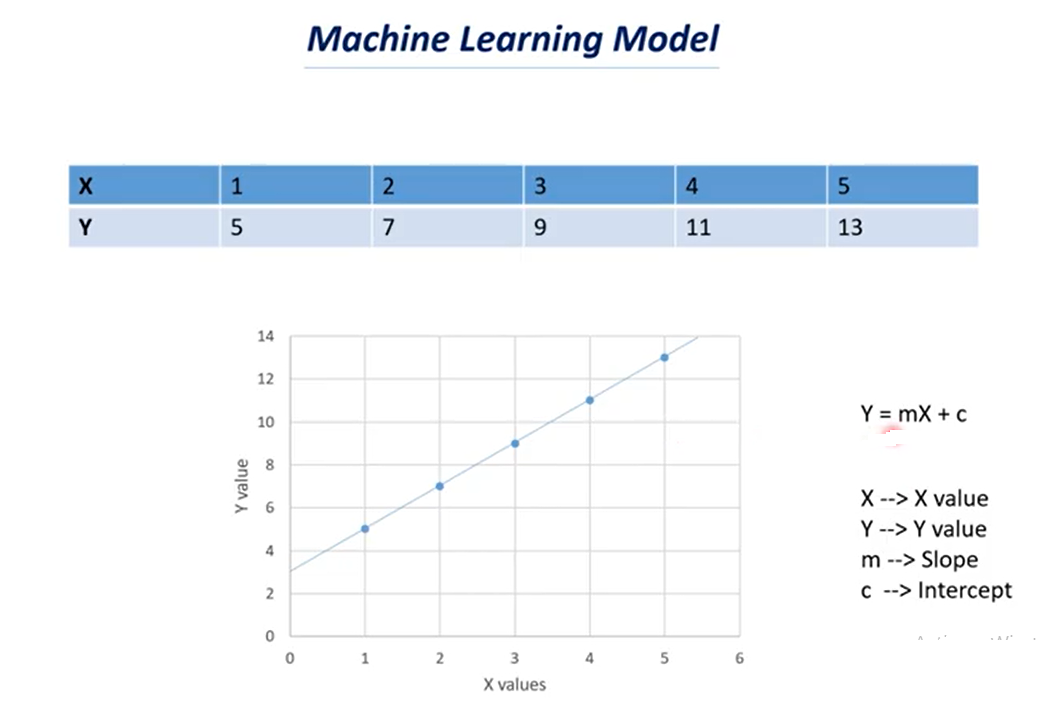
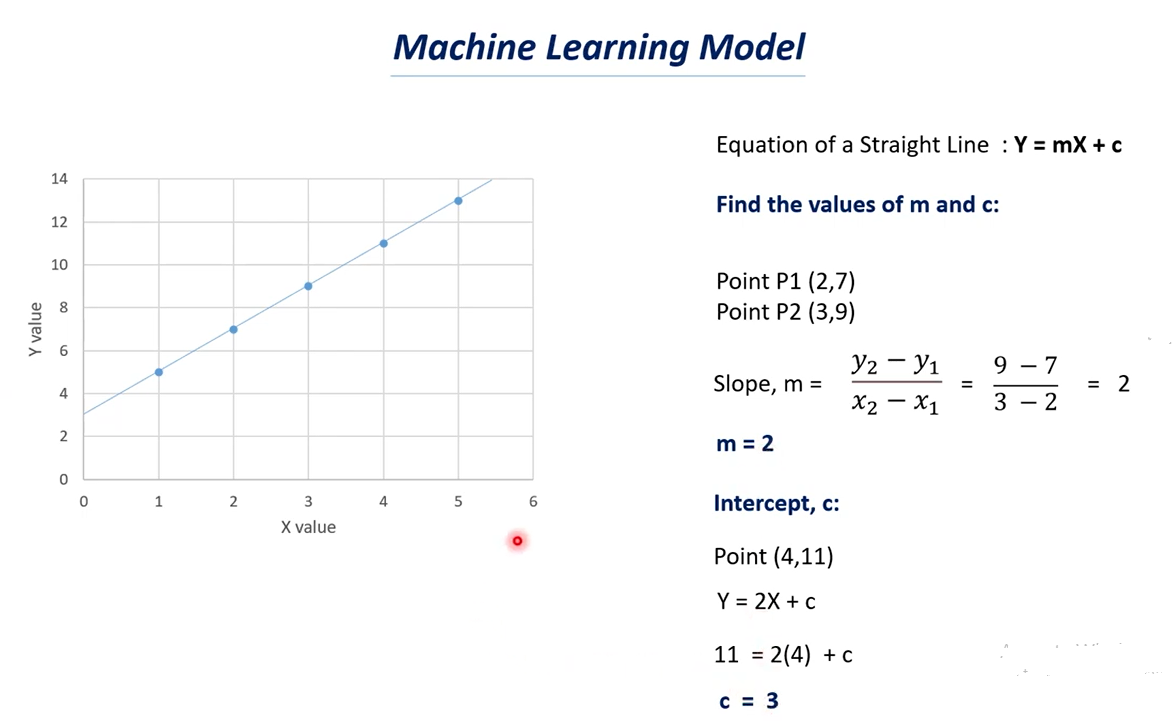
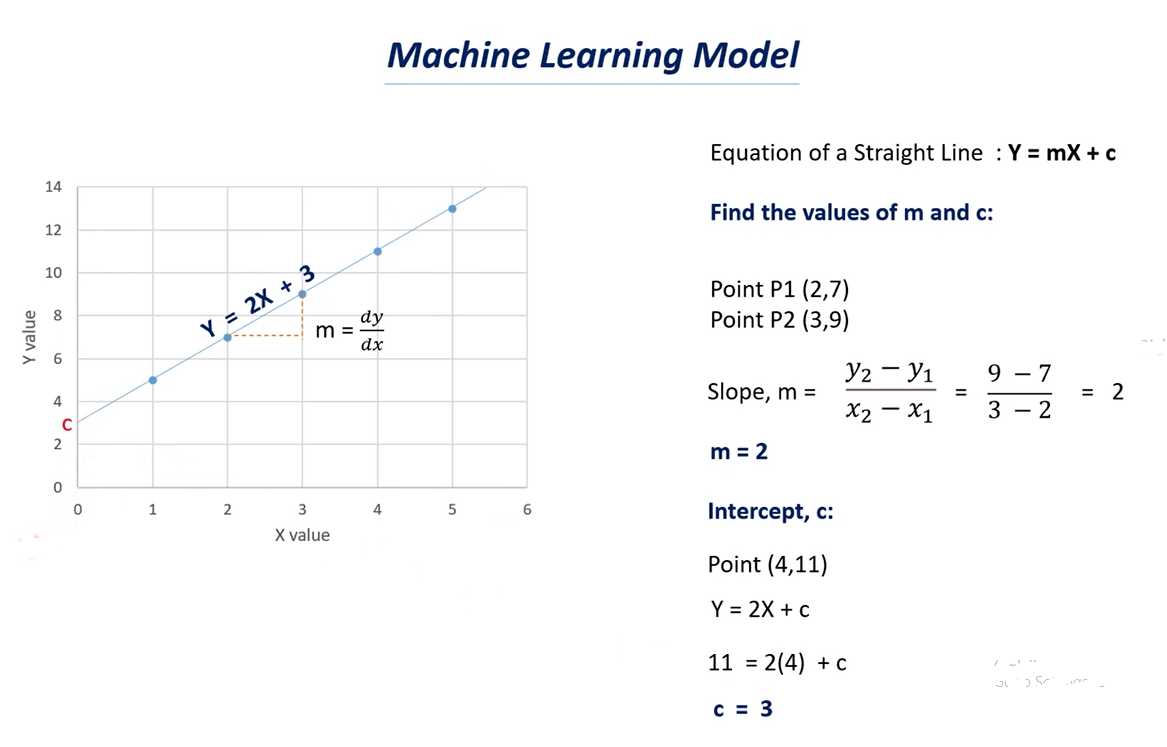
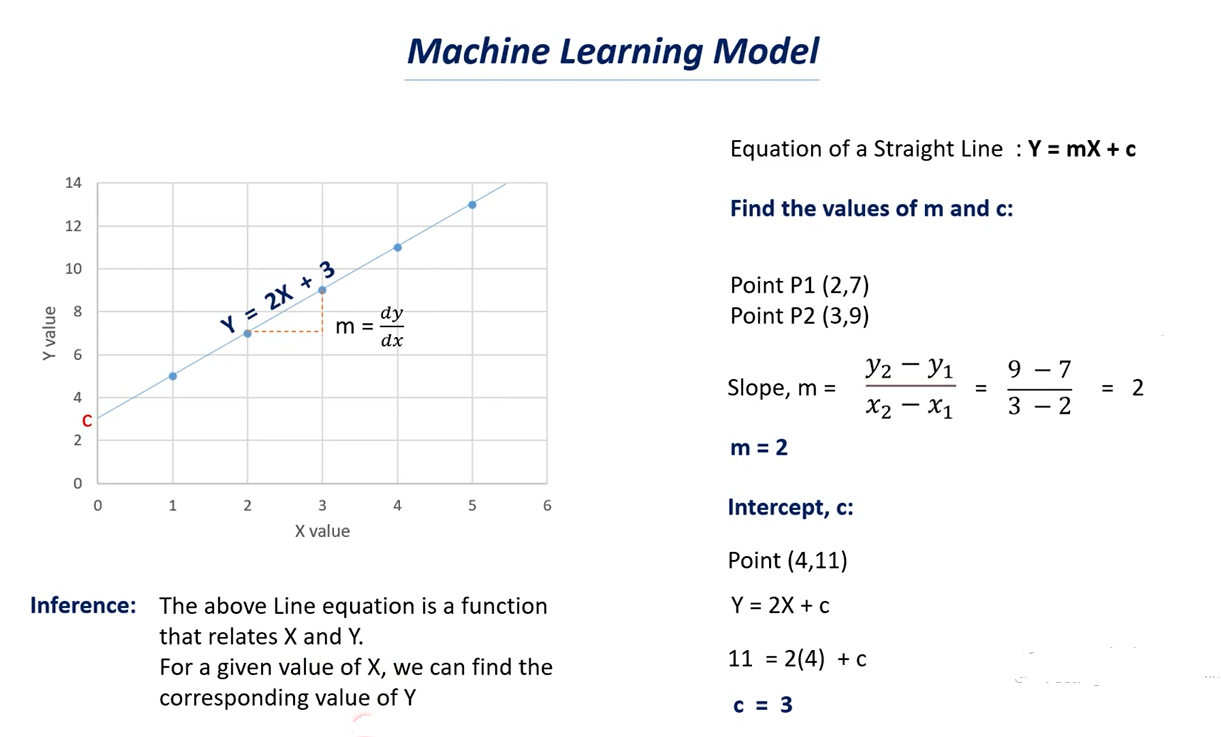
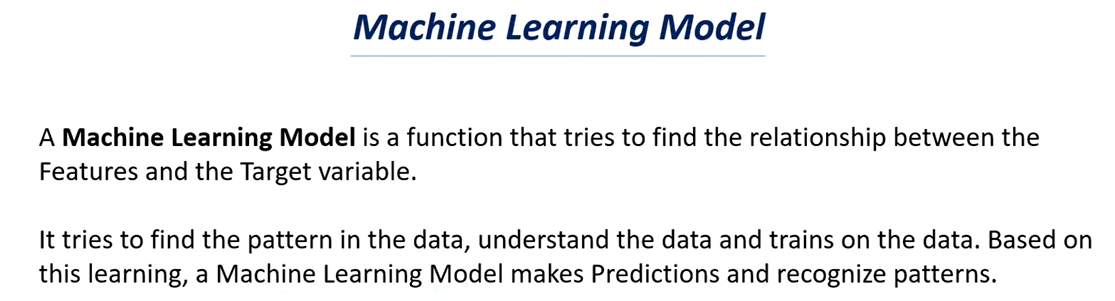
What is a Machine Learning Model? 

