



# XLVIII Coloquio Argentino de Estadística

VI JORNADA DE EDUCACIÓN ESTADÍSTICA "MARTHA DE ALIAGA"

27 al 30 oct 2020

Poster:

## ***Comparación del clasificador RDA con LDA y QDA en imágenes sintéticas***

*Agustina González, Gabriela Palacio, Sabina Bigolin, Susana Ferrero*



Esta obra está bajo una  
Licencia Creative Commons  
Atribución-NoComercial 4.0  
Internacional

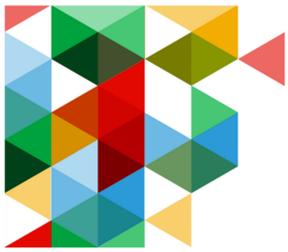


FACULTAD  
DE CIENCIAS  
ECONÓMICAS



Universidad  
Nacional  
de Córdoba





# Comparación del clasificador RDA con LDA y QDA en imágenes sintéticas



González Agustina, Palacio Gabriela, Bigolin Sabina, Ferrero Susana  
 Universidad Nacional de Río Cuarto-*mariagustina2419@gmail.com*

## Objetivos

El objetivo de este trabajo es analizar el desempeño del clasificador Análisis Discriminante Regularizado (RDA) comparándolo con Análisis Discriminante Lineal (LDA) y Cuadrático (QDA). Para ello, mediante simulación, se generan imágenes sintéticas, donde se representa una escena formada con diferentes grupos o clases con distribución normal multivariada y se consideran cuatro casos modificando los parámetros de la distribución.

## RDA

Friedman (1989) propone una nueva estimación para  $\Sigma$  incorporando dos parámetros,  $\lambda$  y  $\gamma$ :

$$\hat{\Sigma}_k(\lambda, \gamma) = (1 - \gamma)\hat{\Sigma}_k(\lambda) + \gamma \frac{\text{tr}[\hat{\Sigma}_k(\lambda)]}{n} I$$

donde  $\hat{\Sigma}_k(\lambda) = (1 - \lambda)\hat{\Sigma}_k + \lambda\hat{\Sigma}_{pool}$ ,  $\lambda, \gamma \in (0, 1)$  y  $n$  es la cantidad de clases. Para cada par  $(\lambda, \gamma)$  se obtiene una regla de clasificación diferente.

\*Si  $(\lambda = 1, \gamma = 0) \rightarrow \hat{\Sigma}_k(0, 0) = \hat{\Sigma}_{pool}$  (LDA)

\*Si  $(\lambda = 0, \gamma = 0) \rightarrow \hat{\Sigma}_k(0, 0) = \hat{\Sigma}_k$  (QDA)

## Optimización de los parámetros

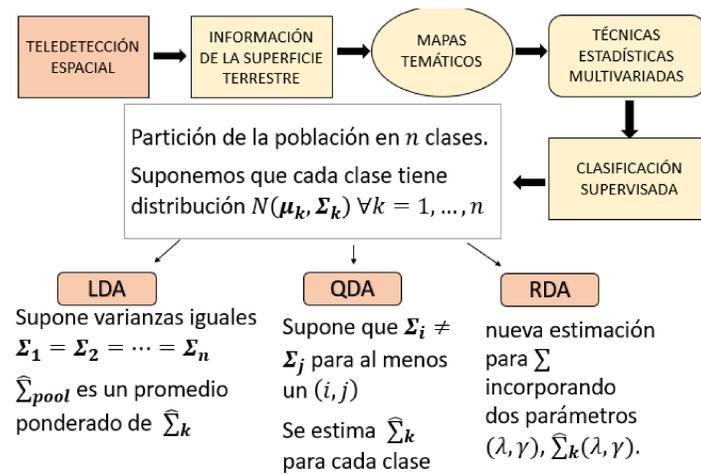
Para la optimización de los parámetros se consideran tres casos:

\*  $\lambda = 0,2$  (RDA2)

