**EJERCICIOS DE PROGRAMACIÓN USANDO PCA**

A continuación se plantean una serie de ejercicios. Por favor, responde a las preguntas que plantean. A parte de ello, no olvides adjuntar el código empleado para producir las respuestas a cada ejercicio.

**1 (2.5 puntos).** Carga el archivo de datos “Ingredient\_data.xlsx”, el cual contiene información en 5 variables diferentes representadas en sus 5 columnas (aluminato tricálcico, silicato tricálcico, aluminoferrito tetracálcico, silicato beta-dicálcico, calor durante endurecimiento a lo largo de 180 días) para 12 ingredientes diferentes. Aunque el conjunto de datos está caracterizado por 5 variables, representa los observaciones en un scatterplot de 3 dimensiones caracterizado por las 2 primeras variables y la última: aluminato tricálcico (columna 1), silicato tricálcico (columna 2) y calor durante endurecimiento a lo largo de 180 días (columna 5). Haz que en el gráfico los diferentes ejes tengan el nombre correspondiente: “aluminato tricálcico”, “silicato tricálcico” y “calor”.

**2 (2.5 puntos).** Normaliza los datos de tal manera que su media sea 0 y su desviación típica 1. Indica los valores de la media y la desviación típica de los datos resultantes de la normalización.

Seguidamente, realiza un análisis de componentes principales en el que extraigas los 3 principales componentes mediante los cuales podrías llevar a cabo una reducción de la dimensionalidad. Indicar los autovectores correspondientes a cada uno de estos 3 componentes e indica también el autovalor de cada uno.

**3 (2.5 puntos).** Haz un PCA en el que extraigas los 2 componentes principales. Seguidamente, haz un DataFrame en el que para cada uno de los ingredientes se especifique su puntuación en el componente principal 1 y en el componente principal 2. Llama a las columnas de este DataFrame “componente principal 1” y “componente principal 2” respectivamente, y muestra muestra los valores numéricos de dicho DataFrame (para ello, puedes simplemente generar el DataFrame e imprimirlo en la ventana de comandos con la función “print”. Seguidamente, basta con que pegues una captura de pantalla del output en tu respuesta).

¿Qué porcentaje de la varianza total de los datos explican respectivamente el componente principal 1 y el componente principal 2?

**4 (2.5 puntos).** Representa gráficamente los datos en función de los 2 componentes principales identificados. En la representación gráfica, llama a los ejes x e y “componente principal 1” y “componente principal 2” respectivamente.