Intro to Linear Regression

By

Amritansh

Regression

- Regression refers to correlation between dependent and independent variables.

  Consider eqn of line: $\mathbf{Y = mx + c}$

  or, $\mathbf{y = \Theta_0 + \Theta_1 x}$ (*linear regression*)

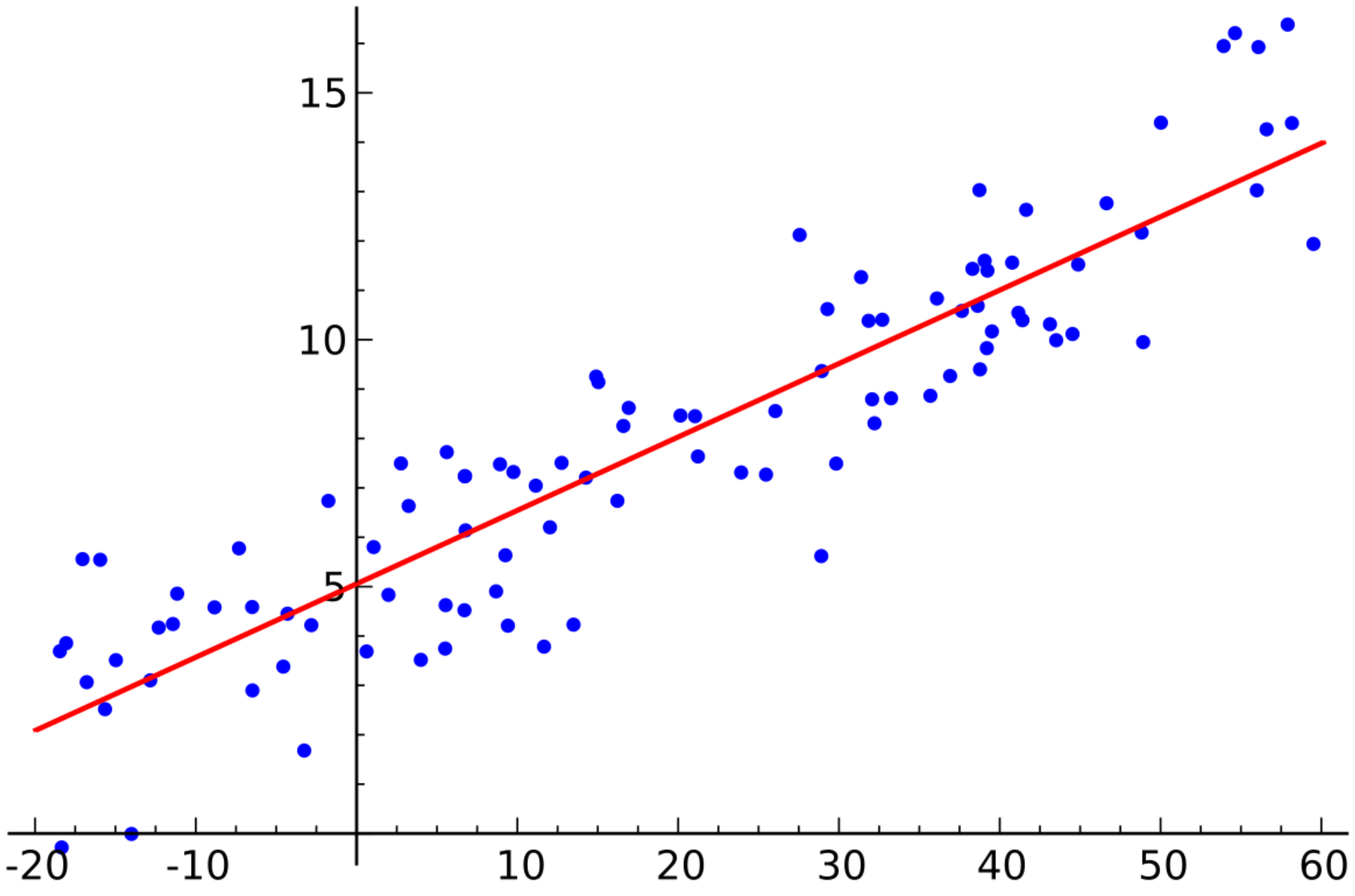
  or, $\mathbf{y = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \ldots\ldots\ldots + \Theta_n x_n}$ (*polyvariate regression*)

  In statistics regression is defined as a measure of the relation between the mean value of one variable (e.g. output) and corresponding values of other variables (e.g. time and cost).

