Homework 2 report

Name: Reza Sayyari Student Number:0079105

The defined function first reads the input data. The input data is the price and living square of 200 houses. Then, it deletes the NaN values. We specify the price as input and square living as the target value of our code.

I have used "scipy.stats" for calculating t-static. Then, the code can calculate a 95% credible interval.

We need to normalize our input data in order to avoid overflow. For this reason, the input data is divided by its maximum value. After that, our input data will be between zero and one.

We need to compute the hypothesis by evaluating the linear relationship between X and y. The idea is, if we plot the simple line on data that has less deviation or error from the actual values, then it can be used to predict the future value with very minimal error. Here we know the value for x, y from sample data; using that, we have to compute optimal theta0 and theta1, which has minimal error cost to plot the linear fit.

First, we define a function to calculate the cost.

Appending a term x0 in our existing matrix X for mathematical convenience,x0 should be having values as '1' and assign some initial theta as 0. Our goal is to further reduce this cost J(theta) value to achieve the optimal linear fit for our data. We define Gradient descend to find the optimal parameter 'theta' using the given parameters.

x — Input values

y — output values

Initial_theta — in most cases NULL theta

alpha — the rate at which gradient pointer descending to the optimal value

iteration — setting how many iterations it should take

Finally, the code plots our line on data to see how well it fits it.