

* Linear Regression

- Supervised Algorithm
- It attempts to model the relationship between dependent and independent variables by fitting a linear equation to observed data.
- Independent (explanatory) and dependent variables represent the respective feature of dataset.
- Represented by, (for one independent and dependent)

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \mu$$

where

- $h_{\theta}(x)$ is the actual value that we want to predict
- θ_0 is intercept of $h_{\theta}(x)$ means where line intersect $h_{\theta}(x)$ axis
- θ_1 is slope or gradient of line means how steep is line
- x is independent variable value
- μ is random error or noise
- We can generalize that previous equation for multiple independent and a dependent variable.

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n + \mu$$

where n is total data points in our given dataset.

- \hat{y} is predicted value same as $(h_{\theta}(x))$, y is actual value.

- Regression for one dependent and one independent variables is called simple linear regression.
- Regression for one dependent and multiple independent variables is called multiple linear regression.

- This equation represents a straight line passing through our data points.

- Our aim is to draw a line or to fit this line in our data points such that cost function is minimum so it generalizes the whole dataset and predict the value of $h_{\theta}(x)$ for given x_n accurately.

- This line is called best fit line or regressor line.

- The process of fitting this line is called Training of the model and is completely hit and trial procedure.

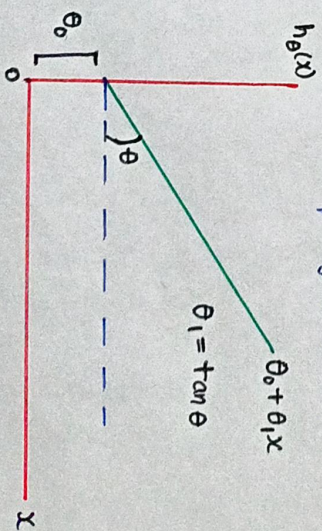
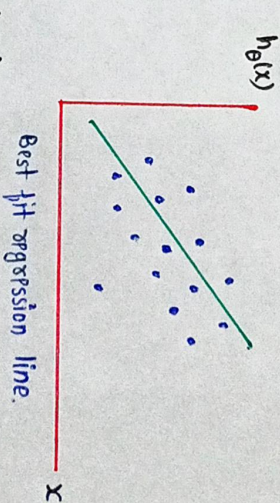
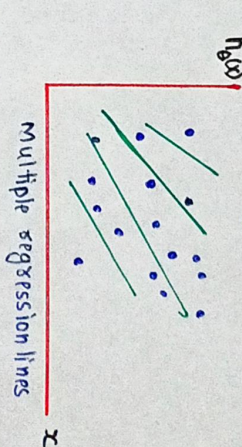
- To adjust the regressor line we tweak the intercept (θ_0) and slopes (θ_n), which in turn minimizes the cost function.

→ Loss function

- Loss: is the difference between a predicted value and actual value. (Residual)
- Loss function quantifies loss for all data points in a single number.
- Used to refer to errors for single training dataset.

→ Cost function: used to refer to an average of loss function over entire dataset.

- Here, in our Regression we can see,



- When we tweak slope and intercept in training process to get best-fit regression line is hit and trial method. To make it more efficient we use Gradient descent algorithm.