

Sistemas de Apoyo a la Decisión



2. Examen Teórico (2 ptos)

Nombre y Apellidos: Xabier Galdeano Berriamendi

1. Enunciado

1. Preguntas teóricas (1 pto, 15 minutos)

a) Explica (0.25 pto):

- 1) Indica cuál de los siguientes supuestos se refiere a una tarea de clasificación o clustering:

1.- Supongamos que disponemos de un conjunto de pacientes caracterizados por 100 atributos numéricos como la presión arterial, si toman o no anticoagulantes, peso, altura, si tienen antecedentes familiares de cardiopatías, y sabemos si han sufrido o no problemas cardíacos. Queremos construir un modelo que sea capaz de identificar los pacientes que muestran una alta probabilidad de tener problemas cardíacos. (Clustering o Clasificación).

Escribe tu respuesta: Clasificación

Razonala:

2.- Somos una empresa de venta de libros online. Disponemos de una muestra de clientes caracterizados por su edad, su sexo, si tienen familia o no, los años que llevan siendo clientes, ella facturación anual que realizan con nosotros, su código postal, si leen novela, o ensayo, o biografías. Queremos realizar una campaña de marketing y queremos personalizar dicha campaña. (Clustering o Clasificación).

Escribe tu respuesta: Clustering

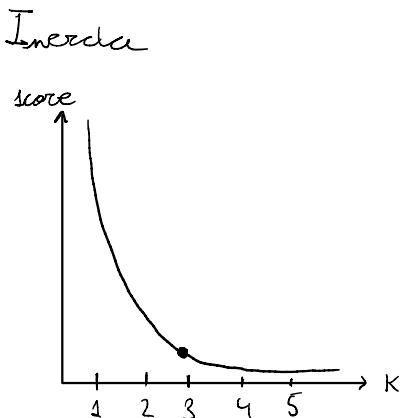
Razonala:

3.- Somos una empresa de venta masiva online y cuando un cliente compra un producto queremos mostrarle productos que le podrían interesar. Eso lo haremos partiendo de los históricos de compras que hacen nuestros clientes e intentando identificar grupos de productos asociados a tendencias de compra. (Clustering o Clasificación).

Escribe tu respuesta: *Clustering*.....

Razonala:

- b) (0.25) Elige una de las métricas, coherencia o la inercia, inventa un gráfico coherente de la métrica elegida y describe su uso para decidir el k apropiado y en qué algoritmo de clustering se suele emplear cada una.



La inercia mide la suma de los distancias de los elementos a sus centroides. Esto mide cuan dispersos están los puntos dentro de los clusters. Se usa en algoritmos llamados como K-means.

Para calcular la K óptima se usa el método Σ de los codos que trata de coger el punto donde la curva deje de disminuir de forma significativa.

- c) (0.25) En Kmeans, explica formas de mejorar el algoritmo modificando la inicialización de los centroides al azar.

K-means ++ es la versión mejorada de Kmeans. La idea es la en una vez elegido el primer centroide de forma aleatoria los siguientes se eligen también aplicando aleatoriedad pero con probabilidad menor mayor probabilidad aquellos que estén más lejos.

- d) (0.25) Explica las diferencias entre el soft y el hard clustering e indica a cuál de los grupos pertenecen Kmeans, Nearest Neighbours y LDA.

Hard - Los items están en un único cluster - Kmeans, Nearest Neighbours

Soft - Los items pueden estar en varios clusters - LDA

2. Ejercicio (1 pto, 30 minutos):

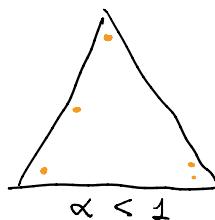
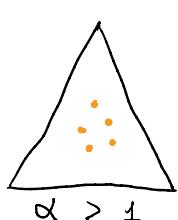
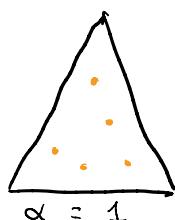
- a) (0.75) Aplicamos el topic modeling en el dominio de la industria minorista de comestibles, donde los temas son distribuciones sobre un surtido fijo de productos y las tendencias de consumidores son descritas como mezclas de temas.