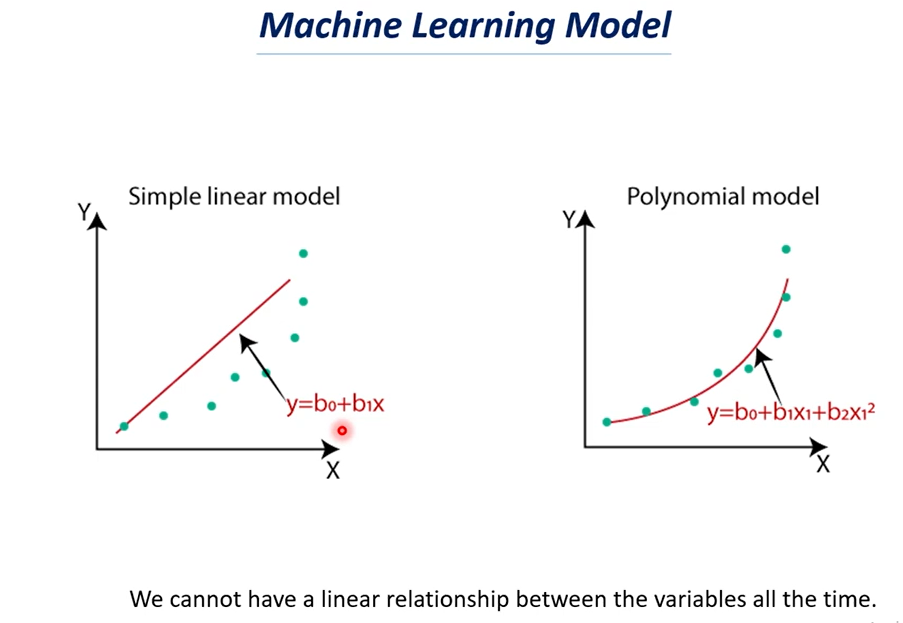


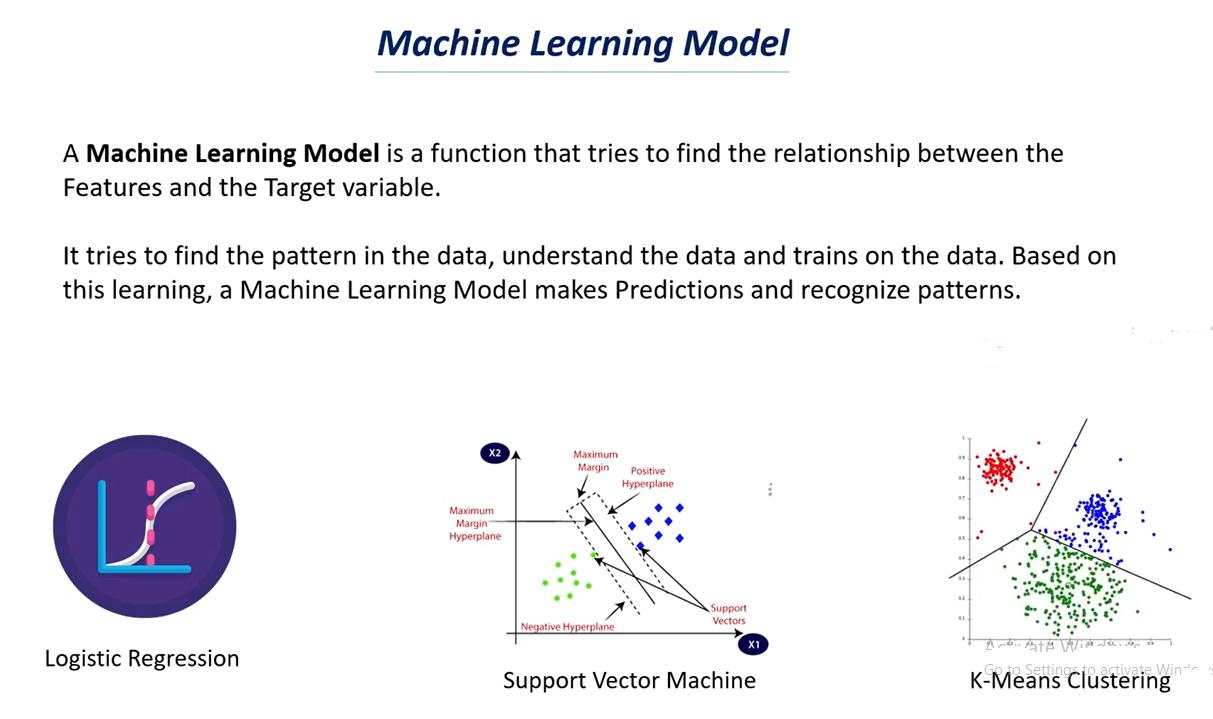












# Supervised Learning Models | Supervised Learning

# 

# 

# 

# 

# Unsupervised Learning Models | Unsupervised Learning

# 

# 

# 

# 

# 

# How to choose the right Machine Learning Model ,Model Selection ,Cross Validation

# 

# 

# 

# 

# 

# 

# 

# 

# Over fitting in Machine Learning:

# The Model over trained on the training data points.

# We have X values and Y Values and we have these many datapoints.

# We try to fit the curve to these data points.

# It is an optimal Model.

# It will try to find the common trends in it,

# If the model is over fitted we will get a curve something like this above.

# Here the curve try’s to join all the data points.

# In first case it is the regular curve.

# In second case the curve try’s to join all the data points.

# The over fit model try’s to model the data too well which means it tries to fit to all the data points, Here we can’t have a regular curve.

# Note: when we don’t get a regular curve we can’t make a good prediction out of it.

# In case of optimal model we have generalized curve or generalized mode.

# In case of over fit model we don’t have a generalized curve as we observed we have rises and dips in the curve

# If you see the data points in both the graphs are same but curves are different i.e over trained of the data.

# In case of over fit when a model learns the details and noise in the training dataset.

# Noises can be outliers and some data points that don’t make sense.

# Note: The good model try’s to ignore these noises that are presents in the dataset.

# In case of optimal it tries to find a generalized value.

# In case of over fit it tries to pick all the data points which can be noise.

# Sign that the model has over fit:

# The accuracy on the training data will be very high and the accuracy on the test data will be very low.

# That why we try to find the accuracy score for both training data and test data as well.

# If the accuracy of training data & test data is the same let’s say accuracy of the training data is 85% and the accuracy of the test data is 83% its almost similar.

# In this case we can say the model is optimal.

# Suppose the accuracy of the training data is 95% and the accuracy of the test data is 30% or 40% then we say that the model is over fitted.

# When the model tries to fit to all the data points it can’t give generalization and it can’t be reliable to make prediction out of it.

# 

# 

# 

# As we have discuss less data causes over fitting and if we have large dataset then the model ignores the outliers most of the time.

# It can understand the data better if it have more data so it is very important concept in machine learning and deep learning as well.

# More the data better the performance of the model is.

# If you have more data than the chances of over fitting is very less.

# Reduce the number of layers in the neural network as we discuss increase of layers in the neural network that causes over fitting so we need to reduce the number of layers.

# Early stopping: Early stopping is a technique that is used in machine learning what happened in machine learning is that in machine learning we iterate the data multiple times so the model tries to learn from the data multiple times this is called as iteration because it does the same thing again and again.

# So when we do early stopping technique the model tries to stop learning once it is over fitted.

# If over fit start it stop the training part this is called early.

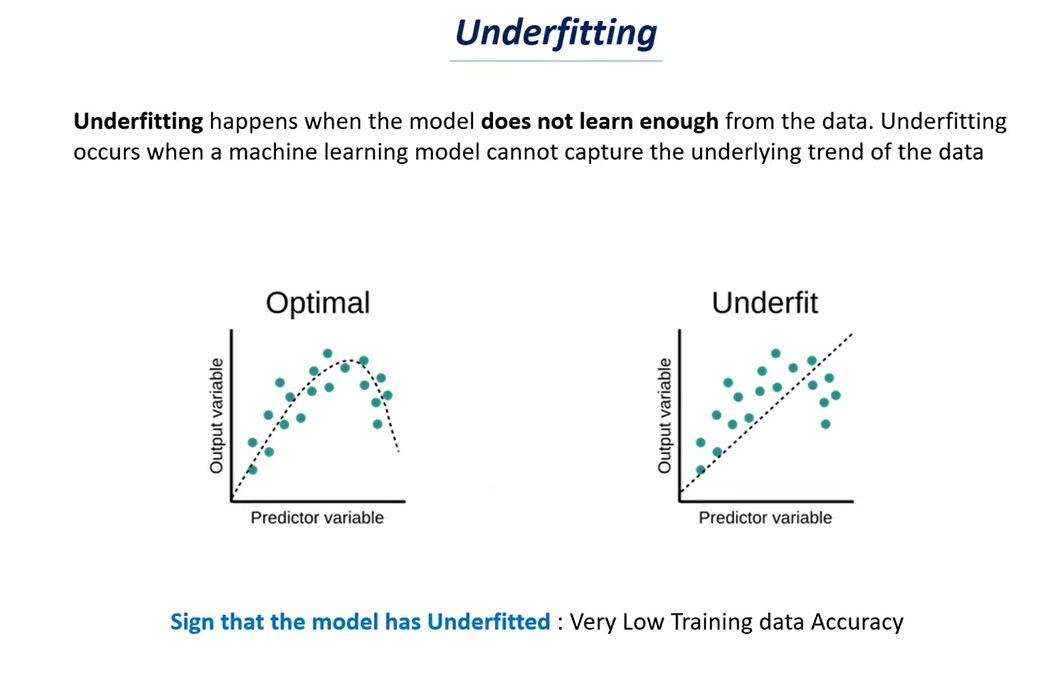
# Bias – Variance tradeoff it is the most important topic in machine learning .

