Siddhardhan

What is a Machine Learning Model?

Machine Learning

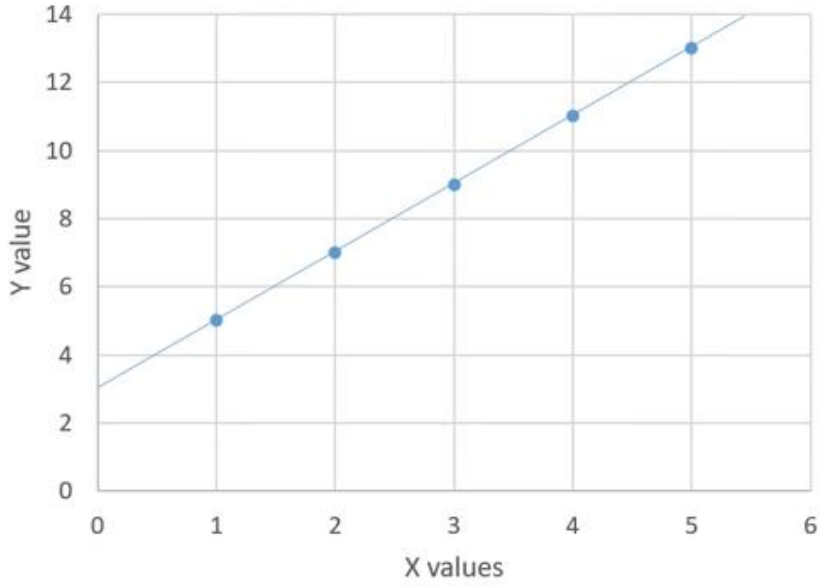
Siddhardhan

Data

Machine Learning Models

Machine Learning Model

X	1	2	3	4	5
Y	5	7	9	11	13

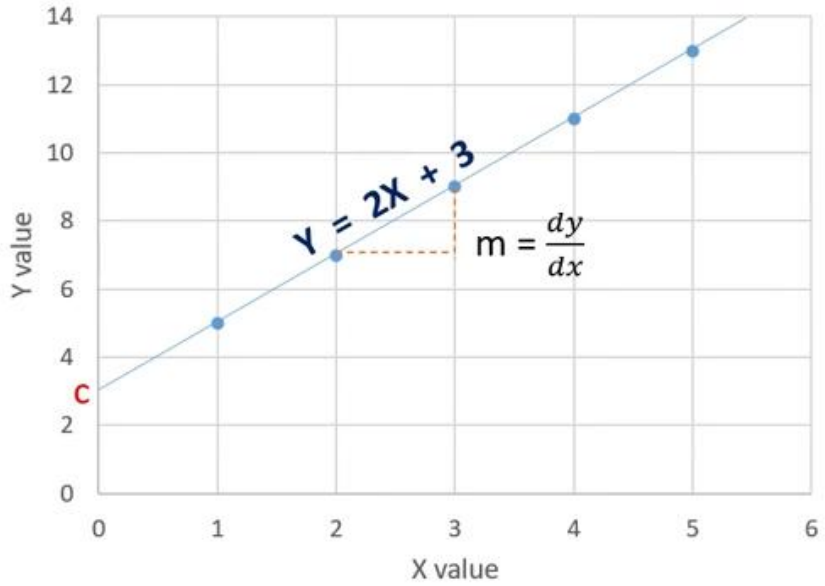


Siddhardhan

$Y = mX + c$

X --> X value
Y --> Y value
m --> Slope
c --> Intercept

Machine Learning Model



Inference: The above Line equation is a function that relates X and Y.
For a given value of X, we can find the corresponding value of Y

Equation of a Straight Line : **Y = mX + c**

Find the values of m and c:

Point P1 (2,7)
Point P2 (3,9)

Slope, m = $\frac{y_2 - y_1}{x_2 - x_1} = \frac{9 - 7}{3 - 2} = 2$

m = 2

Intercept, c:

Point (4,11)

