

Examen final
Inteligencia artificial, período 2018–1.
 Profesor: Julio Waissman Vilanova.

Nombre: _____

1. Responde a los siguientes enunciados como falso o verdadero.
 - (a) ___ Una red bayesiana con n variables aleatorias, cada una de ellas con m valores, *siempre* requerirá establecer menos de m^n parámetros.
 - (b) ___ Al estar basado en la teoría estadística del aprendizaje, las máquinas de vectores de soporte *siempre* son mejores que el método de naive Bayes, pero naive bayes se puede aplicar a problemas más grandes al ser más sencillo su algoritmo de entrenamiento.
 - (c) ___ A diferencia del aprendizaje supervisado, el aprendizaje no supervisado no implica un método de optimización.
 - (d) ___ Cuando la cantidad de datos es muy grande, los resultados obtenidos utilizando un SVM o un método de *naive bayes* son muy similares.
 - (e) ___ Cuando la cantidad de datos es muy grande, los resultados obtenidos utilizando una red neuronal profunda o un método de *naive bayes* son muy similares.
 - (f) ___ Existen problemas en los cuales el método de *naive bayes* es el mejor método de aprendizaje posible, aún mejor que algoritmos muy sofisticados.
 - (g) ___ Las redes bayesianas no son más que una forma gráfica de representar una distribución de probabilidad conjunta de n variables.
 - (h) ___ Los métodos de muestreo para inferencia en redes bayesianas son preferibles, ya que el cálculo es más sencillo y el resultado más exacto.
 - (i) ___ Todo problema modelado como red bayesiana se puede expresar como un modelo oculto de Markov (HMM).
 - (j) ___ El profe de la materia de Inteligencia Artificial es el más barco de toda la Licenciatura.
2. Asumamos que tenemos un filtro antispam para correos electrónicos basado en el método de *naive bayes* y bolsa de palabras (BOW). La idea básica es que, sobre un conjunto predeterminado de palabras, se calcula la probabilidad que una palabra w se encuentre en un correo *una o más veces en alguna parte del texto*, si sabemos si es *spam* o *ham* ($P[w|spam]$, $P[w|ham]$ respectivamente). Igualmente, se calcula la probabilidad *a priori* de que un correo sea *spam* o *ham* ($P[spam]$, $P[ham]$). A partir de esto, es posible calcular la probabilidad que un correo sea *spam* utilizando la simplificada ecuación

$$P[spam|w_1, w_2, \dots, w_n] = \frac{P[spam] \prod_{i=1}^n P[w_i|spam]}{P[spam] \prod_{i=1}^n P[w_i|spam] + P[ham] \prod_{i=1}^n P[w_i|ham]}$$

Contesta lo siguiente:

- (a) Asumamos la siguiente tabla después del aprendizaje:

w	mira	viagra	dinero	Cancún	todos
<i>spam</i>	1/6	1/8	1/4	1/4	1/8
<i>ham</i>	1/8	1/3	1/4	1/12	1/12

y recibimos un correo electrónico que dice «mira a todos bailar».

Selecciona los valores de $P[spam]$ para los cuales el correo tiene más probabilidades de ser *spam* que *ham*. 0.0 0.2 0.4 0.6 0.8 1.0

(b) Supongamos ahora que solo tenemos 3 correos en el conjunto de aprendizaje:

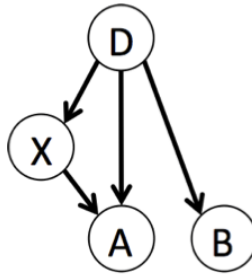
1. «Gran oportunidad para ganar mucho dinero sin hacer nada.»
2. «Hey, tengo mucho calor, vamos por unas cheves.»
3. «Nos vemos mañana a las 12:00»

A partir de estos, calcula los siguientes parámetros utilizando el modificador de Laplace:

- (a) $P[nada|spam]$ _____
- (b) $P[mucho|spam]$ _____
- (c) $P[cerveza|ham]$ _____
- (d) $P[ham]$ _____

3. Considere la siguiente red bayesiana:

$P(A D, X)$			
+d	+x	+a	0.9
+d	+x	-a	0.1
+d	-x	+a	0.8
+d	-x	-a	0.2
-d	+x	+a	0.6
-d	+x	-a	0.4
-d	-x	+a	0.1
-d	-x	-a	0.9



$P(D)$	
+d	0.1
-d	0.9

$P(X D)$		
+d	+x	0.7
+d	-x	0.3
-d	+x	0.8
-d	-x	0.2

$P(B D)$		
+d	+b	0.7
+d	-b	0.3
-d	+b	0.5
-d	-b	0.5

Responda a las siguientes preguntas:

- (a) $P[+d, +a] =$ _____.
- (b) $P[-d, +a] =$ _____.
- (c) $P[+d | +a] =$ _____.
- (d) $P[+d | +b] =$ _____.
- (e) $P[+d] =$ _____.

4. Considere el conjunto de datos siguiente:

	x_1	x_2
$x^{(1)}$	0	0
$x^{(2)}$	-1	2
$x^{(3)}$	-2	1
$x^{(4)}$	-1	-2
$x^{(5)}$	3	3
$x^{(6)}$	1	1

Vamos a probar con dos métodos de *clustering* muy similares: las K-medias y las K-medianas. Existen dos diferencias fundamentales entre las K-medias y las K-medianas, las cuales son las siguientes:

- Las K-medias utiliza a la distancia euclidiana para definir los *clusters*, mientras que las K-medianas se basan en la distancia de *Manhattan*
- La función de actualización de las K-medias es obteniendo la media de los puntos pertenecientes a una clase para calcular el nuevo prototipo, mientras que para las K-medianas, se calcula la mediana de los elementos pertenecientes a la clase para el nuevo prototipo. Como podrá verse, no se quebraron mucho la cabeza con los nombres.

Ahora asumamos que tenemos dos prototipos (dos *clusters*, o lo que es lo mismo, $K = 2$). Los clusters están dados por $C_1 = (-3, 0)$ y $C_2 = (2, 2)$.

Contesta las siguientes preguntas:

- (a) ¿Cuáles variables pertenecen a C_1 si estamos utilizando las K-medias? _____.
- (b) Realiza un único paso del algoritmo de las k-medias y escribe el valor de C_1 y de C_2 .

- (c) Con los valores de la tabla, ¿Cuáles variables pertenecen a C_1 si estamos utilizando las K-medianas? _____.
- (d) Realiza un único paso del algoritmo de las k-medianas y escribe el valor de C_1 y de C_2 .

5. En la figura siguiente tenemos un grafo, el cual le faltan los sentidos a las aristas para que sea dirigido. Selecciona la dirección de cada una de las aristas, tal que el grafo represente una distribución de probabilidad conjunta tal que:

1. $D \perp\!\!\!\perp G$,
2. $D \not\perp\!\!\!\perp A$,
3. $D \perp\!\!\!\perp E$,
4. $H \perp\!\!\!\perp F$.

