**Tak 2 : My notes**

1.1

Machine learning is a subset of AI, Deep Learning is a subset of Machine Learning.

Ai: is a branch of Computer Science that is concerned with building smart & intelligent Machines.

Machine Learning: is a technique to implement AI that can learn from DATA by themselves without being explicitly programmed.

Deep Learning: is a subfield of ML that uses Artificial Neural Networks to learn from the data.

1.2

Machine Learning is a technique to implement AI that can learn from the data by themselves without being explicitly programmed.

Types of Machine Learning:

* Supervised Learning: the ML algorithm learns from **Labelled** Data
* Unsupervised Learning: the ML algorithm learns from **Unlabeled** Data
* Reinforcement Learning: is an area of ML concerned with how intelligent agents take actions in an environment to maximize its rewards
  1. Environment (chess board)
  2. Agent (computer)
  3. Action
  4. Reward (each step that come closer to win)

1.3

Types of Supervised Learning (2 main types)

* Classification: is about predicting a class or discrete values (Male or Female; True or False)
  + Algorithm:
    1. Decision Tree Classification
    2. Random Forest Classification
    3. K-nearest Neighbor
* Regression: is about predicting a quantity or continuous values (Salary; Age; price)
  + Algorithm:
    1. Logistic Regression
    2. Polynomial Regression
    3. Support Vector Machines

1.4

Types of Unsupervised Learning

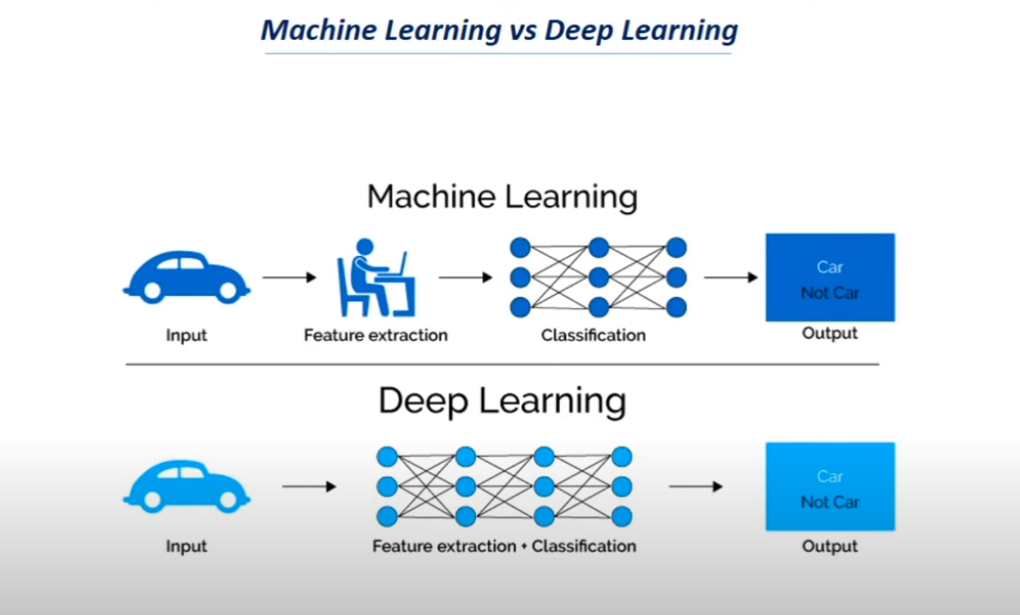
* Clustering: is an unsupervised task which involves grouping the similar data points
* Association: is an unsupervised task that is used to find important relationship between data points

Algorithm:

* K-means Clustering
* Hierarchical Clustering
* Principal Component Analysis (PCA)
* Apriori
* Eclat

1.5

Deep Learning is a subfield of Machine Learning that uses Artificial Neural Networks to learn from the data.

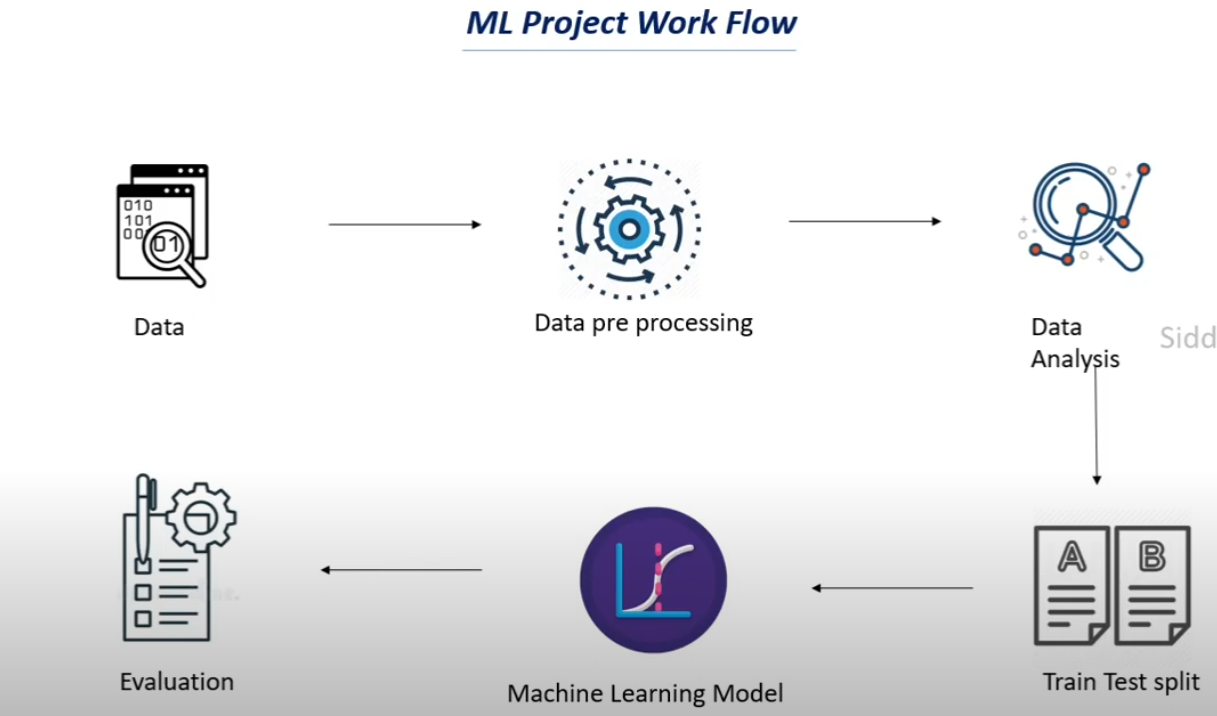


4.1

Where to collect Data?

* **Kaggle**
* **UCI Machine Learning Repository**
* **Google Dataset Search**

4.6



4.7

The mapping from textual data to real valued vectors is called feature extraction.

**Bag Of Words (BOW):** list od unique words in the text corpus

**Term Frequency-Inverse Document Frequency (TF-IDF):** To count the number of times each word appears in a document.

Term Frequency (TF) = (Number of times term t appears in a document) / (Number of terms in the document)

Inverse Document Frequency (IDF) = log-N/n), where N is the number of documents and n is the number of documents a term t has appeared in.

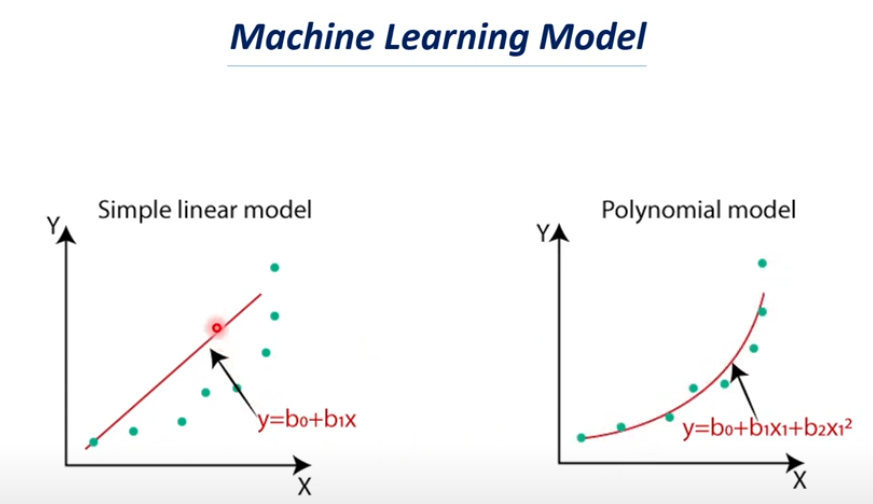
The IDF value of a rare word is high, whereas the IDF of a frequent word is low.