Y = 2X + c

11 = 2(4) + c

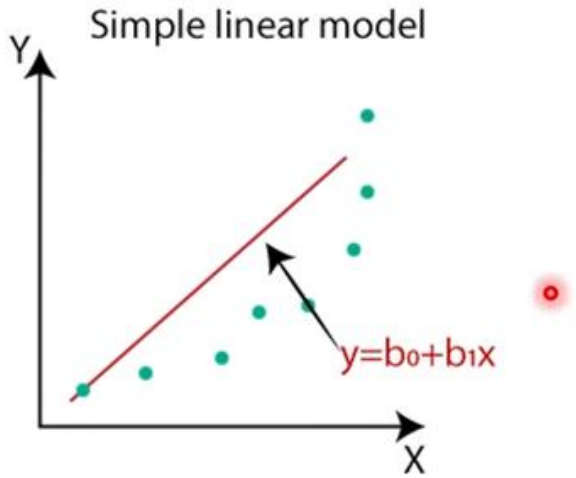
c = 3

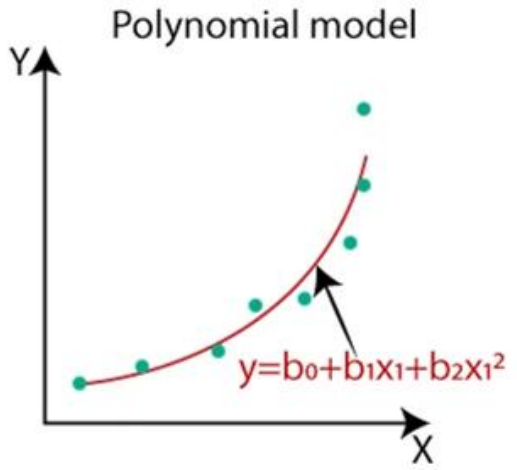
Machine Learning Model

A **Machine Learning Model** is a function that tries to find the relationship between the Features and the Target variable.

It tries to find the 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.

Machine Learning Model





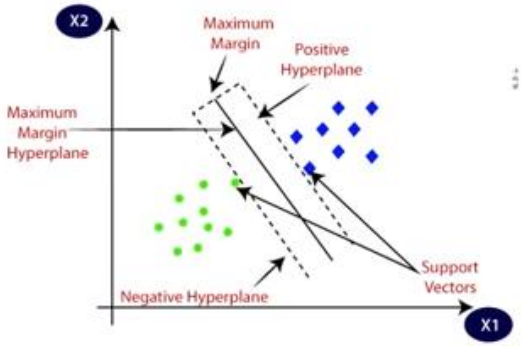
Machine Learning Model

A **Machine Learning Model** is a function that tries to find the relationship between the Features and the Target variable.

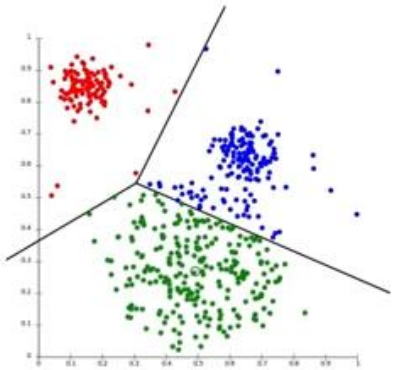
It tries to find the pattern in the data, understand the data and trains on the data. Based on this learning, a Machine Learning Model makes Predictions and recognize patterns.

