

NOTA: Esta Tarea se puede resolver de manera individual o en parejas.

1. Obtener la función $p = f(u)$ a partir de la función: $u = \ln\left(\frac{p}{1-p}\right)$.
2. Los siguientes incisos se basan en los datos de los archivos que se encuentran en Canvas: *datatrain.csv*, y *datatest.csv*. Los datos fueron generados aleatoriamente a partir de la función cúbica $y = 5x^3 - 12.5x^2 + 7.5x$ incluyendo ruido gaussiano con media 0 y desviación estándar 0.5.
 - a. Obtener la gráfica de dispersión de los datos de entrenamiento (train), x vs y .
 - b. Obtener la gráfica de dispersión de los datos de prueba (test), x vs y .
 - c. Incluir además la gráfica de la función cúbica $y = 5x^3 - 12.5x^2 + 7.5x$ en las dos gráficas anteriores.
 - d. Contestar los siguientes incisos:
 - i. Usando los datos de entrenamiento (train), obtener el modelo de regresión lineal $\hat{y} = \beta_0 + \beta_1 x$. Interpreta el valor de significancia de cada uno de los coeficientes del modelo obtenidos.
 - ii. Incluir la gráfica de dispersión de los datos de entrenamiento, junto con la gráfica de la recta de regresión \hat{y} obtenida.
 - iii. Usando los datos de prueba (test), obtener el valor del error cuadrático medio MSE y el coeficiente de determinación ajustado R^2 .
 - iv. Repite los incisos i, ii, iii para obtener el modelo de regresión cuadrático:
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
 - v. Repite los incisos i, ii, iii para obtener el modelo de regresión cúbico:
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$
 - vi. Repite los incisos i, ii, iii para el modelo de regresión de grado cuatro:
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$
 - vii. Incluye tus conclusiones finales a partir de los resultados de los modelos anteriores.
 - e. Los siguientes incisos se resolverán incluyendo un valor extremo (*outlier*) tanto en el conjunto de entrenamiento, como en el conjunto de prueba. Para ello, deberás cambiar únicamente el valor de y del primer dato de los conjuntos de entrenamiento y prueba como sigue: En el conjunto de entrenamiento (*datatrain*) el primer punto es (1.41414141, 0.3476383) y deberás cambiarlo a (1.41414141, 4). En el conjunto de prueba (*test*) el primer punto es (0.36363636, 1.42458487) y deberás cambiarlo a (0.36363636, 5). Usando estos nuevos conjuntos de entrenamiento y prueba contesta los siguientes incisos:

- i. Obtener el modelo de regresión cúbico con los datos de entrenamiento y los niveles de significancia de sus coeficientes. Obtener el coeficiente de determinación con los datos de prueba y la gráfica de dispersión de los datos de prueba junto con la gráfica de la función cúbica de regresión obtenida. Compara tus resultados con el modelo de los datos sin los outliers.
 - ii. Con los datos de prueba calcular el valor del error cuadrático medio MSE y la suma de los cuadrados de los errores SSE.
 - iii. Con los datos de prueba calcular el valor del error absoluto medio MAE y la suma de los valores absolutos de los errores.
 - iv. Con los datos de prueba calcular el valor del error con la función de costo Huber para $\delta = 1$.
 - v. Con los datos de prueba calcular el valor del error con la función de costo Log-Cosh, así como la de su valor promedio.
 - vi. Con los datos de prueba calcular el valor del error del logaritmo cuadrático medio, MSLE. NOTA: indica la constante que en dado caso le sumes a la fórmula para obtener el valor buscado.
 - vii. Con los datos de prueba calcular el valor del error del porcentaje absoluto medio, MAPE. NOTA: indica la constante que en dado caso le sumes a la fórmula para obtener el valor buscado.
- f. Incluye tus conclusiones finales de esta actividad.