



## Sistemas de Apoyo a la Decisión

### 2. Examen Teórico

(2 ptos)

Nombre y Apellidos:.....

---

## 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: .....

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: .....

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: .....

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.

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

*d)* (0.25) Explica las diferencias entre el soft y el har clustering e indica a cuál de los grupos pertenecen Kmeans, Nearest Neighbours y 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 platanos o 2 kg se considerará como platanos. 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
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)		

Suponiendo  $\alpha=0.1$  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 demuestran.

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