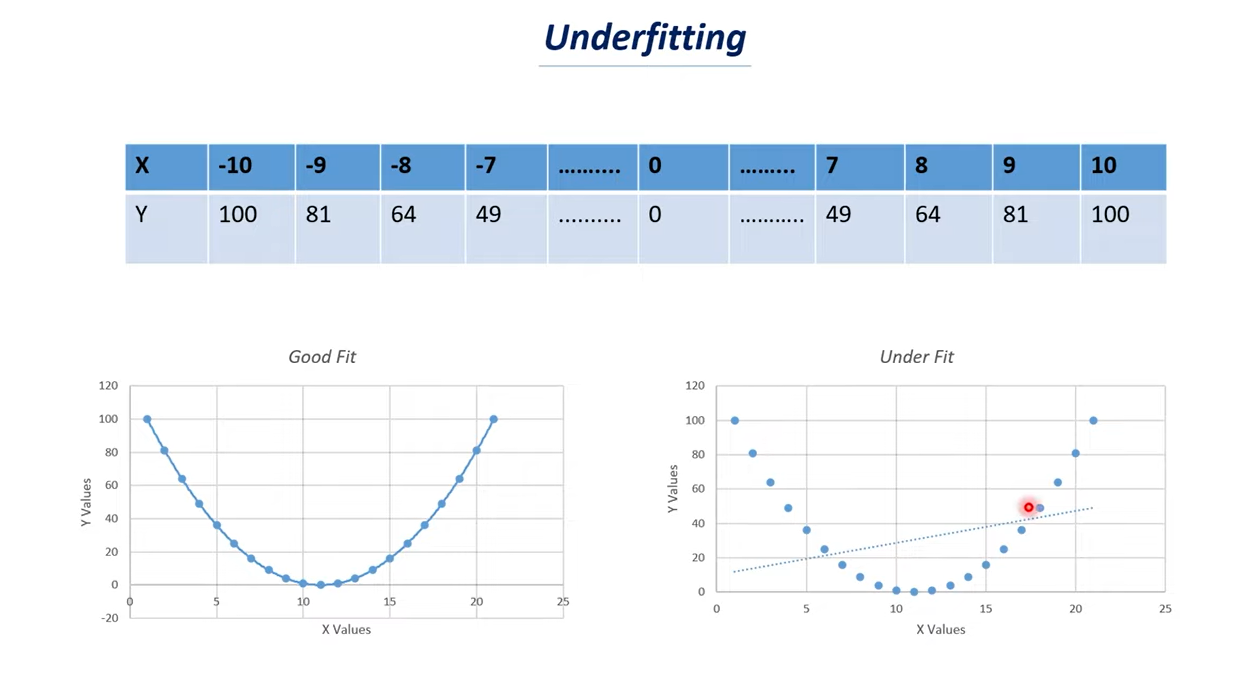
# Bias-variance tradeoff is used to find the optimal model for a dataset.

# This is used for a model optimization technique.

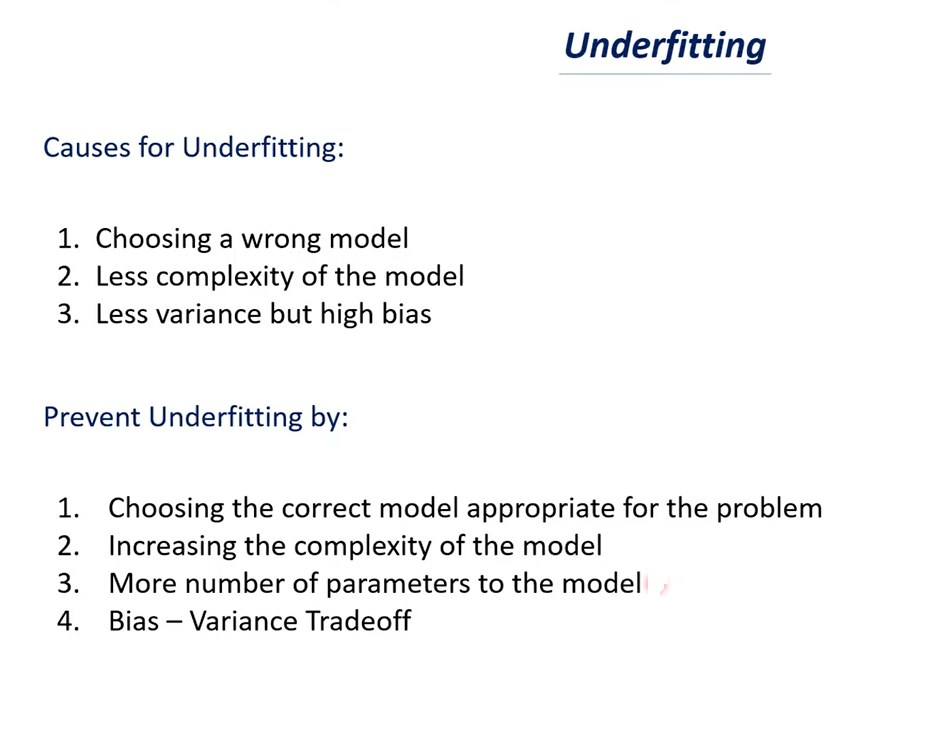
# Dropouts: if we have dropouts some neurons will be dropout randomly and some neurons will be turned on randomly and this stop the problem of over fitting.

# As we have less neuron the complexity of the model reduced.





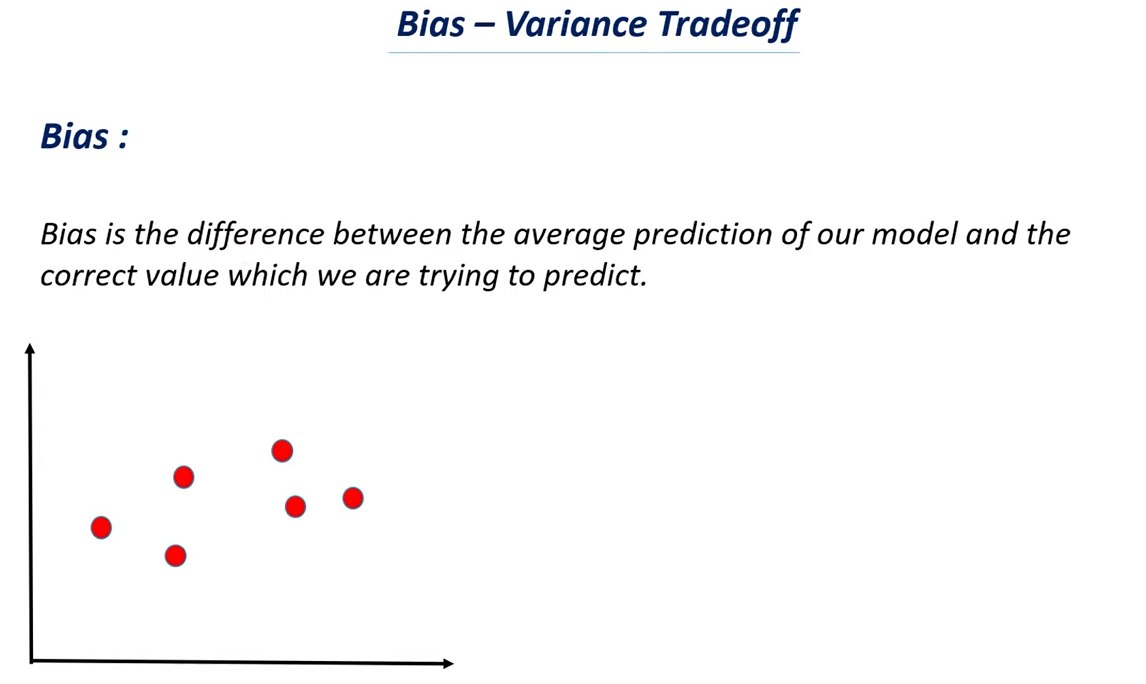
* y value depends on the value of x.
* If you see the graph it represents the parabola if you join all the data points.
* It tries to fit the data in a line but it is impossible to fit all these data points in a linear line.
* We can’t get a proper fit because it is a parabolic relationship but the model tries to fit in a simple line.
* Here the model fails to find the trends present in the dataset and it can’t fit all the data points in the curve.
* Whereas good fit tries to fit all the data points in a curve and it can find a trend.
* If we observed the values first decreases for a certain values of x and y increases if the value of x increases.



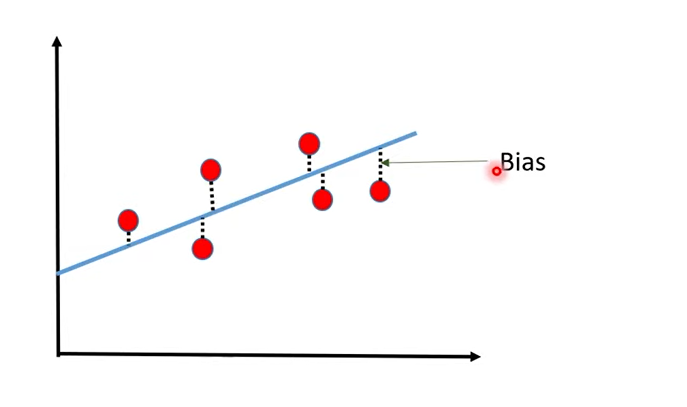
**Causes of Under fitting:**

* Choosing a **wrong model** - So in the previous case we have to choose a parabola model and if we have choose a linear model it can’t fit this particular data.
* So choosing a correct model is very important.
* If you choose a wrong model that will be one of the cause of under fitting.
* Having a less complex model is another example of under fitting.
* When you compare the parabola and linear model. Linear model is a very simple model whereas parabola is more complex model compare to parabola.
* Complexity of parabola is more comparing to a line.
* If you have very less complex model then there is a problem of under fitting.
* **Less Variance but high bias:**
* Bias variance Tradeoff is something that deals with to find the optimal model for our dataset.
* For this we need to understand the variance and bias.
* Bias is nothing but approximation that our model make for a target function so in machine learning we have features and targets.
* In machine learning model is something that finds the function that relates the features and target.
* Here features are nothing but x values and the target is nothing but y values.
* Bias is nothing but approximation function.
* Variance is how your model changes or how your function changes if you use different training data.
* If you have less variance and high bias values you will have the issue of under fitting.

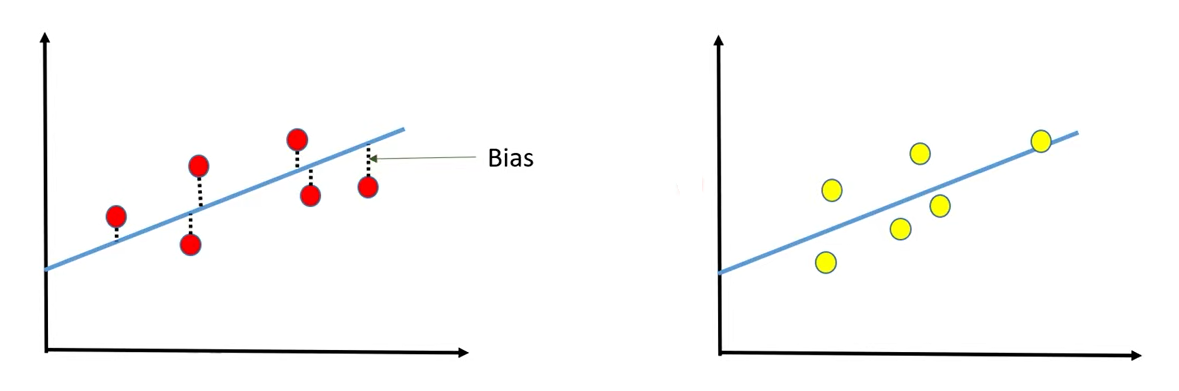
This are the few important issues faced in machine learning model.