Siddhardhan

Logistic Regression



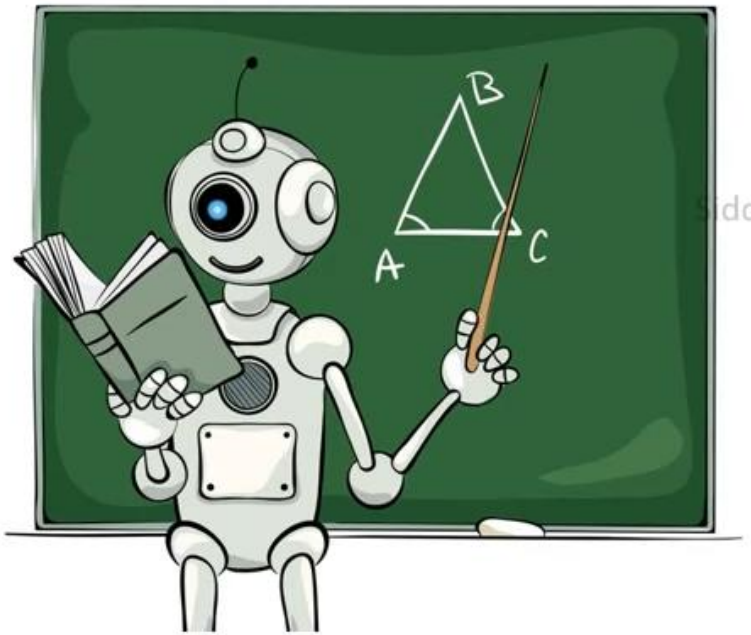
Support Vector Machine



K-Means Clustering

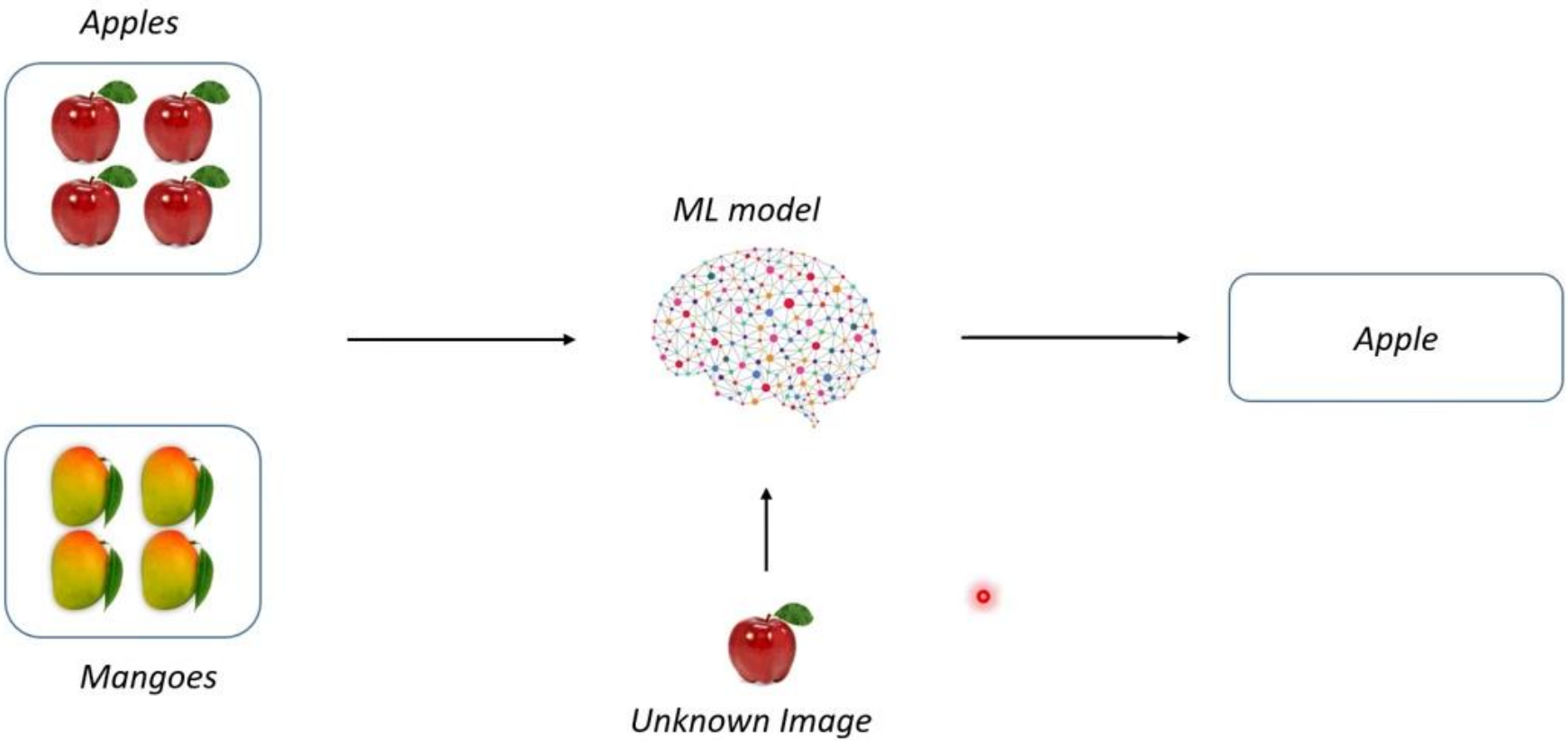
Siddhardhan

Supervised Learning Models



Supervised Learning

In Supervised Learning, the Machine Learning algorithm learns from ***Labelled Data***



Types of Supervised Learning

Classification

Classification is about predicting a class or discrete values
Eg: Male or Female; True or False

Regression

Regression is about predicting a quantity or continuous values
Eg: Salary; age; Price.

