

Assignment 2

SUBMIT ASSIGNMENT

Due Saturday by 11:59pm

Points 16

Submitting a file upload

Housing Price

Univariate Linear Regression is a supervised learning algorithm where we train a hypothesis function so that it could predict a single output value from a single input feature. Here, we assume that a straight line in two dimensions is sufficient to fit the data well.

Let f denotes the hypothesis function, θ_0 and θ_1 denotes the modelling parameters and x denote the input feature. Then, the hypothesis function becomes, $f(x) = \theta_0 + \theta_1 x$.

In a vectorized form, this can be written as, $f(X) = X^T \theta$, where $X = [1 \ x]^T$, $\theta = [\theta_0 \ \theta_1]^T$.

A management firm wants to buy a house from a landlord. The firm wants to estimate an appropriate price of the house that they can offer to the landlord. They hire you as a machine learning engineer to do the job. You have to make an estimation based on the total land area in square feet.

To assist you, you are given data of 97 other houses with their corresponding land area and the cost of the house. This data can be found in the file 'HousingData.txt' in the 'Files' Folder. The first column represents the area of the land in 100s of square feet and the second column represents the cost of the house in \$100,000s.

- Import the data into the workspace and plot it with area on x-axis and cost on y-axis. (4 points)
- Use linear least squares to find the optimal values of θ . (Hint: if data can be written in the form $y = H\theta$ where y is the vector of cost of the house, then the parameter vector $\theta_{LS} = (H^T H)^{-1} H^T y$). (6 points)
- Plot the hypothesis function $f(x)$ with the data. (2 points)
- Suppose if the area of the house is 700 sq. feet. Find the estimated price of the house in dollars. (4 points)