TF-IDF value of a term = TF \* IDF

6.1

A ML model is a function that tries to find the relationship between the Features and the Target variable.

It tries to find a pattern in the data, understand the data and trains on the data. Based on this learning, a Machine Learning Model makes Predictions and recognize patterns.



6.4

How to choose the right ML model (Model Selection)?

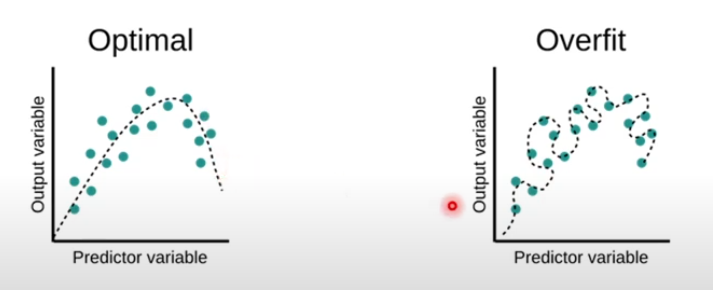
Model Selection in ML is the process of choosing the best suited model for a particular problem. Selecting a model depends on various factors such as the dataset, task, nature of the model, etc.

Models can be selected based on:

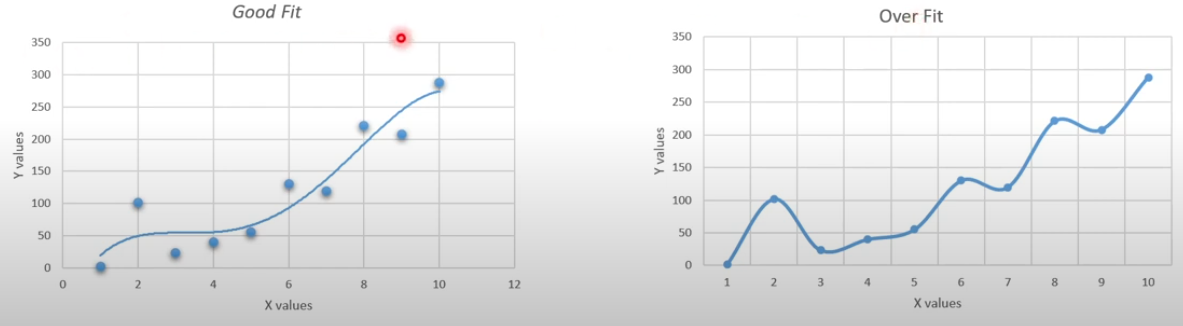
1. Type of Data available:
   1. Image and Videos – CNN
   2. Text data or Speech data – RNN
   3. Numerical data – SVM, Logistic Regression, Decision trees, etc.
2. Based on the test we need to carry out:
   1. Classification tasks – SVM, Logistic Regression, Decision trees, etc.
   2. Regression tasks – Linear regression, Random Forest, polynomial regression, etc.
   3. Clustering tasks – K Means Clustering, Hierarchical Clustering

6.5

**Overfitting** refers to a model that models the training data too well. Overfitting happens when a model learns the details and noise in the training dataset to the extent that is negatively impacts the performance of the model.



Sign that the model ha Overfitted: High training data Accuracy and very low Test data Accuracy.



Causes of Overfitting:

1. Less Data
2. Increased Complexity of the model
3. More number of layers in Neural Network

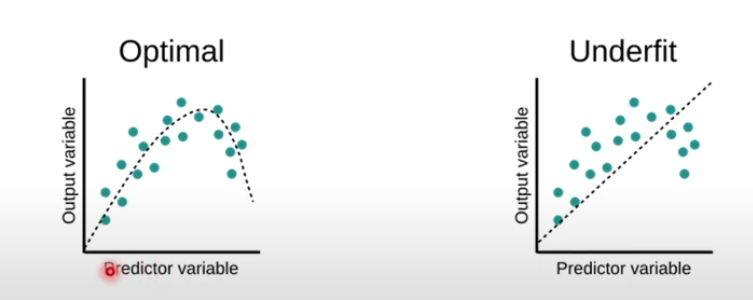
Preventing Overfitting by:

1. Using more data
2. Reduce the number of layers in the Neural network
3. Early Stopping
4. Bias – Variance Tradeoff
5. Use Dropouts

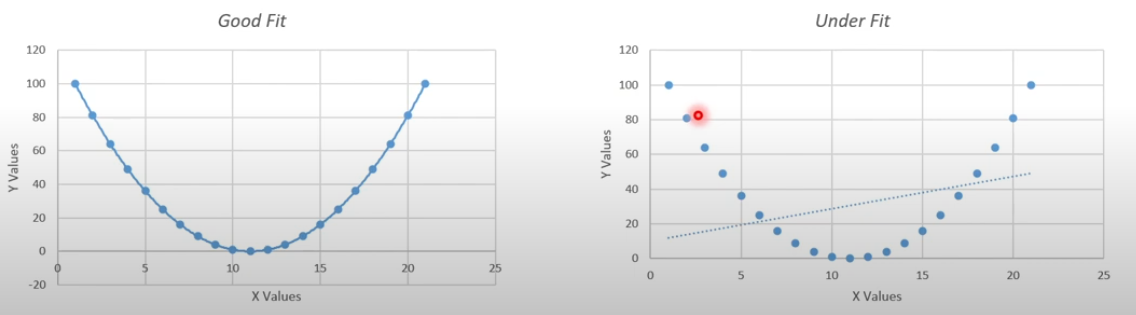
6.6

**Underfitting in ML**

Underfitting happens when the models does not learn enough from the data. Underfitting occurs when a machine learning model cannot capture the underlying trend of the data.



Sign that the model has Underfitted: Very low Training data Accuracy



Cause for Underfitting:

1. Choosing a wrong model
2. Less complexity of the model
3. Less variance but high bias

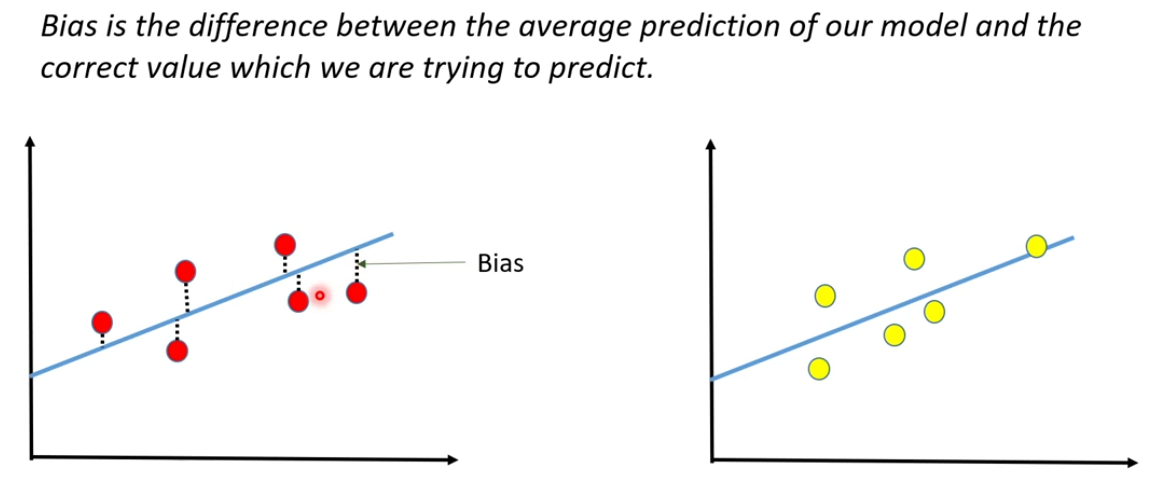
Prevent Underfitting by:

1. Choosing a correct model appropriate for the problem
2. Increasing the complexity of the model
3. More number of parameters to the model
4. Bias – Variance Tradeoff