Types of Supervised Learning

Classification:

Dog

Cat

Siddhardhan

Regression:

Temperature

Rainfall in cm

Supervised Learning Models

Classification:

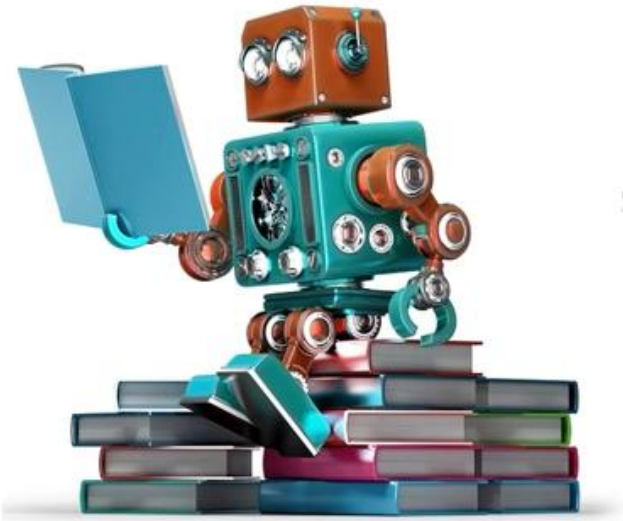
1. *Logistic Regression*
2. *Support Vector Machine Classifier*
3. *Decision Tree*
4. *K-Nearest Neighbors*
5. *Random Forest*
6. *Naïve Bayes Classifier*

Regression:

1. *Linear Regression*
2. *Lasso Regression*
3. *Polynomial Regression*
4. *Support Vector Machine Regressor*
5. *Random Forest Regressor*
6. *Bayesian Linear Regressor*

Siddhardhan

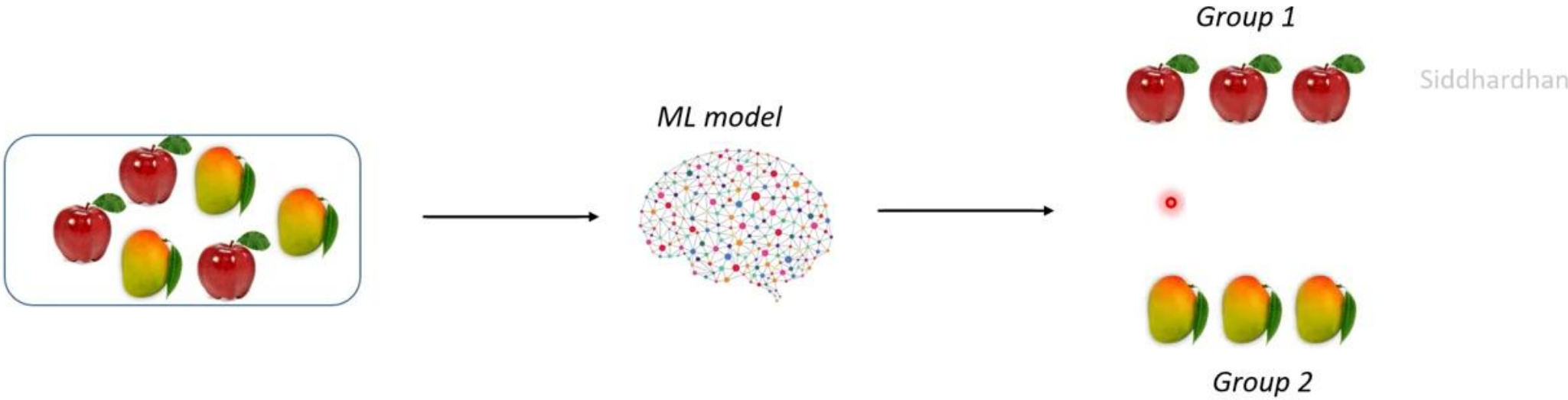
Unsupervised Learning Models



Siddhardhan

Unsupervised Learning

In Unsupervised Learning, the Machine Learning algorithm learns from ***Unlabelled Data***