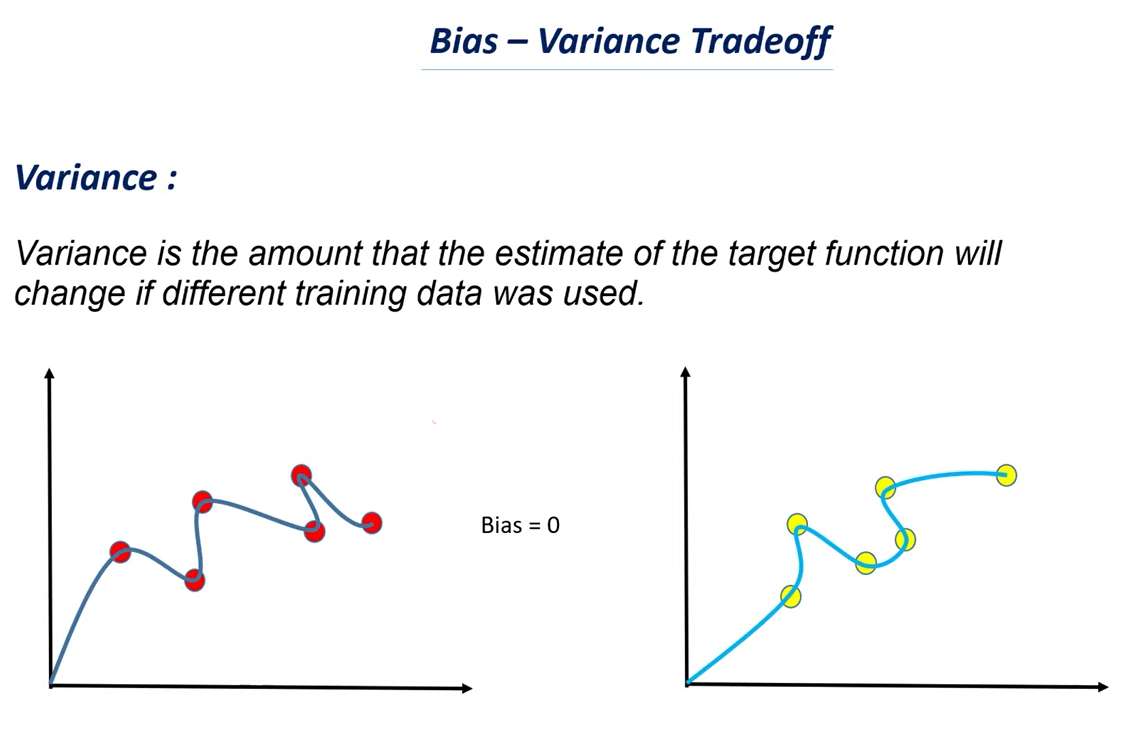
* Before going to Bias – Variance Tradeoff let’s understand Bias.
* This is the formal definition of Bias above.
* Let’s say we are making machine learning model to predict the salary of a person given on their experience.
* Here we are using experience data and we feed this to a machine learning model and let’s say the model has predicted that person earns a salary of about 28000 rupees per month.
* So in this case the value predicted by our model is 28000 rupees and the current value or actual value is 30000.
* So the difference between the two values is 2000 rupees and the difference between the predicted value and correct value is called as Bias.
* Let’s say that the first data point represent one year of experience which is on x axis and the corresponding value on y-axis is 20,000.
* Similarly we have several data points and we need to find the model in order to fit the particular data.
* If you look at the data we can say there is a linear relationship between them.
* And there is a positive correlation between them i.e. if one value increases the other value also increases i.e. the number of years of experience of a person increases there salary also increases.
* So when you have such kind of relationship between x-axis and y-axis or x- values and y-values we can fit the data by using normal straight line.
* This line signifies that if you take any particular point let’s say it represent 10 years of experience and if you take the corresponding values 80,000 rs.
* This is how we try to understand the data and the model which we are fitting this is a straight line which means there is a linear relationship between them when one value increases the other value also increases.
* Now it is impossible to draw a straight line which touches all these data points **so we try to build our model that close to all these data points.**
* **Note**: it is impossible for this particular model to touch all this data points so you can’t use straight line for that you need to use curve.
* Straight line can’t touch all the data points.

****

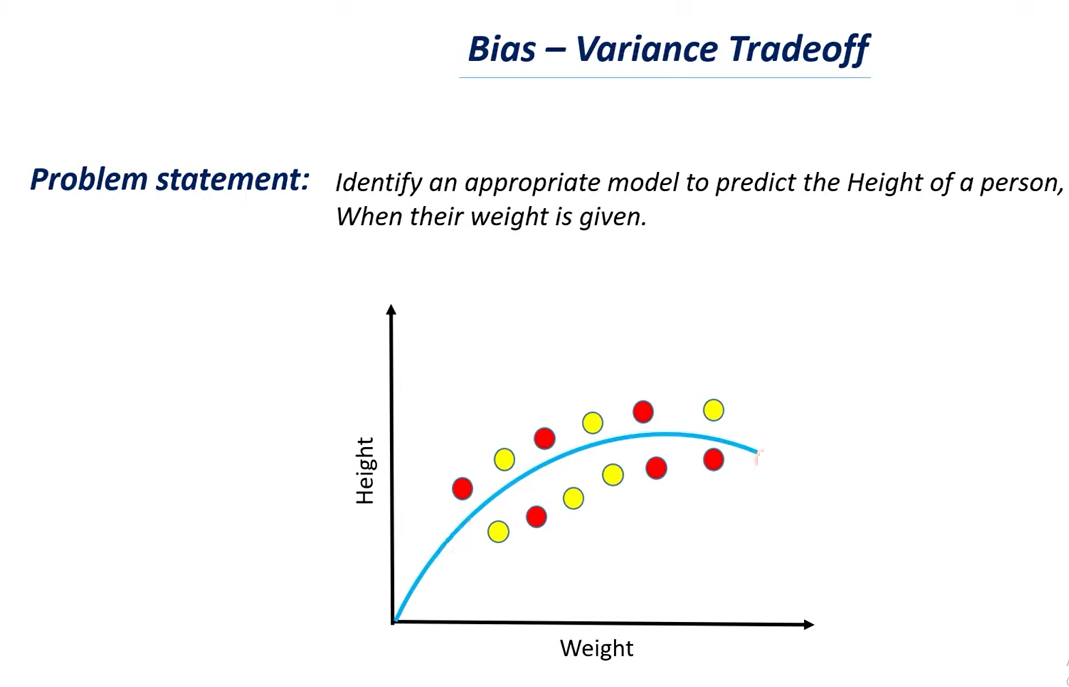
* The distance between there data points and the model is Bias.
* The blue color line is called the model which is an example of linear model or linear regression and we know the equation of the line is y=mx+c i.e. this equation of that particular line.
* We try to find the distance between data points and the model is called as Bias.



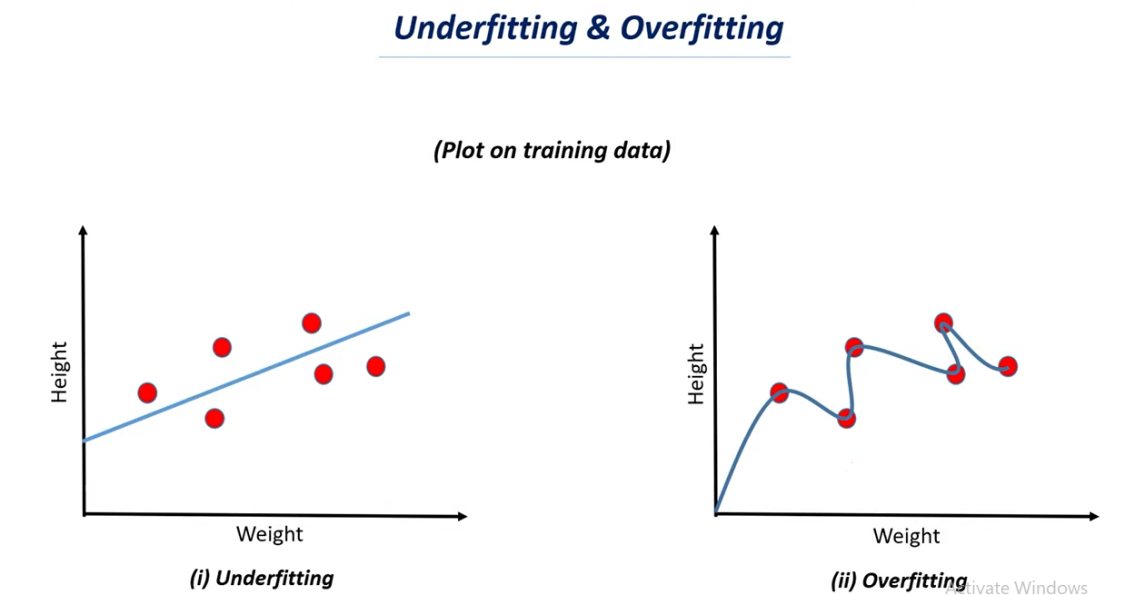
* In case we have some set of training data and now we are changing the training data.
* So this is one set of data and another set of data.
* Both these lines are different with respect to slope and intercept because we are using different kind of data but they are almost similar to each other.
* There is huge bias in both cases.
* So there is bias the distance between data points and the model in both the cases.
* But there is not much variation to the model because these two model are very similar to each other so there is a very less variation between them we called these variation as variance.
* In this particular case as we can see very high Bias and very low variance that is the influence which we are getting.