Linear Regression

- Simple linear regression is a statistical method that allows us to summarize and study relationships between two real variables
- In simple words to find a line to describe the linear function that best represents the given data

$$y = \Theta_0 + \Theta_1 x$$



Case in point

Housing price prediction

Size in Sq ft (x)	Price in 1000 $ (y)
2104	460
1416	232
1534	315
852	178
.	.
.	.
.	.
.	.

The idea is that we are given training data with sample inputs. We have input(x) as Sq footage and output(y) as the prices of house

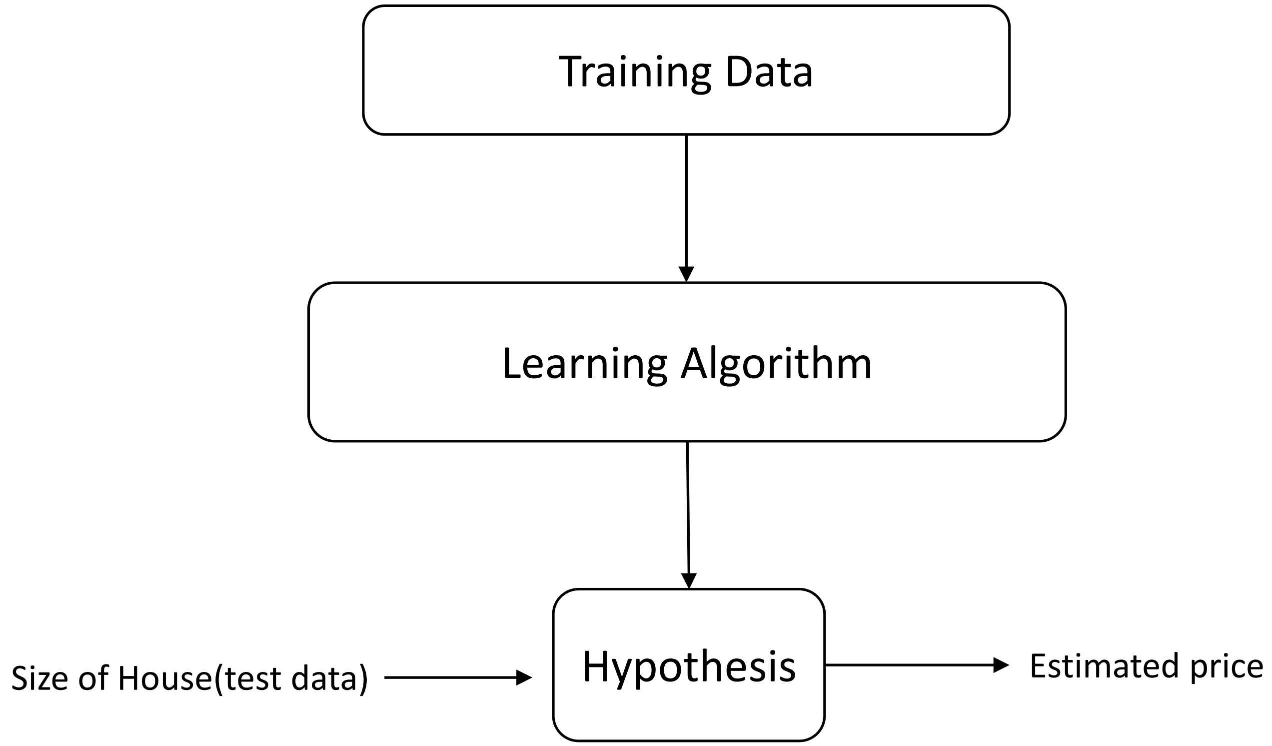
Now we need to estimate the correct features to approximate the line that represents the data

m = no. of data points

x = input variables/features

y = output variables/features

Hypothesis



The hypothesis(h) maps $x_{(i)}$ to $y_{(i)}$

Cost function

$$\mathbf{h_0(x) = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}_1 x}$$

where, Θ_i = *Parameters of hypothesis*

The Biggest question is how to choose the correct parameters ?