Types of Unsupervised Learning

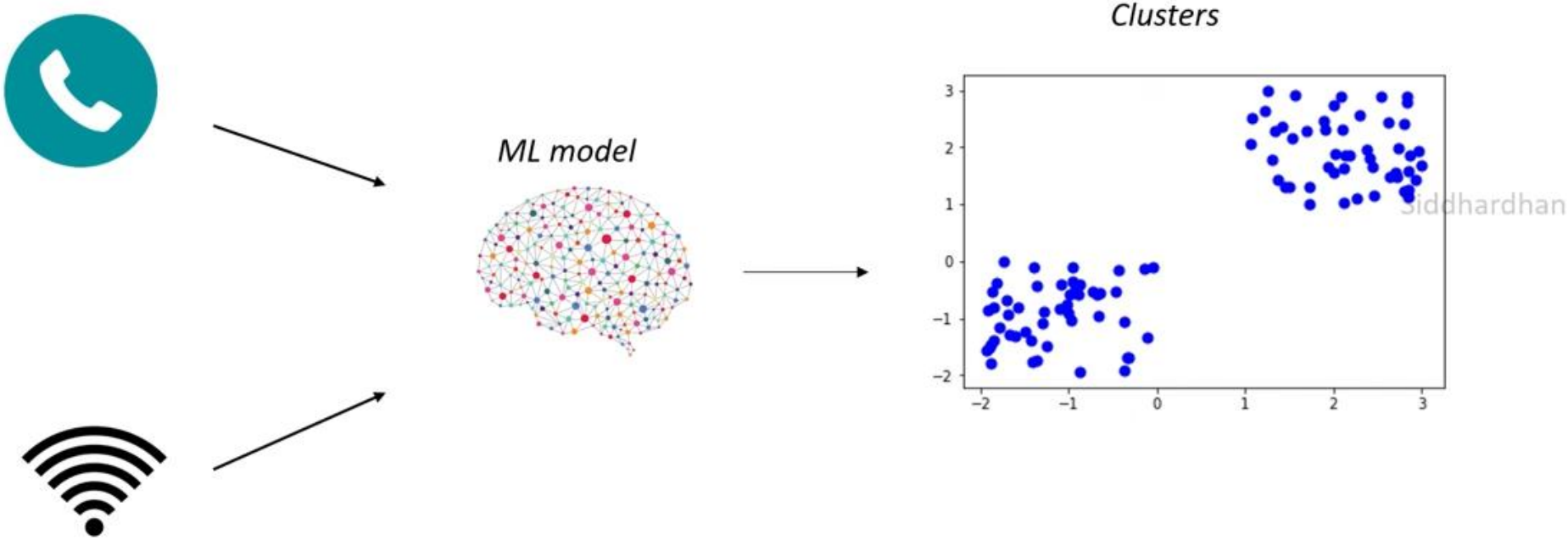
Clustering

Clustering is an unsupervised task which involves grouping the similar data points.

Association

Association is an unsupervised task that is used to find important relationship between data points

Clustering



Association

Customer 1

- *Bread*
- *Milk*
- *Fruits*
- *wheat*

Customer 2

- *Bread*
- *Milk*
- *Rice*
- *Butter*

Customer 3

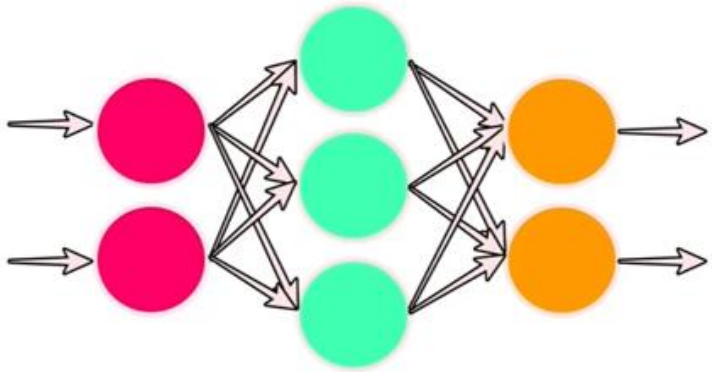
Now, when customer 3 goes and buys bread, it is highly likely that he will also buy milk.

Unsupervised Learning Models

1. *K-Means Clustering*
2. *Hierarchical Clustering*
3. *Principal Component Analysis (PCA)*
4. *Apriori*
5. *Eclat*

Siddhardhan

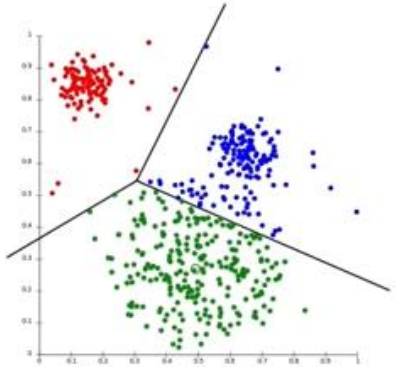
How to choose the right Machine Learning Model? (Model Selection)



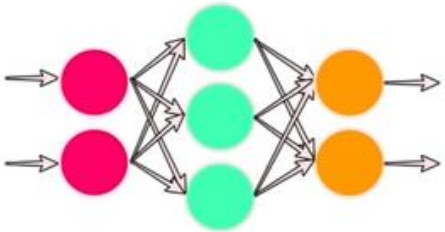
Model Selection

Model Selection in Machine Learning 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.