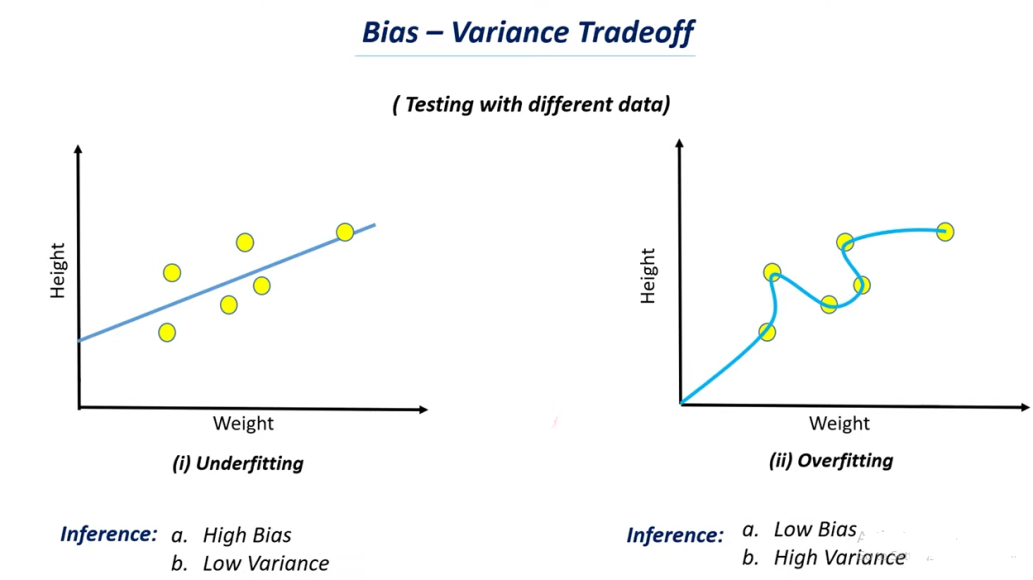
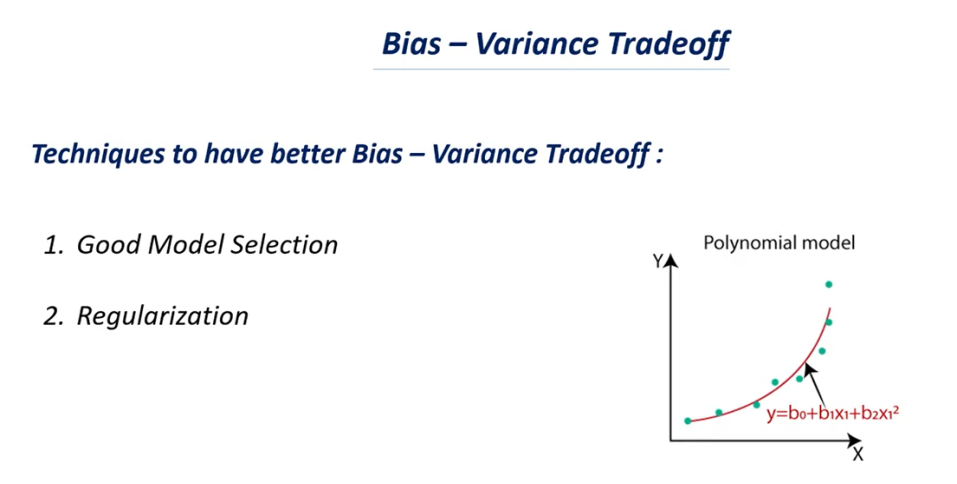


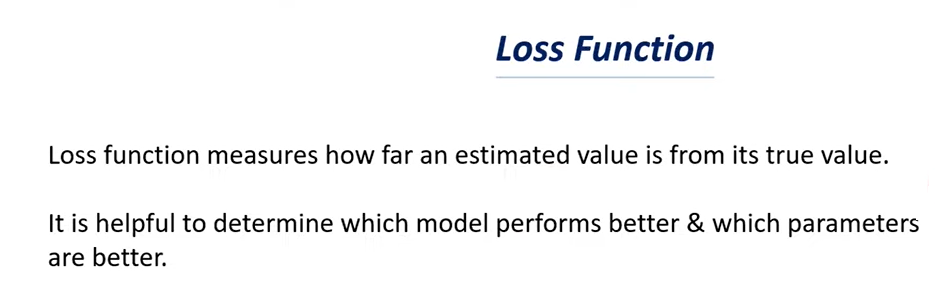
* If the change is huge we called this particular case as high variance.
* In this case there is a zero bias as it touches all the data points and there is no distance between the data points and the model.
* When you use similar kind of model as straight line most of the cases you will have a very high bias but there wouldn’t be any much variance in it.
* When we use more complex model the bias will be zero but there will be huge variance in it.



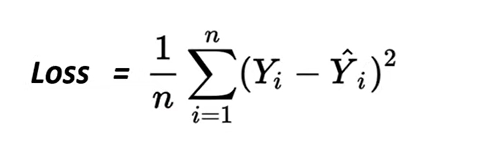
* We have a problem to identify an appropriate model to predict the height of a person when there weight is given.
* We need to analyzes this and we need to find weather we need a simple linear model or we need a more complex curve.
* Let’s say we have this data, weight in the x-axis and height in the y-axis and these are all the data points we have.
* If you want to fit optimized curve to this particular data we will get a curve as above.



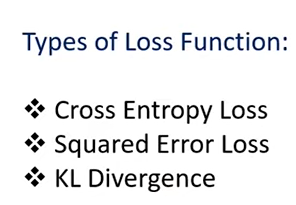
* ****
* ****

****

* **Here we are taking two values one is the actual value or the true value and other one is the estimated value.**
* **Let’s say we are using machine learning model to predict the average blood sugar level in a person.**
* **Here the estimated value or the predicted value is 140 milligram and the true value is 160 milligram so there is a difference of 20 milligram.**
* **This error of 20 is called as lost.**
* **We are going to find a function or a relation that calculate this particular loss value.**
* **So the main purpose is to find out how much difference is there between the estimated value and the true value.**
* **You can consider this as error made by our model.**
* **Loss function helpful to determine which model performs better and which parameters are better.**
* **Let’s say that there are two models SVM and random forest.**
* **In machine learning we have several kind of models in this particular case let’s consider this two model’s SVM and random forest.**
* **We want to find which model performs well in this particular case.**
* **Here we trained both of these models with this particular dataset and we try to find the loss value of each of these model.**
* **Which model has the lower loss value that model will make better prediction.**

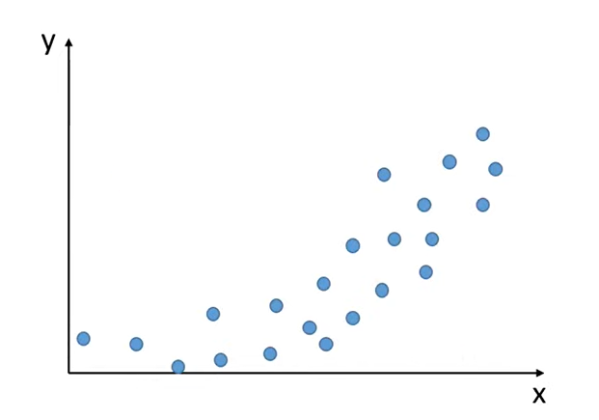
****

* **Low loss value means there is no distance between the estimated value and true value or there is no difference between the estimated value and true value.**
* **So the loss function be less for a model to perform better.**
* **Different model will have different parameter.**
* **Let’s say that we have hundred data points we need to predict the values for all this 100 data points and each of these data points will have some error.**
* **So all the predicted values will have error we need to find all those error and divided by n.**
* **In this case n is 100 i.e the number of data points that we are taking.**
* **So this is the formula for loss function.**

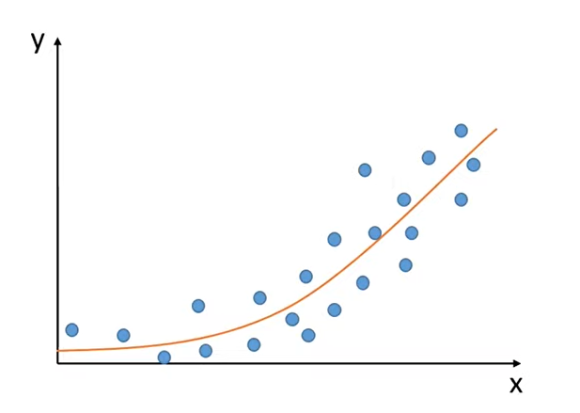
****

****

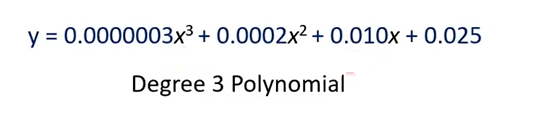
Let’s say that we have x axis and y axis we are taking some data.



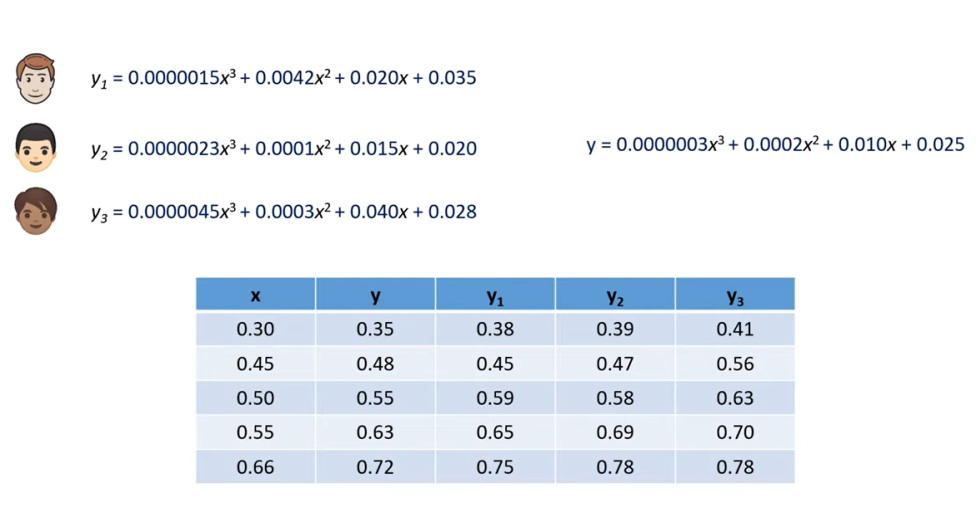
* This is the dataset we have and we have plotted the values for x axis and y axis
* We need to find the model or curve which can fit this dataset.
* It is not possible to draw a line or curve which can pass through all the data points.
* We can draw the curve like this that can almost pass through all the data points.



* We can draw the curve that can almost pass through all the data points.
* So this is the trend we are seeing in this particular line.
* So here the value doesn’t increase much.
* When x value increases y value doesn’t increase much in first case it is almost the same and after certain points the y value increases as the x value increases.
* So this is the trend we are getting for this particular dataset.