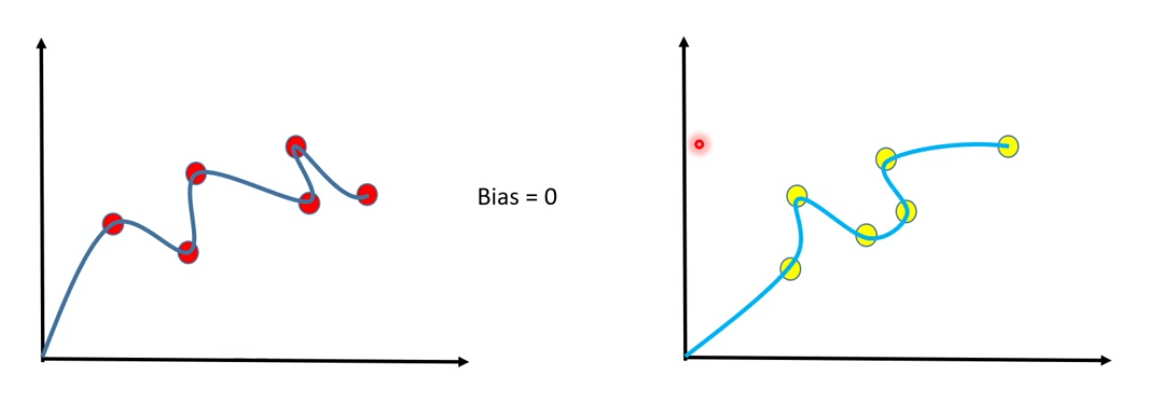
6.7

Bias – Variance Tradeoff in ML

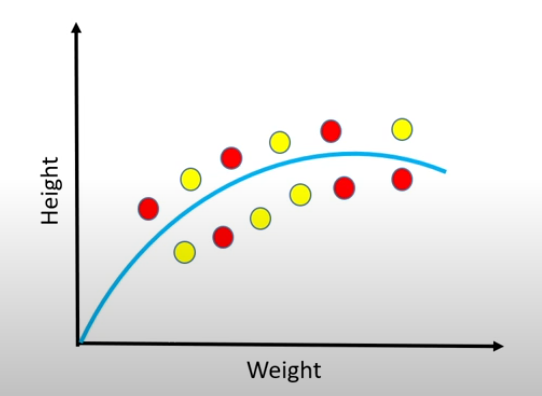
**Bias** is the difference between the average prediction of our model and the correct value which we are trying to predict.

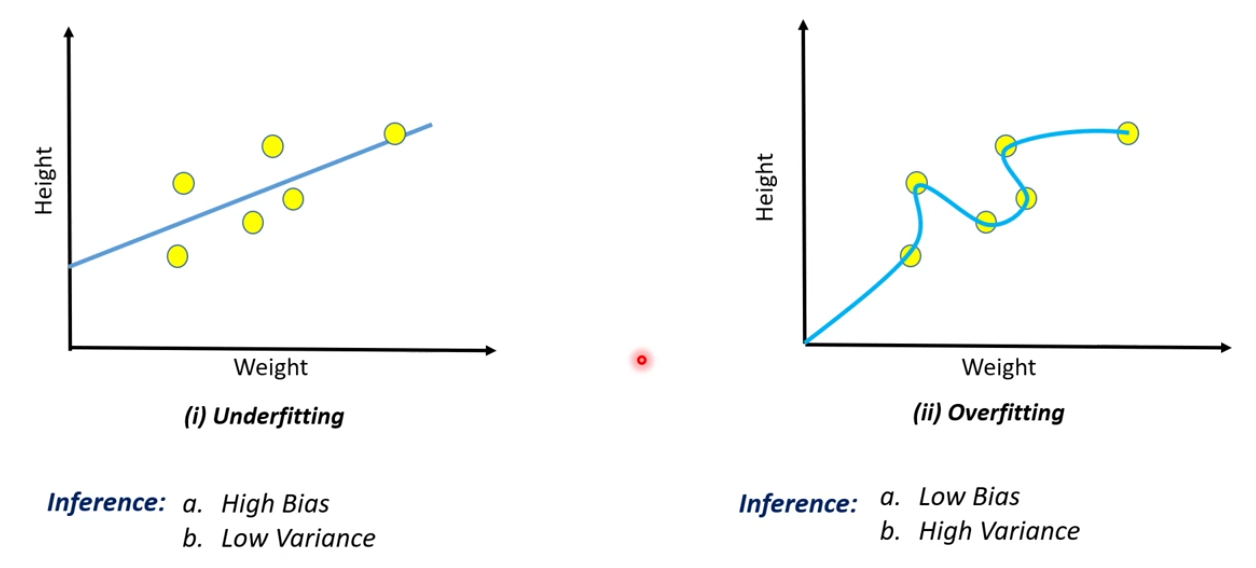


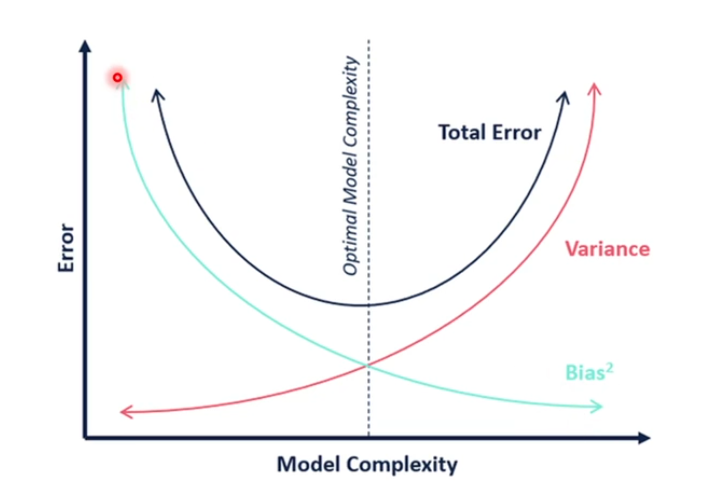
**Variance** is the amount that the estimate of the target function will change if different training data was used.



**EXEMPLE:** Problem statement: Identify an appropriate model to predict the Height of a person, when their weight is given.







Techniques to have a better Bias – Variance Tradeoff:

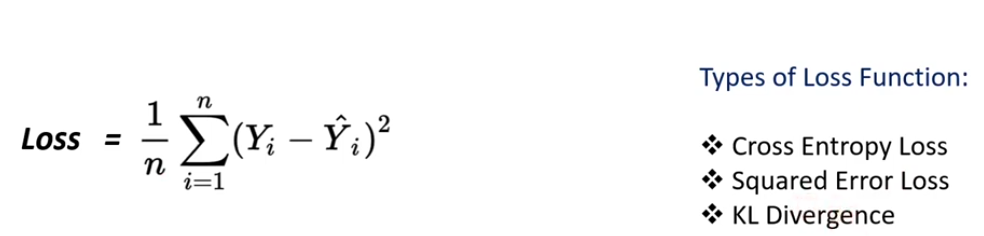
1. Good Model Selection
2. Regularization
3. Dimension Reduction
4. Ensemble methods

