluyen los parámetros asociados a la función del kernel $k(\mathbf{x}, \mathbf{x}')$, que se denota como $\boldsymbol{\phi}$, y las varianzas asociadas a cada anotador. $\{\sigma_j^2\}_{j=1}^R$. Luego $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \{\sigma_j^2\}_{j=1}^R\}$. Los valores de este vector de parámetros se estiman al minimizar el logaritmo negativo de la evidencia (1.15). Para realizar esta minimización, es necesario determinar la derivada del logaritmo negativo de la evidencia respecto a los parámetros σ_j (ver Apéndice C).

Capítulo 2

Materiales y Métodos

En este capítulo se describen los procedimientos usados en la implementación de las técnicas de clasificación de múltiples anotadores para la valoración automática de la calidad de voz. Las señales de voz se caracterizan por medio de los coeficientes cepstrales en la escala de frecuencias mel (MFCC). En la etapa de clasificación se comparan dos técnicas de clasificación multiclase, una de ellas basado en Procesos Gaussianos, mientras que la otra se basa en un modelo de Regresión logística multiclase que mide el desempeño de los anotadores en términos de la sensibilidad y la especificidad. La comparación de las dos técnicas de clasificación se realiza en términos de la precisión y de las curvas ROC.

2.1. Bases de datos

Para este trabajo se usaron dos bases de datos. La primera se denomina "Iris Plant Database", la segunda es una colección de muestras de voz evaluada en su calidad a partir del protocolo GRBAS.

2.1.1. "Iris Plant Database"

La base de datos "Iris Plant Database" consta de 150 muestras de cuatro dimensiones que corresponden a tres clases (50 muestras por cada clase), cada clase corresponde a una variedad de la planta iris (iris virginica, iris versicolor e iris setosa). Es necesario resaltar que esta base de datos sólo se usa con el fin de validar el funcionamiento de los algoritmos de clasificación para múltiples anotadores implementados en este trabajo.

2.1.2. Base de datos de voz

La base de datos voz usada fue la distribuida por la compañía Kay Elemetrics, la cual posee alrededor de 700 grabaciones. Las muestras acústicas consisten en la fonación sostenida de la vocal /a/ (las fonaciones tuvieron una duración entre $1s$ y $3s$), dichas muestras acústicas fueron tomadas tanto de personas sanas, como de personas con gran diversidad de desórdenes del habla en diferentes etapas (de fases iniciales hasta crónicas). Las grabaciones de la voz, se realizaron en un ambiente controlado y usando frecuencias de muestreo de 25KHz o 50KHz, se realizó un submuestreo a las señales de voz muestreadas a 50kHz, con el fin de normalizar todas las frecuencias de muestreo a 25KHz. La base de datos fue segmentada siguiendo el criterio explicado en [14]. Se tomó un conjunto de 218 muestras, de las cuales 51 muestras corresponden a personas normales y las 167 muestras restantes pertenecen a personas con diferentes tipos de desórdenes de la voz. Este conjunto de 218 muestras, fue valorado en su calidad por cuatro especialistas siguiendo el protocolo GRBAS (ver Apéndice A.3). La Tabla 1 se muestra la distribución de muestras por clase de ésta base de datos.

		Características				
		G	R	B	A	S
Clase	0	83	89	131	208	198
	1	30	41	32	1	2
	2	43	49	23	3	5
	3	62	39	32	6	13
Total		218	218	218	218	218

Tabla 1: División de muestras por clase para la base de datos de voz usada. Las filas corresponde a cada una de las clases dentro de la base de datos, por su parte las columnas corresponden a cada una de las características evaluadas en el protocolo GRBAS.

2.2. Extracción de características para señales de voz

La extracción de características es un proceso en el cual se obtiene un conjunto de datos (que se conocen como vector de características) a partir de un conjunto de variables [15] (Para este estudio, el conjunto de variables representa una serie de tiempo que

corresponde a una señal voz). La extracción de características tiene como objetivos principales: reducir la dimensionalidad y relazar aspectos de la señal que contribuyan a realizar procesos posteriores [16] [17] (en este caso realzar patrones asociados con algún tipo de patología). En el análisis de voz, existe un diverso número de métodos para extracción de características. En este estudio se reconocen dos grupos: los modelos que generan características de representación, las cuales no tienen significado físico relacionado con la producción de voz [18]; y los modelos que generan características acústicas las cuales sí poseen un sentido físico relacionado con la producción de la voz [15] [16]. En el primer grupo se encuentran los coeficientes cepstrales derivados de un análisis de predicción lineal (LPCC por su nombre en inglés Linear Predictive Cepstrum Coefficients), vectores de análisis de predicción perceptual (PLP por sus siglas en inglés Perceptual Linear Predictive) y los coeficientes cepstrales sobre la escala de frecuencias Mel (MFCC por su nombre en inglés Mel-Frecuency Cepstrum Coefficients) [15], por su parte en el segundo grupo se encuentran representaciones como el pitch, el jitter o la relación ruido-armónico (HNR por sus siglas en inglés Harmonic Noise Ratio) [15] [16].

2.2.1. Coeficientes cepstrales sobre la escala de frecuencias Mel (MFCC)

Se ha mostrado que la señal de voz contiene información acerca de la forma de onda de excitación [3]. Una de las finalidades de este estudio, es diseñar un sistema automático que permita valorar la calidad de voz, por dicha razón, es necesario usar un método de parametrización que tenga la capacidad de modelar los efectos que producen las patologías sobre el sistema de excitación del aparato fonador, específicamente sobre los pliegues o cuerdas vocales.

Los MFCC se consideran los más apropiados para este estudio debido principalmente a que demuestran una gran capacidad para modelar tanto movimientos irregulares de los pliegues vocales como anomalías en el cierre debido a masas alojadas sobre los pliegues o a cambios en las propiedades de los tejidos que recubren las cuerdas vocales. En las bandas bajas de los MFCC, se caracterizan las alteraciones producidas por masas alojadas sobre las cuerdas vocales. Mientras que en las bandas altas se modelan las componentes ruidosas generadas por las alteraciones en el cierre de las cuerdas vocales [3]. La caracterización por MFCC se realiza siguiendo los pasos mostrados en el Apéndice B.

2.3. Validación

2.3.1. Base de datos "Iris Plant Database"

La base de datos "Iris Plant Database" datos posee 150 muestras, donde cada muestra se compone de 4 características y una etiqueta verdadera. Es claro que esta base de datos no posee múltiples anotaciones, por lo que es necesario generar dichas anotaciones de manera sintética. Para el algoritmo de clasificación basado en regresión logística multiclase, las anotaciones se generan a partir de una distribución de probabilidad multinomial, donde la probabilidad de un suceso está dada por el vector de parámetros multinomiales α_{ck}^{j} (ver Ecuación 1.8). Ahora para el algoritmo de clasificación basado en Procesos Gaussianos para regresión, las anotaciones se generan añadiendo ruido Gaussiano, $\mathcal{N}(0, \sigma_j^2)$, a las etiquetas verdaderas. El parámetro σ_j^2 corresponde a la varianza asociada a cada anotador. A partir de del vector de parámetros multinomiales y de la varianza, se puede modificar el nivel de experiencia asociado a cada uno de los anotadores.

2.3.2. Base de datos de voz

Las señales de voz son segmentadas y ventaneadas usando ventanas Hamming con duración una duración de 40 ms cada una. Cada una de estas ventanas es caracterizada a partir de los MFCC, Dicha caracterización alimenta la etapa de clasificación. En esta etapa de clasificación la valoración automática de la calidad de voz usando el protocolo GRBAS (ver Apéndice A.3), se toma como cinco problemas de clasificación, es decir, se toma por separado la información de cada una de las características que componen dicho protocolo (Grado de Disfonía, Aspereza, Respiración dificultosa, Astenia y Voz forzada), y se obtiene la clasificación de cada una de ellas, mediante el uso de modelos de clasificación multiclase para múltiples anotadores. Cada clase corresponde a uno los valores en el rango $[0 - 3]$, el cual como ya se dijo, corresponde a la valoración que da el especialista a cada una de las características del protocolo GRBAS al escuchar una señal de voz determinada.

2.4. Algoritmos

La clasificación en este estudio se lleva a cabo usando las dos técnicas de clasificación que consideran múltiples anotadores estudiadas en el capítulo 1. La primera de estas técnicas se basa en Procesos Gaussianos para regresión, la segunda técnica se basa en un modelo de regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad. A continuación, se muestra los algoritmos utilizados para la implementación de los esquemas de clasificación comentados.

2.4.1. Procesos Gaussianos para regresión con múltiples anotadores

Originalmente este modelo basado en Procesos Gaussianos está diseñado para tareas de regresión, sin embargo en este trabajo se usa como clasificador. Para realizar la clasificación a partir de los resultados obtenidos con el sistema de regresión, se siguen los siguientes pasos.

- Se calcula la media del Proceso Gaussiano para la probabilidad posterior sobre una nueva observación (ver ecuación (1.13)).
- Ahora se calcula la distancia euclidiana entre el valor de la media del Proceso Gaussiano y cada una de las clases involucradas en el problema de clasificación.
- Por último se asigna la nueva observación a la clase que presente menor distancia euclidiana con respecto a la media del Proceso Gaussiano.

Algoritmo 1 Implementación del algoritmo de clasificación basado en procesos Gaussianos para regresión con múltiples anotadores

Entrada: El conjunto de características para entrenamiento $XTrain$, el conjunto de características para prueba $XTest$ y el conjunto de anotaciones realizadas por cada anotador $YTrain$.

Salida: Estimación probabilística de que un ejemplo de $XTest$ pertenezca a cada una de las clases k, donde $k = [1, \ldots, K]$.

1: Inicialización de parámetros: Vector de varianzas asociadas a cada anotador $\sigma_j^2 \leftarrow 0$. Vector de parámetros asociados a la función del kernel $\boldsymbol{\phi} \leftarrow 0$.

2: **repetir**

3: Se estiman los valores de los parámetros σ_j^2 y $\boldsymbol{\phi}$.