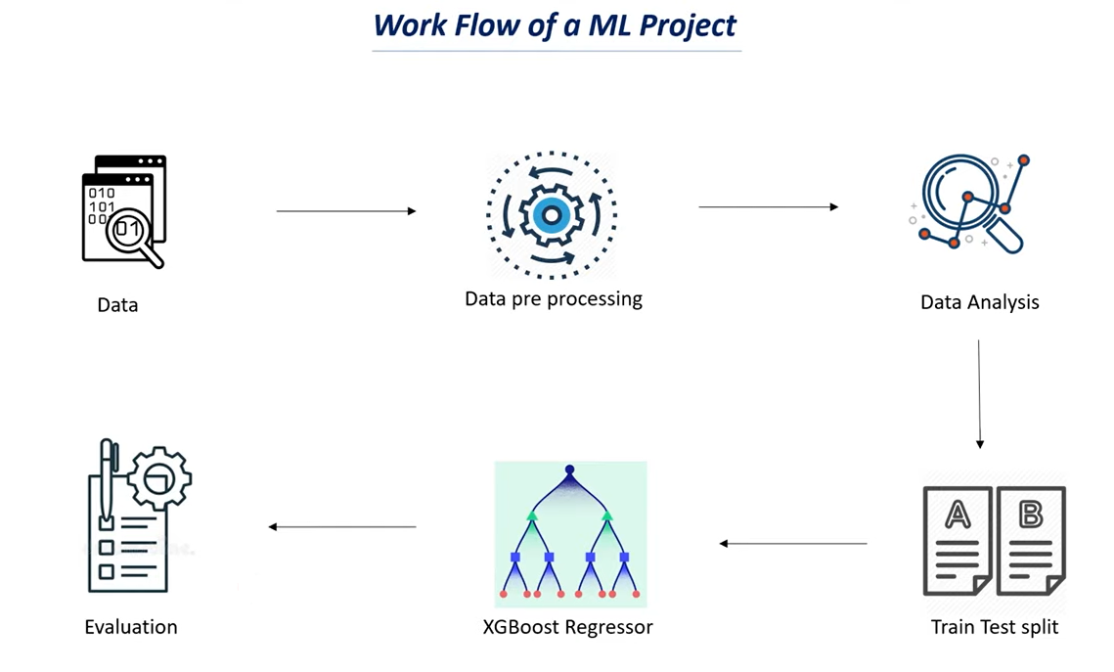
* This is how the equation of this particular line may look like.
* This is not the exact equation but I am giving an example of how the equation will be for a line like this or a curve like this.
* This is an example for degree 3 polynomial as we can see we have x3, x2 & x.
* Degree 4 polynomial will have values as x4,x3,x2,x and a constant.
* This particular values are coefficients i.e the values before x3,x2,x are the coefficient.
* We can also call this as parameters of the model.
* This is the model we have found for this particular dataset.





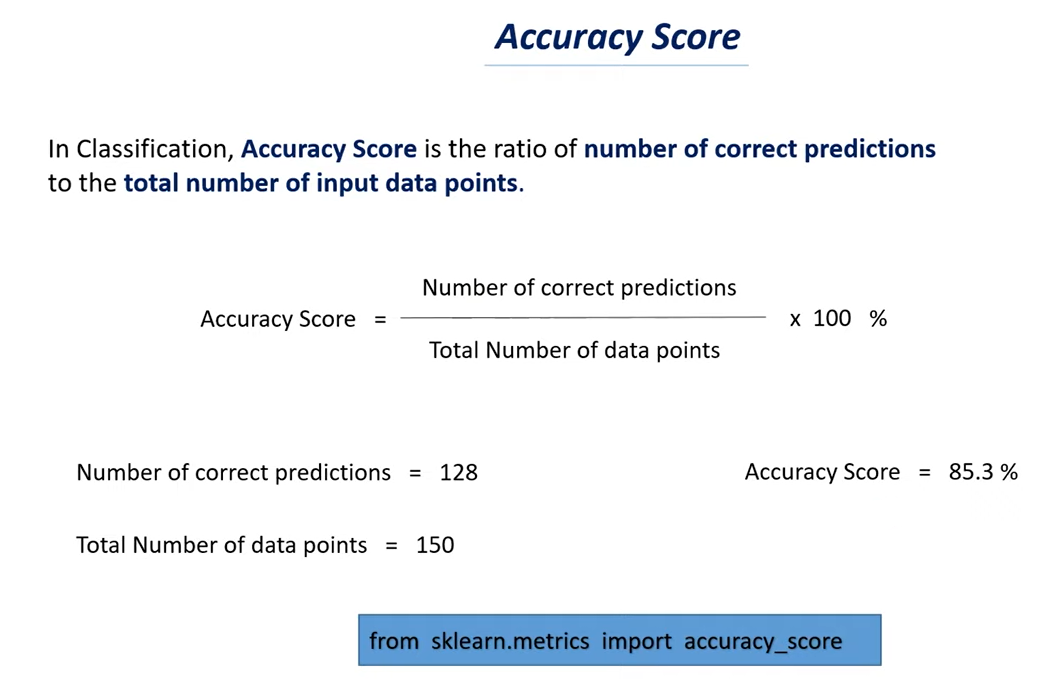
* As we know that low loss value means high accuracy.
* So we want our loss value to be as close to zero.
* If the loss value is zero that means that the true value and the predicted value are the same.
* Which means our model have 100% of accuracy.

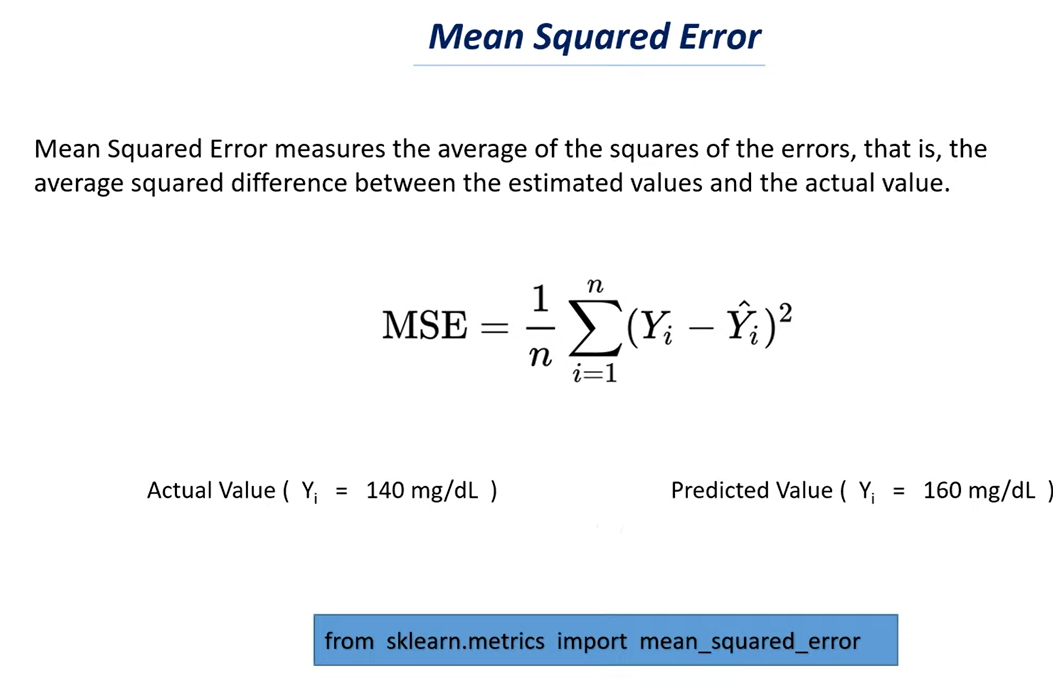




* Steps -1 finding appropriate data for our model.
* Step -2 we need to do some pre-processing on this data.
* The raw data can’t be feed to machine learning model.
* So we need to process this data in order to make it suitable to feed to a machine learning model.
* Step-3 We often does data analysis to find some insights from this data. So that we can use features better for a machine learning model.
* We split our dataset into training and testing.
* Step 4 : model selection based on our problem statement.
* Step5 – this is the evaluation matrix.
* This evaluation matrix is use to find performance of our model.
* It is use to check how well our model is performing how many correct prediction our model is making.
* You can’t just train a model use to do prediction.
* First we need to understand what is the accuracy of the model and how well our model is performing .
* Only if the accuracy is good we can use that particular model for our prediction so this step is called as evaluation.
* We know that there are two main types of supervised learning.



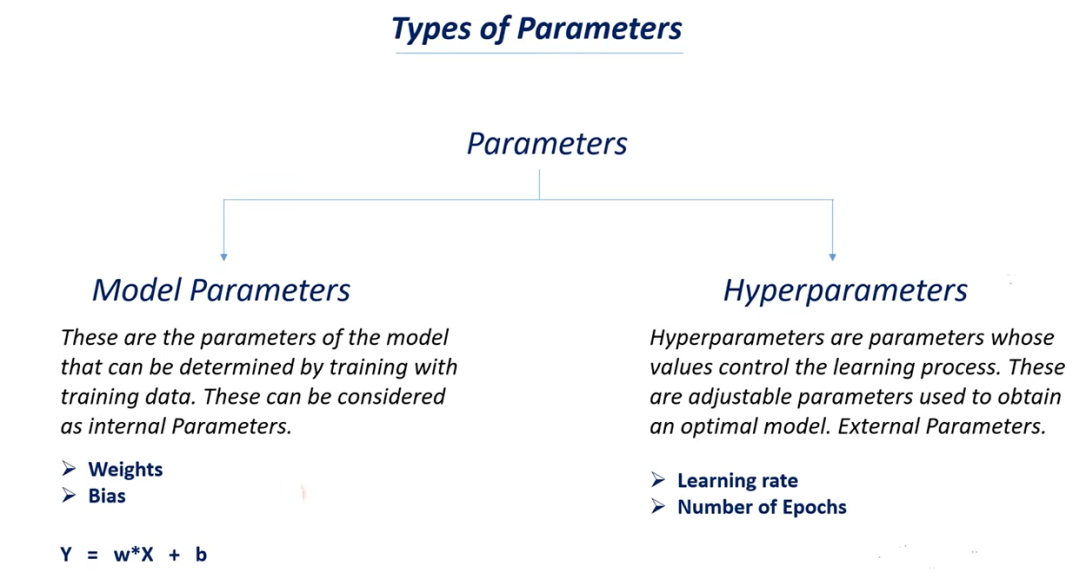




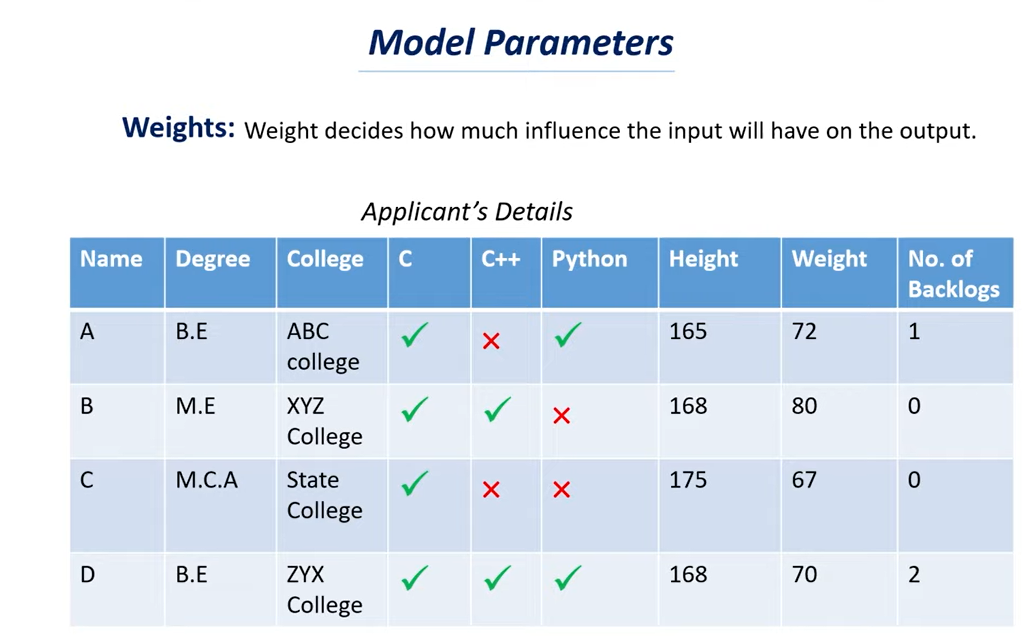
**Model Parameter and Hyperparameters :**