6.8

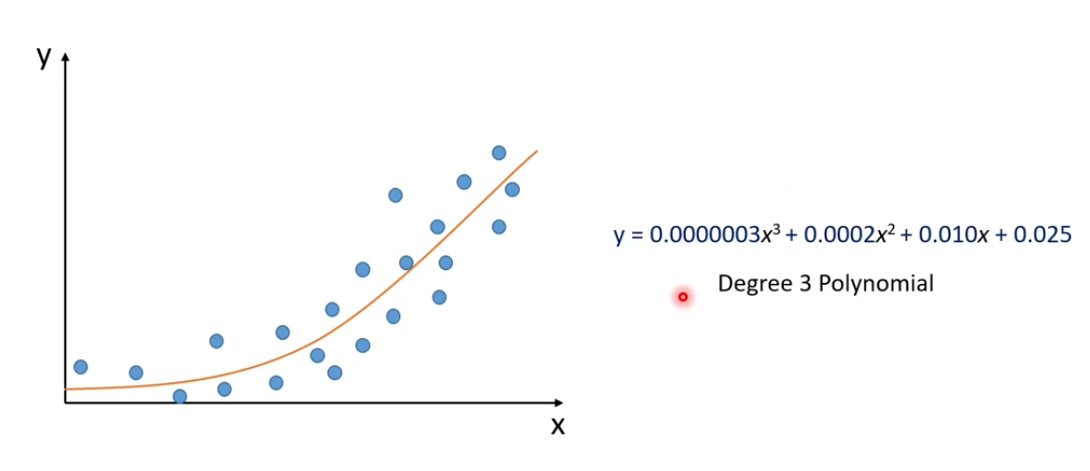
Loss Function in ML

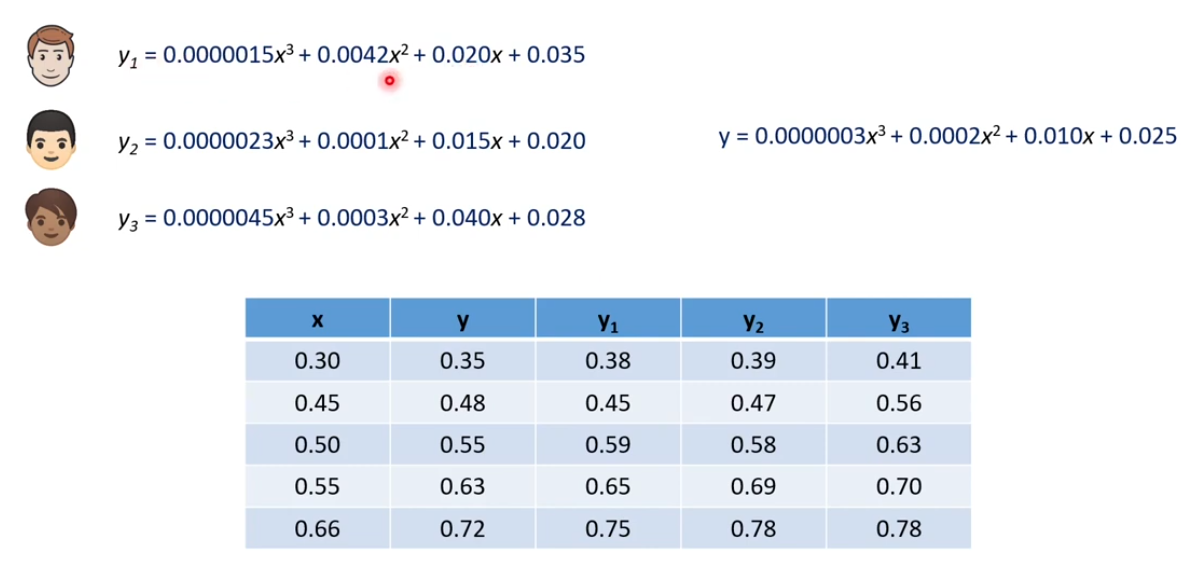
Loss function measures how far an estimated value is from its true value.

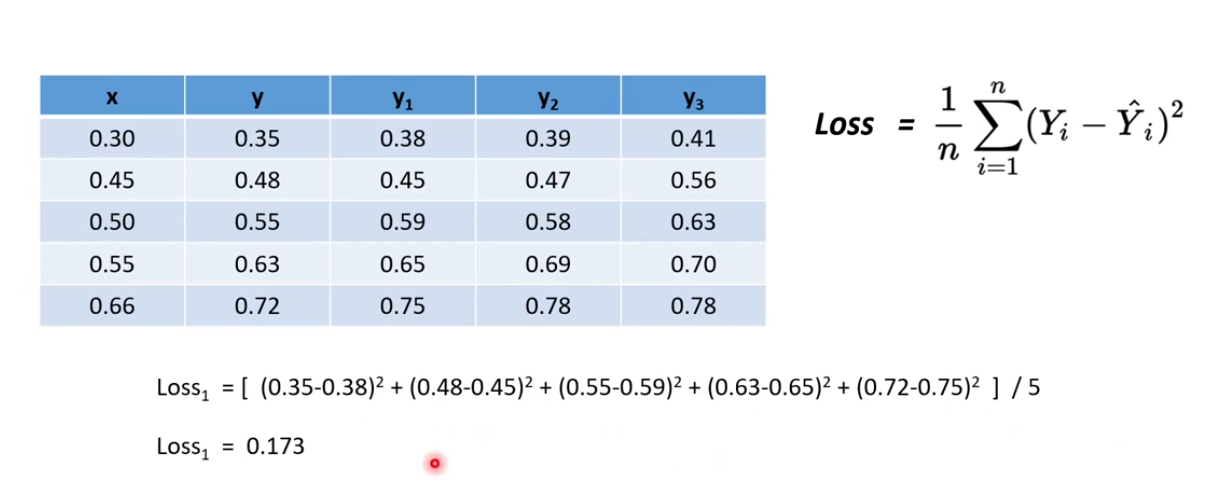
It is helpful to determine which model performs better and which parameters are better.



Example:







**Low Loss Value -> High Accuracy**

6.9

Model Evaluation in ML

Workflow of ML Project:

1. Get Data
2. Data Pre-Processing
3. Data Analysis
4. Train-Test splitting Data
5. XGBoost Regressor
6. Evaluation

For Classification: Accuracy score

For Regression: Mean Absolute Error

Accuracy Score is the ratio of the number of correct predictions to the total number of input data points.

Une image contenant texte, Police, capture d’écran, ligne

Le contenu généré par l’IA peut être incorrect.

Example:

Une image contenant texte, Police, capture d’écran, blanc

Le contenu généré par l’IA peut être incorrect.

Mean Square Error measures the average od the squares of errors, that is, the average squared difference between the estimated values and the actual value.

Une image contenant texte, capture d’écran, Police, ligne

Le contenu généré par l’IA peut être incorrect.

6.10

Model Parameters and Hyperparameters

Une image contenant texte, capture d’écran, Police

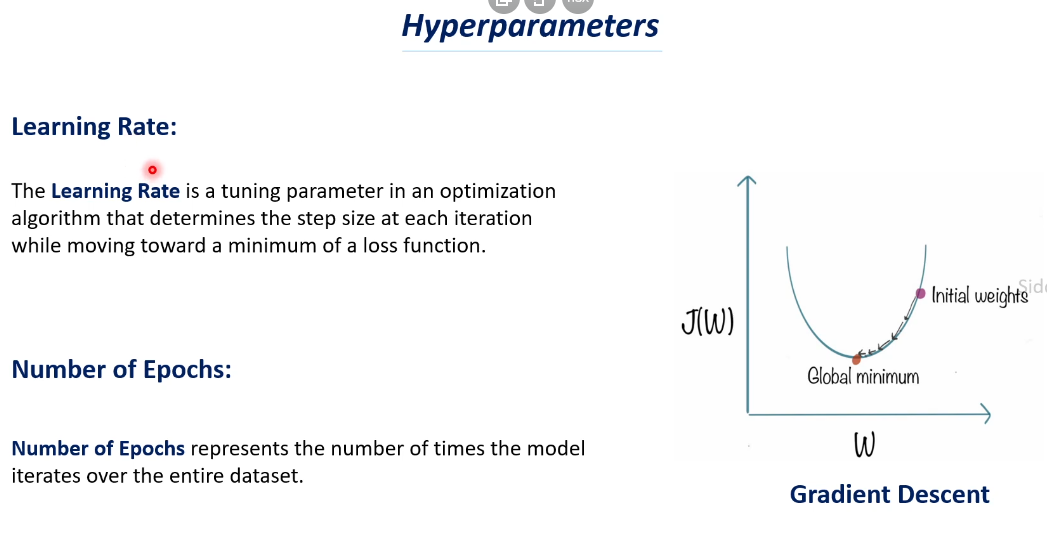
Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, capture d’écran, Police, nombre

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, capture d’écran, Police

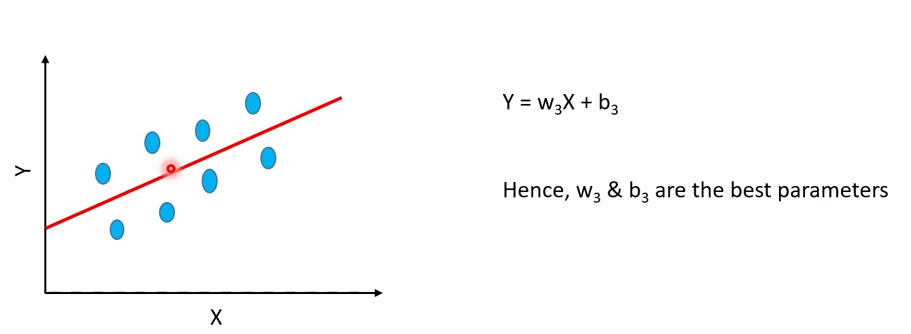
Le contenu généré par l’IA peut être incorrect.



