Linear Regression With One Variable

(Univariate Linear Regression)

Linear regression predicts a real-valued output based on an input value. We discuss the application of linear regression to housing price prediction, present the notion of a cost function, and introduce the gradient ~~descent~~ descent method for learning.

Model and Cost Function

Notations:—

$m \rightarrow$ Size of training set

$x \rightarrow$ Input variable / feature

$y \rightarrow$ Output variable / target

$x^{(i)} \rightarrow$ i^{th} input variable (i is index)

$y^{(i)} \rightarrow$ i^{th} output

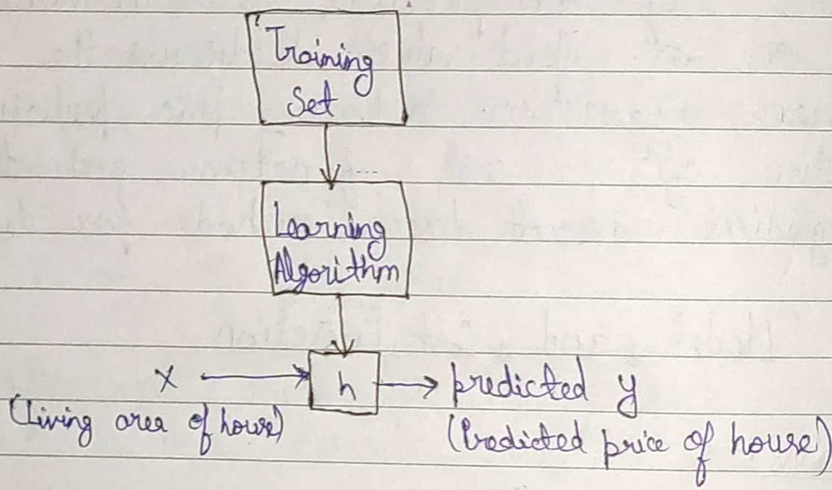
$X \rightarrow$ Space of input values

$Y \rightarrow$ Space of output ~~vari~~ values

$(x^{(i)}, y^{(i)}) \rightarrow$ Training example of i^{th} index

- To describe this ~~that~~ supervised learning problem more formally, our goal is, given a training set, to learn a function $h: X \rightarrow Y$ so that $h(x)$ is a "good" predictor for the corresponding value of y.
  For historical reasons, this function h is called a hypothesis.

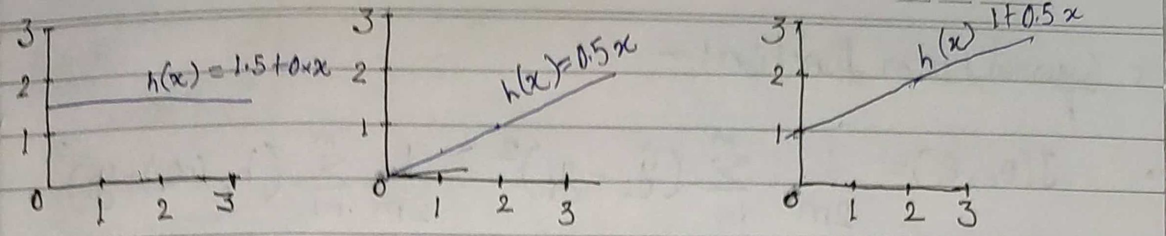
- Process is like this:—



- $h_\theta(x) = \theta_0 + \theta_1 x$

  $h_\theta(x)$ is also written as $h(x)$.

- θ_i s (θ_0, θ_1 etc.) are parameters.



$\theta_0 = 1.5$ $\theta_1 = 0$ | $\theta_0 = 0$ $\theta_1 = 0.5$ | $\theta_0 = 1$ $\theta_1 = 0.5$

★ How to get θ_0 and θ_1 (parameters) ? :—

- Idea → Choose θ_0, θ_1 so that $h_\theta(x)$ is at least close to y for training examples (x, y).

★ Cost ★ Function :—

- We can measure the accuracy of our hypothesis function by using a cost function
- This takes an average difference (fancier version of average) of all results of the hypothesis with input from x's and actual output y's.
- There different types of cost functions. One of the most commonly used is Squared Error Function.

Premium

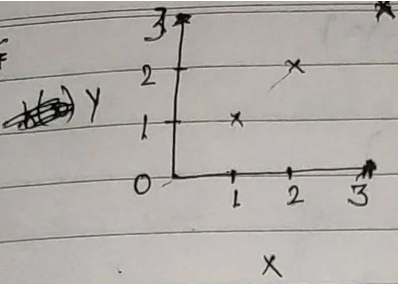
* Squared Error Function:—

- $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$
- $h_\theta(x_i) = \theta_0 + \theta_1 x_i$
- It is $\frac{1}{2} \bar{x}$, where $\bar{x}$ is the mean of the squares of $h_\theta(x_i) - y_i$, or the difference between the predicted value and the actual value.
- Its also called Mean Squared Error Function or Squared Error Cost Function.
- The mean is halved ($\frac{1}{2}$) as a convenience for the computation of the gradient descent, as the derivative term of the square function will cancel out the $\frac{1}{2}$ term.
- Goal → /Optimisation Objective $\underset{\theta_0, \theta_1}{\text{minimize}}\ J(\theta_0, \theta_1)$

* Example:—

Ques → Suppose we have a training set with $m = 3$ examples, plotted below. Our hypothesis is $h_\theta(x) = \theta_1 x$, with parameter θ_1. The cost function $J(\theta_1)$ is $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$. What is $J(0)$?

~~Ans~~



Ans →

$J(0) = ?$

$\theta_1 = 0$

~~$h(0) = 0 \times x = 0$~~

~~$h(1) = 1 \times x = x$~~

~~So, $h(x) = 1 \times x = x$~~

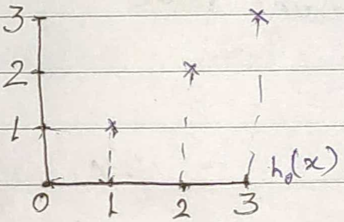
So, $h(0) = \theta_1 x = 0 \times 0 = 0$

$h(1) = 0 \times 1 = 0$

$h(2) = 0 \times 2 = 0$

$h(3) = 0 \times 3 = 0$

Lets plot this :—



$$J(0) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i \right)^2$$

$$J(0) = \frac{1}{2 \times 3} \times \left[(0-1)^2 + (0-2)^2 + (0-3)^2 \right]$$

Premium

$$J(0) = \frac{1}{6}\left((-1)^2 + (-2)^2 + (-3)^2\right)$$

$$J(0) = \frac{1}{6} \times (1+4+9)$$

$$J(0) = \frac{14}{6}$$

Ans

- Similary $J(1) = 1$, $J(0.5) = 0.58$ and so on.

Contour Plot :-

A contour plot is a graph that contains many contour lines. A contour line of a two variable function has a constant value at all points of same line.

- Eg :-