4: A partir del valor actual de los parámetros de varianza σ_j^2, se estiman los valores para la etiqueta verdadera $\hat{\mathbf{y}}$, y los valores para la matriz de covarianza $\hat{\boldsymbol{\Sigma}}$ (ver ecuación (1.14)).

5: Se evalúa el logaritmo negativo de la evidencia dado por (1.15).

6: **hasta que** se minimice el logaritmo de negativo de la verosimilitud marginal (1.15)

7: Por último se calcula la media del Proceso Gaussiano de la probabilidad posterior dada por la ecuación (1.13), a partir de esta media se asigna cada una de las muestras del conjunto de prueba $XTest$, a la clase k que presente menor distancia con respecto a la media del Proceso Gaussiano previamente calculada.

2.4.2. Regresión logística multiclase para múltiples anotadores

Algoritmo 2 Implementación del algoritmo de clasificación para múltiples anotadores usando regresión logística multiclase

Entrada: El conjunto de características para entrenamiento $XTrain$, el conjunto de características para prueba $XTest$ y el conjunto de anotaciones realizadas por cada anotador $YTrain$.

Salida: Estimación probabilística de que un ejemplo de $XTest$ pertenezca a cada una de las clases k, donde $k = [1, \ldots, K]$.

Inicialización de parámetros: Vector de ponderación $\mathbf{w} \leftarrow 0$. Vector de parámetros multinomiales $\alpha_{ck}^{j} \leftarrow 0$. Estimación etiqueta a partir de las anotaciones verdadera usando majority voting (el más votado), $\mu_{ic} \leftarrow \begin{cases} 1 & \text{Si} \quad 1/R \sum_{j=1}^{R} y_i^j > 0{,}5 \\ 0 & \text{Si} \quad 1/R \sum_{j=1}^{R} y_i^j < 0{,}5 \end{cases}$. La probabilidad que la muestra i pertenezca a la c, p_{ic}, éste parámetro se calcula según la ecuación (1.9).

2: **repetir**

Se calcula el gradiente de la función de costo usando la expresión (1.11).

4: Se determinan los valores de la Matriz Hessiana usando (1.12).

Usando el gradiente y la matriz Hessiana, se calcula el valor para el vector de ponderaciones $\mathbf{w}$ usando la ecuación (1.10)

6: Se estima la probabilidad que la muestra $\mathbf{x}_i$ pertenezca a cada una de las clases, a partir de la expresión (1.9).

Se calculan los valores del vector de parámetros multinomiales α_{ck}^{j} usando la expresión (1.8), para medir el rendimiento de los expertos con respecto al valor actual de μ_{ic}.

8: Se estima el valor de la etiqueta verdadera para la muestra i, μ_{ic}, a partir de la ecuación (1.7).

Se calcula el valor del logaritmo de la verosimilitud dado por (1.5).

10: **hasta que** se maximice el logaritmo de la verosimilitud dado por la ecuación (1.5)

Por último se calcula la probabilidad posterior de cada clase k dado el conjunto de prueba $XTest$ y el modelo aprendido en la fase de entrenamiento.

2.5. Medidas de desempeño

El objetivo de las técnicas que evalúan el rendimiento de los clasificadores, consiste en estimar la probabilidad del error de clasificación, probando la respuesta correcto/falso del clasificador usando un conjunto finito de N vectores de características [15]. Considérese un problema de clasificación con K clases, que tiene N_i vectores de características por clase, con $\sum_{i=1}^{K} N_i = N$. Sea P_i la correspondiente probabilidad de error para la clase w_i. Si se asume la "independencia entre los vectores de características" [19], se puede demostrar que el estimado para la probabilidad de error total está dado por

$$\hat{P} = \sum_{i=1}^{K} p(w_i) \frac{k_i}{N_i},$$

donde k_i es el número de vectores mal calificados que corresponden a la clase w_i es la probabilidad de ocurrencia de la clase w_i. Es posible demostrar [19] que $\hat{P}$ es un estimador de la verdadera probabilidad de error, y que es asintóticamente consistente cundo la cantidad de de muestras por clase tiende a infinito $N_i \to \infty$.
En la práctica, se cuentan con conjunto finitos de características para entrenar y validar el clasificador. Por lo tanto se sugieren algunos métodos para evaluar el rendimiento de un clasificador entre los que se encuentran: validación simple, validación cruzada usando k divisiones o validación Leave-One-Out. Para este trabajo se usó el método Leave-One-Out debido al reducido número de muestras de voz que pertenecen a algunas clases.

Validación Leave-One-Out

En este tipo de validación se utiliza un conjunto de entrenamiento conformado por $N-1$ muestras y la validación se lleva a cabo con la muestra excluida. Si ésta muestra se califica de manera errónea, se cuenta un error. Este procedimiento se realiza N veces excluyendo una muestra diferente en cada realización.
La estimación de la probabilidad del error, se lleva a cabo con el total de errores obtenidos de las N realizaciones. A diferencia de otros métodos como el de validación simple, la probabilidad de error estimada, no depende de las muestras que se usen en

los conjuntos de entrenamiento y prueba. Sin embargo su desventaja es su gran carga computacional [15].

Curvas ROC

Existen otros métodos que además de medir el rendimiento de los clasificadores, permiten la comparación con otros clasificadores. Entre estos métodos se encuentra principalmente las curvas ROC, que son una herramienta que permite evaluar el rendimiento de un clasificador en términos de la sensibilidad contra la especificidad, considerando todos los valores de umbral posibles. Una métrica derivada de las curvas ROC es el área bajo la curva (AUC), la cual puede ser usada como un estimado de la probabilidad que un clasificador asigne un mayor rango a un ejemplo positivo elegido al azar que a un ejemplo negativo elegido de manera aleatoria (Ver Apéndice D).

Capítulo 3

Resultados y Discusión

En el presente capítulo se exponen los resultados obtenidos al aplicar los algoritmos desarrollados en este trabajo, sobre las bases de datos disponibles. Se inicia reportando los resultados obtenidos con la base datos "Iris Plant Database". Se finaliza mostrando los resultados obtenidos con la base de datos de voz.

3.1. Resultados obtenidos sobre la base de datos "Iris Plant Database"

A continuación se presentan los resultados obtenidos al aplicar los esquemas de clasificación para múltiples anotadores desarrollados en este trabajo sobre la base de datos "Iris Plant Database". Es necesario aclarar que estos datos sólo sirven para validar el funcionamiento de los esquemas de clasificación estudiados.

3.1.1. Clasificador basado en regresión logística multiclase para múltiples anotadores

Para la evaluación de este clasificador se realizaron tres experimentos donde se generaron etiquetas sintéticas para tres anotadores (las etiquetas sintéticas se generan según lo descrito en la sección 2.3.1). En cada uno de los experimentos, se varía los valores del vector de parámetros multinomiales asociados a cada anotador. La validación para estos experimentos se realiza usando el método Leave-One-Out. Se reporta la precisión y el área bajo la curva ROC (AUC).

Experimento 1

Para este experimento se consideran tres anotadores. El primer anotador tiene una capacidad de acierto alta (experto), El segundo anotador se considera que tiene un nivel de acierto medio, por último, el tercer anotador tiene una capacidad de acierto baja (novato). A estos anotadores se les asignan los siguientes vectores de parámetros multinomiales

$$\alpha_{ck}^{1} = \begin{pmatrix} 0{,}8 & 0{,}1 & 0{,}1 \\ 0{,}1 & 0{,}8 & 0{,}1 \\ 0{,}1 & 0{,}1 & 0{,}8 \end{pmatrix}, \quad \alpha_{ck}^{2} = \begin{pmatrix} 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}2 & 0{,}6 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{pmatrix}, \quad \alpha_{ck}^{3} = \begin{pmatrix} 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 \\ 0{,}3 & 0{,}3 & 0{,}4 \end{pmatrix}.$$

Experimento 2

Para este experimento se consideran tres anotadores. El primer y segundo anotador tienen una capacidad de acierto alta (experto), por último el tercer anotador tiene una capacidad de acierto tan baja que se considera anotador malicioso. Los vectores de parámetros multinomiales asignados a cada uno de los anotadores son

$$\alpha_{ck}^{1} = \begin{pmatrix} 0{,}9 & 0{,}05 & 0{,}05 \\ 0{,}05 & 0{,}9 & 0{,}05 \\ 0{,}05 & 0{,}05 & 0{,}9 \end{pmatrix}, \quad \alpha_{ck}^{2} = \begin{pmatrix} 0{,}8 & 0{,}1 & 0{,}1 \\ 0{,}1 & 0{,}8 & 0{,}1 \\ 0{,}1 & 0{,}1 & 0{,}8 \end{pmatrix}, \quad \alpha_{ck}^{3} = \begin{pmatrix} 0{,}3 & 0{,}3 & 0{,}4 \\ 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}.$$

Experimento 3

Al igual que para los anteriores experimentos, se consideran tres anotadores. El primer anotador tiene una capacidad de acierto media, el segundo anotador tiene una capacidad de acierto baja y por último el tercer anotador tiene una capacidad de acierto muy baja (malicioso). A estos anotadores se les asigna los siguientes vectores de parámetros multinomiales

$$\alpha_{ck}^{1} = \begin{pmatrix} 0{,}6 & 0{,}2 & 0{,}2 \\ 0{,}2 & 0{,}6 & 0{,}2 \\ 0{,}2 & 0{,}2 & 0{,}6 \end{pmatrix}, \quad \alpha_{ck}^{2} = \begin{pmatrix} 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 \\ 0{,}3 & 0{,}3 & 0{,}4 \end{pmatrix}, \quad \alpha_{ck}^{3} = \begin{pmatrix} 0{,}3 & 0{,}3 & 0{,}4 \\ 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}3 & 0{,}4 & 0{,}3 \end{pmatrix}.$$

En la Tabla 2, se registran los valores de Precisión y AUC obtenidos para los experimentos descritos anteriormente.

	Precisión	AUC
Experimento 1	0,9133	0,9861
Experimento 2	0,9667	0,9863
Experimento 3	0,6933	0,8567

Tabla 2: Resultados obtenidos al aplicar el esquema de clasificación para múltiples anotadores basado en el modelo de Regresión logística multiclase sobre la base de datos "Iris Plant Database" según los experimentos descritos. En la columna denominada Precisión se reporta el rendimiento del clasificador en términos de la precisión. En la columna AUC se muestra el área bajo la curva ROC.

