

Universidad Francisco Marroquín

Facultad de Ciencias Económicas

Ingeniería en Computer Science

Catedrático: Ing. Samuel Chávez

Contacto: samuelchavez@ufm.edu

Auxiliar: Paulo Mejía

Contacto: paulomejia@ufm.edu



Laboratorio 4

Regresión lineal polinomial

Instrucciones: Usted usará Python a través de un Jupyter Notebook para llevar a cabo este ejercicio. Recuerde utilizar comentarios para describir lo que está haciendo en cada parte del proceso. Estará usando el juego de datos proporcionado dentro del portal. Al finalizar recuerde subir al portal un link a su repositorio en el que se pueda correr su notebook, usando <https://mybinder.org/>.

El juego de datos proporcionado es parte de la plataforma Kaggle, dentro del cual se muestran los precios de casas en King County, Seattle. Este juego de datos incluye los precios de las casas vendidas entre mayo 2014 y mayo 2015. Nótese que el precio es dependiente de varias características como el número de habitaciones, número de baños, metros cuadrados de la sala, pisos, etc.

Característica	Descripción
id	Identificación única para cada casa vendida
date	Fecha de la venta de la casa
price	Precio de cada vivienda vendida
bedrooms	Número de habitaciones
bathrooms	Número de baños, donde .5 representa una habitación con inodoro, pero sin ducha
sqft_living	Pies cuadrados del espacio habitable interior de los apartamentos
sqft_lot	Pies cuadrados del espacio terrestre
floors	Número de pisos
waterfront	Una variable ficticia para determinar si el apartamento tenía vistas al mar o no
view	Un índice de 0 a 4 de qué tan buena era la vista de la propiedad
condition	Un índice de 1 a 5 sobre el estado del apartamento
grade	Un índice de 1 a 13, en el que 1 a 3 un nivel bajo de calidad de construcción y diseño de edificios, 7 tiene un nivel medio de calidad de construcción y diseño y 11 a 13 tiene un alto nivel de calidad de construcción y diseño.
sqft_above	Pies cuadrados del espacio interior de la vivienda que están a nivel del suelo
sqft_basement	Pies cuadrados del espacio interior de la vivienda que están por debajo del nivel del suelo
yr_built	El año en que se construyó la casa inicialmente

yr_renovate d	El año de la última renovación de la casa.
zipcode	En qué área de código postal se encuentra la casa
lat	Latitud
long	Longitud
sqft_living15	Los pies cuadrados de espacio habitable interior de la vivienda para los 15 vecinos más cercanos
sqft_lot15	Los pies cuadrados de los terrenos de los 15 vecinos más cercanos

Para este ejercicio se le pide que proporcione un modelo de regresión polinomial que muestre la relación precio y pies cuadrados del espacio habitable interior de los apartamentos que se muestra en el juego de datos proporcionado. A continuación, se mencionan las generalidades de los pasos sugeridos a realizar.

1. Leer el archivo CSV proporcionado (kc_house_data.csv) y almacenarlo en un **np.array** para ser trabajado en el notebook.

Ajustar un modelo polinómico (regresión lineal) en base al juego de datos cargado de forma matricial que relacione las variables de precio con los pies cuadrados del espacio habitable interior de los apartamentos (price – sqft_living).

2. Utilice la implementación vectorial del algoritmo de regresión lineal (descenso del gradiente visto en clase).
3. Usando *cross-validation* determine el grado del polinomio que mejor describe la nube de puntos (encuentre el mejor balance entre apego a los datos de entrenamiento y generalización para datos previamente no observados).
4. Haga un análisis sobre sus hallazgos.