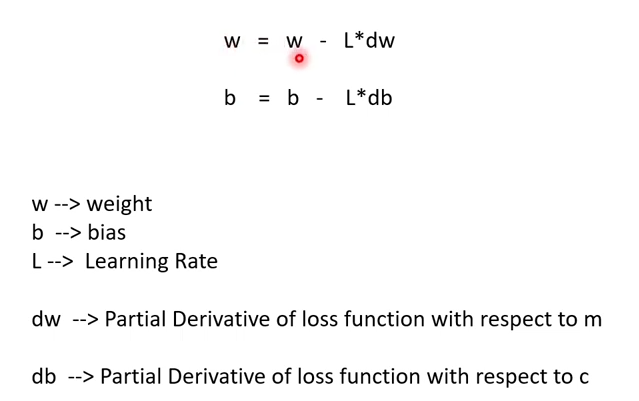
6.11

Gradient Descent in ML

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



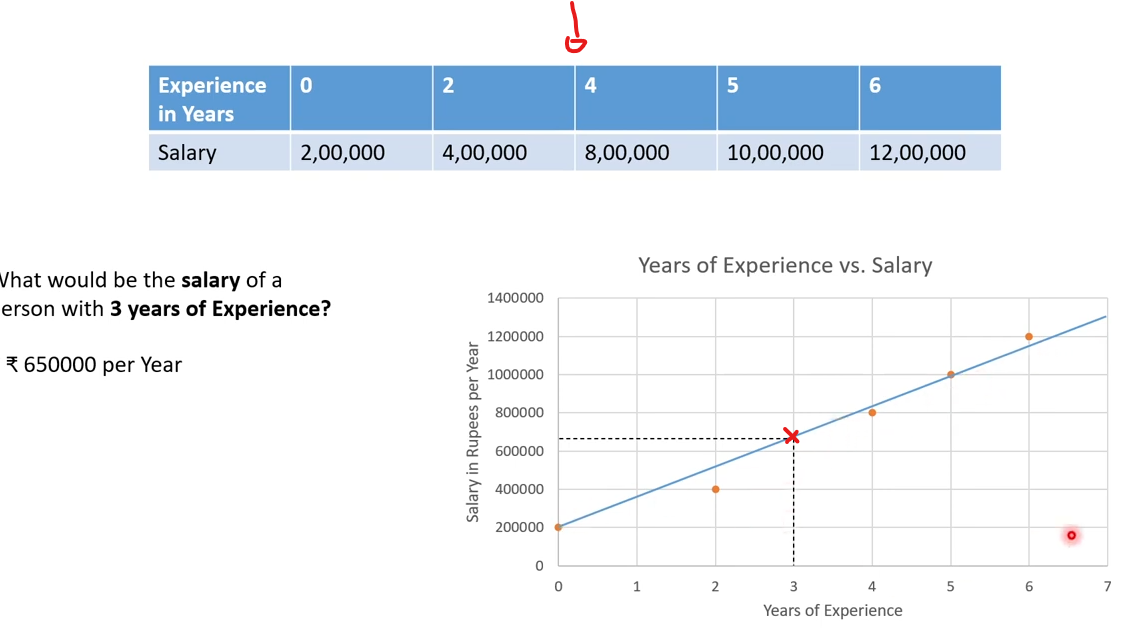
Gradient Descent is an optimization algorithm used for minimizing the loss function in various ML algorithms. It is used for updating the parameters of the learning model.

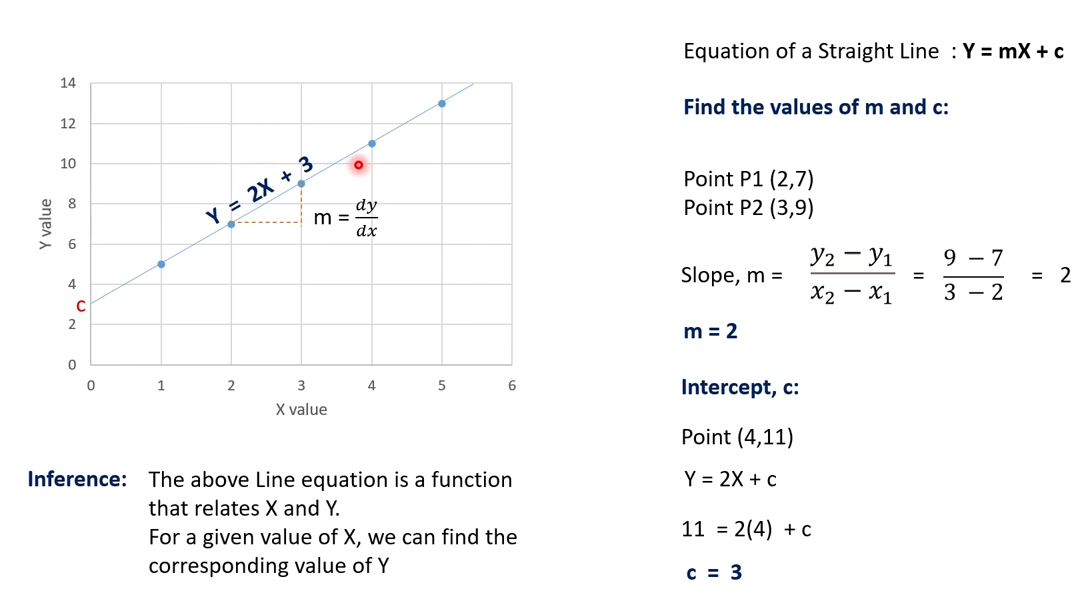


7.1.1

Linear Regression – intuition

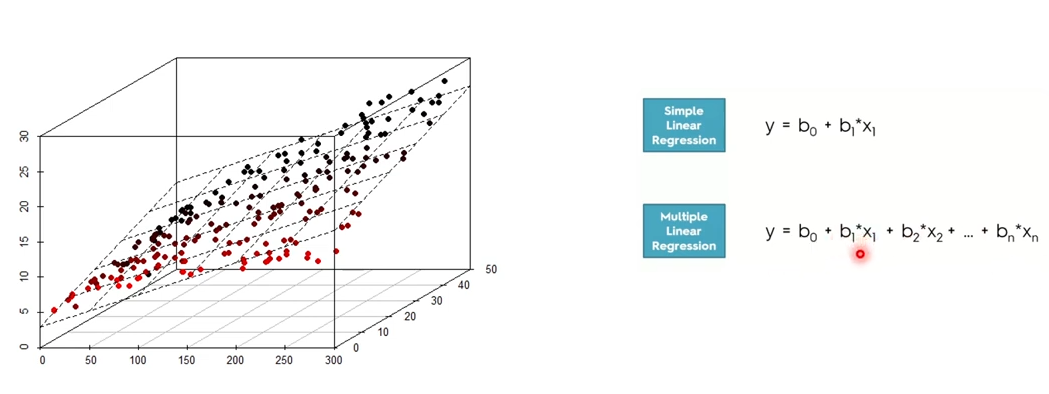
Examples:





What if there are more than **2** variables?

Use -> Multiple Linear Regression is a model for predicting the value of one dependent variable based on two or more independent variables.



