

Proyecto 1

Ruben Cuadra

Dr. Victor de la Cuva

Aprendizaje Automatico

August 21, 2017

Regresión lineal (recta) simple (de una variable)

Implementación en python del caso mas simple de regresión, el cual es cuando los datos se aproximan por medio de una recta y se busca una relación solo entre dos variables, una independiente **x** y una dependiente **y**. El script requiere un archivo csv que posea 2 columnas que representen los vectores x/y.

Requisitos:

Libreria matplotlib (Graficas)

Libreria numpy (Operaciones matemáticas)

Python 2.7 o 3.5 (Instalar las librerías correctamente usando pip o pip3)

Estructura:

main.py: Archivo principal, contiene 6 funciones y el código para ejecutar

getDataFromFile(filename): Reciben como parámetro un archivo separado por comas y regresa una matriz de 2 columnas representando los valores x,y que posee en archivo.

getThetas(x,y): Recibe dos vectores como parámetro e implementa la formula $\theta = A^{-1} \cdot Y$, regresa el vector thetas con dos posiciones.

Proyecto 1

h_t(x, thetas): Es la hipótesis, recibe como parámetro un número x y un vector de 2 posiciones que contiene θ_0 y θ_1 , los evaluar usando

$$h(x) = \theta_0 + \theta_1 x$$

graficarDatos(x,y,thetas): Recibe 2 vectores con datos en x y en y además de un tercer vector que contiene un vector de 2 posiciones con valores de θ s. La función grafica los puntos que representan los primeros dos vectores y genera una recta usando h_t y el vector θ s que fue pasado, no regresa nada.

gradienteDescendente(x,y,theta=[0,0],alpha=0.01,iteraciones=1500): Es la función encargada de iterar y obtener los valores de θ aproximándose α lugares, los primeros dos parámetros son obligatorios y representan los datos x,y , los siguientes parámetros están definidos por default pero pueden ser modificados si se desea iniciar con una θ diferente o si se requiere un mayor número de iteraciones/ α , regresa 1 vector de 2 posiciones representando α_0 y α_1

calculaCosto(x,y,theta): Recibe los valores x,y y posteriormente vector de θ s, usa la fórmula de calcular costo para dar un resultado numérico.

datos.csv: Archivo que contiene valores numéricos x,y ; Son dos columnas separadas por comas

Si se desea correr el código basta con escribir: **python main.py**

Siempre teniendo un archivo **datos.csv** junto a él y que contenga 2 columnas