

Práctica 14 de mayo

Trabajo con datos multivariados

- Se presenta un conjunto de datos que contiene características de casas, en Ames-Iowa, y sus respectivos precios.
- El objetivo de esta práctica es analizar qué variables en el conjunto de datos pueden ser útiles para predecir el precio de las casas mediante un modelo lineal. Para esto se debe hacer uso de técnicas de visualización de datos multivariados presentadas en clase.
- Tenga en cuenta que puede ser necesario limpiar los datos; además de un análisis de datos atípicos, campos con porcentaje alto de datos faltantes, posiblemente normalización de rangos de datos, análisis de correlación/covarianzas que le permita explorar las correlaciones con la variable de interés y multicolinealidad entre variables explicativas, identificación de distribuciones de las variables, posiblemente reducción dimensional...

Regresión lineal

- Una vez tenga seleccionado un conjunto de variables, puede usar regresión lineal para construir un modelo que permita estimar el precio de las casas.
- Puede usar [statsmodels](#).
- Un ejemplo práctico lo puede consultar en este [enlace](#).
- En ésta última hay una sección (*Código para implementar los estimadores*) que da un ejemplo de cómo usar **scikit-learn** para seleccionar los rasgos o características útiles para generar un buen modelo. Puede usar este selector de rasgos para verificar sus hallazgos en la primera parte de la práctica.