ML Week 1

Start Date - 08/03/2020

Finish Date -

Course Name - Machine Learning by Stanford University

Instructor - Andrew Ng

Platform - Coursera

--- Week 1 --

Machine learning is the process of getting computers to learn without being explicitly programmed

Regression means producing continuous value outputs

In Supervised learning, we give the dataset with the correct answers. There is a relationship between the input and the output and the answers are known. It has two forms - Classification and regression

Classification gives a categorial answer. we try to find a discrete output

Regression gives a numerical answer. Regression prediction problems are usually quantities or sizes. we find a continuous function

Support Vector Machine - algo help the computer to deal with an infinite number of attributes

Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

One form of Unsupervised Learning is Clustering , we make cluster or groups on the basis of particular properties of input data

Linear regression with one variable (univariate linear regression) eq => y = mx + c

    the eq is the hypothesis or function

    Cost function : $J(\theta_0, \theta_1) = 1/2m \{\sum(i=1,m)[h(x)-y]^2\}$, minimize J   [aka Squared Error function]

The mean is halved as a convenience for the computation of the gradient descent, as the derivative term of the square function will cancel out the 1/2 term.

If θ-0 is set to 0 then the cost function minimizes at θ-1 = 1 for univariate linear regression

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 the same line.

Gradient descent is used to minimize functions , we start taking value of variables and solving the function. then we update the value and again solve the function with updated values, till we get the minimum value of function with a particular set of values of variables. we start with 0

So we have our hypothesis function and we have a way of measuring how well it fits into the data. Now we need to estimate the parameters in the hypothesis function. That's where gradient descent comes in.

Only square matrices have inverse.
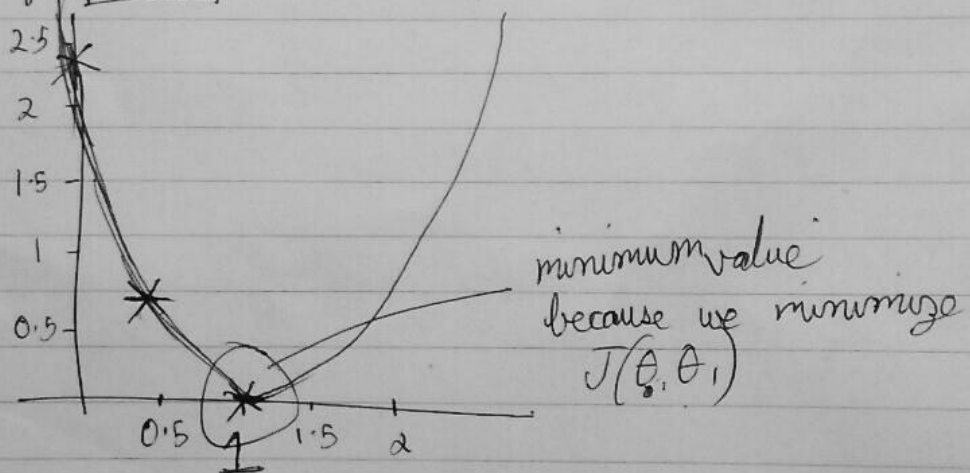
Machine Learning (Week-1)

⇒ Linear Regression with one variable

Cost $f^n \rightarrow J(\theta_0, \theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i \right)^2$

minimize $J(\theta_0, \theta)$

also called Squared Error f^n.

⇒ $J(\theta_1)$ if $\boxed{\theta_0 = 0}$, the curve is

minimum value because we minimize $J(\theta_0, \theta_1)$

⇒ Gradient Descent Algorithm

→ repeat until convergence {

$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$ for $j=0$ & $j=1$

learning rate

$\Rightarrow$ correct : Simultaneous update

only
after
solving
these :
$$\begin{cases} \text{temp } 0 = \theta_0 - \alpha \dfrac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) \\[3mm] \text{temp } 1 = \theta_1 - \alpha \dfrac{\partial}{\partial \theta_1} (J(\theta_0, \theta_1) \end{cases}$$

then $\longrightarrow$ $\theta_0 = \text{temp } 0$, $\theta_1 = \text{temp } 1$

$\Rightarrow$ If we are already at a local minimum, derivative becomes 0, so we dont get moved, ie. $\alpha = 0$

$\Rightarrow$ As we reach towards minimum, gradient descent well automatically decrease even with a fixed learning rate (α).

$\Rightarrow$ On using gradien descent on linear regression

$\Rightarrow$ repeat until convergence {

$$\theta_0 = \theta_0 - \alpha \times 1 \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)$$

$$\theta_1 = \theta_1 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)} - y^{(i)} \right) \cdot x^{(i)}$$

update θ_0 & θ_1 simultaneously.

}

$\hookrightarrow$ this algo is called 'Batch' Gradient Descent.

$\Rightarrow$ It has a convex shape or bow shape