En este contexto, cada ticket de la compra es un documento que puede contener varias tendencias de consumidor, por ejemplo, tendencia a comprar productos pre-cocinados, o tendencia a comprar productos bio, o tendencia a comprar productos vegetales. Como os podéis imaginar un producto puede pertenecer a más de una tendencia o categoría, por ejemplo un gazpacho bio podría encuadrarse en bio y producto precocinado. Suponiendo que abstraemos de las diferencias en cantidades, es decir, 1 kg de plátanos o 2 kg se considerará como plátanos. Aplicar Gibbs Sampling a los siguientes tickets de la compra. Supondremos 4 tópicos (v: verde, az: azul, r: rosa, a: amarillo) y se muestra la inicialización al azar (mirar la Tabla del final):

Carro 1	Carro 2	Carro 3	Carro 4
plátanos (v)	manzanas (v)	pizza (r)	hamburguesa (r)
aceglas (v)	naranjas(r)	pan (a)	pizza (v)
tofu (a)	plátanos (a)	hamburguesa (r)	chocolate (az)
hummus (a)	plátanos (v)	hot-dog (a)	napolitanas (a)
pan (v)	tofu (v)	chocolate (a)	lassagna (r)
pizza (r)	pan (az)	napolitanas (r)	croquetas (r)
lentejas (r)	lechuga (v)	plátanos (r)	
lechuga (v)	chocolate (v)		

Suponiendo alfa=0.1 y Beta=0.01, implementar el Gibbs sampling **hasta que algún producto cambie la tendencia** que se le ha asignado al azar. Indicar claramente QUÉ producto y en QUÉ cesta ha cambiado y presentar TODAS las operaciones que lo demuestren.

- b) Explica cuál sería la incidencia de un alfa y un beta > 1 en el ejemplo (0.25)



Carro 1	Carro 2	Carro 3	Carro 4
/ platanos (v)	/ manzanas (v)	/ pizza (r)	/ hamburguesa (r)
/ acelgas (v)	/ naranjas (r)	/ pan (a)	/ pizza (v)
/ tofu (a)	/ platanos (a)	/ hamburguesa (r)	/ chocolate (az)
/ hummus (a)	/ platanos (v)	/ hot-dog (a)	/ napolitanas (a)
/ pan (v)	/ tofu (v)	/ chocolate (a)	lassagna (r)
/ pizza (r)	/ pan (az)	/ napolitanas (r)	croquetas (r)
/ lentejas (r)	/ lechuga (v)	/ platanos (r)	
/ lechuga (v)	/ chocolate (v)		

Carro → Documento → d
 Color → Término → k
 Comida → Palabras → w

Paso 1 - A que Término pertenece cada palabra de los documentos

		v	a	r	az	
1	plátano	2	1	1	0	4
2	aceitunas	1	0	0	0	2
3	lechuga	1	1	0	0	2
4	humus	0	1	0	0	3
5	pan	1	1	0	0	3
6	pizza	2	0	2	0	3
7	lechuga	0	0	1	0	2
8	lechuga	2	0	0	0	2
9	manzana	2	0	0	0	1
10	naranjas	0	0	1	0	3
11	chocolate	1	1	0	1	3
12	hamburguesa	0	0	2	0	2
13	hot-dog	0	1	0	0	4
14	napolitanas	0	1	0	0	2
15	lassagna	0	0	1	0	1
16	croquetas	0	0	1	0	1
		10	7	10	2	29

Paso 2.1 - Evaluar plátano (?)

$$\frac{3}{7} + \alpha \cdot \frac{1}{3} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{2}{7} + \alpha \cdot \frac{1}{3} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{2}{7} + \alpha \cdot \frac{1}{3} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{0}{7} + \alpha \cdot \frac{0}{3} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right.$$

plátano (?) → plátano (v)

Paso 2.2 - Evaluar acelga (?)

$$\frac{3}{7} + \alpha \cdot \frac{0}{0} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{2}{7} + \alpha \cdot \frac{0}{0} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{2}{7} + \alpha \cdot \frac{0}{0} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{0}{7} + \alpha \cdot \frac{0}{0} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right.$$

Paso 2.3 - Evaluar lechuga (?)

$$\frac{4}{7} + \alpha \cdot \frac{1}{2} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{1}{7} + \alpha \cdot \frac{0}{2} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right. \quad \frac{0}{7} + \alpha \cdot \frac{0}{2} + B \quad \left| \begin{array}{c} v \\ a \\ r \\ az \end{array} \right.$$