Two main important parameter when it comes to  **training a machine learning model.**

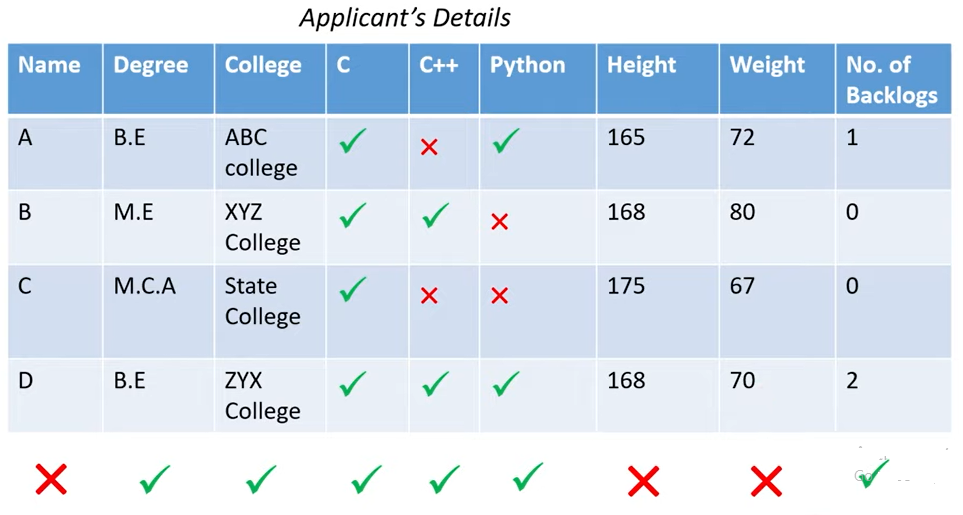
1. **Model parameter**
2. **Hyperparameters**



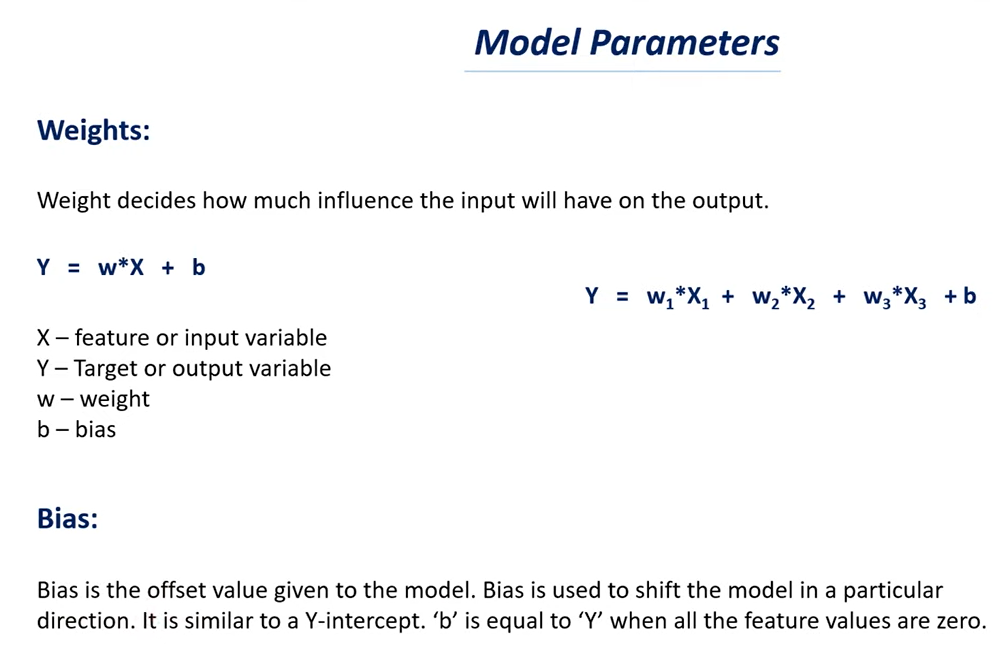
* What is meant by weight of a model?
* Weight is a value.
* Weight decides how much influence the input will have on the output.
* In a dataset we may have several features and we can have target variable as well.

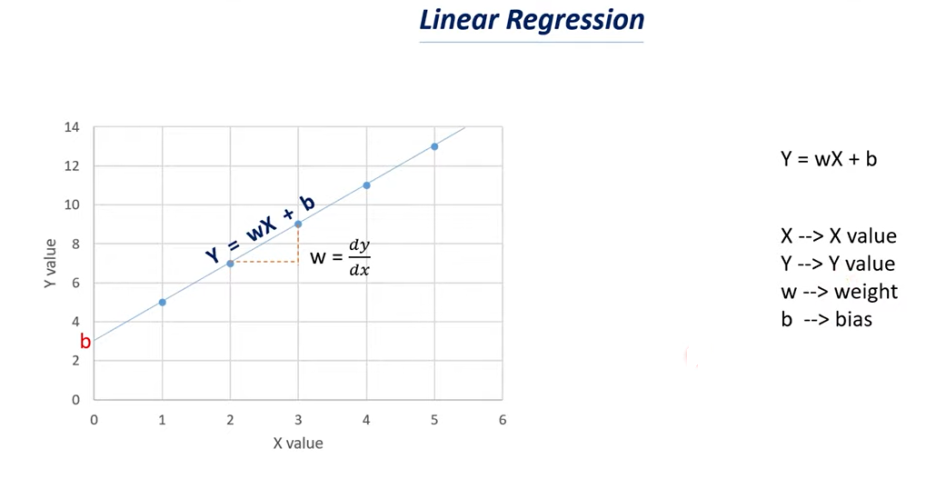


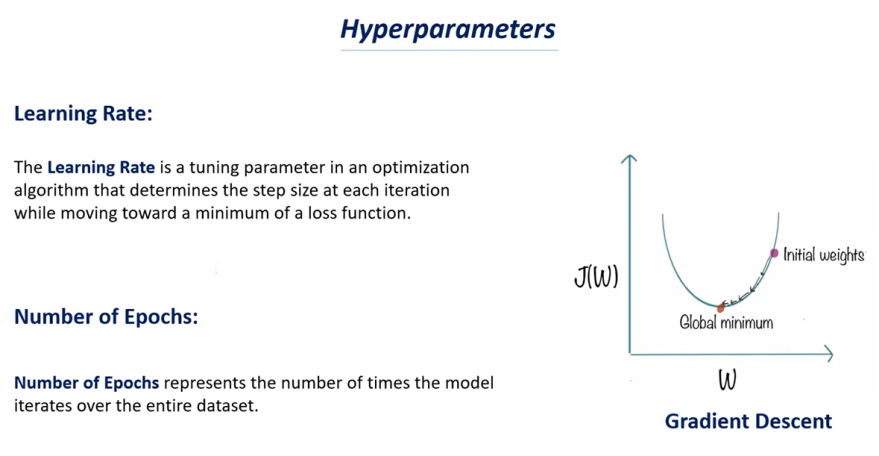
* Let’s consider this particular dataset.
* In this we have data regarding applicant for a job.
* Let’s say we are building machine learning system that can predict whether a person can get a job or not based on some parameters i.e. based on features.
* Here we have totally 9 features in this particular dataset.
* Each column represents 1 feature and the target variable will be whether a person will get a job or not.
* Now we need to find whether a person will get a job or not by analyzing this particular data.
* All this columns are not important for us and some columns are important where as some are not.



* Name of the person is nothing to do whether they will get a job or not.
* Person can have any name and the name doesn’t influence whether a person get a job or not.
* Name column wouldn’t be that much important for us.
* Degree/College/Skills /backlogs are important
* Height and weight is not important for us.
* We need to give numerical values to emphasis the importance of this column and this is where weight will help us.
* For name/height/weight we can give in this case weight as 0.
* Different columns will have different weight values and it is not the common value.
* If there are about 9 features you will have 9 different weight values and this weight values is not constant for different columns because we know that each column has separate importance.







**Gradient Descent in Machine Learning:**

* Each model has their own weight and bias associated to it.
* If you take a simple line let’s consider a simple linear regression model.
* In case of linear regression we just use simple line as our model and we know that the equation of the line is y=mX+C.
* Instead of writing y=mX+C we can write y=wX+b where w represents the weight or you can also called as slope and b represent as bias or we can also called as intercept.
* Here x is the input features and y is the output features.
* This is a simple linear regression model to find salary of a person based on work experience.
* X- features or input variable

Y – Target or output variable

w – Weight

b – Bias

* When it is a line you can consider weight as slope and bias as intercept.
* Different models have their own weight and bias associated to it.
* Y=w1\*X1+w2\*X2+w3\*X3+b
* If you see this particular equation this is the equation of a multiple linear regression.
* In this case instead of a line we will be using plan in order to find this relationship between y= w1\*X1+w2\*X2+w3\*X3+b
* Here x1 represent the number of year of experience of a person and x2 represent the number of programming language or skills that the person knows x3 represents the college that person studied.
* Y represents may be the person would get a job or not.
* All these x represent the input features whereas y represent the target we want to find out.
* This equation is a multiple linear regression as it has multiple features in it.
* Where as in single linear regression we will have one feature and one target variable.
* In this case we have 3 weights w1,w2,w3 and we have only 1 bias.
* If the weight value and the bias value are not proper then you would not get a proper prediction.
* You may get a wrong prediction.
* Weight and bias are depending upon the dataset we have.
* If we fit different dataset to a particular model that model will have different weight depending upon the dataset.
* So in this case what we will do is , this is not a single step process and we will try to repeat this process iteratively so that we can get a better weight value and bias value.

This particular process should be repeated.

* When we try to fit the data to a model and we will repeat this process again and again until we reach the best value of weight and bias.
* In each step we will try to change the value of weight and bias to get a better model.
* Learning rate represent the amount of change that we are imparting to the weight and bias or the model parameter that amount of change is called as learning rate.

**Number of Epochs:**

* As we know that it is not a single step process and we will do it in a multiple steps.
* Number of epochs represents the number of times we want our model to go through the data again.
* Let’s say that we want our model to go through the data 100 times in that case the number of epochs will be 100.