Logistic Regression



K-Means Clustering



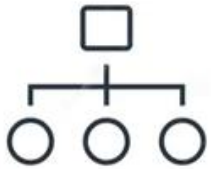
Neural Network

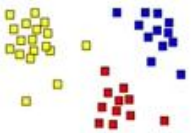
Model Selection

Models can be selected based on :

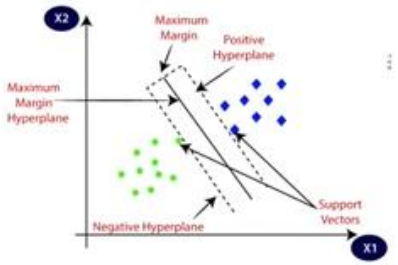
1. Type of Data available:
   a. Images & Videos – CNN
   b. Text data or Speech data – RNN
   c. Numerical data – SVM, Logistic Regression, Decision trees, etc.

2. Based on the task we need to carry out:
   a. Classification tasks – SVM, Logistic Regression, Decision trees, etc.
   b. Regression tasks – Linear regression, Random Forest, Polynomial regression, etc.
   c. Clustering tasks – K-Means Clustering, Hierarchical Clustering





Cross Validation

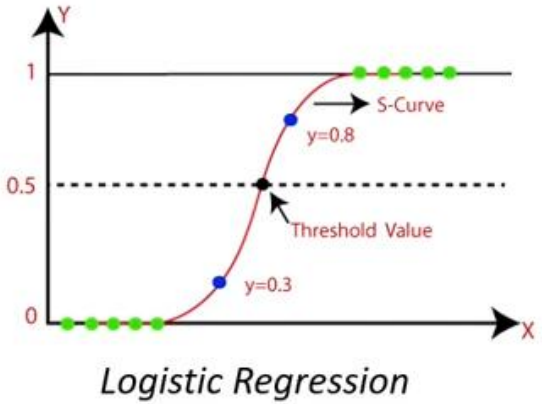


Support Vector Machine

	Dataset					*Accuracy*
Iteration 1	Train	Train	Train	Train	Test	88%
Iteration 2	Train	Train	Train	Test	Train	83%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	81%
Iteration 5	Test	Train	Train	Train	Train	84%

$$\text{Mean Accuracy} = \frac{88 + 83 + 86 + 81 + 84}{5} = 84.4\ \%$$

Cross Validation



Logistic Regression

	Dataset					*Accuracy*
Iteration 1	Train	Train	Train	Train	Test	90%
Iteration 2	Train	Train	Train	Test	Train	88%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	91%
Iteration 5	Test	Train	Train	Train	Train	85%

$$\text{Mean Accuracy} = \frac{90 + 88 + 86 + 91 + 85}{5} = 88\ \%$$

Cross Validation

✓ *Accuracy score for SVM = 84.4 %*

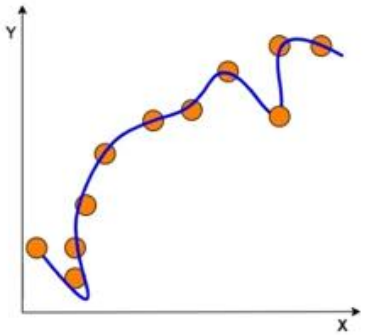
✓ *Accuracy score for Logistic Regression = 88 %*

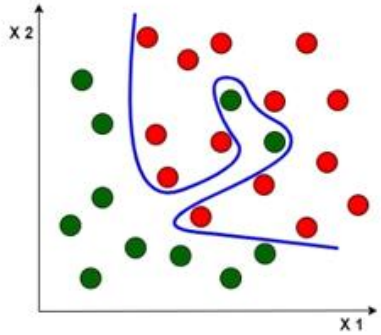
Cross Validation Implementation:

```
>>> from sklearn import datasets, linear_model
>>> from sklearn.model_selection import cross_val_score
>>> diabetes = datasets.load_diabetes()
>>> X = diabetes.data[:150]
>>> y = diabetes.target[:150]
>>> lasso = linear_model.Lasso()
>>> print(cross_val_score(lasso, X, y, cv=3))
[0.33150734 0.08022311 0.03531764]
```

