

Quiz 2 – Minería de datos (Práctico)

1. El dataset Wage (librería ISLR) trabajado en clase tiene otras variables no exploradas como el estado marital (maritl), o la clase de trabajo (jobclass) y otras. Explore las relaciones entre por lo menos estas dos variables con la variable respecto a wage, usando técnicas de ajuste no lineal para ajustar modelos flexibles a los datos. Cree figuras de los resultados obtenidos y describa un resumen de sus hallazgos.
2. Considere el dataset de Auto (librería ISLR) para responder a las siguientes preguntas
 - a. Haga una gráfica de matriz de dispersión que muestre todas las variables del dataset
 - b. Calcule la matriz de correlación entre las variables. Ayuda: Excluya la variable cualitativa de name
 - c. Realice una regresión lineal múltiple con mpg como variable respuesta con el resto de las variables, excluyendo la variable name. Comente resultados sobre esta regresión.
 - i. ¿Hay alguna relación entre los predictores y la respuesta?
 - ii. ¿Cuáles predictores tienen una relación estadísticamente significativa?
 - iii. ¿Qué sugiere el coeficiente de la variable year?
 - d. Obtenga las gráficas de diagnóstico y comente posibles problemas con los datos. Identifique posibles outliers u observaciones que puedan afectar la regresión.
 - e. Realice una nueva regresión contemplando parámetros de interacción. ¿Cómo cambian los resultados?
 - f. ¿Recomendaría usar alguna transformación como $\log(X)$ o X^2 ? Justifique