

# PRACTICA 2

REDES BAYESIANAS

Ángel Igualada Moraga

2018-2019

## EJERCICIO 2.4 A)

En este ejercicio, hemos modificado el número máximo de iteraciones y hemos aumentado el número de muestras a 1000.

### Resultados

Observamos en primer lugar, que no solo nunca se supera el número máximo de iteraciones (1000) sino que no se llega nunca a la iteración 26, por lo que el número máximo de iteraciones no influye en gran medida.

Por el contrario, aumentar el número de muestras, en el caso de datos completos, nos ha hecho ganar mucha precisión a la hora de estimar las variables, y en el caso de los datos incompletos, hemos mejorado en algunos casos, y empeorado en otros.

Dado el gran porcentaje de datos incompletos, los resultados son parcialmente positivos porque en algunos aspectos si reflejan la forma del modelo.

En el caso incompleto, la variable C siempre es falsa y ese es el mayor inconveniente.

A continuación, se adjuntan los resultados obtenidos en los que se puede advertir las diferencias con las probabilidades reales.

PROBABILIDADES MODELO

S	R	$P(W \mid S, R), W =$	
		false	true
false	false	1.0	0.0
true	false	0.1	0.9
false	true	0.1	0.9
true	true	0.01	0.99

C	$P(C)$
false	0.5
true	0.5

C	$P(S \mid C), S =$	
	false	true
false	0.5	0.5
true	0.9	0.1

C	$P(R \mid C), R =$	
	false	true
false	0.8	0.2
true	0.2	0.8

PROBALIDADES ESTIMADAS CON DATOS COMPLETOS  
numMuestras = 100

PROBALIDADES ESTIMADAS PARA EL NODO W

1 1 : 1.0000 0.0000  
2 1 : 0.0556 0.9444  
1 2 : 0.0435 0.9565  
2 2 : 0.0000 1.0000

PROBALIDADES ESTIMADAS PARA EL NODO C

1 : 0.4700  
2 : 0.5300

PROBALIDADES ESTIMADAS PARA EL NODO S

1 : 0.5532 0.4468  
2 : 0.9057 0.0943

PROBALIDADES ESTIMADAS PARA EL NODO R

1 : 0.7234 0.2766  
2 : 0.2264 0.7736

PROBALIDADES ESTIMADAS CON DATOS COMPLETOS  
numMuestras = 1000

PROBALIDADES ESTIMADAS PARA EL NODO W

1 1 : 1.0000 0.0000  
2 1 : 0.1085 0.8915  
1 2 : 0.0905 0.9095  
2 2 : 0.0250 0.9750

PROBALIDADES ESTIMADAS PARA EL NODO C

1 : 0.4860  
2 : 0.5140

PROBALIDADES ESTIMADAS PARA EL NODO S

1 : 0.4979 0.5021  
2 : 0.9066 0.0934

PROBALIDADES ESTIMADAS PARA EL NODO R

1 : 0.8066 0.1934  
2 : 0.1673 0.8327

PROBABILIDADES MODELO

S	R	$P(W \mid S, R), W =$	
		false	true
false	false	1.0	0.0
true	false	0.1	0.9
false	true	0.1	0.9
true	true	0.01	0.99

C	$P(C)$
false	0.5
true	0.5

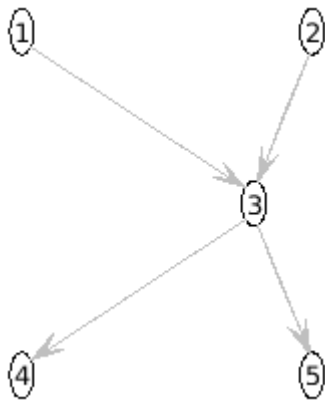
C	$P(S \mid C), S =$	
	false	true
false	0.5	0.5
true	0.9	0.1

