

Espero que ya estudiaron la regresión lineal hasta Lecture 4.7.

Ahora ustedes van a programar la regresión lineal y la van a correr sobre el archivo de datos que yo proporcioné en la tarea anterior.

Mi visión de la estructura del programa:

1. Función de lectura de datos: en la entrada el archivo con datos de precios y características de casas, en la salida la matriz X y el vector y .
2. Función de *feature scaling*: $x_i := \frac{x_i - \mu_i}{s_i}$, donde x_i es i -ésima característica, por ejemplo, `sqft_living`, μ_i es el promedio de los valores de esta característica, s_i es desviación estándar de los valores de esta característica. Tal vez, hay que hacer *feature scaling* antes de la lectura de los datos en el punto 1, pero esto es como convenga a cada uno. En la matriz X esta característica va a estar en i -ésimo renglón, también sería cómodo hacer *feature scaling* a partir de esta matriz.
3. Partición de los datos en dos conjuntos: primero se mesclan (revuelven) los datos aleatoriamente, luego se dividen en dos conjuntos: conjunto de entrenamiento 70% y conjunto de prueba 30%. No olviden que, mesclando los datos, se tiene que preservar la alineación de la matriz y el vector: es decir, en la posición n del vector y tiene que estar el precio de la casa cuyas características están en la columna n de la matriz X .
4. Función del cálculo de predicciones \hat{y} .
5. Función de costo.
6. Algoritmo de descenso de gradiente, vamos a dar valor a *learning rate*=0.1.

Hay que correr las funciones en los puntos 3, 4, 5 in 1000 iteraciones, imprimiendo el valor de la función de costo cada 50ª iteración. Este valor tiene que ir disminuyendo.

Después de la fase de entrenamiento hay que calcular las predicciones para el conjunto de prueba y calcular el valor de la función de costo para estas predicciones.

Ahora me van a entregar hasta el martes 31 de marzo:

El programa en solo un script.

Captura de pantalla: impresión de los valores de la función de costo cada 50ª iteración y el valor de la función de costo para el conjunto de prueba.