En la Tabla 2 se observa que que a pesar de no contar con las etiquetas verdaderas en la etapa de entrenamiento, es posible obtener rendimientos tan altos como un AUC de 0,9863 y una precisión de 0,9667. Estos resultados permiten validar el funcionamiento del clasificador para múltiples anotadores basado en regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad.

Se observa una dependencia entre el rendimiento del clasificador y la calidad de los anotadores, puesto que al disminuir considerablemente la calidad de los anotadores usados en estos experimentos (ver Experimento 3) se obtuvieron valores de rendimiento tan bajos como 0,8567 para AUC y 0,6933 para la precisión.

3.1.2. Clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores

Para la evaluación de este clasificador se realizaron tres experimentos donde se generaron etiquetas sintéticas para tres anotadores (las etiquetas sintéticas se generan según lo descrito en la sección 2.3.2). En cada uno de los experimentos, se varía los valores de varianza σ_j^2 asociados a cada anotador. La validación para estos experimentos se realiza a partir del método Leave-One-Out. Se reporta la precisión y el área bajo la curva ROC (AUC).

Experimento 1

Para el primer experimento se consideran dos anotadores con un nivel de acierto alto (expertos), el anotador restante corresponde a un anotador con una capacidad de acierto media. A continuación, se muestra el vector de varianzas σ^2.

$$\sigma_1^2 = [0{,}1 \quad 0{,}2 \quad 0{,}5].$$

Experimento 2

Para el segundo experimento se consideran dos anotadores con un nivel de acierto medio, el anotador restante corresponde a un anotador con una capacidad de acierto muy baja. El vector de varianzas σ^2 para este experimento es el siguiente.

$$\sigma_2^2 = [0{,}9 \quad 1{,}1 \quad 6{,}5].$$

Experimento 3

Para el segundo experimento se consideran dos anotadores con un nivel de acierto muy bajo, el anotador restante corresponde a un anotador con una capacidad de acierto media. A continuación, se muestra el vector de varianzas σ^2.

$$\sigma_3^2 = [0{,}8 \quad 6{,}2 \quad 6{,}5].$$

En la Tabla 3, se registran los valores de Precisión y AUC obtenidos para los experimentos descritos anteriormente.

	Precisión	AUC
Experimento 1	0,9533	0,9722
Experimento 2	0,9133	0,9691
Experimento 3	0,8600	0,9300

Tabla 3: Resultados obtenidos al aplicar el esquema de clasificación para múltiples anotadores basado en el modelo de Procesos Gaussianos para regresión sobre la base de datos "Iris Plant Database" según los experimentos descritos. En la columna denominada Precisión se reporta el rendimiento del clasificador en términos de la precisión. En la columna AUC se muestra el área bajo la curva ROC.

En la Tabla 3 se observa que que a pesar de no contar con la etiquetas verdaderas en la etapa de entrenamiento, es posible obtener rendimientos tan altos como un AUC de 0,9722 y una precisión de 0,9533. Estos resultados permiten validar el funcionamiento del clasificador para múltiples anotadores basado Procesos Gaussianos para regresión con múltiples anotadores.

Además se observa una clara dependencia entre el rendimiento del clasificador y la calidad de los anotadores, puesto que al disminuir considerablemente la calidad de los anotadores usados en estos experimentos (ver Experimento 3) se obtuvieron valores de rendimiento tan bajos como 0,9300 para AUC y 0,8600 para la precisión.

3.2. Resultados obtenidos sobre la base de datos de voz

3.2.1. Problema de clasificación binaria

Originalmente la base de datos de voz configura un problema de clasificación de cuatro clases. Se efectúa un primer experimento donde se convierte ésta base de datos en un problema de clasificación con dos clases (clasificación binaria), dicha conversión se realiza así: Las muestras de voz pertenecientes a las clases 0 y 1, se les asigna la etiqueta 0 (voz normal); por su parte, a las muestras de voz que pertenecen a las clases 2 y 3, se les asigna la etiqueta 1 (voz patológica). La base de datos modificada es usada por los algoritmos de múltiples anotadores desarrollados en este trabajo. Se realiza un experimento control, que usa la base de datos modificada sobre un esquema

de clasificación basado en regresión logística, donde la etiqueta verdadera se estima usando el procedimiento denominado "Majority Voting", usado en la literatura ([2]) para convertir problemas de clasificación con múltiples anotadores en un problema de clasificación típico.

- La Tabla 4 muestra los resultados de precisión para el problema de clasificación binaria obtenidos con los esquemas de clasificación para múltiples anotadores desarrollados en este trabajo y con el experimento de control que usa un clasificador típico.

- La Tabla 5 muestra los resultados de rendimiento (en términos AUC) obtenidos para el problema de clasificación binaria con el experimento de control y con los esquemas de clasificación para múltiples anotadores desarrollados en este trabajo.

	Características														
Parámetros	G			R			B			A			S		
	Precisión			Precisión			Precisión			Precisión			Precisión		
	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO
MFCC-3	**0,70**	**0,70**	0,68	**0,66**	0,65	**0,66**	0,75	**0,77**	0,75	**0,96**	**0,96**	0,95	**0,92**	0,90	0,91
MFCC-6	0,69	**0,71**	0,68	0,62	**0,67**	0,64	0,75	**0,77**	0,72	0,95	**0,96**	0,91	0,88	**0,92**	0,89
MFCC-12	**0,71**	0,70	0,66	**0,65**	0,62	0,64	0,72	**0,77**	0,71	0,91	**0,96**	0,89	0,79	**0,91**	0,83

Tabla 4: Resultados considerando un problema de clasificación binaria. Se consideran tres experimentos donde en cada uno de ellos se varia el número de parámetros usados en el proceso de caracterización de las señales de voz. Para el experimento MFCC-3 se usan 3 MFCC además de las derivadas de primer y segundo orden para un total de 12 parámetros. Para el experimento MFCC-6 se obtienen un total de 21 parámetros. En el experimento MFCC-12 se usan un total de 39 parámetros. Cada experimento se realiza para cada una de las características que pertenecen al protocolo GRBAS. Se reporta la precisión para cada uno de los experimentos realizados. En las columnas "RL" se reporta la precisión para el clasificador basado en regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad. En las columnas "PG" se reporta los valores de precisión para el clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores. Por último en las columnas denominadas "CO" se reporta la precisión para el experimento de control, el cual usa un clasificador típico basado en Regresión logística.

Al analizar los resultados de precisión obtenidos con los clasificadores usados en el problema de clasificación binaria (clasificar una señal de voz como patológica o sana) y que son mostrados en la Tabla 4, es posible determinar que el mejor clasificador de múltiples anotadores es el basado en Procesos Gaussianos para regresión, con el que se obtuvo una precisión promedio máxima de 0,81, mientras que el clasificador para múltiples anotadores basado en regresión logística multiclase que mide el rendimiento del anotador en términos de sensibilidad y especificidad obtuvo una precisión promedio máxima de 0,80. Claramente no se observan grandes diferencias entre los resultados obtenidos, sin embargo cabe resaltar el rendimiento obtenido por el clasificador basado en Procesos Gaussianos para regresión, ya que originalmente fue diseñado para tareas de regresión.

También se observa que los esquemas de clasificación para múltiples anotadores implementados en este trabajo obtuvieron mejores resultados que el experimento de control el cual implementa un clasificador típico usando "Majority Voting" para calcular la etiqueta verdadera a partir de las anotaciones (con este experimento se obtuvo una precisión promedio máxima de 0,79).

	Características														
Parámetros	G			R			B			A			S		
	AUC			AUC			AUC			AUC			AUC		
	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO
MFCC-3	0,75	**0,77**	0,75	0,66	**0,68**	0,65	0,70	**0,72**	0,70	0,39	**0,48**	0,43	0,44	**0,65**	0,58
MFCC-6	0,74	**0,77**	0,73	0,65	**0,68**	0,63	0,67	**0,71**	0,70	0,41	**0,55**	0,32	**0,62**	0,58	0,60
MFCC-12	0,69	**0,76**	0,70	**0,65**	0,49	0,62	0,66	**0,70**	0,67	0,54	0,48	**0,62**	0,54	**0,60**	0,45

Tabla 5: Resultados considerando un problema de clasificación binaria. Se consideran tres experimentos donde en cada uno de ellos se varia el número de parámetros usados en el proceso de caracterización de las señales de voz. Para el experimento MFCC-3 se usan 3 MFCC además de las derivadas de primer y segundo orden para un total de 12 parámetros. Para el experimento MFCC-6 se obtienen un total de 21 parámetros. En el experimento MFCC-12 se usan un total de 39 parámetros. Cada experimento se realiza para cada una de las características que pertenecen al protocolo GRBAS. Se reporta el área bajo la curva ROC (AUC) para cada uno de los experimentos realizados. En las columnas "RL" se reporta el AUC para el clasificador basado en regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad. En las columnas "PG" se reporta los valores AUC para el clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores. Por último en las columnas denominadas "CO" se reporta el AUC para el experimento de control, el cual usa un clasificador típico basado en Regresión logística.

Analizando los resultados de AUC de los clasificadores usados en el problema de clasificación binaria (clasificar una señal de voz como patológica o sana), mostrados en la Tabla 4, es posible afirmar que el mejor clasificador para múltiples anotadores es el basado en Procesos Gaussianos para regresión, con el que se obtuvo un AUC promedio máximo de 0,66, mientras que el clasificador para múltiples anotadores basado en regresión logística multiclase que mide el rendimiento del anotador en términos de sensibilidad y especificidad obtuvo un AUC promedio máximo de 0,62. Claramente no se observan grandes diferencias entre los resultados obtenidos con los clasificadores, sin embargo cabe resaltar el rendimiento obtenido por el clasificador basado en Procesos Gaussianos para regresión, ya que originalmente fue diseñado para tareas de regresión.

Los esquemas de clasificación para múltiples anotadores implementados en este trabajo obtuvieron mejores resultados que el experimento de control el cual implementa un

clasificador típico usando “Majority Voting” para calcular la etiqueta verdadera a partir de las anotaciones (con este experimento se obtuvo un AUC promedio máximo de 0,62).