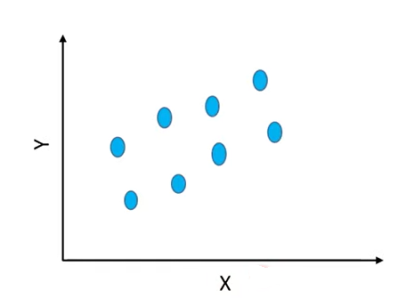
**Loss Function:**

* The value predicted by our model is very close to the true value.
* If you don’t have the best parameters such as weight and bias i.e. if the value of weight and bias are not proper you are most probably going to get a bad prediction and the difference between the true value and the predicted value will be huge.
* This loss function will tell us whether the model is performing well or not and whether the parameters are suitable or not.

**Gradient descent**

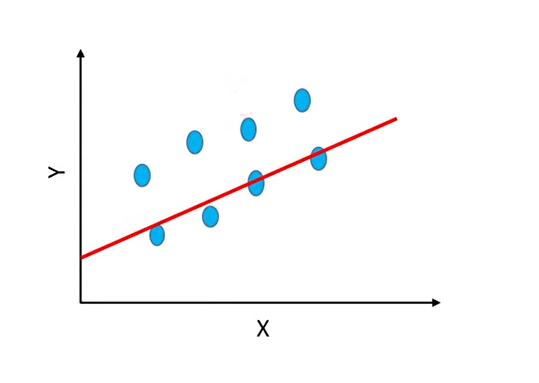
* Gradient descent is one of the model optimization techniques.
* Optimization refers to determining best parameter for a model such that the loss function of the model decreases as a result of which the model can predict more accurately.
* In order to make a better model we have to find a best possible parameter.
* So a best possible parameter is one which has minimum loss function.
* Minimum loss function means accurate prediction.
* So if the model has minimum loss function that means it has better parameters i.e. better weight and bias value.
* Finding this better weight and bias value is called as optimization.
* So we are going to optimize our model to get a best parameter or best values for weight and bias.
* If you have best parameter values your loss function will be very minimum in that case your model will make accurate prediction.
* For this we use algorithm called as gradient descent.

Example:

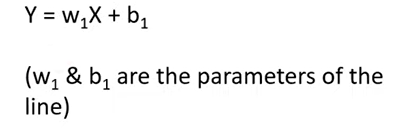


* - Let’s say that we have some dataset x represent the number of rooms in that particular house and y represents the price of the house.
* Here there is direct relationship between number of rooms and the price of the house.
* Higher the number of rooms represents higher the price.
* So there is a linear relationship between them.
* If x increases then y also increases.
* For this particular purpose we can use simple linear regression for this particular dataset.

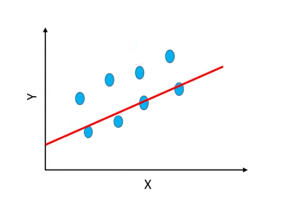
**Let’s try to fit this data to different lines**



* Let’s say that we want to use this particular line in order to fit this data points to this line.
* This is a simple linear regression model and this line will have a particular slope value and intercept value associated to it.
* So we call the slope as weight and intercept as bias.
* If you want to write the equation of this particular line is

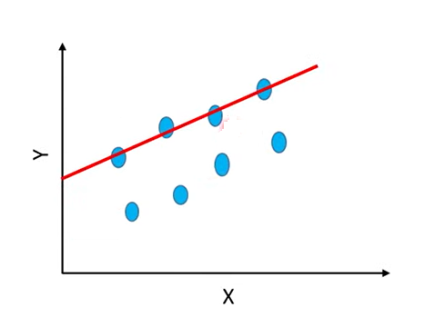


* It is similar to y=mx+c where m represents the slope and c represents intercept.

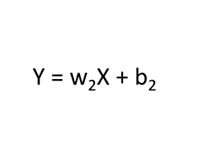


* If you see this particular line there is a large distance between data points and the line i.e model.
* Below 4 points are closer to the line and the above set of 4 points are not closer to the line.
* So can’t say that it is an accurate model because all the data points are not closer to the line.
* So we will try to find the best line for this particular dataset.
* This called best fit .
* If you get a best fit your data points will be closer to the line and this are called the best parameters.
* If you have the best parameter you will have minimum loss function that should be our goal to have minimum loss function that we can make better prediction.

Now we will try to change the values of weight and bias so that we can get a line which can fit to all the data points.

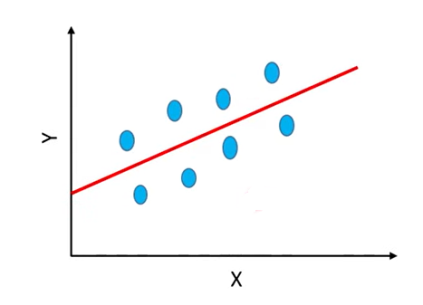


* We have line like this.
* In this case this particular line is closer to the above 4 data points but the distance between below 4 points is very huge.
* This line doesn’t fit with below 4 points.
* Let’s say that this particular line have an equation of

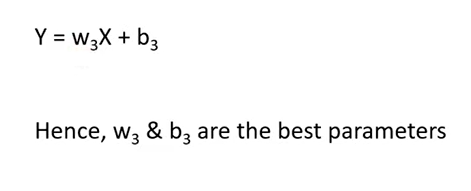


* We can say both of these lines don’t fit to the data well.

Let’s consider another line



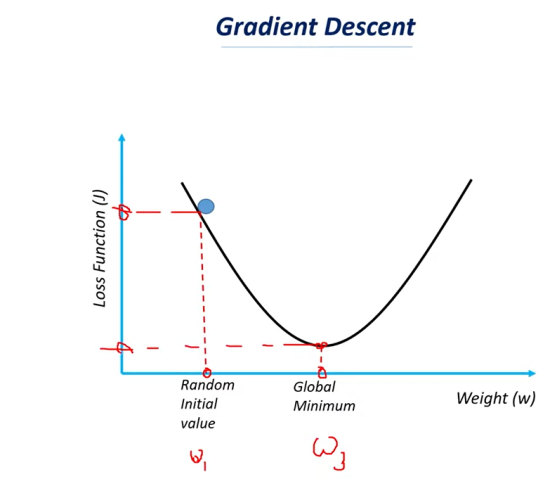
* This line is not very close to the data points but it is almost close to all the data points.
* In previous case only 4 data points are close to first and second line but the 3 rd line is closer to all the data points.
* It is not as close as this particular lines but it is close to all this data points.
* The equation for this line say as



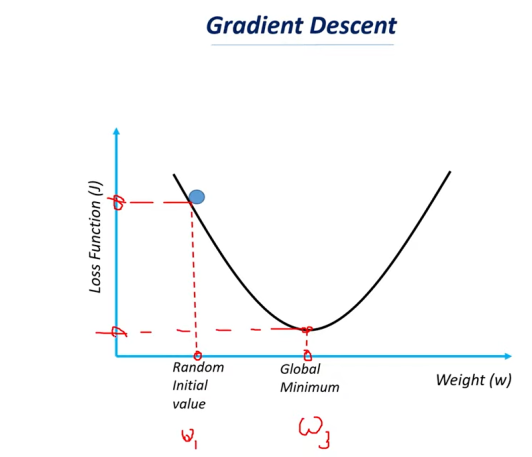
* So we can say that this particular line has the best parameter as w3 & b3.
* We can call this as best fit for our model.
* We have optimized our model as manually in this case.

**Model optimization:**

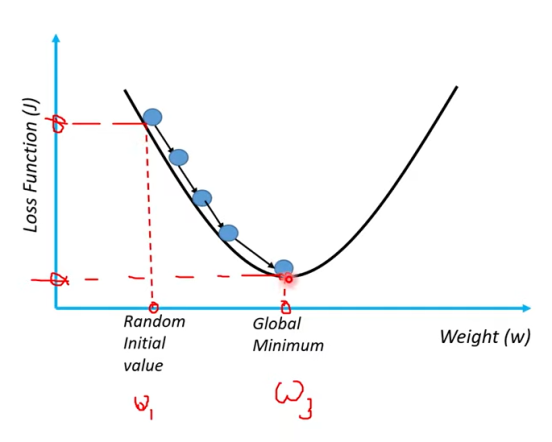
Optimization refers to determining best parameters for a model such that the loss functions of the model decreases, as a result of which the model can predict more accurately.



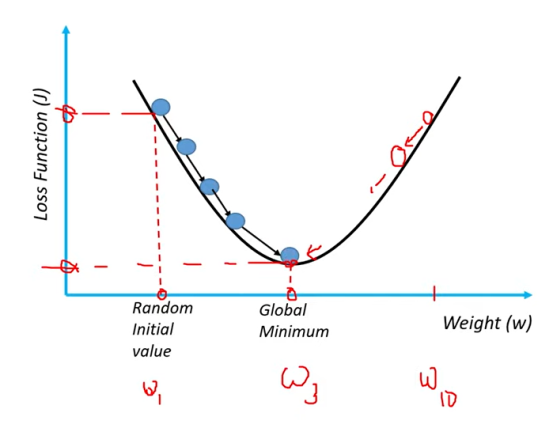
* Machine learning model can’t do it easily for this purpose we are going to use a technique called gradient descent.
* Lets say that we have a curve
* On x-axis weight of the model and on y axis loss function value .
* As we know that loss function is difference of true value and predicted value.
* If the loss function is high that means the model is not making good prediction (or) in other words the difference between the true value and the estimated value is very huge.
* If the loss function is very minimum if it is lower side of the y axis then the model is making better prediction because the difference between the true value and the predicted value is very small.
* On axis we have weight
* If you take different weight values in each case if you try to find the loss function you will get a ’ U’ shape curve.
* Let’s say that we are taking random initial value for the weight so in machine learning what we will do is whenever we are trying to fit the data we will assign random values to weight & bias.

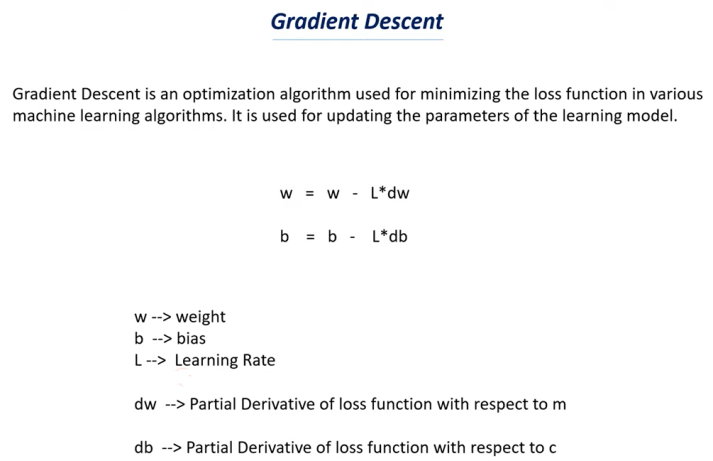


* If you take w3 you will have minimum loss function.
* If you have the weight value at global minimum you will have minimum loss function and your model will be in better prediction.
* We will reiteratively change the values of weight & biases so that we can find which weight & biases has the least loss function and that model will be consider as better model.



* We will try to change this weight values little by little.
* We gradually change the weight & bias values till we reach the global minimum.





In machine learning we use instead of m as w and c as b.