C	$P(R \mid C), R =$	
	false	true
false	0.8	0.2
true	0.2	0.8

PROBALIDADES ESTIMADAS CON DATOS INCOMPLETOS numMuestras = 100 Máximo numero de iteraciones = 100	PROBALIDADES ESTIMADAS CON DATOS INCOMPLETOS numMuestras = 1000 Máximo numero de iteraciones = 1000
PROBALIDADES ESTIMADAS PARA EL NODO W 1 1 : 0.9996 0.0004 2 1 : 0.0340 0.9660 1 2 : 0.0071 0.9929 2 2 : 0.1329 0.8671	PROBALIDADES ESTIMADAS PARA EL NODO W 1 1 : 0.9988 0.0012 2 1 : 0.2952 0.7048 1 2 : 0.1496 0.8504 2 2 : 0.0601 0.9399
PROBALIDADES ESTIMADAS PARA EL NODO C 1 : 1.0000 2 : 0.0000	PROBALIDADES ESTIMADAS PARA EL NODO C 1 : 0.9996 2 : 0.0004
PROBALIDADES ESTIMADAS PARA EL NODO S 1 : 0.6676 0.3324 2 : 0.9839 0.0161	PROBALIDADES ESTIMADAS PARA EL NODO S 1 : 0.6969 0.3031 2 : 1.0000 0.0000
PROBALIDADES ESTIMADAS PARA EL NODO R 1 : 0.4096 0.5904 2 : 0.0000 1.0000	PROBALIDADES ESTIMADAS PARA EL NODO R 1 : 0.4247 0.5753 2 : 0.0000 1.0000

## EJERCICIO 2.4 B)

En este ejercicio, se ha realizado un script que resuelve los tres apartados, En el se ha creado la red bayesiana, y usando dos motores se han inferido las dos probabilidades pedidas. Este ejercicio, no se ha preparado para poder ser lanzado directamente desde bash sino desde la interfaz Matlab.



### Resultados

Resultados ejercicio B.2

PROBABILIDAD DE QUE NO TENGA CANCER DE PULMON 0.998395

Resultados ejercicio B.3

EXPLICACION MAS PROBABLE PARA QUE UN PACIENTE TENGA CANCER:

POLUCION: false

FUMADOR: true

CANCER: true

RAYOS X: true

DISNEA: true

**La explicación más probable es que sea fumador, los rayos x hayan dado positivo y sufra disnea.**

## EJERCICIO 3.2.1 1)

Para resolver este ejercicio y el siguiente, se han creado dos scripts, uno en bash y otro en Matlab.

Se puede ejecutar el .sh como sigue:

USO: analisisMixturaGaussianas.sh [1] [2] [3] [4] [5]

[1] = datos\_entrenamiento

[2] = etiquetas\_entrenamiento

[3] = datos\_test

[4] = etiquetas\_test

[5] = nom\_funcion\_matlab

Ejemplo ejecucion:

