

# Análisis de la caracterización de imágenes de lesiones melanocíticas mediante la regla ABCD usado técnicas de selección de atributos




## **PINV18-1199: Diagnóstico automático de lesiones melanocíticas asistido por computadora**

14/Octubre/2021

Miguel García Torres

Data Science & Big Data Research Lab 

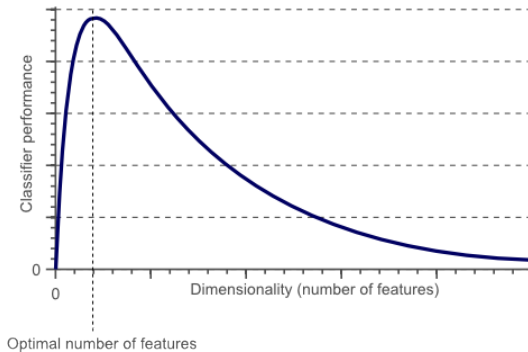
Lenguajes y Sistemas Informáticos, Universidad Pablo de Olavide 

# Outline

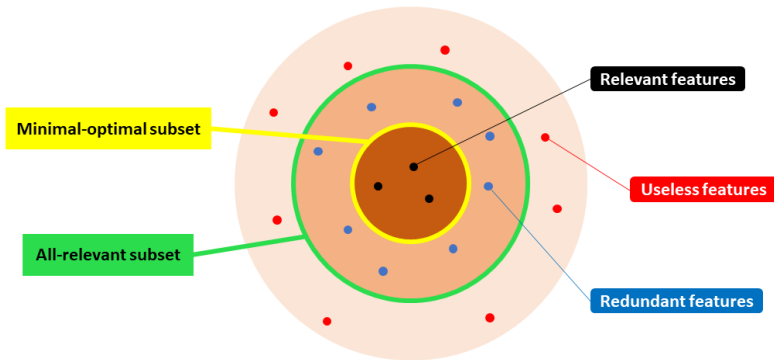
- 1 Feature selection
  - Introduction
  - Feature grouping based Scatter Search
- 2 Melanocytic tumors
- 3 Results
  - Biological datasets
  - Melanoma dataset

# Curse of dimensionality

## Motivation

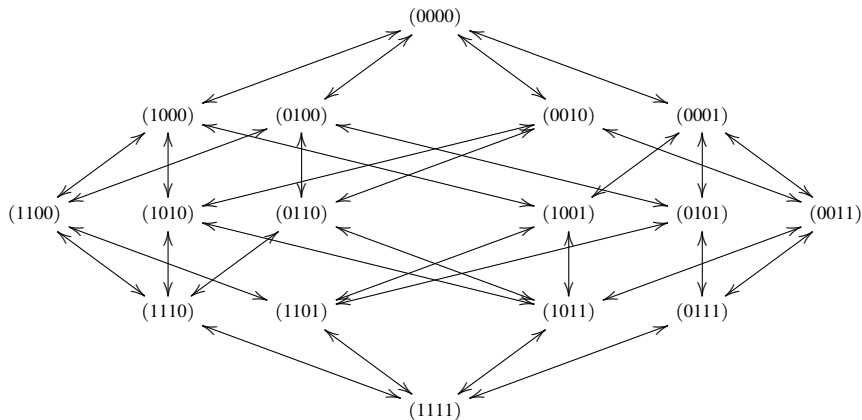


# Feature types



# Feature subset generation

Search space

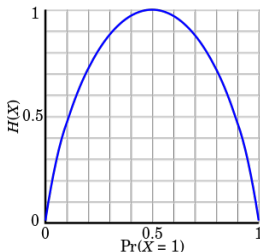


# Entropy

## Entropy

It measures the uncertainty about the value of a random variable  $X$ .

$$H(X) = - \sum_i P(x_i) \log_2(P(x_i)).$$



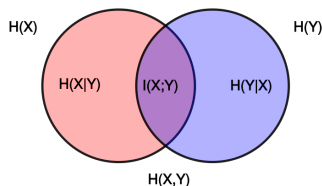
- Feature  $X$  with values  $\{0, 1\}$ .
- Entropy is 0 if there is no uncertainty.

# Entropy

## Information Gain

It measures the reduction in uncertainty about the value of  $X$  given the value of  $Y$

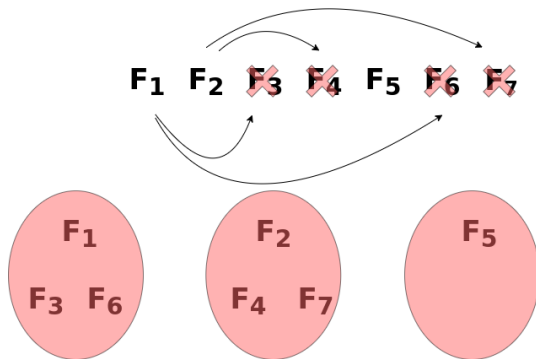
$$IG(X; Y) = H(X) - H(X|Y).$$



- $H(X) \equiv$  circle on the left (red and violet).
- $H(Y) \equiv$  circle on the right (blue and violet).
- $H(X, Y) \equiv$  area contained by both circles.
- $H(X|Y) \equiv$  red.
- $H(Y|X) \equiv$  blue.
- $I(X; Y) \equiv$  violet.