3.2.2. Problema de clasificación multiclase

La base de datos de voz configura un problema de clasificación con cuatro clases, donde cada una de estas clases corresponde a la valoración emitida por el especialista sobre una muestra de voz determinada siguiendo el protocolo GRBAS. Esta base de datos es usada sobre los esquemas de clasificación desarrollados en este trabajo. Además se realiza un experimento de control, que usa la base de datos de voz sobre un esquema de clasificación basado en regresión logística multiclase, donde la etiqueta verdadera se estima usando el procedimiento denominado “Majority Voting”, usado comúnmente en la literatura ([2]) para convertir problemas de clasificación con múltiples anotadores en un problema de clasificación típico.

- La Tabla 6 muestra los resultados de precisión para el sistema de valoración automática de la calidad de voz (clasificación multiclase) obtenidos con los esquemas de clasificación para múltiples anotadores desarrollados en este trabajo y con el experimento de control que usa un clasificador típico.

- La Tabla 7 muestra los resultados de rendimiento en términos de AUC para el sistema de valoración automática de la calidad de voz (clasificación multiclase) obtenidos con los esquemas de clasificación para múltiples anotadores desarrollados en este trabajo y con el experimento de control que usa un clasificador típico.

Parámetros	Características														
	G			R			B			A			S		
	Precisión			Precisión			Precisión			Precisión			Precisión		
	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO
MFCC-3	**0,50**	0,22	0,40	0,41	0,27	**0,49**	0,57	0,32	**0,67**	**0,96**	**0,96**	0,87	0,87	0,80	**0,90**
MFCC-6	**0,44**	0,25	0,40	0,42	0,25	**0,45**	0,55	0,32	**0,61**	0,85	**0,96**	0,86	0,78	0,80	**0,83**
MFCC-12	**0,45**	0,25	0,35	0,39	**0,91**	0,41	0,45	**0,94**	0,55	0,85	**0,97**	0,89	0,70	**0,82**	0,77

Tabla 6: Resultados del sistema de valoración automática de la calidad de voz (problema de clasificación multiclase). Se consideran tres experimentos donde en cada uno de ellos se varia el número de parámetros usados en el proceso de caracterización de las señales de voz. Para el experimento MFCC-3 se usan 3 MFCC además de las derivadas de primer y segundo orden para un total de 12 parámetros. Para el experimento MFCC-6 se obtienen un total de 21 parámetros. En el experimento MFCC-12 se usan un total de 39 parámetros. Cada experimento se realiza para cada una de las características que pertenecen al protocolo GRBAS. Se reporta la precisión para cada uno de los experimentos realizados. En las columnas "RL" se reporta la precisión para el clasificador basado en regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad. En las columnas "PG" se reporta los valores de precisión para el clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores. Por último en las columnas denominadas "CO" se reporta la precisión para el experimento de control, el cual usa un clasificador típico basado en Regresión logística.

Al analizar los resultados de precisión obtenidos con los clasificadores usados en el sistema de valoración automática de la calidad de voz (problema multiclase) mostrados en la Tabla 6, es posible determinar que el mejor clasificador para múltiples anotadores es el basado en Procesos Gaussianos para regresión, con el que se obtuvo una precisión promedio máxima de 0,78, mientras que el clasificador para múltiples anotadores basado en regresión logística multiclase que mide el rendimiento del anotador en términos de sensibilidad y especificidad obtuvo una precisión promedio máxima de 0,67.

También se observa que los esquemas de clasificación para múltiples anotadores implementados en este trabajo obtuvieron mejores resultados que el experimento de control el cual implementa un clasificador típico usando "Majority Voting" para calcular la etiqueta verdadera a partir de las anotaciones (con este experimento se obtuvo una precisión promedio máxima de 0,67).

Se observa una disminución en los rendimientos para tareas de valoración de la calidad de voz (clasificación multiclase), al contrastarlos con los obtenidos al evaluar el rendimiento en tareas de detección de voces patológicas (clasificación binaria). Este comportamiento se justifica debido a que el número de muestras que pertenecen a cada una de las clases, están desbalanceadas en gran proporción (ver Tabla 1).

	Características														
Parámetros	G			R			B			A			S		
	AUC			AUC			AUC			AUC			AUC		
	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO	RL	PG	CO
MFCC-3	**0,58**	0,57	**0,58**	**0,56**	**0,56**	**0,56**	0,50	**0,59**	**0,59**	**0,57**	0,50	0,41	0,52	0,52	**0,60**
MFCC-6	0,59	0,58	**0,60**	**0,57**	0,55	0,55	0,53	**0,58**	0,56	0,38	**0,56**	0,41	0,46	**0,67**	0,50
MFCC-12	0,58	0,55	**0,59**	0,54	0,50	**0,57**	0,53	0,47	**0,57**	0,37	0,38	**0,43**	0,54	**0,57**	0,50

Tabla 7: Resultados del sistema de valoración automática de la calidad de voz (problema de clasificación multiclase). Se consideran tres experimentos donde en cada uno de ellos se varia el número de parámetros usados en el proceso de caracterización de las señales de voz. Para el experimento MFCC-3 se usan 3 MFCC además de las derivadas de primer y segundo orden para un total de 12 parámetros. Para el experimento MFCC-6 se obtienen un total de 21 parámetros. En el experimento MFCC-12 se usan un total de 39 parámetros. Cada experimento se realiza para cada una de las características que pertenecen al protocolo GRBAS. Se reporta el área bajo la curva ROC (AUC) para cada uno de los experimentos realizados. En las columnas "RL" se reporta el AUC para el clasificador basado en regresión logística multiclase que mide el rendimiento de los anotadores en términos de sensibilidad y especificidad. En las columnas "PG" se reporta los valores AUC para el clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores. Por último en las columnas denominadas "CO" se reporta el AUC para el experimento de control, el cual usa un clasificador típico basado en Regresión logística.

los resultados de AUC para los clasificadores usados en el sistema de valoración automática de la calidad de voz (problema multiclase) mostrados en la Tabla 6, determinan que el mejor clasificador para múltiples anotadores es el basado en Procesos Gaussianos para regresión, con el que se obtuvo un AUC promedio máximo de 0,59, mientras que el clasificador para múltiples anotadores basado en regresión logística multiclase que mide el rendimiento del anotador en términos de sensibilidad y especificidad obtuvo un

AUC promedio máximo de 0,55. Claramente no se observan grandes diferencias entre los resultados obtenidos con los clasificadores, sin embargo cabe resaltar el rendimiento obtenido por el clasificador basado en Procesos Gaussianos para regresión, ya que originalmente fue diseñado para tareas de regresión.

Los esquemas de clasificación para múltiples anotadores implementados en este trabajo obtuvieron mejores resultados que el experimento de control el cual implementa un clasificador típico usando "Majority Voting" para calcular la etiqueta verdadera a partir de las anotaciones (con este experimento se obtuvo un AUC promedio máximo de 0,55).

Se observa una disminución en los rendimientos para tareas de valoración de la calidad de voz (clasificación multiclase), al contrastarlos con los obtenidos al evaluar el rendimiento en tareas de detección de voces patológicas (clasificación binaria). Este comportamiento se justifica debido a que el número de muestras que pertenecen a cada una de las clases, están desbalanceadas en gran proporción (ver Tabla 1).

La Tabla 6 y la Tabla 7, reflejan una diferencia considerable entre los resultados de rendimiento obtenidos en términos precisión y AUC, éste fenómeno se debe al gran nivel de desbalance de muestras por clase que presenta la base de datos de voz usada en este trabajo (ver Tabla 1).

Capítulo 4

Conclusiones

De lo observado se puede concluir.

- Los resultados obtenidos con los experimentos realizados sobre la base de datos "Iris Plant", permiten validar el funcionamiento del algoritmo de clasificación para múltiples anotadores basado en regresión logística multiclase con sensibilidades y especificidades por anotador.
- Los resultados del experimento realizado sobre la base de datos "Iris Plant" verifican el buen funcionamiento del algoritmo de clasificación para múltiples anotadores basado en Porcesos Gaussianos para regresión.
- El mejor esquema de clasificación para múltiples anotadores que es aquel basado en Procesos Gaussianos para regresión. A pesar que las diferencias de rendimiento con el clasificador para múltiples anotadores basado en regresión logística multiclase no son significativamente grandes,cabe resaltar que el clasificador basado en Procesos Gaussianos, fue diseñado originalmente para tareas de regresión.
- Según los resultados obtenidos con la base de datos de voz usada en este trabajo, es viable usar esquemas de clasificación para múltiples anotadores en tareas de valoración automática de la calidad de voz, puesto que los valores de rendimiento obtenidos con estos fueron levemente superiores a los valores obtenidos con esquemas de clasificación típicos, donde se estimaba la etiqueta verdadera usando "Majority Voting" [2].

4.1. Trabajo Futuro

A partir de este trabajo, es posible determinar algunos aspectos que se pueden implementar en el futuro

- En este trabajo para cada experimento fue necesario realizar cinco procesos de clasificación (uno para cada característica del protocolo GRBAS). Sería posible usar técnicas de aprendizaje automático multitarea con el fin de aprovechar la posible correlación existente entre las anotaciones suministradas por los anotadores para cada una de las características en la escala GRBAS.

- Implementar un algoritmo de clasificación multiclase basado en Procesos Gaussianos que considere el caso de múltiples anotadores. Al usar este clasificador, se esperaría mejorar el rendimiento obtenido con el clasificador basado en Procesos Gaussianos para regresión con múltiples anotadores.

- Diseñar esquemas de clasificación para múltiples anotadores que puedan mejorar el rendimiento ante casos donde el número de muestras por clase en una base de datos no esté equilibrada.

- Aplicar las técnicas desarrolladas en este trabajo a problemas en otras áreas como lo puede ser la minería de datos o aplicarlo a otros campos de la medicina como el estudio de imágenes diagnósticas.

Bibliografía

[1] P. Groot, A. Birlutiu, y T.Heskes: *Learning from multiple annotators with Gaussian Processes*. In Proc. of the 21st International Conference on Artificial Neural Networks, páginas 159–164, 2011.