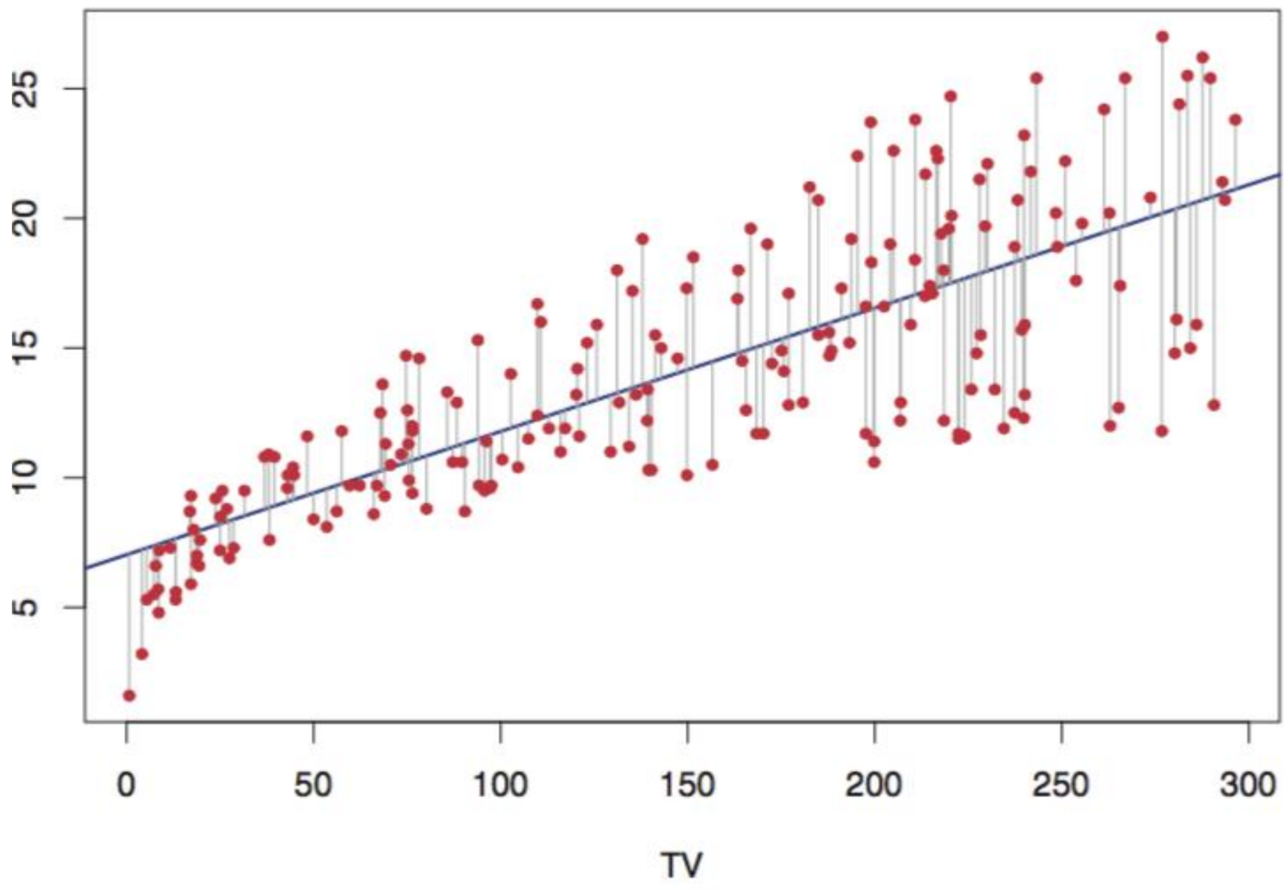
The need to minimize the difference between actual and predicted output is all the fuss is about.

minimize $(\boldsymbol{\Theta}_{0,} \boldsymbol{\Theta}_1) = y_{(i)} - h_{(i)}(x)$

or, minimize $j\ (\boldsymbol{\Theta}_{0,} \boldsymbol{\Theta}_1) = (y_{(i)} - h_{(i)}(x))^2$

Intuition behind cost function

- The idea is to reduce the error/loss between predicted output and actual output.
- Thus to determine the best approximation of a line we need to take average of squared distances of all points provided in training set.
- Then optimize each parameter of hypothesis function in order to get closer to the best approximation.