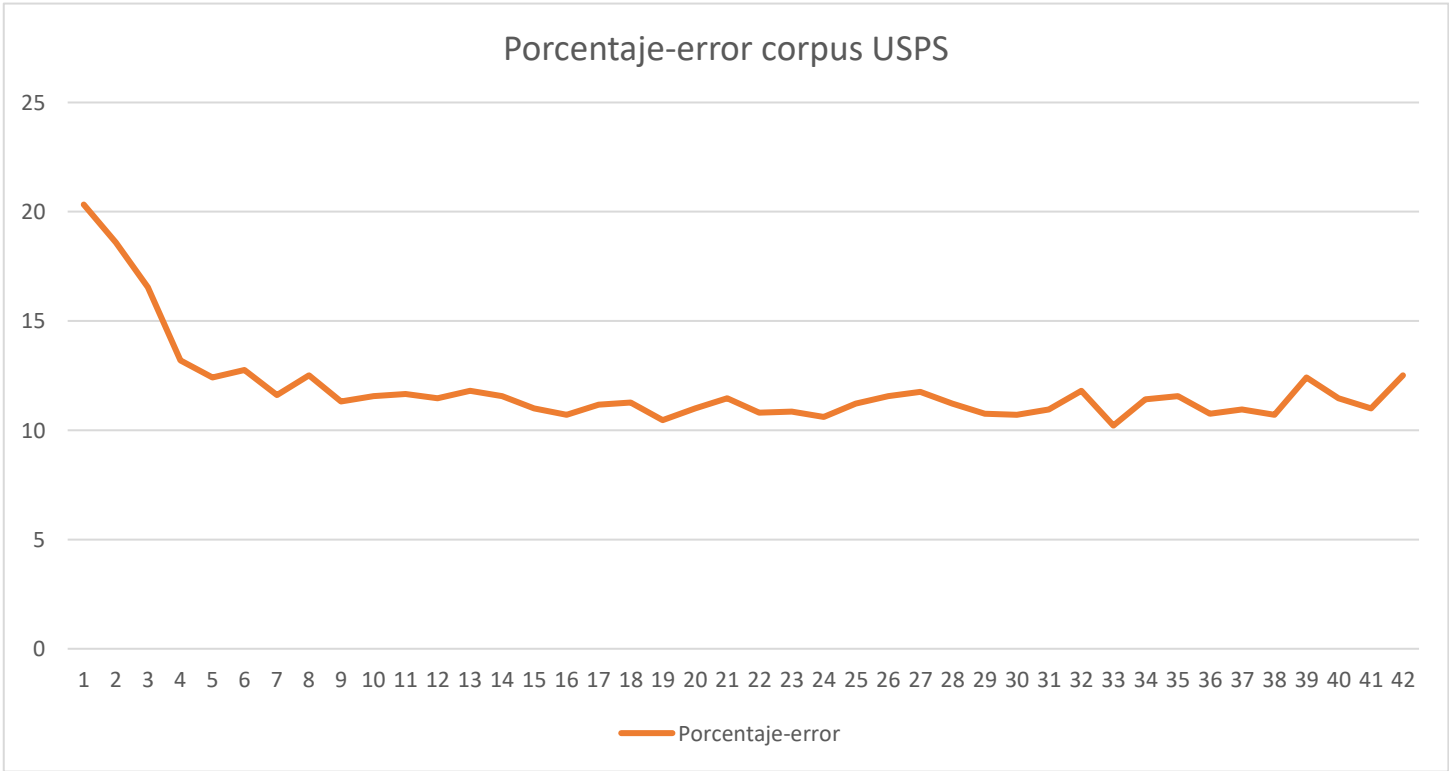
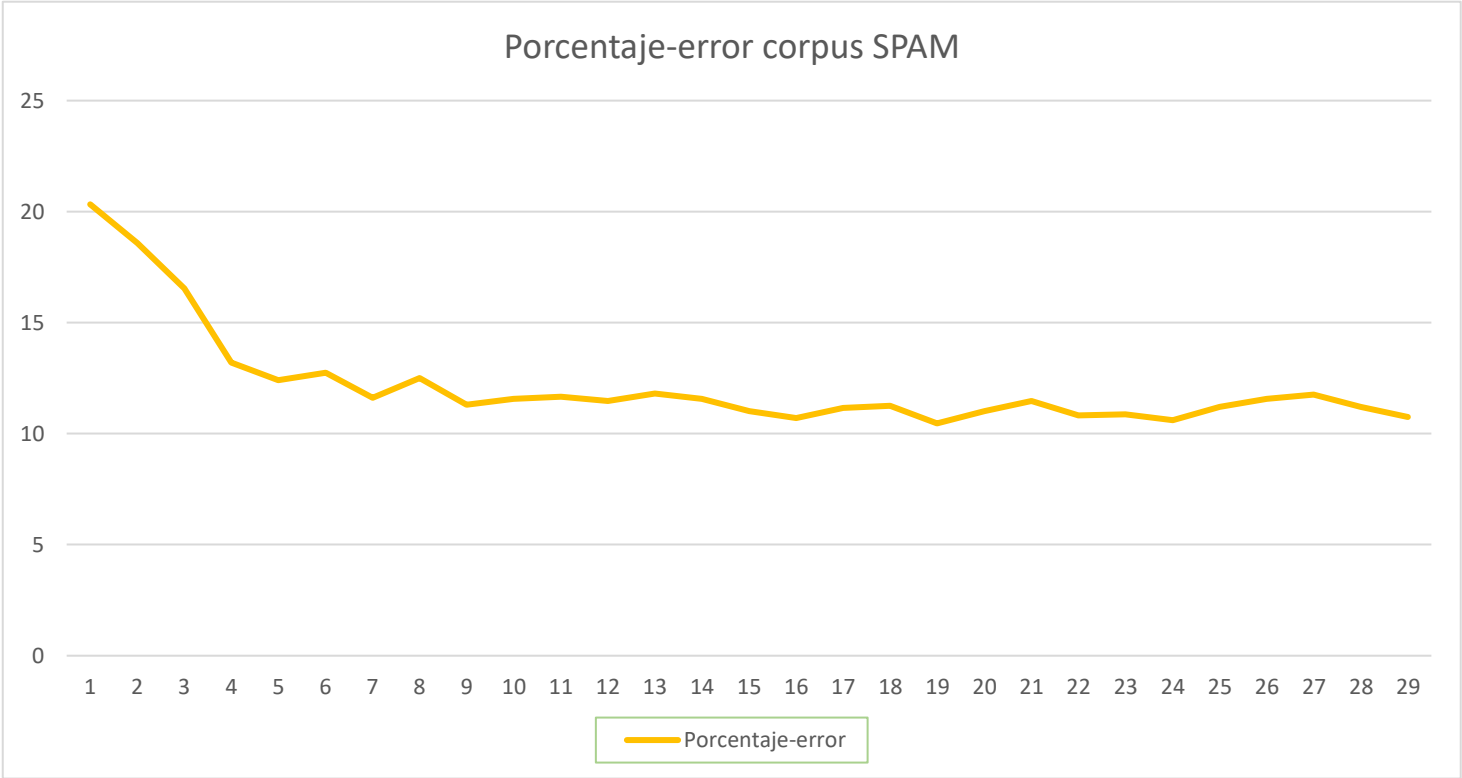
./analisisMixturaGaussianas.sh ./data/spam/tr.dat