[2] V.C. Raykar, S. Yu, L.H. Zhao, G. Hermosillo-Valadez, C. Florin, L. Bogoni, y L. Moy : *Learning from crowds*. JMLR, 11, 12971322, 2010.

[3] J.I. Godino-Llorente, P. Gómez-Vilda, y M. Blanco-Velasco : *Dimensionality Reduction of a pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short Term Cepstral Parameters*. IEEE Transaction on Biomedical Engineering, páginas 1943-1953, 2006.

[4] N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, M. Blanco-Velasco y F. Cruz-Roldán: *Automatic Assessment of Voice Quality According to the GRBAS Scale*. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, páginas 2478 - 2481, 2006.

[5] M. P. Karnell, S. D. Melton, J. M. Childes, T. C. Coleman, S. A. Dailey, y H. T. Hoffman: *Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders*, páginas 576–90, 2007.

[6] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, y J.Ma Gutiérrez-Arriola: *Automatic GRBAS Assessment Using Complexity Measures and a Multiclass GMM-BASED Detector*, 2011.

[7] A. P Dawid, y A. M. Skene: *Maximum Likelihood Estimation of Observer Errorrates using the EM Algorithm*. In Applied Statistics, páginas 20–28, 1979.

[8] P. Smyth, U. Fayyad, M. Burl, P. Perona, y P. Baldi: *Inferring Ground Truth from Subjective Labelling of Venus Images*. In Advances in Neural Information Processing Systems 7, páginas 1085–1092, 1995.

[9] S. R. Cholleti et. al: *Veritas Combining expert opinions without labeled data*. In International Journal on Artificial Intelligence Tools, páginas 633-651, 2009.

[10] O. Dekel, y O. Shamir: *Vox populi Collecting high-quality labels from a crowd*. 2009.

[11] O. Dekel, C. Gentile, y K. Sridharan: *Selective sampling and active learning from single and multiple teachers*. In Journal of Machine Learning Research, páginas 2655-2697, 2012

[12] A. P. Dempster, N. M. Laird y D. B. Rubin: *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society: páginas 1938, 1977.

[13] C. Bishop : *Pattern recognition and machine learning*. New York, Springer, 2006.

[14] V. Parsa, y D. G. Jamieson: *Identification of pathological voices using glottal noise measures*. Journal of Speech, Language and Hearing Research, 2000.

[15] M. A. Álvarez-López: *Reconocimiento de Voz sobre Diccionarios Reducidos Usando Modelos Ocultos de Markov*. Tesis Pregrado, Universidad Nacional sede Manizales, 2004.

[16] C. Duque-Sánchez, M. Morales-Pérez: *Caracterización de voz empleando análisis tiempo-frecuencia aplicada al reconocimiento de emociones*. Tesis Pregrado, Universidad Tecnológica de Pereira, 2007.

[17] F. A. Sepúlveda: *Feature Extraction of Speech Signals using Time-Frecuency Analysis*. requisito parcial para optar al título de Magister, 2004.

[18] C. F. Ojeda: *Extracción de características usando transformada wavelet en la identificación de voces patológicas*, 2003.

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2003.

[20] R. Echeverri-Rivera : *Identificación de parámetros prosódicos en señales de voz, mediante el uso de dispositivos de hardware reconfigurable (FPGA)*. Tesis Pregrado, Universidad Tecnológica de Pereira, 2008.

[21] M. C. Melandi, A. Jackson : *La Voz Patológica*. Ed. Médica Panamericana, 2002.

[22] M. Hirano : *Clinical Examination of Voice*. Springer-Verlag, 1981.

[23] J. Revis, A. Giovanni, y F. Wuyst : *Comparison of different types of vowel fragments for the evaluation of voice quality*. In Proceedings of Voicedata '98, páginas 80-85, 1998.

[24] S. Bedoya-Jaramillo et al : *Automatic Emotion Detection in Speech Using Mel frecuency Cepstral Coefficients*. In Symposium of Image, Signal Processing, and Artificial Vision (STSIVA), páginas 62-65, 2012.

[25] F. Z. Chelali, y A. Djeradi : *MFCC and vector quantization for Arabic fricatives speech/speaker recognition*. In 2012 International Conference on Multimedia Computing and Systems (ICMCS), páginas 284-289, 2012.

[26] M. R. Hasan, M. Jamil, G. Rabbani, y S. Rahman : *Speaker identification using Mel frequency cepstral coefficients*. 2004

[27] S. K. Mitra: *Digital Signal Processing: A computer-Based Approach*, 2nd ed. Singapore: McGraw-Hill, 2001.

[28] K. B. Petersen, y M. S. Pedersen : *The Matrix Cookbook*. Disponible en internet. 2007.

[29] J. A. Swets : *Measuring the accuracy of diagnostic systems*. Science, páginas 1285-1293, 1988.

[30] T. Fawcet : *An Introduction to ROC Analysis*. In Pattern Recognition Letters, páginas 861-874, 2005.

[31] T. Fawcett : *ROC graphs: Notes and practical considerations for researchers*. Machine Learning, páginas 1-38, 2004.

[32] D. J. Hand y R. J. Till: *A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems*. In Machine Learning, páginas 171-186. 2001.

[33] F. Provost y P. Domingos: *Well-trained PETs: Improving probability estimation trees*. CeDER Working Paper. 2001.

Apéndice A

Aparato Fonador

A.1. Anatomía

En el proceso de producción de la voz intervienen órganos del sistema digestivo y del sistema respiratorio los cuales son controlados por el sistema nervioso central [16]. Esencialmente la voz se produce mediante la vibración de las cuerdas vocales (también conocidas como pliegues vocales), que se propaga a través de la laringe y de las cavidades bucal y nasal. El aparato fonador se compone de tres partes fundamentales [16] [20].

- **Sistema de Generación:**
  Los músculos abdominales y torácicos incrementan la presión en los pulmones lo que produce un exceso en el flujo de aire, este flujo sale por los bronquios y la tráquea hasta llegar a la faringe donde excita el sistema de vibración [16] [20].

- **Sistema de Vibración:**
  Este sistema está conformado básicamente por los pliegues vocales, los cuales se dividen en dos pares, superiores e inferiores. En el caso de la respiración, las cuerdas vocales se abren hacia las paredes de la laringe permitiendo que el aire fluya libremente. En el caso de la producción de voz la cuerdas vocales se tensan y se unen, así el aire choca contra ellas produciendo sonidos que son articulados por el sistema resonante [20].

- **Sistema Resonante:**
  Se compone de tres cavidades articulatorias: Cavidad faríngea, Cavidad oral y Cavidad nasal. Los sonidos producidos por el sistema de vibración se desplazan

hasta los orificios nasales y la boca, la articulación de las cavidades modifica y amplifica los sonidos que son expulsados al exterior [16] [20].

Los órganos que intervienen en el proceso de fonación, se pueden dividir en tres grupos bien delimitados: Cavidades infra-glóticas, Cavidad Laríngea y Cavidades supra-glóticas.

A.1.1. Cavidades infra-glóticas

Las cavidades infra-glóticas están compuestas por los órganos encargados de la respiración (pulmones, bronquios y tráquea), que representan a fuente de energía para el proceso de la producción de voz. En el momento de la inspiración, los pulmones toman aire bajando el diafragma y expandiendo la cavidad torácica. Al momento de la fonación, la espiración provocada por la contracción del diafragma, aporta la energía necesaria para generar la onda de presión acústica que se convertirá en voz al atravesar los órganos fonadores superiores [16] [20].

A.1.2. Cavidad Laríngea

En el proceso de fonación, la cavidad laríngea es la responsable de modificar el flujo de aire generado por los pulmones en una señal susceptible de excitar adecuadamente las posibles configuraciones de las cavidades supra-glóticas. El último cartílago de la tráquea, el cricoides, forma la base de la laringe, cuyo principal órgano son los pliegues vocales que son dos pares de repliegues compuestos de ligamentos y músculos. El par inferior son las llamadas cuerdas vocales verdaderas, que pueden juntarse o separarse mediante la acción de los músculos crico-aritenoides lateral y posterior, y que están protegidas en su parte anterior por el cartílago tiroides [16].

A.1.3. Cavidades Supra-glóticas

Las cavidades supra-glóticas están constituidas por la faringe, la cavidad oral y la cavidad nasal. Su rol en el proceso de fonación es modificar adecuadamente el flujo de aire procedente de la laringe para así generar la señal acústica emitida a través de la nariz y la boca [16]

A.2. Patologías

Úlcera de Contacto o Granuloma: Es una patología poco frecuente, producida por irritación crónica y formación de tejido de granulación en la parte superior de la cuerda vocal. Se observa en pacientes que realizan esfuerzo excesivo para hablar, con reflujo gastroesofágico hiperacidez gástrica y tos crónica. Por lo general los granulomas son bilaterales siendo más común en hombres de 40 a 60 años, que consumen alcohol o tabaco. Los pacientes a quienes se les diagnostica esta patología, presentan molestias para tragar y hablar, sensación de cuerpo extraño, carraspeo por la necesidad de aclarar la voz. [21].

Carcinoma de Laringe: Es una tumoración maligna que afecta a personas con edades cercanas a los sesenta años, este tipo de patología se produce principalmente por el consumo de tabaco, además también se puede producir por exposición prolongada a radiaciones, aunque también se ha encontrado una fuerte relación con el síndrome de Plummer-Vinson [21].

Nódulos Vocales: Los nódulos vocales constituyen uno de los trastornos más comunes en las personas que abusan de su voz. Si bien pueden presentarse a cualquier edad, son más frecuentes entre niños varones y mujeres adultas. Suele ser el temor de los cantantes, aunque muchas veces no interfieren en la producción de la voz. Los nódulos suelen presentarse con relativa frecuencia en profesores, actores, telefonistas, entrenadores, cantantes, etc. El síntoma más común es la disfonía, ronquera con voz áspera, tendencia a tonos graves y fatiga vocal con el correr del día. Algunos cantantes refieren incapacidad para elevar el tono de la voz y sensación de realizar mayor esfuerzo al cantar [21].