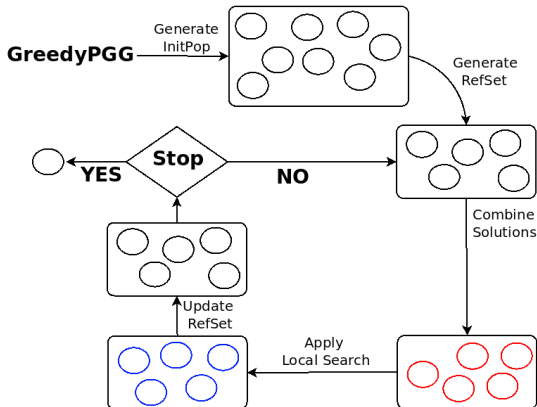
# Feature grouping heuristic

Greedy Predominant Group Generator (GreedyPGG)

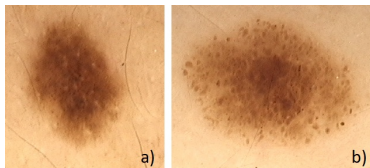
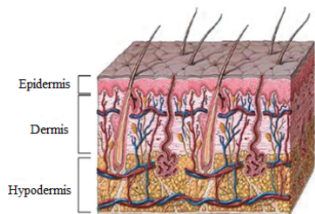




# Scatter Search workflow



# Melanocytic tumors



Dermatoscopic images of (a) a nevus, and (b) a melanoma.

# Image preprocessing

From image to data

- Step 1: Preprocessing (Hair removal).
- Step 2: Segmentation.
- Step 3: Feature extraction.

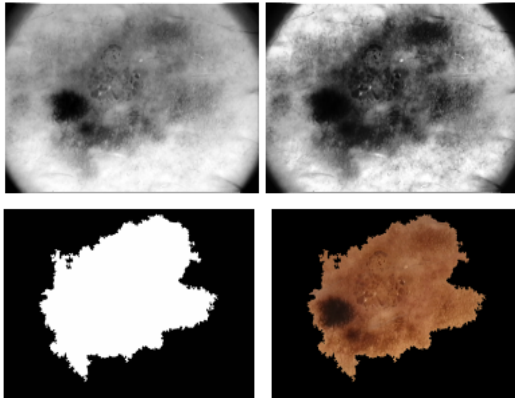
# Image preprocessing

## Step 1: Preprocessing (Hair removal)



# Image preprocessing

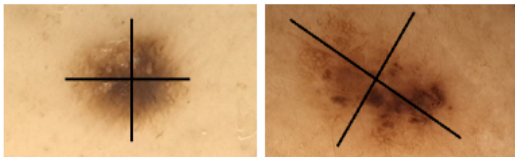
## Step 2: Segmentation



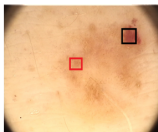
# Image preprocessing

## Step 3: Feature extraction

**Asymmetry** is the pattern generated by the uncontrolled growth of the lesion.



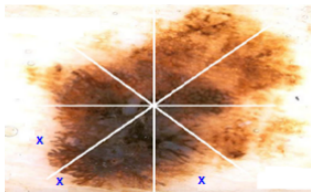
**Color** is related to the excess of melanin under the surface of the lesion.



# Image preprocessing

## Step 3: Feature extraction

**Borders** are patterns associated with abnormal terminations of color that melanocytic lesions have. The area of interest is divided into 8 segments.



## Biological datasets

Dataset	Id	#Inst.	#Feat.	Labels	#Inst./label
colon	cln	62	2000	normal/tumor	22 / 40
lymphoma	lym	77	2647	diffuse/follicular	58 / 19
breast	bcg	168	2905	good/poor	111 / 57
lung	lng	181	12533	MPM/ADCA	31 / 150
breast	brc	118	22215	positive/negative	75 / 43
breast/colon	bco	104	22283	breast/colon	62 / 42
crohn	cro	127	22283	normal/colitis/crohn	42 / 26 / 59

Id	Naive Bayes			#Features	
	Baseline	SS	PGSS	SS	PGSS
cln	54.87 ± 25.00	82.56 ± 16.50	84.10 ± 13.15	25.80 ± 5.90	13.20 ± 3.90
lym	81.83 ± 10.39	92.25 ± 6.80	90.83 ± 5.83	49.80 ± 3.56	32.20 ± 5.17
bcg	70.21 ± 7.61	74.92 ± 11.79	71.96 ± 8.00	75.00 ± 6.04	32.80 ± 2.17
lng	97.78 ± 3.62	96.67 ± 3.04	95.02 ± 2.34	19.00 ± 11.58	6.2 ± 2.40
brc	85.65 ± 8.32	82.25 ± 10.93	82.25 ± 6.85	104.60 ± 7.10	71.60 ± 11.01
bco	69.24 ± 7.91	94.19 ± 5.31	93.24 ± 4.34	18.20 ± 15.16	7.80 ± 2.17
cro	74.06 ± 8.06	87.45 ± 3.06	85.11 ± 6.30	137.40 ± 12.28	69.80 ± 6.30
avg	76.23	87.18	86.07	61.40	33.37
pval	0.156	0.142		0.017	



# Melanoma dataset

Criterion	Feature	Id	Description
asymmetry	index of asymmetry	<i>as</i>	#pixels into irregular disjoint areas
borders	segment 1-8	<i>b1</i> – <i>8</i>	variation of colors from center pixel to border pixels
colors	white light brown dark brown black	<i>wh</i> <i>lb</i> <i>db</i> <i>bk</i>	#pixels with tis color
dermatoscopic structures	linear branches irregular pigment network structureless areas dots and globules	<i>lr</i> <i>ip</i> <i>ne</i> <i>sa</i> <i>dg</i>	variation of distance from center to border number of unconnected pixels micro-regions number of pixel into the area number of dots and globules

Performance measure	Baseline	SS	PGSS
Accuracy	$78.86 \pm 7.92$	$82.81 \pm 7.88$	$81.81 \pm 6.05$
$ S $	-	$4.60 \pm 0.55$	$4.20 \pm 0.84$