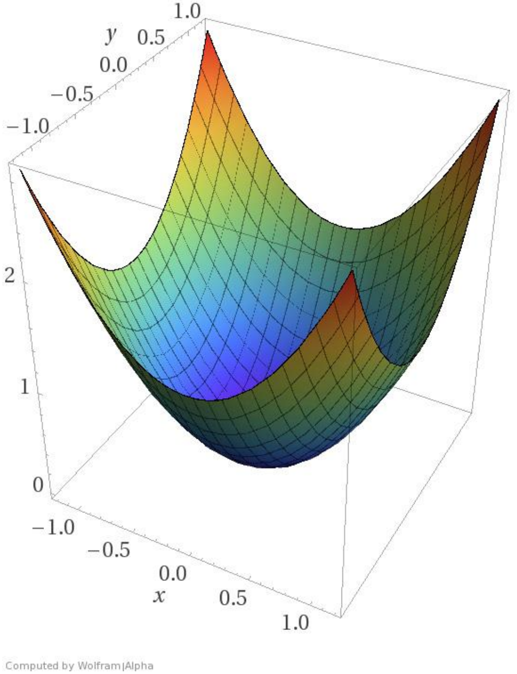


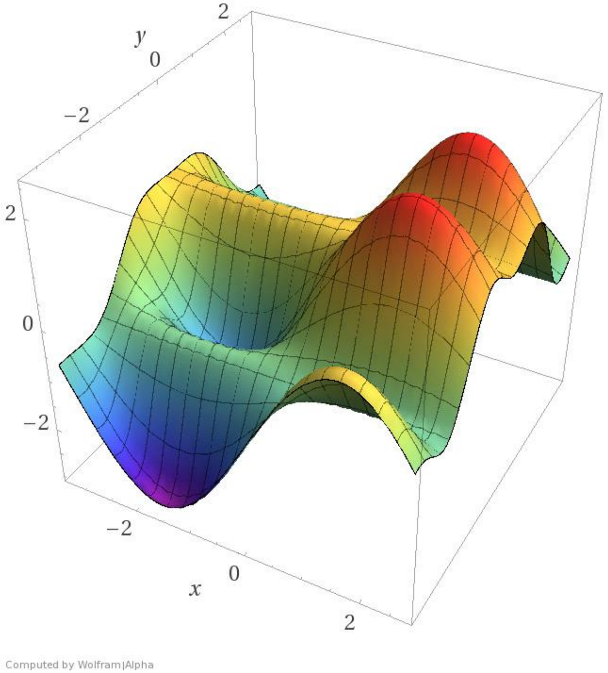
To summarize cost function

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

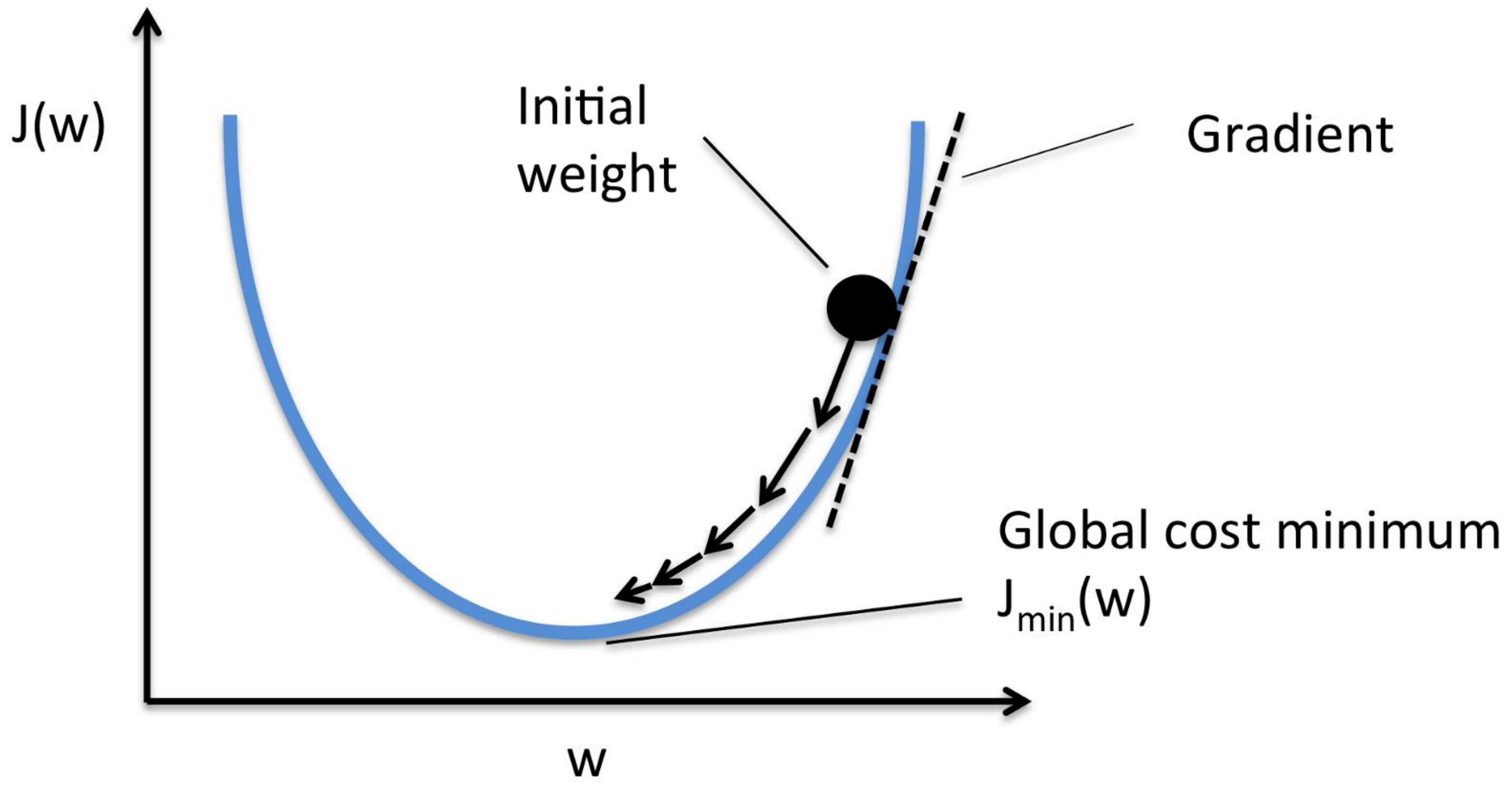
Gradient descent

The idea is to move towards the correct direction while approximating the parameters (θ_i)





Intuition behind Gradient descent



Gradient descent algorithm

repeat until convergence {

$$\theta_0 := \theta_0 - \alpha \overbrace{\boxed{\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)}}^{\frac{\partial}{\partial\theta_0}J(\theta_0,\theta_1)}$$

$$\theta_1 := \theta_1 - \alpha \underbrace{\boxed{\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)\cdot x^{(i)}}}_{\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1)}$$

}

update θ_0 and θ_1 simultaneously

Steps in nutshell

1. Separate the data into training and testing if not done already
2. Initialize random real numbers for parameters of hypothesis
3. Initialize a small number for learning rate (0.001)
4. Calculate cost/loss function for all data points and average them
5. Then shift the value of parameters towards local optimum step by step
6. Repeat till you hit optimum point.
7. If it takes too much time initialize the experiment with different values and run again.

Linear Regression with multiple variables

Area (sq ft)	# Bedrooms	# Bathrooms	Years of home	# of Floors	Price ($1000)
2140	5	3	10	2	460
1416	3	3	5	2	330
1534	3	2	7	1	300
852	2	1	3	1	252
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Here, hypothesis function looks like,

$$\mathbf{h_0(x) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \Theta_3 x_3 + \Theta_4 x_4 + \Theta_5 x_5}$$

Making it bit general :

$$\mathbf{h_0(x) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \ldots\ldots\ldots + \Theta_n x_n}$$

$$\mathbf{Or,\ h_0(x) = \Theta_0\, x_0 + \Theta_1 x_1 + \Theta_2 x_2 + \ldots\ldots\ldots + \Theta_n x_n}$$

Cost function (again)

- No Change. Its the same.