Quiste Intracordal: Existen dos tipos de quiste intracordal. El primero se debe a la obstrucción de una glándula con retención de material mucoso. El segundo es un quiste de tipo epitelial. Ambos se localizan en el espacio de Reinke, justo por debajo del epitelio escamoso y raramente en el músculo. Los quistes intracordales son difíciles de diferenciar de los pólipos o nódulos pequeños. Presentan ronquera y, a medida que crecen, aparece diplofonía. Refieren disminución de capacidad vocal. Por lo general son

unilaterales, pero casi siempre se observa un edema asociado en la cuerda vocal opuesta [21].

Edema de Reinke: El edema de Reinke ha sido asociado con fumadores y a veces con personas que abusan de su voz. Otros consideran que uno de los síntomas del reflujo gastroesofágico es el edema de las cuerdas vocales. Es bilateral, raramente se observa en una cuerda vocal y ocurre más en varones mayores de 40 años. Se puede observar también en el hipotiroidismo. El paciente refiere disfonía crónica, voz con tono bajo, tanto en el hombre como en la mujer y en algunas ocasiones puede producir obstrucción respiratoria. Las mujeres se quejan de voz masculinizada y los cantantes de disminución del registro vocal [21].

A.3. Protocolo para la valoración de la calidad de voz (Escala GRBAS)

La escala GRBAS es un protocolo propuesto por [22], usado en el análisis perceptivo de la calidad de voz. Este protocolo es aceptado como estándar por la sociedad Japonesa de Fonoaudiología y Foniatrías, y por el grupo Europeo de la laringe. La escala GRBAS comprende cinco características cualitativas, las cuales son, (G, *Grade of Disphony*) Grado de Disfonía , (R, *Roughness*) Aspereza , (B, *Breathiness*) Respiración Dificultosa, (A, *Asthenicity*) Astenia o fatiga vocal y (S, *Strainess*) Voz forzada [5], a cada una de estas características se le asignan un valor en el rango $[0-3]$, donde 0, corresponde a una voz saludable, 1, alteración manifiesta, 2 alteración moderada y 3, alteración severa.

La severidad de la ronquera se cuantifica en el parámetro G. Se pueden identificar dos componentes asociados a la ronquera: Respiración dificultosa (B) que es la sensación audible de la fuga de aire turbulento a través de un cierre glotal insuficiente, lo que puede inducir a momentos áfono cortos (segmento sin voz), y la rugosidad (R) que es una impresión acústica de impulsos glóticos irregulares [4]. El parámetro Astenia se relaciona con la hipofunción de las cuerdas vocales y poca energía en la voz emitida por último la tensión (S) se asocia con el esfuerzo vocal generado por aumento en la aducción glótica.

La evaluación GRBAS es normalmente llevada a cabo mediante una conversación continua, sin embargo esta evaluación también es efectuada mediante vocales sostenidas, sin embargo en estudios realizados se ha observado que el uso de vocales sostenidas desestima el nivel de disfonía [23]; además también se ha visto gran variabilidad entre cada una de las cinco características GRBAS. Se encontró que el parámetro más consistente es el grado de disfonía (G) y que las características A y S son las que presentan más variabilidad debido a que esos conceptos son más complejos para evaluar incluso por expertos [4].

Apéndice B

Parametrización usando coeficientes cepstrales en la escala de frecuencia Mel

Los coeficientes cepstrales sobre la escala Mel, han sido ampliamente utilizados en aplicaciones que incluyen el procesamiento de señales de voz debido principalmente a su capacidad para caracterizar la energía de la señal en bandas de frecuencia de acuerdo con la escala auditiva humana [24], además porque se ha demostrado que este esquema de parametrización no depende de las estimaciones de tono [4]. La Figura 1 muestra el diagrama de bloques para la técnica MFCC. Para obtener una representación paramétrica usando los MFCC, se deben seguir los siguientes pasos: [15] [25] [26].

Segmentación

En este paso la señal de voz $x[n]$ es dividida en bloques de N muestras, con tramas adyacentes de M muestras. Denotando el bloque l-ésimo como $x_l[n]$ y además considerando que la señal de voz se puede dividir en L bloques, se tiene [15] [25].

$$x_l[n] = x[Ml + n], \qquad n = 0, 1, 2, \ldots, N-1 \qquad l = 0, 1, 2, \ldots, L-1.$$

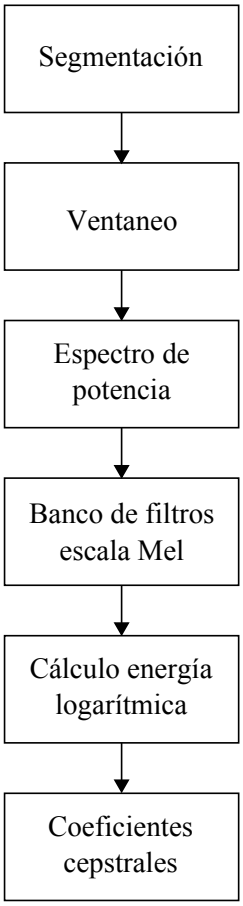

Figura 1: Diagrama de bloques para el proceso de caracterización por MFCC

Ventaneo

La segmentación realizada en el paso anterior, genera discontinuidades al inicio y fin de cada trama. Este paso busca eliminar dichas discontinuidades multiplicando cada bloque por una ventana adecuada con el fin de disminuir paulatinamente hasta cero los valores de inicio y fin de cada trama [15] [27]. La ventana se define como $w(n)$, $0 \leq n \leq N-1$, así la señal de voz ventaneada estará dada por

$$\hat{x}_l[n] = x_l[n]w[n].$$

Una de las ventanas más usadas es la ventana Hamming que tiene la forma [15]

$$w[n] = 0{,}54 - 0{,}46\cos\left(\frac{2\pi n}{N-1}\right).$$

Espectro de Potencia

El espectro de potencia se obtiene a partir de la transformada discreta de Fourier (DFT por sus siglas en inglés Discrete Fourier Transform) sobre cada una de las tramas ventaneadas. A partir del espectro obtenido con la DFT, se obtiene la potencia usando la siguiente expresión [15]

$$P(w) = \mathbb{R}[S(w)]^2 + \mathbb{I}[S(w)]^2,$$

donde $\mathbb{R}[S(w)]^2$ y $\mathbb{I}[S(w)]^2$ corresponden respectivamente a la parte real y la parte imaginaria del espectro de Fourier $S(w)$.

Banco de filtros escala Mel

Se define el banco de filtros mostrado en la Figura 2 el cual cuenta con M filtros ($m = 1, 2, 3, \ldots, M$), donde m corresponde a un filtro triangular dado por

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \dfrac{2(k - f[m-1])}{(f[m+1] - f[m-1])(f[m] - f[m-1])} & f[m-1] \leq k \leq f[m] \\ \dfrac{2(f[m+1] - k)}{(f[m+1] - f[m-1])(f[m+1] - f[m])} & f[m-1] \leq k \leq f[m] \\ 0 & \geq f[m+1]. \end{cases}$$

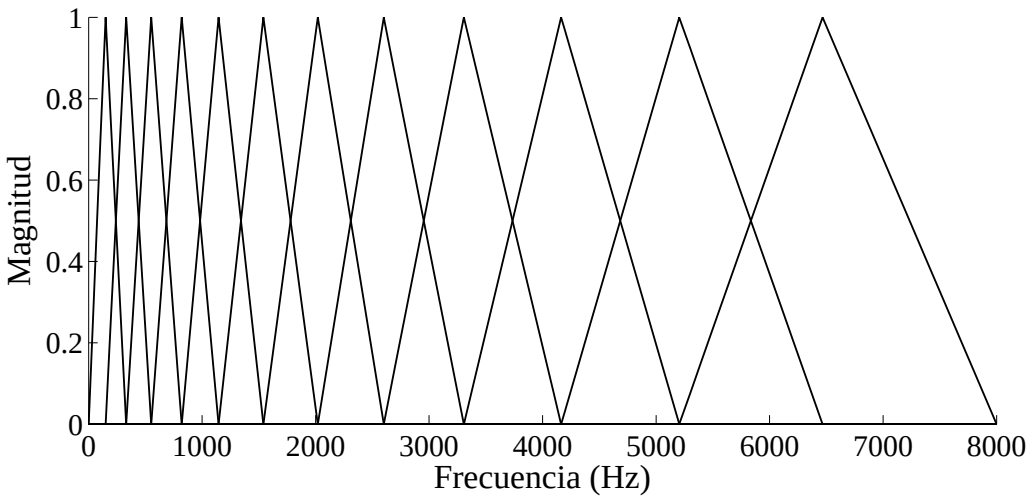

Figura 2: Banco de filtros en la escala de frecuencias Mel

Este banco de filtro calcula el espectro promedio alrededor de cada frecuencia central con anchos de banda que crecen progresivamente [15].

La escala de frecuencias Mel, es una escala logarítmica. Dada una frecuencia f expresada en Hertz (Hz), es posible calcular la frecuencia Mel $B(f)$ a partir de la siguiente expresión [25] [15]

$$B(f) = 1125 \ln \left(1 + \frac{f}{700} \right)$$

Cálculo de la energía logarítmica

Con el objetivo de simular la relación logarítmica existente entre la intensidad emitida y percibida, se calcula la energía logarítmica a la salida de cada filtro

$$Y[m] = \ln \left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k] \right], \qquad 0 < m \leq M$$

donde $X_a[k]$, corresponde a la DFT de la señal de voz original $x[n]$.

Coeficientes Cepstrales

Por definición los coeficientes cepstrales se calculan a partir de la TDF inversa de la energía logarítmica $Y[m]$, debido a que $Y[m]$ es una función impar, es posible usar la transformada inversa del coseno dada por [15] [3]

$$c_n = \sum_{m=0}^{M-1} Y[m] \cos\left(\frac{\pi n(m-0{,}5)}{M}\right), \qquad 0 \leq n \leq M$$

Se ha mostrado que la señal de voz contiene información acerca de la forma de onda de excitación [3]. Una de las finalidades de este estudio, es diseñar un sistema automático que permita valorar la calidad de voz, por dicha razón, es necesario usar un método de parametrización que tenga la capacidad de modelar los efectos que producen las patologías sobre el sistema de excitación del aparato fonador, específicamente sobre los pliegues o cuerdas vocales.