Linear Regression:

Advantages -

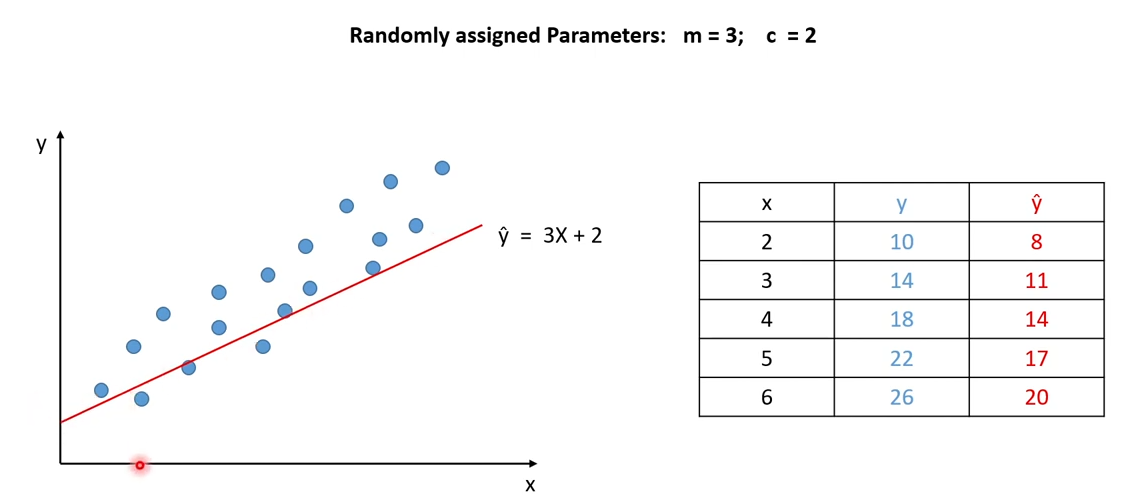
1. Very simple to implement
2. Performs well on data with linear relationship

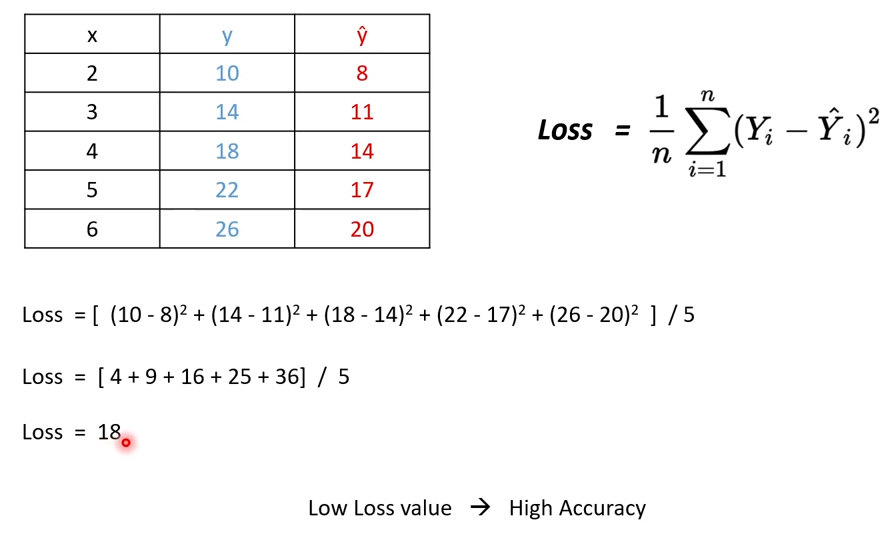
Disadvantages –

1. Not suitable for data having non-linear relationship
2. Underfitting issue
3. Sensitive to Outliers

7.1.2

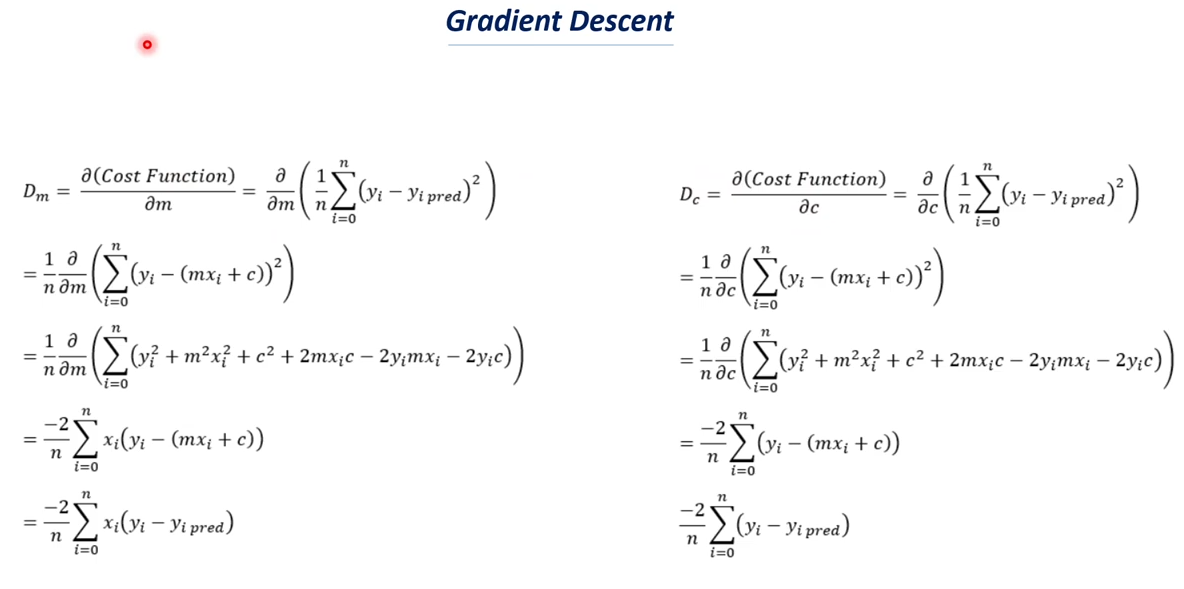
Linear Regression – Mathematical Understanding



