**EJERCICIOS DE PROGRAMACIÓN USANDO GMM**

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 puntos).** Carga el archivo de datos “ClimaticData\_GMM.xls” y representa sus observaciones gráficamente en un scatterplot que tenga por título “Climatic data”, que su eje x tenga la etiqueta “Pressure” y su eje y la etiqueta “Temperature”.

**2 (2 puntos).** Usa k-medias para clasificar las observaciones en 2 clústers diferentes y representa el mismo scatterplot que en el ejercicio anterior, esta vez con las observaciones clasificadas en cada clúister.

Comenta la agrupación realizada por k-medias. ¿Te parece una agrupación correcta de los datos?

**3 (2 puntos).** Usa ahora GMM para realizar una agrupación correcta de los datos. Representa el scatterplot resultante, haciendo, además, que cada observación tenga un tamaño proporcional a su probabilidad de pertenencia a un grupo.

**4 (2 puntos).** Hemos visto la clasificación hecha por GMM. Representa las funciones Guassianas del modelo ajustadas a cada clúster de observaciones.

**5 (2 puntos).** Ahora utilizarás 3 modelos GMM cada uno con un tipo de covarianza: “diag”, “spherical” y “full”. Utiliza cada uno de estos 3 modelos para aprender a partir de los datos su distribución. Seguidamente, usa cada uno de los modelos aprendidos para generar 1000 puntos de datos totalmente nuevos que sigan la distribución aprendida. Representa estos nuevos datos en 3 scatterplots diferentes (uno para cada tipo de covarianza de GMM) que tengan por título “New Climatic data”, que su eje x tengan la etiqueta “Pressure” y su eje y la etiqueta “Temperature”. A partir de las diferentes gráficas podrás deducir que hay un tipo de covarianza que crea un mal ajuste a los datos. ¿De cuál se trata?