B.1. Características dinámicas

Es posible obtener una mejor representación del comportamiento dinámico de la señal de voz, introduciendo información acerca de las derivadas temporales de los parámetros a través de los segmentos vecinos y de la energía de cada segmento (la energía es calculada como la suma cuadrática de las amplitudes y dividida por el tamaño de la ventana). Para este estudio se usaron las primeras derivadas (Δ) y las segundas derivadas ($\Delta\Delta$) de los parámetros. Para introducir un orden temporal en la representación de los parámetros se define el m-ésimo coeficiente en el tiempo t como $c_n(t)$ [4], así la primera derivada de $c_n(t)$ se define por:

$$\frac{\partial c_n(t)}{\partial t} = \Delta c_n(t) \approx \mu \sum_{k=-K}^{K} k c_n(t+k), \tag{B.1}$$

donde μ es una constante de normalización y $(2k+1)$ es el número de segmentos sobre los que se realiza el cálculo.

Las primeras y segundas derivadas proveen información acerca de las variaciones temporales que tengan los parámetros MFCC. Estas características pueden ser consideradas importantes debido a que en voces patológicas, existe una estabilidad muy baja en la señal del habla, por lo que se esperara obtener grandes variaciones temporales de los parámetros al compararlos con voces normales [3], [4].

Ahora para cada segmento t, el resultado del análisis por MFCC, es un vector de L coeficientes cepstrales $c_n(t)$, L primeras derivadas de los coeficientes $\Delta c_n(t)$, L segundas derivadas de los coeficientes $\Delta\Delta c_n(t)$, la energía $E(t)$, la derivada de la energía $\Delta E(t)$ y la segunda derivada de la energía $\Delta\Delta E(t)$ [3]. Así, el vector de características está dado por:

$$\begin{aligned} x_i(t) =& (E(t), c_1(t), c_2(t), \ldots, c_L(t), \\ & \Delta E(t), \Delta c_1(t), \Delta c_2(t), \ldots, \Delta c_L(t), \\ & \Delta\Delta E(t), \Delta\Delta c_1(t), \Delta\Delta c_2(t), \ldots, \Delta\Delta c_L(t)), \end{aligned}$$

donde $x_i(t)$ es el vector de características con dimensionalidad $D = 3L + 3$.

Apéndice C

Derivadas para los Parámetros de los Modelos de clasificación para múltiples anotadores.

C.1. Regresión Logística Multiclase

C.1.1. Clasificación Binaria

Para maximizar la expresión (1.2), se debe igualar a cero el gradiente de esperanza condicional. La esperanza condicional para clasificación binaria puede ser escrita como

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{anterior}) = \sum_{i=1}^{N} \mu_i \ln(a_i) + \mu_i \ln(p_i) + (1-\mu_i) \ln(b_i) + (1-\mu_i) \ln(1-p_i). \quad \text{(C.1)}$$

Ahora el gradiente de (C.1) está dado por

$$\frac{\partial}{\partial \alpha^j} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{anterior}) = \frac{\partial}{\partial \alpha^j} \left[\sum_{i=1}^{N} \mu_i \ln(a_i) + \mu_i \ln(p_i) + (1-\mu_i) \ln(b_i) + (1-\mu_i) \ln(1-p_i) \right].$$

Efectuando el gradiente e igualando a cero se tiene

$$\begin{aligned}\frac{\partial}{\partial\alpha^j}Q(\boldsymbol{\theta},\boldsymbol{\theta}^{anterior}) &= \frac{\partial}{\partial\alpha^j}\left[\sum_{i=1}^{N}\mu_i\sum_{j=1}^{R}\left[y_i^j\ln(\alpha^j)+(1-y_i^j)\ln(1-\alpha^j)\right]\right]=0\\ &=\sum_{i=1}^{N}\mu_i\left[\frac{y_i^j}{\alpha^j}+\frac{(1-y_i^j)}{1-\alpha^j}(-1)\right]=0\\ &=\sum_{i=1}^{N}\mu_i\left[y_i^j(1-\alpha^j)+\alpha^j(y_i^j-1)\right]=0\\ &=\sum_{i=1}^{N}\mu_i\left[y_i^j-y_i^j\alpha^j+y_i^j\alpha^j-\alpha^j\right]=0\end{aligned}$$

Despejando α^j, se obtiene la expresión para la sensibilidad

$$\alpha^j=\frac{\sum_{i=1}^{N}\mu_i y_i^j}{\sum_{i=1}^{N}\mu_i}.$$

A partir de la expresión de la sensibilidad se obtiene la ecuación para la especificidad β^j

$$\beta^j=\frac{\sum_{i=1}^{N}(1-\mu_i)(1-y_i^j)}{\sum_{i=1}^{N}(1-\mu_i)}.$$

C.1.2. Clasificación Multiclase

Para maximizar la expresión (1.6), se debe igualar a cero el gradiente de esperanza condicional. La esperanza condicional para clasificación multiclase puede ser escrita como

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^{anterior})=\sum_{i=1}^{N}\sum_{c=1}^{K}[\mu_{ic}\ln(a_{ic})+\mu_{ic}\ln(p_{ic})],$$

donde

$$\ln(a_{ic})=\sum_{j=1}^{R}\sum_{k=1}^{K}Z_k^{i,j}\ln(\alpha_{ck}^j).$$

Recordando que para un c determinado, $\sum_{k=1}^{K} \alpha_{ck}^{j} = 1$, luego el problema es un problema de restricciones y se deben usar multiplicadores de Lagrange. Se debe definir así, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{anterior}) + \lambda(\sum_{k=1}^{K} \alpha_{ck}^{j} - 1)$. Por lo tanto se tiene

$$\frac{\partial}{\partial \alpha_{ck}^{j}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{anterior}) = \frac{\partial}{\partial \alpha_{ck}^{j}} \left[\sum_{\forall i} \sum_{\forall c} \mu_{ic} \sum_{\forall j} \sum_{\forall k} Z_{k}^{i,j} \ln(\alpha_{ck}^{j}) \right] + \lambda$$

$$= \sum_{\forall i} \frac{\mu_{ic} Z_{k}^{i,j}}{\alpha_{ck}^{j}} + \lambda = 0 \tag{C.2}$$

Despejando λ de la expresión (C.2), se tiene

$$\lambda = -\frac{1}{\alpha_{ck}^{j}} \sum_{\forall i} \mu_{ic} Z_{k}^{i,j}.$$

De igual forma se tiene

$$\lambda \alpha_{ck}^{j} = -\sum_{\forall i} \mu_{ic} Z_{k}^{i,j}.$$

Sumando a ambos lados $\forall k$

$$\sum_{\forall k} \lambda \alpha_{ck}^{j} = -\sum_{\forall i} \mu_{ic} \sum_{\forall k} Z_{k}^{i,j}.$$

$$\lambda = -\sum_{\forall i} \mu_{ic} \tag{C.3}$$

Reemplazando el valor de λ obtenido en (C.3) en la ecuación (C.2) y despejando α_{ck}^{j} se obtiene

$$\alpha_{ck}^{j} = \frac{\sum_{\forall i} \mu_{ic} Z_{k}^{i,j}}{\sum_{\forall i} \mu_{ic}}$$

C.2. Procesos Gaussianos

Se define el vector de parámetros como $\boldsymbol{\theta}$. El vector de parámetros, incluye los parámetros asociados a la función del Kernel $k(\mathbf{x}, \mathbf{x}')$, los cuales se denotan como $\boldsymbol{\phi}$, y las varianzas asociadas a cada anotador $\{\sigma_j^2\}_{j=1}^R$. Así $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \{\sigma_j^2\}_{j=1}^R\}$. Ahora se hallan las derivadas de $\hat{\sigma}_i^2$ y $\hat{y}_i$ con respecto a σ_j. Se tiene que.

$$\hat{\sigma}_i^2 = \left(\sum_{j \sim i} \sigma_j^{-2} \right)^{-1}.$$

La derivada de $\hat{\sigma}_i^2$ con respecto a σ_j sigue la forma

$$\frac{d\hat{\sigma}_i^2}{d\sigma_j} = -\left(\sum_{j \sim i} \sigma_j^{-2} \right)^{-2} \frac{d\sigma_j^{-2}}{d\sigma_j} = 2\left(\sum_{j \sim i} \sigma_j^{-2} \right)^{-2} \sigma_j^{-3}. \tag{C.4}$$

El término $\hat{y}_i$ está dado como.

$$\hat{y}_i = \hat{\sigma}_i^2 \sum_{j \sim i} y_i^j \sigma_j^{-2}.$$

Luego la derivada de $\hat{y}_i$ con respecto a σ_j se escribe como

$$\frac{d\hat{y}_i}{d\sigma_j} = \frac{d\hat{\sigma}_i^2}{d\sigma_j} \left(\sum_{j \sim i} y_i^j \sigma_j^{-2} \right) - 2\hat{\sigma}_i^2 y_i^j \sigma_j^{-3}. \tag{C.5}$$

