



Escuela de Ciencias
Maestría en Matemáticas Aplicadas
Técnicas robustas y no paramétricas
Taller 2

NOTA: _____

NOMBRE: _____ CÓDIGO: _____

GRUPO: _____ PROFESOR: Henry Laniado Fecha de entrega _____

1. Abra un fichero word y punto por punto escriba los resultados y pegue las gráficas en el fichero. El fichero returns.txt contiene rentabilidades de 10 activos durante los últimos 1100 meses
 - a) [10 %] Tome los últimos 900 meses y pase un test no paramétrico de rangos para identificar qué activos provienen de distinta distribución. Estime y gráfique las densidades en un mismo plano para un par de activos que pasen el test. Estime y grafique la densidad en un mismo plano para un par de activos que no pasen el test. Observa congruencia entre el resultado del test y las gráficas de las densidades?. Explique el tipo de kernel y el ancho de banda utilizado para las estimaciones.
 - b) [10 %] Considere un par de activos que según el punto anterior vienen de distinta distribución. Sobre ese par de activos vuelva y tome los 1100 datos y suponga que la variable x es el activo de menor índice de columna, y la variable y el otro.
 - 1) Contruya un modelo de regresión y explique la significancia del modelo y su ajuste.
 - 2) Utilice varias medidas de profundidad estadística para eliminar el 25 % de los datos menos profundos. Realice las gráficas de visualización de la región central. Sobre la región central contruya el modelo de regresión y comente los resultados.
 - 3) Con el método de eliminación de registros con mayor valor en su residual, elimina el 25 % de los meses con mayor valor en su residuo (uno a uno, no todos de una). Grafica la región central y los outliers según este procedimiento y sobre la región central construya el modelo de regresión. Comente los resultados.
 - 4) Construya varios modelos de regresión lineal no paramétrica y robusta, modificando la matriz de covarianzas por una versión robusta o no paramétrica o combinación de ellas. (Kendall, Spearman, fastMCD, máxima kurtosis, shrinkages) y haga inferencia bootstrap para los parámetros.
 - 5) Realice un análisis sobre los procedimientos anteriores y concluya cual sería el del mejor y peor performance.

- c) Suponga que $x^a = (x_1^a, \dots, x_{n_a}^a)$ son n_a realizaciones de una uniforme en el intervalo $[6, 8]$ y $x^b = (x_1^b, \dots, x_{n_b}^b)$ son n_b realizaciones de una uniforme en el intervalo $[2, 10]$. Considere la siguiente mixtura de normales

$$Y \sim 0.2N(-2x^a + 10, 1) + 0.8N(2x^b + 4, 1).$$

- 1) Genere 15000 pares ordenados del modelo anterior y resuelva todos los 5 apartados del punto anterior. ($n_a + n_b = 15000$)
- 2) Con las realizaciones (x^a, x^b) estime y grafique la densidad de la mixtura

$$X \sim 0.2U[6, 8] + 0.8U[2, 10].$$

Haga lo mismo con las realizaciones de Y para estimar su densidad.

- d) Con el par de activos distintos mas correlados, suponga que la variable x es el activo de menor índice de columna, y la variable y el otro. Realice regresiones no paramétricas (estimación Nadaraya-Watson) para explicar en términos de x a y^2 , y^5 , $y^{\frac{1}{5}}$, $\cos(y)$
- e) Tome las primeras 5 columnas del fichero de tiempos de vida con todos los datos. Divida la muestra en dos conjuntos de datos de 5 columnas cada uno (puede ser con diferente tamaño). Realice un test de hipótesis basado en profundidad estadística para determinar si ambas muestras vienen de la misma población. (esto es haciendo un $D - D$ plot y haciendo una regresión lineal sobre esos valores del $D - D$ plot. Si observas que se rechaza la hipótesis nula que las muestras vienen del mismo modelo, entonces vuelve y haz otra división hasta encontrar un par de muestras que vengan de la misma población. Justifica bien por qué crees que se rechaza o no la hipótesis.
- f) Realiza lo mismo anterior con el fichero completo *return.txt*
- g) Con base en los dos puntos anteriores explica, para esta entrega, una metodología para identificar si dos conjuntos de imágenes vienen de la misma población. Intenta implementar la técnica para un par de conjuntos de imágenes de cáncer o las de face94. (esto último, la implementación, no para esta entrega, hazlo para tu formación).