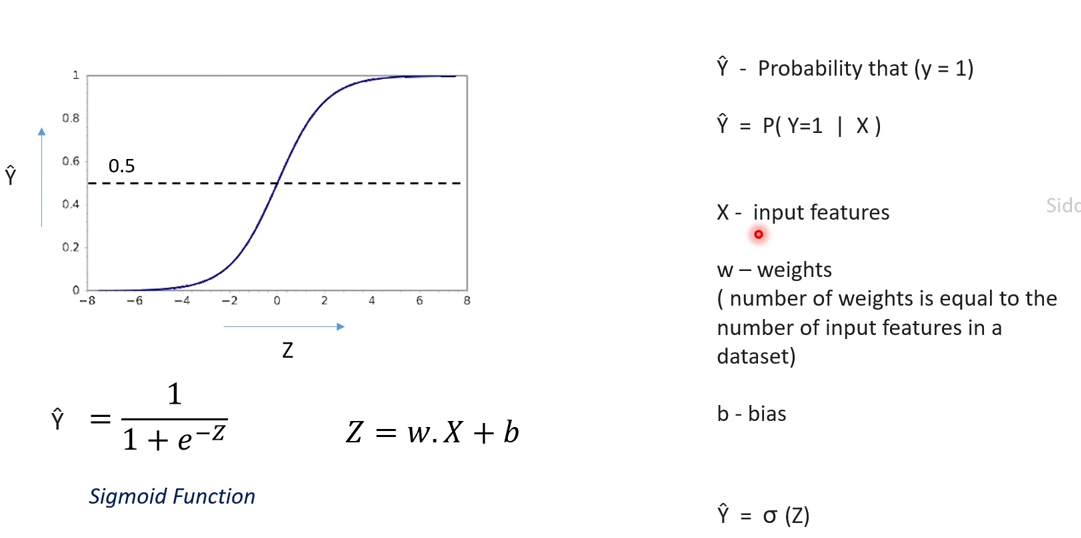


7.1.3

Gradient Descent

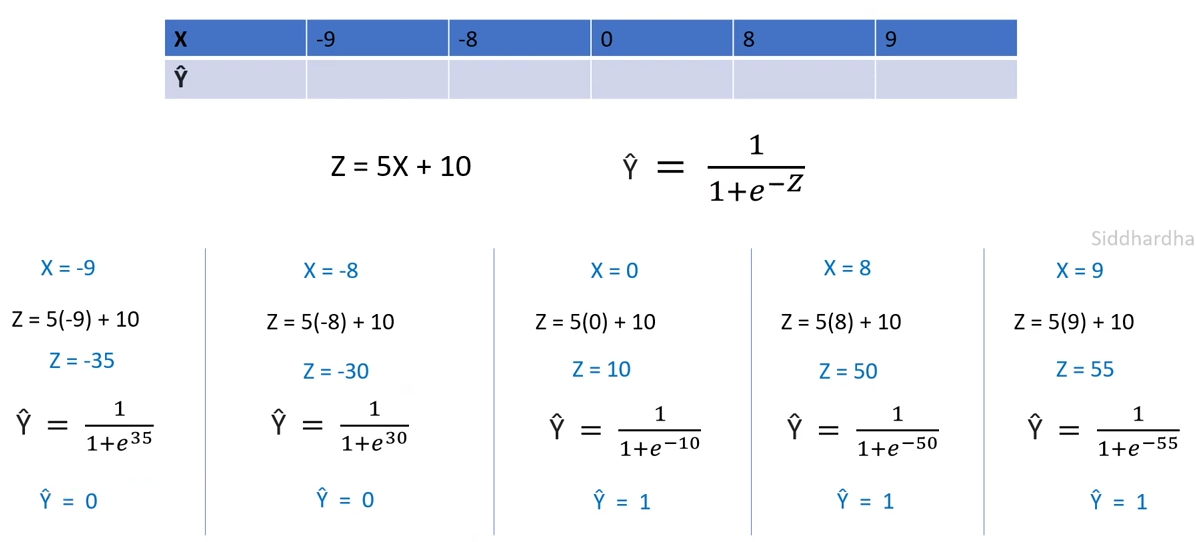


7.2.1



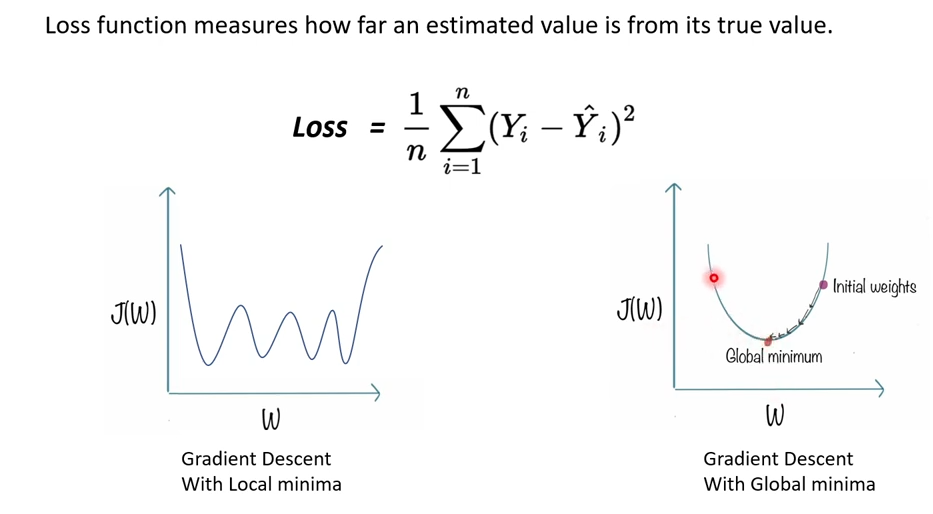
7.2.2

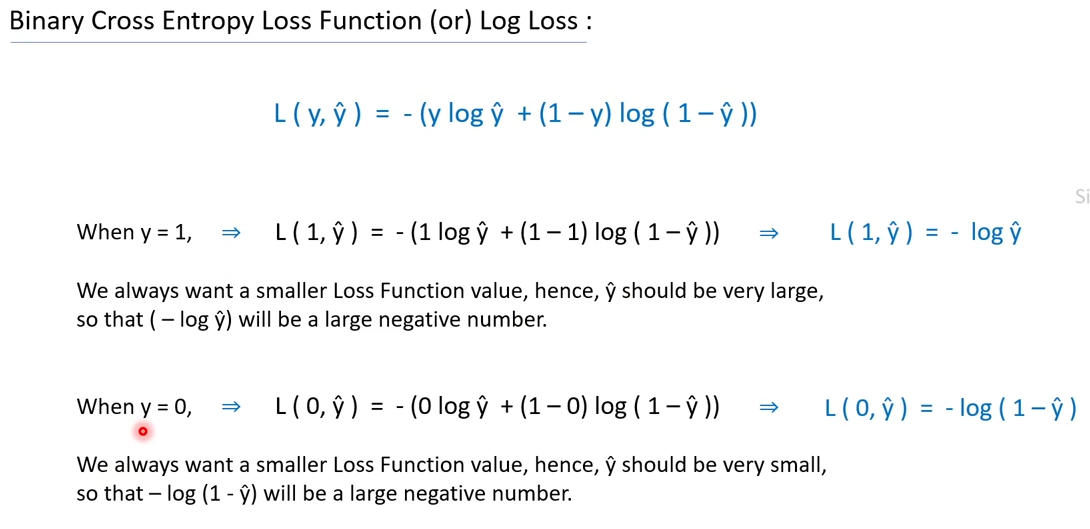
Math behind Logistic Regression



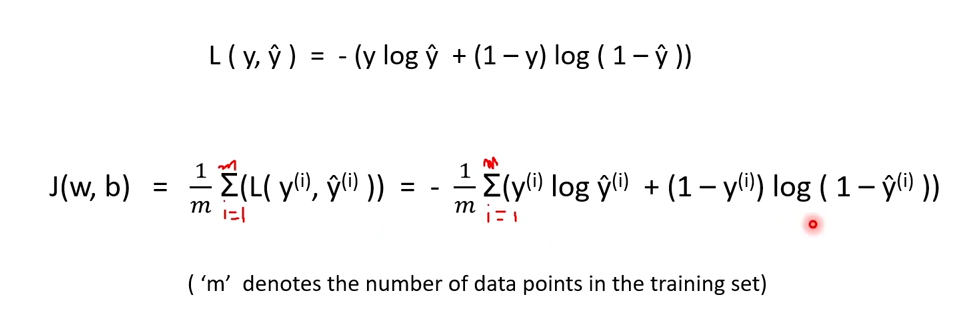
7.2.3

Loss Function and Cost Function for Logistic Regression



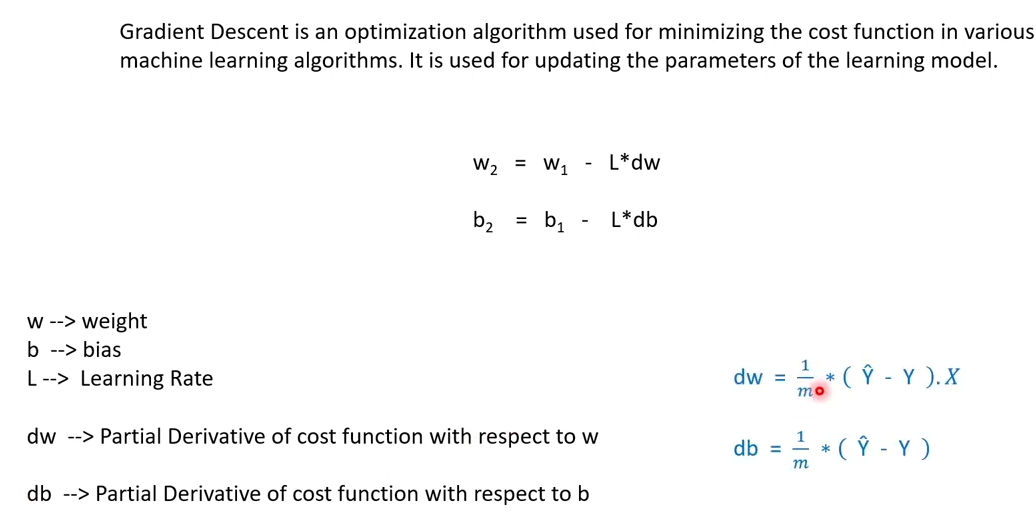


**Loss function ( L )** mainly applies for a single training set as compared to the **cost** **function ( J )** which deals with a penalty for a number of training sets or the complete batch.



