Parte I

1. La principal diferencia entre los métodos supervisados (I) y no supervisados (II) es que:

a) (I) requieren que el usuario especifique algunos hiperparámetros.
b) (II) no tienen restricciones y/o supuestos.
c) (I) usan la variable respuesta para entrenar el modelo.
d) (II) se aplican a problemas autónomos.

2. Considere las siguientes afirmaciones:
(i) PCA es un método no supervisado.
(ii) Todos los componentes principales de un PCA son ortogonales entre si.
(iii) PCA busca las direcciones en las que los datos tienen la mayor varianza.
(iv) El número máximo de componentes principales es menor o igual al número de variables.

Elija la opción con el mayor número de ítems correctos:
a) (i) e (iii).
b) (ii) e (iii).
c) (i), (ii) e (iii).
d) (i), (ii), (iii) e (iv).

3. Como parte de un análisis de datos se analizaron 11 indicadores económicos y sociales de 96 países. Las variables observadas son:
X1: Tasa anual de crecimiento de la población,
X2: Tasa de mortalidad infantil por cada 1000 nacidos vivos,
X3: Porcentaje de mujeres en la población activa,
X4: PNB en 2005 (en millones de dólares),
X5: Producción de electricidad (en millones kW/h),
X6: Líneas telefónicas por cada 1000 habitantes,
X7: Consumo de agua per ćapita,
X8: Proporción de la superficie del país cubierta por bosques,
X9: Proporción de deforestación anual,
X10: Consumo de energía per cápita,
X11: Emisión de CO2 per cápita.

Dada la gran cantidad de variables se aplicó un análisis de componentes principales, utilizando la matriz de correlación, donde los vectores de carga de las dos primeras componentes son:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Y_1	−0.314	−0.392	0.116	0.295	0.259	0.446	0.092	0.006	−0.244	0.415	0.375
Y_2	0.348	−0.041	−0.583	−0.177	−0.174	−0.027	0.321	−0.457	−0.154	0.233	0.292

A partir de la información anterior, se puede concluir que:

a) El porcentaje de variabilidad explicado por las dos primeras componentes es 63.45%.
b) Asumiendo las condiciones necesarias sobre las componentes no reportadas, entonces es posible que las variables X2, X6, X10 y X11 son las que más contribuyen en la primera componente principal.
Estas variables son las que más contribuyen en la primera componente principal, ya que son las variables con mayores contribuciones (mayores valores absolutos) según la tabla.

c) Asumiendo las condiciones necesarias sobre las componentes no reportadas, entonces es posible que las variables X2, X6, X10 y X11 son las que ḿas contribuyen en la segunda componente principal.
d) No es posible interpretar los resultados anteriores debido a que es un error haber utilizado la matriz de correlación y en su lugar se debería haber utilizado la matriz de covarianzas.

4. Considere las siguientes observaciones:

i	1	2	3	4	5	6	7	8
x_i	10	8	34	9	46	68	80	50
y_i	4	99	44	50	77	30	25	35
z_i	50	31	86	57	75	14	40	60

Sin escalar las variables, describa tres iteraciones del algoritmo K-means para k = 2. Use los centroides C1 = (47.5, 37.5, 21.8) y C2 = (53.2, 22.4, 75.3).
El desarrollo de este ejercicio está en archivo adjunto: Pregunta 4 _Parte I

5. Enuncie al menos tres diferencias entre el análisis factorial y el método de componentes principales.

1. El análisis factorial busca factores que expliquen la mayor parte de la varianza en común o covarianza. En cambio el PCA se caracteriza por analizar la varianza total del conjunto de variables observadas
2. El análisis factorial supone que existe un factor común subyacente a todas las variables, el análisis de componentes principales no asume esto.
3. En el análisis factorial los factores explican las variables y en el PCA las variables explican los factores.
4. El análisis factorial permite interpretar los factores latentes que se extraen y cómo se relacionan con las variables observadas. En el PCA la interpretación de los componentes principales a menudo se basa en la contribución relativa de las variables originales a cada componente

6. ¿Qué significa que el método de clusterización sea jerárquico?

Agrupa los datos en una estructura jerárquica de forma de árbol o dendrograma. Esto significa que los elementos o puntos de datos se agrupan en niveles sucesivos, donde los grupos más pequeños se combinan para formar grupos más grandes.