A partir de la ecuación (1.15), se necesita obtener la derivada del logaritmo negativo de la evidencia con respecto a σ_j. Esta derivada se hace término a término que aparece en la ecuación. Para el término A, se tiene

$$\begin{aligned}
\frac{1}{2} \frac{d}{d\sigma_j} \log |\mathbf{K} + \hat{\boldsymbol{\Sigma}}| &= \frac{1}{2} \operatorname{trace} \left\{ \left[\frac{d}{d\hat{\boldsymbol{\Sigma}}} \log |\mathbf{K} + \hat{\boldsymbol{\Sigma}}| \right]^\top \frac{d\hat{\boldsymbol{\Sigma}}}{d\sigma_j} \right\} \\
&= \frac{1}{2} \operatorname{trace} \left\{ \left[\left(\mathbf{K} + \hat{\boldsymbol{\Sigma}} \right)^\top \right]^{-1} \frac{d\hat{\boldsymbol{\Sigma}}}{d\sigma_j} \right\} \\
&= \frac{1}{2} \operatorname{trace} \left\{ \left(\mathbf{K} + \hat{\boldsymbol{\Sigma}} \right)^{-1} \frac{d\hat{\boldsymbol{\Sigma}}}{d\sigma_j} \right\},
\end{aligned}$$

donde se usó [28]

$$\frac{d}{d\mathbf{X}} \log |\mathbf{X}| = \left(\mathbf{X}^\top\right)^{-1}.$$

Además, $\mathbf{X}^\top = \mathbf{X}$ es la matriz simétrica. Nótese que $\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}$ es una matriz diagonal con elementos dados por la ecuación (C.4).
La derivada del término B en la ecuación, (1.15) puede ser escrita como

$$\begin{aligned}
\frac{1}{2}\frac{d}{d\sigma_j}\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}} &= \frac{1}{2}\left\{\text{trace}\left[\left(\frac{d}{d\hat{\mathbf{y}}}\left(\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\right)\right)^\top \frac{d\hat{\mathbf{y}}}{d\sigma_j}\right]\right\} \\
&\quad + \frac{1}{2}\left\{\text{trace}\left[\left(\frac{d}{d\hat{\mathbf{\Sigma}}}\left(\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\right)\right)^\top \frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right]\right\} \\
&\quad + \frac{1}{2}\left\{\text{trace}\left[\left(\frac{d}{d\hat{\mathbf{y}}}\left(\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\right)\right)^\top \frac{d\hat{\mathbf{y}}}{d\sigma_j}\right]\right\} \\
\frac{1}{2}\frac{d}{d\sigma_j}\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}} &= \text{trace}\left[\left((\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\right)^\top \frac{d\hat{\mathbf{y}}}{d\sigma_j}\right] \\
&\quad - \frac{1}{2}\left\{\text{trace}\left[\left((\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\hat{\mathbf{y}}^\top(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\right)\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right]\right\},
\end{aligned}$$

donde la entradas de $\frac{d\hat{\mathbf{y}}}{d\sigma_j}$ están dadas por la ecuación(C.5).
La derivada para el término C se da por

$$\frac{1}{2}\frac{d}{d\sigma_j}\log|\hat{\mathbf{\Sigma}}| = \frac{1}{2}\,\text{trace}\left[\hat{\mathbf{\Sigma}}^{-1}\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right].$$

La derivada del término D está dada por

$$\frac{1}{2}\frac{d}{d\sigma_j}\sum_i\sum_{j\sim i}(y_i^j)^2(\sigma_j)^{-2} = -\sum_i(y_i^j)^2(\sigma_j)^{-3}.$$

La derivada del término E es

$$\frac{1}{2}\frac{d}{d\sigma_j}\sum_i \hat{y}_i\hat{\sigma}_i^{-2} = \frac{1}{2}\sum_i\left(\frac{d\hat{y}_i}{d\sigma_j}\hat{\sigma}_i^{-2} - \hat{y}_i(\hat{\sigma}_i^2)^{-2}\frac{d\hat{\sigma}_i^2}{d\sigma_j}\right).$$

Finalmente, el término F tiene la siguiente derivada

$$\frac{d}{d\sigma_j}\sum_j\sum_{i\sim j}\log\frac{1}{\sigma_j} = -\frac{N_j}{\sigma_j}.$$

colocando todos los términos juntos, se obtiene

$$
\begin{aligned}
-\frac{d}{d\sigma_j}\log p(\mathbf{y}) = \frac{1}{2}\,\mathrm{trace}\left\{\left(\mathbf{K}+\hat{\mathbf{\Sigma}}\right)^{-1}\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right\} + \left((\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\right)^{\top}\frac{d\hat{\mathbf{y}}}{d\sigma_j} \\
-\frac{1}{2}\left\{\mathrm{trace}\left[\left((\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\hat{\mathbf{y}}\hat{\mathbf{y}}^{\top}(\mathbf{K}+\hat{\mathbf{\Sigma}})^{-1}\right)\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right]\right\} - \frac{1}{2}\,\mathrm{trace}\left[\hat{\mathbf{\Sigma}}^{-1}\frac{d\hat{\mathbf{\Sigma}}}{d\sigma_j}\right] \\
-\sum_i (y_i^j)^2(\sigma_j)^{-3} - \frac{1}{2}\sum_i\left(\frac{d\hat{y}_i}{d\sigma_j}\hat{\sigma}_i^{-2} - 2\hat{y}_i\hat{\sigma}_i^{-3}\frac{d\hat{\sigma}_i}{d\sigma_j}\right) + \frac{N_j}{\sigma_j}.
\end{aligned}
$$

Apéndice D

Curvas ROC

Una curva de características de funcionamiento del receptor o curva ROC (Receiver Operating Characteristics) por su nombre en inglés, es una técnica utilizada para la visualización, organización y selección de sistemas de clasificación basándose en su rendimiento. El análisis de las curvas ROC ha sido principalmente usado en el área de la medicina con el fin de analizar el comportamiento de sistemas diseñados para el diagnóstico de enfermedades [29]. Recientemente, el análisis de las curvas ROC han incrementado su uso en el área del aprendizaje de máquina, debido principalmente a que a partir de estos análisis se pueden obtener métricas que permiten la evaluación del rendimiento de un modelo de clasificación [30].

D.1. Rendimiento del Clasificador

Se considera un problema de clasificación binaria, en el cual cada una de los ejemplos I, es asignado a un elemento del conjunto $\{\mathbf{p}, \mathbf{n}\}$ de clases positivas y negativas. Un clasificador se considera como una asignación de los ejemplos a clases predichas $\{\mathbf{Y}, \mathbf{N}\}$. Ahora dado un clasificador y un ejemplo, existen 4 posibles salidas [31].

1. Si el ejemplo pertenecía a la clase positiva ($\mathbf{p}$) y se clasifica como positiva ($\mathbf{Y}$), se cuenta como *verdadero positivo*.

2. Si el ejemplo pertenecía a la clase positiva ($\mathbf{p}$) y se clasifica como negativa ($\mathbf{N}$), se cuenta como *falso negativo*.

3. Si el ejemplo pertenecía a la clase negativa (**n**) y se clasifica como negativa (**N**), se cuenta como *verdadero negativo*.

4. Si el ejemplo pertenecía a la clase negativa (**n**) y se clasifica como positiva (**Y**), se cuenta como *falso positivo*.

Estas cuatro salidas pueden ser explicadas de mejor forma a través de la matriz de confusión mostrada en la Figura 3.

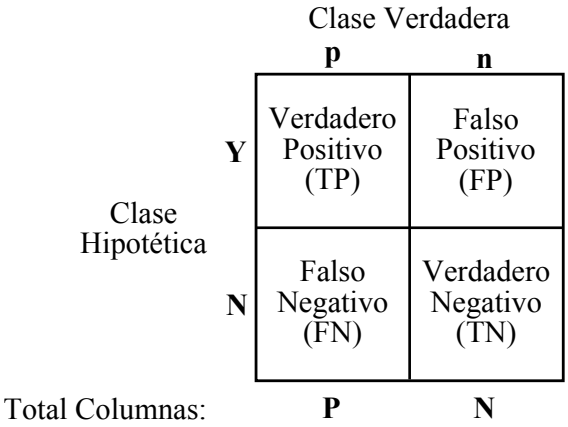

Figura 3: Matriz de Confusión. Explica el comportamiento de un sistema de clasificación a partir de un conjunto de muestras (el conjunto de prueba).

Ahora, a partir la matriz mostrada en la Figura 3, es posible obtener una serie de métricas útiles para estimar el rendimiento del clasificador, entre dichas métricas se encuentran:

La proporción de verdaderos positivos (también conocida como *recall* o *sensibilidad*) de un clasificador, se calcula a partir:

$$\text{Sensibilidad} = \frac{TP}{P}$$

La Proporción de falsos positivos, FA, (conocida también como la proporción de falsa alarma) de un clasificador es:

$$FA = \frac{FP}{N}$$

Otros términos relacionados con las curvas ROC son:

$$\text{especificidad} = \frac{TN}{FP + TN}$$

$$\text{exactitud} = \frac{TP}{TP + FP}$$

$$\text{precisión} = \frac{TP + TN}{P + N}$$

D.2. Área Bajo la Curva ROC (AUC, Area Under Curve).

Las curvas ROC, es una descripción bidimensional del rendimiento de un clasificador. Ahora para comparar los clasificadores, se quisiera reducir el rendimiento ROC, a un escalar que represente el rendimiento de un determinado clasificador. Un método común es el del área bajo la curva, este tiene una propiedad estadística muy importante, puesto que, el AUC de un clasificador es equivalente a la probabilidad que un clasificador asigne un mayor rango a un ejemplo positivo escogido de manera aleatoria que a un ejemplo negativo elegido aleatoriamente [30].

D.2.1. AUC Multiclase.

El AUC es una métrica calculada sobre dos clases. Para un problema de dos clases el AUC es un simple valor, pero para múltiples clases aparece el problema de combinar los múltiples valores de AUC provenientes de cada par de clases. Estos problemas, se explican de manera más detallada en [32].

Un enfoque para calcular el AUC en el caso de múltiples clases es el usado en [33]. Este enfoque calcula los AUCs generando la curva ROC para clase de referencia y a su vez midiendo el área bajo la curva. Luego sumando las AUC ponderadas por su prevalencia en los datos.

$$\text{AUC}_{\text{total}} = \sum_{c_i \in C} \text{AUC}(c_i) p(c_i).$$

Por otro lado, [32] toma un diferente enfoque para el cálculo del AUC para múltiples clases. Este enfoque se basa en el hecho de que el AUC es equivalente a la probabilidad que un clasificador asigne un mayor rango a un ejemplo positivo escogido de manera aleatoria que a un ejemplo negativo elegido aleatoriamente. La medida propuesta por este enfoque es equivalente a.

$$\mathrm{AUC}_{\mathrm{total}} = \frac{2}{|C|(|C|-1)} \sum_{\{c_i,c_j\}\in C} \mathrm{AUC}(c_i, c_j),$$

donde C es el número de clases, y $\mathrm{AUC}(c_i, c_j)$ es el área bajo la curva ROC que involucra a las clases c_i y c_j