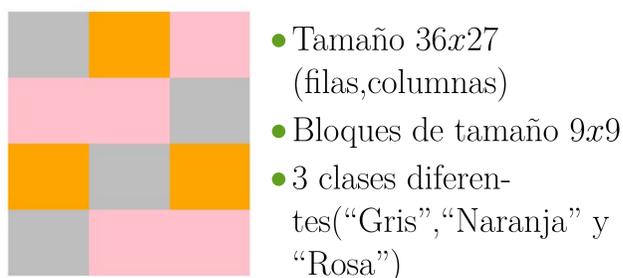
\*  $\lambda = 0,5$  (RDA5)

\*  $\lambda = 0,8$  (RDA8)

Para cada caso se obtiene  $\gamma$  aplicando *K-Fold Cross-Validation* o Validación Cruzada de  $K$ -iteraciones, considerando  $K=10$ .



## Imagen Sintética



## Metodología

Se simulan imágenes en donde cada pixel es un vector aleatorio con distribución  $N(\mu_k, \Sigma_k)$   $k = 1, \dots, 3$  y se analizan las siguientes situaciones:

\***Primer Caso**: Todas las clases tienen igual  $\Sigma$  y  $\mu$  similares.

\***Segundo Caso**: Todas las clases tienen igual  $\Sigma$ , dos clases con  $\mu$  cercanos y una con  $\mu$  distante.

\***Tercer Caso**: Todas las clases tienen distintos  $\Sigma$  y  $\mu$  similares.

\***Cuarto Caso**: Todas las clases tienen distintos  $\Sigma$ , dos clases con  $\mu$  cercanos y una con  $\mu$  distante.

## Resultados

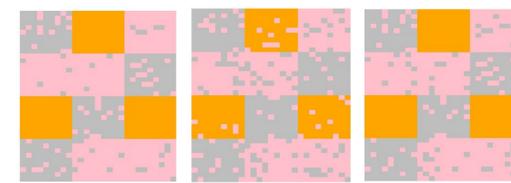


Figura 1: Mapas temáticos para el segundo caso.



Figura 2: Mapas temáticos para el tercer caso.

	Primer	Segundo	Tercer	Cuarto
LDA	96.7	91.8	94	99.1
RDA2	97.3	87.6	95.5	96.8
RDA5	98.1	90	94.3	98.2
RDA8	97.9	90.5	96.1	98.4
QDA	89.1	86.1	91	96.3

Cuadro 1: Coeficientes  $F$  (%) de los clasificadores para cada caso

\* Para el primer y tercer caso los RDA obtuvieron  $F$  y  $\kappa$  mayores que los de QDA y LDA, siendo para cada caso el mayor RDA5 y RDA8 respectivamente.

\* Para el segundo y cuarto caso el clasificador con mayor  $F$  y  $\kappa$  es LDA.

\* En todos los casos el clasificador con menores  $F$  y  $\kappa$  fue QDA.

\* En los mapas temáticos que se observan en la Figura 1 correspondientes al segundo caso se puede observar que LDA se confunde menos que RDA y QDA es el de peor desempeño.

\* En la Figura 2, correspondiente al tercer caso, el mapa temático producido por RDA8 tiene menos errores que el producido por LDA, y el que más se equivoca es QDA.

## Conclusiones

- Cuando los vectores de medias son cercanos RDA se desempeña mejor, tanto para  $\Sigma$  iguales o distintas.
- Cuando al menos una de las clases tiene su vector de medias distante el clasificador con mejor performance es LDA, aún con  $\Sigma$  distintas.
- Desde el punto de vista teórico QDA debería tener mejor desempeño en los casos donde las clases tienen  $\Sigma$  diferentes. Sin embargo, en nuestros resultados eso no se cumple.

## Referencias

- \* J. Friedman. Regularized Discriminant Analysis. Journal of the American Statistical Association, March 1989.
- \* James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R.
- \* Equipo RStudio. RStudio: Desarrollo integrado para R. RStudio, PBC, Boston, 2020.