Siddhardhan

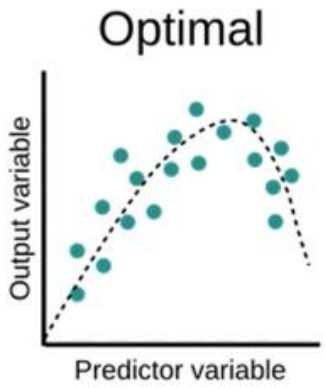
Overfitting in Machine Learning

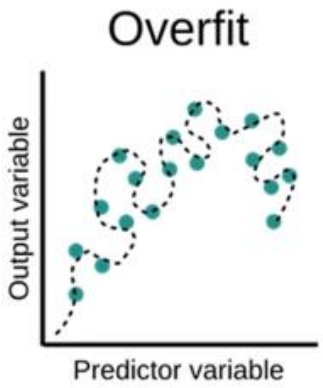




Overfitting

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

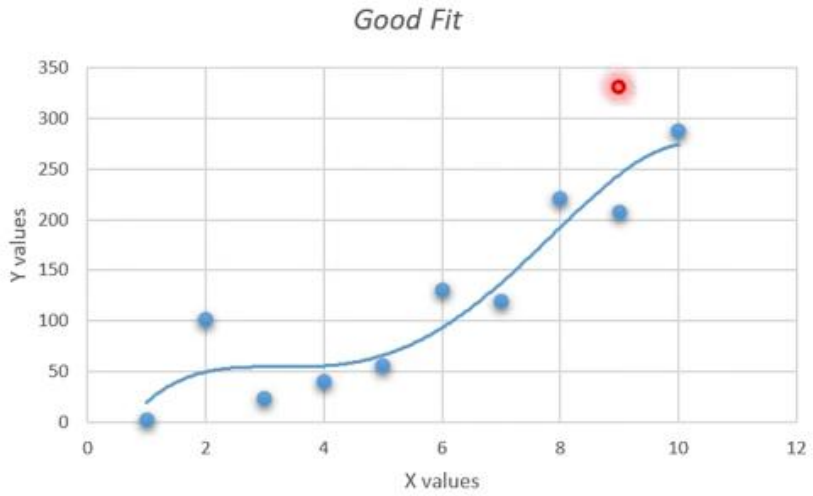


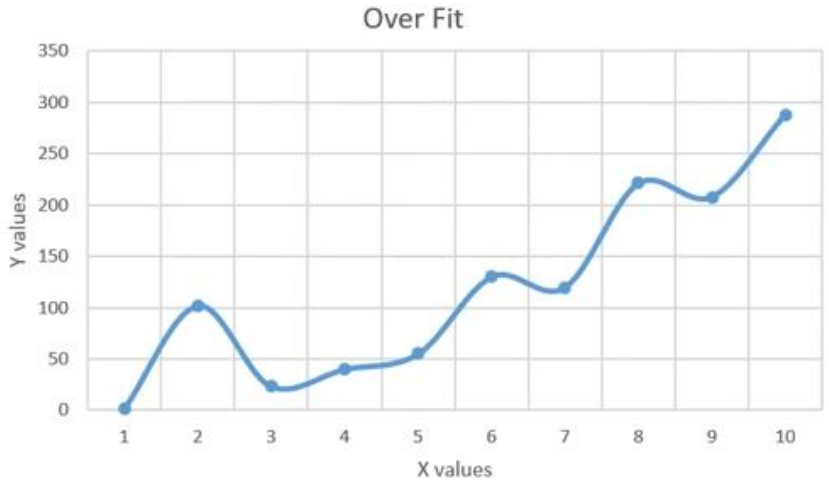


Sign that the model has Overfitted : High Training data Accuracy & very low Test data Accuracy

Overfitting

X	1	2	3	4	5	6	7	8	9	10
Y	1.38	101.41	23.34	39.89	55.23	129.91	119.33	221.09	207.43	287.80





Overfitting

Causes for Overfitting:

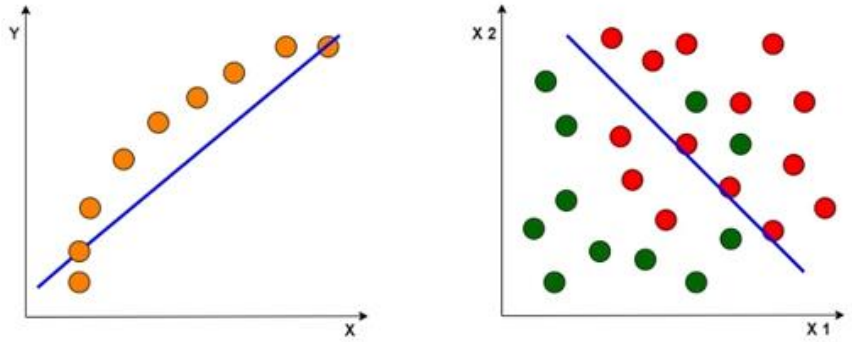
1. Less Data
2. Increased Complexity of the model
3. More number of layers in Neural Network

