

# Modelos Gráficos probabilísticos

Marcos Esteve Casademunt Enric Bonet Cortés

Diciembre 2018

## 1. Ejercicio 2.4 A

En este ejercicio se pretende estudiar la evolución de las probabilidades estimadas al incrementar el número de muestras generadas aleatoriamente y entrenar una nueva Red Bayesiana.

W				S			R		
1	1	1	0	1	0.5532	0.4468	1	0.7234	0.2766
2	1	0.0556	0.9444	2	0.9057	0.0943	2	0.2264	0.7736
1	2	0.0435	0.9565						
2	2	0.0000	1.0000						

C	
1	0.4700
2	0.5300

Cuadro 1: número de muestras: 100

W				S			R		
1	1	1.0000	0.0000	1	0.4979	0.5021	1	0.8066	0.1934
2	1	0.1085	0.8915	2	0.9066	0.0934	2	0.1673	0.8327
1	2	0.0905	0.9095						
2	2	0.0250	0.9750						

C	
1	0.4860
2	0.5140

Cuadro 2: número de muestras: 1000

Como podemos observar, si se usan 100 muestras el resultado de las probabilidades del modelo difieren bastante con las del modelo original . Solo hace falta mirar en W el caso para  $S = 2$  y  $R = 1$  para ver que la probabilidad estimada baja hasta incluso la mitad de la original. En cambio,

esta diferencia con el modelo original se mejora con el uso de 1000 muestras y se consigue reducir la diferencia entre el modelo estimado y el original significativamente. Por lo que podemos concluir que a cuantas mas muestras se empleen para estimar el modelo, mas se acercara este al original

## 2. Ejercicio 2.4 B

Se pide calcular la probabilidad de que un paciente no fumador no tenga cáncer de pulmón si la radiografía ha dado un resultado negativo pero sufre disnea. insertando las evidencias y estudiando cáncer:

$$P(C = n \mid F = n, X = n, D = p) = 0,9989$$

Es decir, con un 99,89 % de probabilidades no tiene cancer y con un 0.11 % sufre cancer.

La explicación más probable de que un paciente sufra cancer es: Polucion **baja**, Fumador **si** cancer **si**, Rayos x **positivo**, Disnea **positivo**

## 3. Experimentación con SPAM y USPS

En la implementación se ha realizado una función que dado como parámetros las matrices de Train y Test y el número de gaussianas se estima el modelo y devuelve la tasa de error. Una vez implementada la función se implementa una función "lanzadera" que invoca a la función anteriormente comentada con distintos número de gaussianas (escala logarítmica). De esta forma se pretende estudiar la variación del error al variar el número de gaussianas. Además se ha fijado la semilla aleatoria a 0 para permitir la repetición del experimento. El código se adjunta junto con la memoria.

Gaussianas	error (%)
1	0.579
2	0.217
5	0.145
10	0.145
20	0.217

Cuadro 3: Error al incrementar el número de gaussianas en el corpus SPAM

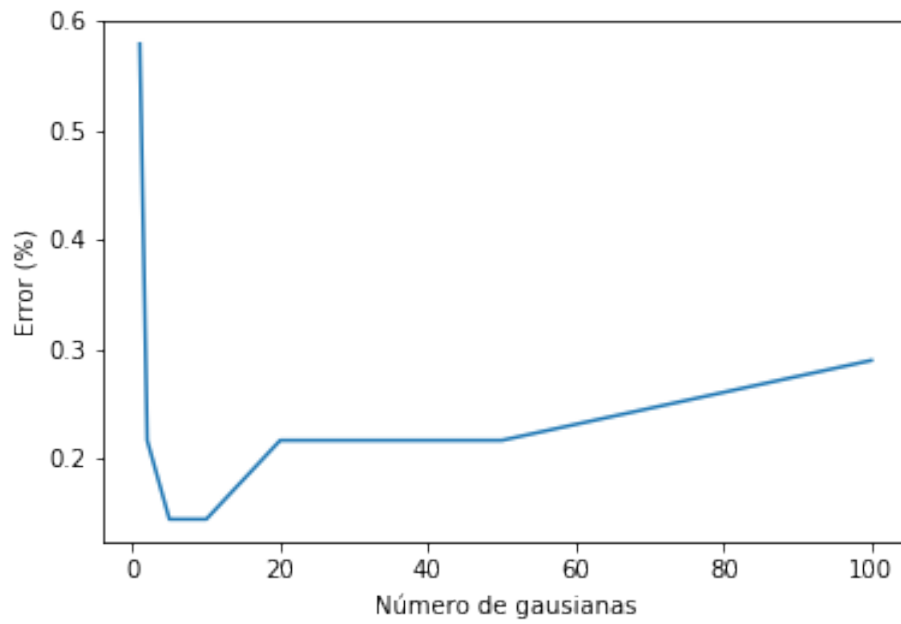


Figura 1: Evolución del error en el corpus SPAM

Como podemos observar en la tabla, el mejor resultado se obtiene con 5 o 10 gaussianas. Consiguiendo reducir el porcentaje de error a 0.145 %

Gaussianas	error ( %)
1	20.329
2	17.588
5	12.108
10	11.908
20	10.812
50	11.659

Cuadro 4: Error al incrementar el número de gaussianas en el corpus USPS

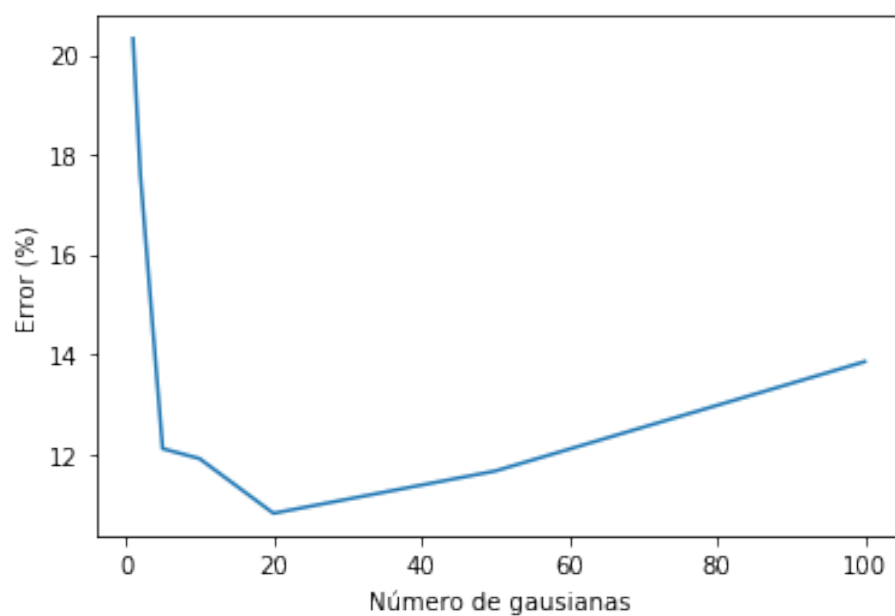


Figura 2: Evolución del error en el corpus USPS

Como se observa en la imagen, el mejor resultado se obtiene con 20 gaussianas con un error de 10.812 %