7.2.4

Gradient Descent for Logistic Regression

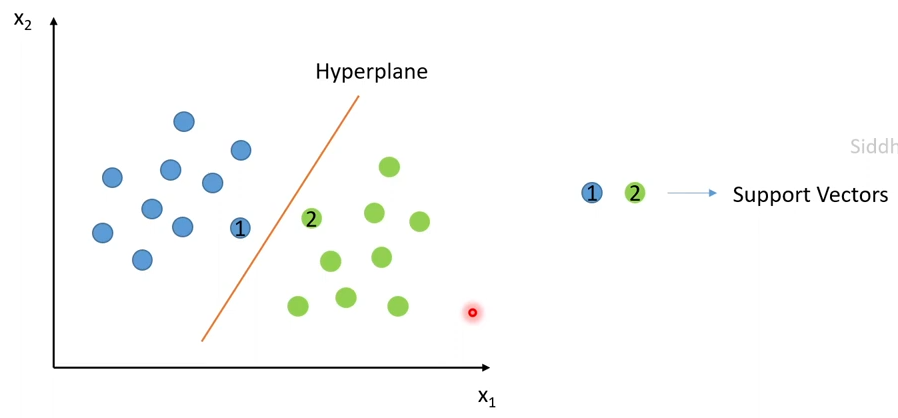


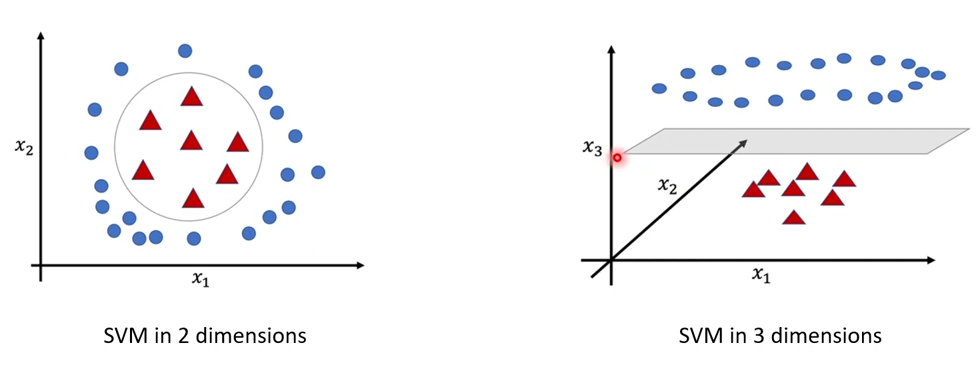
7.3.1

Support Vector Machine ( SVM ) Classifier – intuition

About Support Vector Machine model:

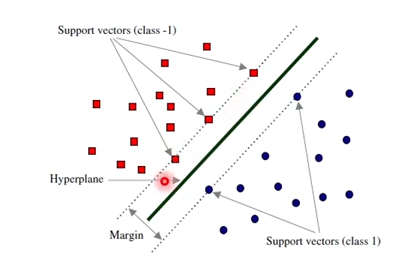
1. Supervised Learning Model
2. Both Classification and Regression
3. Hyperplane
4. Support Vectors





Hyperplane is a line (in 2d space) or a plane that separate the data points into 2 classes.

Support Vectors are the data points which lie nearest to the hyperplane. If theses data points changes, the position of the hyperplane changer.



Of SVM:

Advantages:

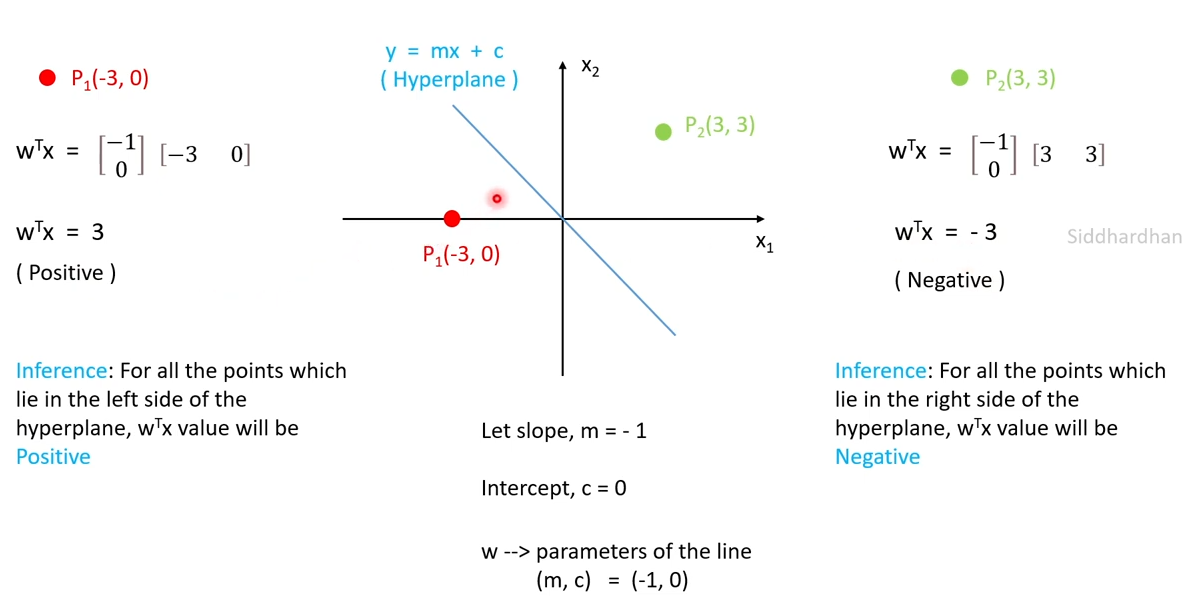
1. Work well with smaller datasets
2. Works efficiently when there is a clear margin of separation
3. Works well with high dimensional data

Disadvantages:

1. Not suitable for large datasets as the training time is higher
2. Not suitable for noisier datasets with overlapping classes

7.3.2

Math behind SVM Classifier

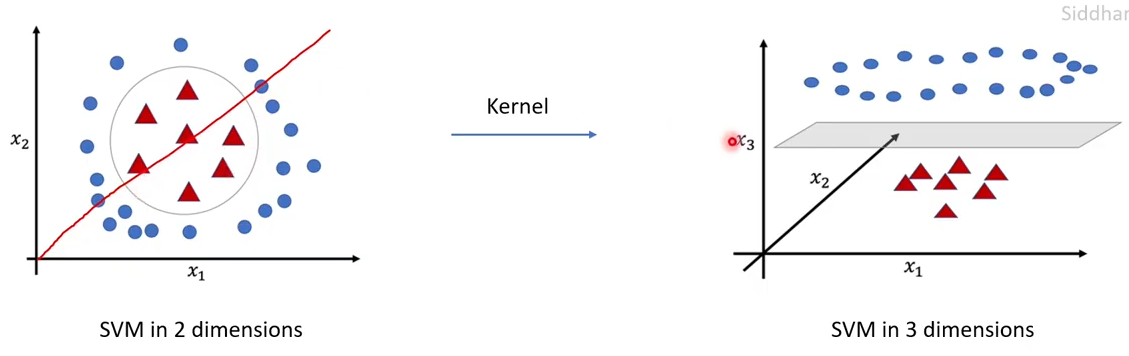




7.3.3

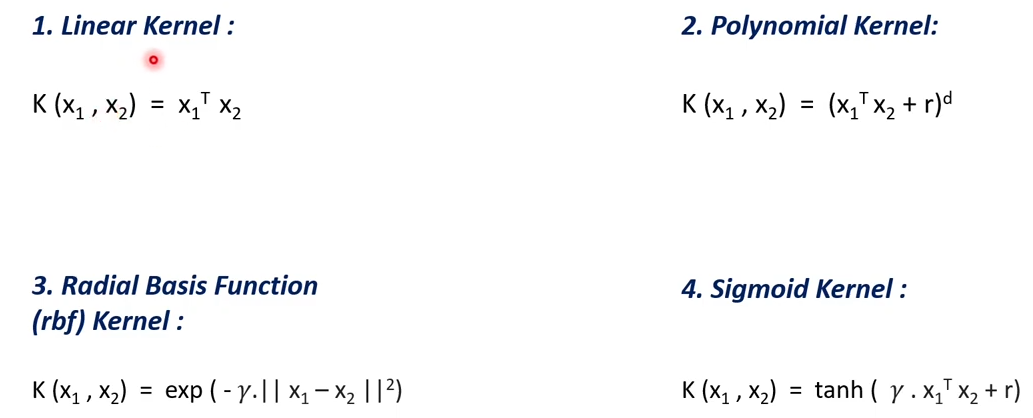
SMV – Kernels

Kernel Function generally transforms the training set of data so that non-linear decision surface can be transformed to a linear equation in a higher number of dimension spaces. It returns the inner product between two points in a standard feature dimension.



Types of SVM Kernels:

1. Linear
2. Polynomial
3. Radial Basis Functions (rbf)
4. Sigmoid



7.3.4

Loss Function for Support Vector Machine Classifier - Hinge Loss

Hinge Loss is one of the types of Loss Function, mainly used for maximum margin classification models.

Hinge Loss incorporates a margin or distance from the classification boundary into the loss calculation. Even if new observations are classified correctly can incur a penalty if the margin from the decision boundary is not large enough.

