Escuela de Ciencias

Maestría en Matemáticas Aplicadas

Técnicas robustas y no paramétricas

Examen

NOTA: ______________

NOMBRE: ____ CÓDIGO: _________________________

GRUPO: ____ PROFESOR: Henry Laniado Fecha

Abra un fichero word y punto por punto escriba los resultados y pegue las gráficas en el fichero. El fichero return.txt contiene rentabilidades de 12 activos. Elimine la primera columna que es la fecha. Construya una etiqueta (1) si media de la fila es positiva y (-1) si la media de la fila no es positiva.

1. [40 %]

   a) [8 %] Pase un test no paramétrico de rangos para identificar qué activos de etiqueta (-1) provienen de distinta distribución. Realice una matriz de 12X12 con el valor 1 si se rechaza ´H_0 y valor 0 si no se rechaza H_0. Estime y gráfique las densidades 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 %] Genere 10000 datos sintéticos que vengan de la misma población de los datos crudos con etiqueta (-1). Realice el punto anterior con los datos simulados. Hay diferencias en la matriz 12X12 con del punto anterior.

   c) [8 %] A cada columna del fichero completo, realice la transformación $\frac{x-min}{max-min}+j/12$. Donde j es el índice de columna. Sobre los datos transformados, realice la gráficas en un mismo plano de la distribución empírica. Explique lo que observa. Pase un test de normalidad sobre la que según la distribución empírica tiene mejores rendimientos. Asi no pase el test, estime los parámetros de la normal y grafique en un mismo plano la empírica y la teórica, comente los resultados.

   d) [8 %] Determine estadísticamente si la muestra conjunta original de los datos con etiqueta (-1) vienen de la misma población de la muestra conjunta de los datos con etiqueta (1). Explique con detalle el procedimiento.

   e) [8 %] Considere los datos originales y suponga que la variable x es la columna 12 y la variable $y = \sin(5x) + 0.1 * columna10$. Realice regresiónes no paramétricas (estimación Nadaraya-Watson) para explicar la variable y en términos de x

   f) [8 %] Considere los datos originales, calcule un intervalo de confianza bootstrap para estimar el parámetro $\theta = (asimetria + 1)/(kurtosis + 1)$. de la columna 12. Cómo usaría este intervalo para concluir normalidad o no?. Estime el sesgo del estimador.

2. [30 %] Sobre el fichero original, considere como variable respuesta la columna 12, y como explicativas las otras 11. (sin etiqueta).

   a) [5 %] Contruya un modelo de regresión lineal habitual y explique la significancia del modelo y su ajuste. Exponga el modelo sólo con las variables significativas.

   b) [5 %] Con todas las variables, utilice la profundidad de mahalanobis para eliminar uno a uno el 20 % de los datos menos profundos. Sobre la región central, realice el punto anterior. Comente los resultados

   c) [5 %] Con todas las variables, con el método de eliminación de registros con mayor valor en su residual, elimine registros uno a uno hasta obtener un ajuste al menos del 90 %. Sobre la región central construya de nuevo el modelo sólo con variables significativas. Comente los resultados

   d) [10 %] Construya el modelo de regresión lineal robusta, cambiando la matriz de covarianza por aquella que sale del procedimiento de cambiar la correlación de pearson por la correlación $\rho = \frac{Spearman+Kendall+Pearson}{3}$. Haga inferencia bootstrap sobre $\beta_1 0$.

   e) [5 %] Explique con argumentos con qué método de los 4 anteriores se quedaría.

3. [20 %]

   a) [10 %] Pruebe que si $X_1, X_2, \ldots, X_n$ son variables aleatorias Bernoully de parámetro p, entonces

$$\sum_{i=1}^{n} \sqrt{X_i} \text{ converge en probabilidad a } np$$

   b) [10 %] Calcule la distribución del primer estadístico ordenado para una muestra de n variables aleatorias exponenciales de parámetro 1. Calcule el sesgo teórico de $X_{[1]}$. Genere 1000 variables exponenciales de parámetro 1 y estime el sesgo Jacknife de $X_{[1]}$ y cómparelo con el sesgo teórico