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

Gradient descent (again)

- We just update multiple feature with the same rule as previously done.

Repeat until convergence {

$$\theta_1 \leftarrow \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(x_i) - y_i)$$

$$\theta_2 \leftarrow \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(x_i) - y_i).x_i$$

}

Pros and Cons of Linear Models

Pros

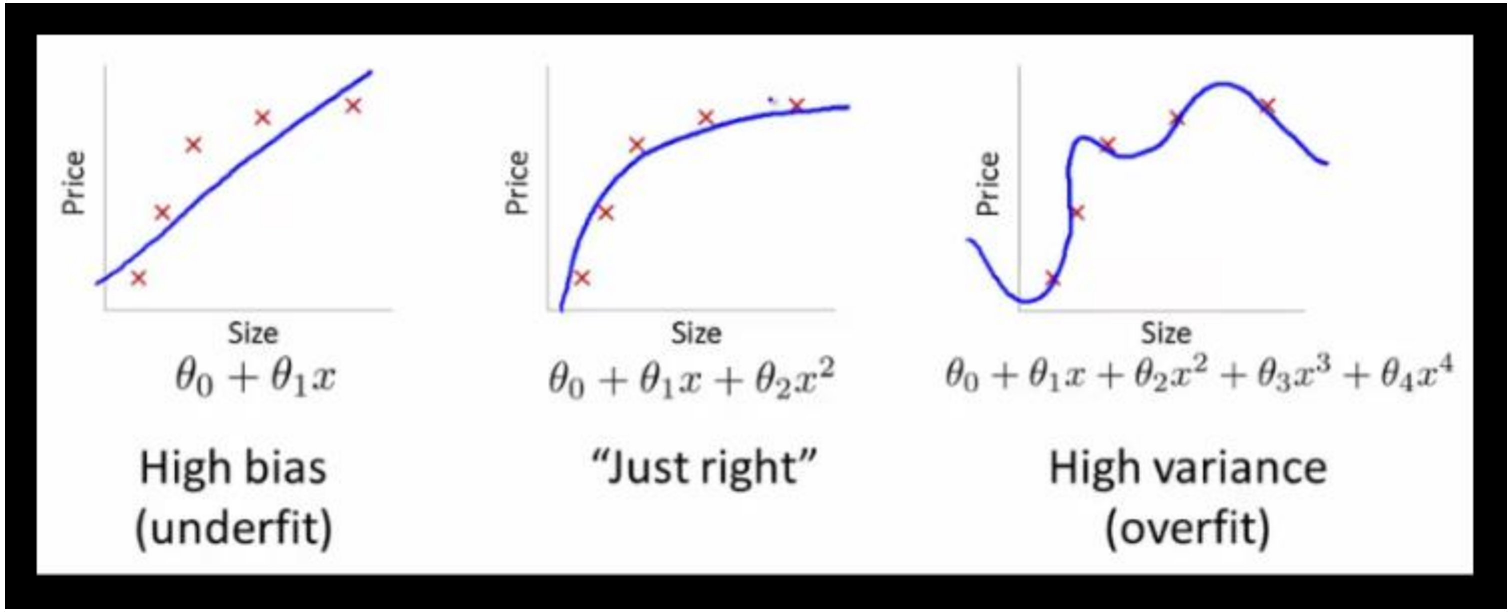
- Easy to calculate
- Quick to Converge
- Easy to interpret

Cons

- Low accuracy
- Does very bad with large feature set
- Doesn’t works with non linear data representations
- Won’t scale up very well

Model Interpretation

- Simple models are easy to interpret
- More the features in a model, its complexity increases and so does its interpretability
- Its pays to remember Occam’s Razor : Best Explanation Is the Simplest.



Use cases

- To assess risk in financial services or insurance domain
- To determine population in a given state
- To determine future sales
- Recommender systems
- To determine drug dosage in large patient cohorts
- Determine temperature and weather patterns
- To determine outcome of elections