./data/spam/trlabels.dat ./data/spam/ts.dat ./data/spam/tslabels.dat p3

Tomando esos parámetros, llamará al script octave y nos creará un .csv con los resultados pedidos, nos dará el número de gaussianas utilizadas, el número máximo de iteraciones, el porcentaje de error y el intervalo de confianza, con estos datos, llegamos a que para el corpus de spam, con 4 gaussianas se puede llegar al 0% de error y normalmente los mejores resultados se obtienen con dos iteraciones.

En cuanto al corpus USPS, necesitamos llegar a las 9 gaussianas para obtener un valor cercano al mejor ~ 10,5 %. Para obtener los mejores resultados se necesitan realizar 6 iteraciones normalmente. No se puede bajar del 10% de error.

Gráficas relación número de gaussianas y porcentaje de error.



CORPUS SPAM			
Numero-gaussianas	Número máximo iteraciones	Porcentaje-error	+/- Intervalo-Confianza
1	2	0,57929	0,40026
2	4	0,14482	0,20057
3	2	0,21723	0,24556
4	4	0	0
5	2	0	0
6	2	0	0
7	2	0,14482	0,20057
8	2	0,072411	0,14187
9	2	0,072411	0,14187
10	2	0,072411	0,14187
11	4	0,072411	0,14187
12	2	0,072411	0,14187
13	2	0	0
14	2	0,072411	0,14187
15	2	0,072411	0,14187
16	2	0,072411	0,14187
17	2	0,072411	0,14187
18	2	0,072411	0,14187
19	2	0,072411	0,14187
20	2	0,072411	0,14187
21	2	0	0
22	2	0,072411	0,14187
23	2	0,072411	0,14187
24	2	0,072411	0,14187
25	2	0,072411	0,14187
26	2	0,072411	0,14187
27	2	0,072411	0,14187
28	2	0,072411	0,14187
29	2	0,072411	0,14187



CORPUS USPS			
Numero-gaussianas	Número máximo iteraciones	Porcentaje-error	+/- Intervalo-Confianza
1	2	20,329	1,7607
2	8	18,585	1,7018
3	8	16,542	1,6256
4	8	13,204	1,4811
5	5	12,407	1,4423
6	8	12,755	1,4595
7	5	11,609	1,4015
8	8	12,506	1,4472
9	6	11,31	1,3857
10	8	11,56	1,3989
11	4	11,659	1,4041
12	8	11,46	1,3936
13	8	11,809	1,4119
14	8	11,56	1,3989
15	6	11,011	1,3695
16	5	10,713	1,3531
17	4	11,161	1,3776
18	5	11,261	1,383
19	6	10,463	1,3391
20	6	11,011	1,3695
21	3	11,46	1,3936
22	6	10,812	1,3586
23	6	10,862	1,3613
24	6	10,613	1,3475
25	6	11,211	1,3803
26	6	11,56	1,3989
27	6	11,759	1,4093
28	6	11,211	1,3803
29	6	10,762	1,3558
30	6	10,713	1,3531
31	6	10,962	1,3668
32	6	11,809	1,4119
33	6	10,214	1,3249
34	6	11,41	1,391
35	6	11,56	1,3989
36	6	10,762	1,3558
37	6	10,962	1,3668
38	6	10,713	1,3531
39	6	12,407	1,4423
40	6	11,46	1,3936
41	6	11,011	1,3695
42	6	12,506	1,4472