

Proyecto 1

Regresión lineal (recta) simple (de una variable)

Descripción

El caso más simple de regresión es cuando los datos se aproximan por medio de una recta y se busca una relación sólo entre dos variables, una independiente x y una dependiente y .

Hacer un programa en Python que reciba como entrada un conjunto de ejemplos de la forma **(entrada,salida)** y regrese los valores de Theta (θ_0 y θ_1) que mejor ajusten al conjunto de datos. La recta debe ser calculada por el algoritmo de **regresión lineal**.

Datos

Los datos están en un archivo (ex1data1.txt) de 2 columnas. El archivo deberá ser separado en dos vectores:

1. El vector **X** conteniendo TODAS las entradas. Se refieren a la cantidad de feet² de la casa.
2. El vector **y** conteniendo TODAS las salidas. Se refieren al valor de la casa en USD.

Notas

- Recuerde que para facilitar la vectorización es necesario agregar una columna de 1's a la matriz de entradas X.
- Para poder vectorizar tiene que contar con operaciones con matrices por lo que puede usar una librería que las proporcione.

Funciones

Se deberán programar 3 funciones

1. **graficaDatos(X,y,theta)**. Recibe los datos de entrada y los grafica como puntos en un plano **(x,y)**. Además, recibe un vector theta y grafica sobre los datos la recta que resulte de esos valores.
2. **gradienteDescendente(X,y,theta,alpha,iteraciones)**. Recibe:
 - a. Datos de entrada ya separados en vectores **X** y **y**.
 - b. Vector **theta** = $[\theta_0, \theta_1]$ inicial. Dicho vector puede estar inicializado en 0 para ambos valores.
 - c. Razón de aprendizaje **alpha**. Se recomienda probarlo con un valor $\alpha = 0.01$.
 - d. El número de iteraciones **iteraciones**. El número de iteraciones que va a realizar el algoritmo. Se recomienda probarlo con $\text{iteraciones} = 1500$.

La función debe regresar el valor del vector theta final.

3. **calculaCosto(X,y,theta)**. Recibe las entradas y un vector theta y debe regresar la función de costo $J(\theta_0, \theta_1)$ que resulta.

Resultados de prueba

Una vez entrenado el sistema, es decir, una vez encontrados los valores de los parámetros para la recta de regresión, se pueden hacer las siguientes pruebas:

Preducion1 = $[1, 3.5] * \theta$, lo cual debe regresar un valor de 4519.767868

Prediccion2 = $[1, 7] * \theta$, lo cual debe regresar un valor de 45342.450129

Recuerde que $h_{\theta}(x) = \theta^T X = \theta_0 + \theta_1 x_1$ donde X representa en este caso el vector de ejemplos completo, es decir, con las parejas ordenadas (entrada, salida). Por eso, x_1 se refiere al valor de la entrada (los índices en Python inician en 0), que en realidad se refiere a la entrada. Por esta razón, las predicciones se obtienen multiplicando una entrada **[x,y]** por el vector theta (que es un vector columna).

Pueden también imprimir el costo al inicio, antes de llamar a **gradienteDescendente**, con el vector inicial de theta igual a **[0,0]**. Luego lo imprimen después de correr la función y el vector debe disminuir casi hasta 0.

Vectorización

VECTORICE TODAS las operaciones que pueda vectorizar, utilizando alguna librería para manejo de matrices.