Preventing Overfitting by:

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

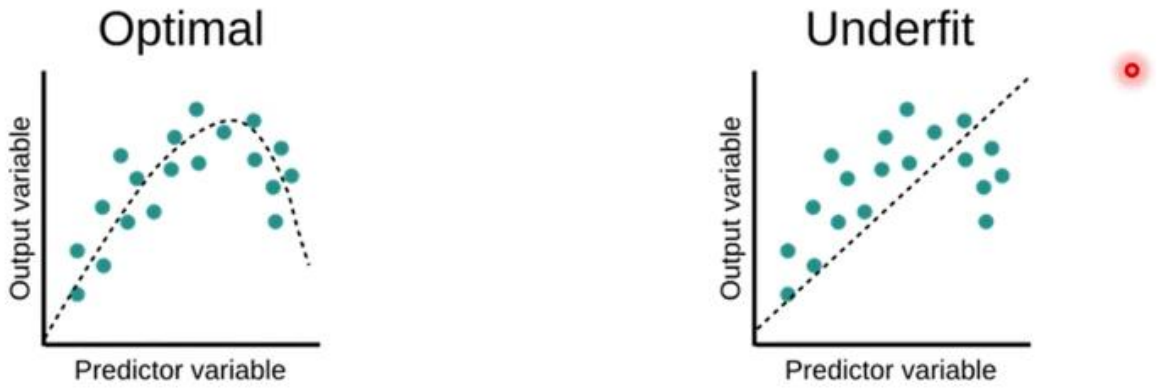
Siddhardhan

Underfitting in Machine Learning



Underfitting

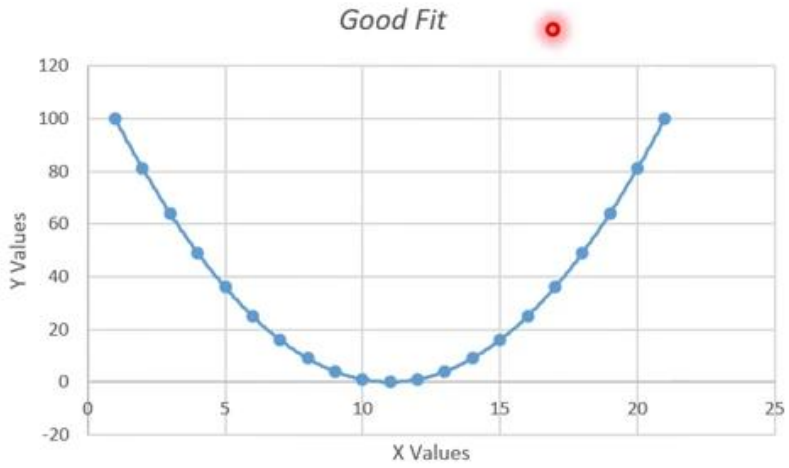
Underfitting happens when the model **does not learn enough** from the data. Underfitting occurs when a machine learning model cannot capture the underlying trend of the data

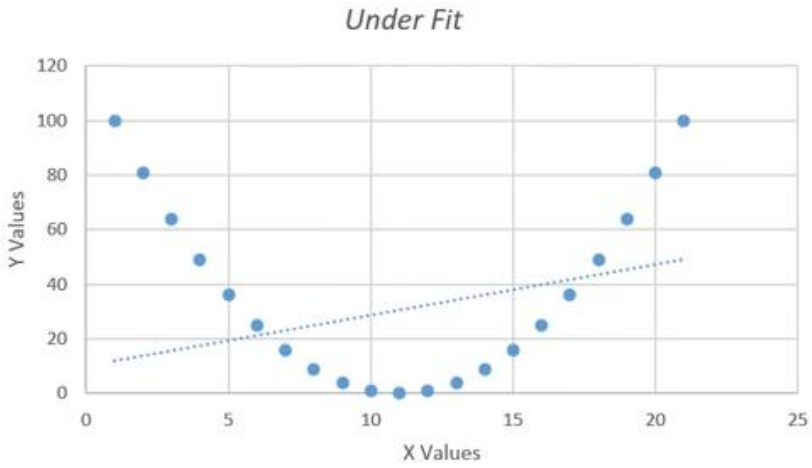


Sign that the model has Underfitted : Very Low Training data Accuracy

Underfitting

X	-10	-9	-8	-7	.....	0	.....	7	8	9	10
Y	100	81	64	49	.....	0	.....	49	64	81	100





Underfitting

Causes for Underfitting:

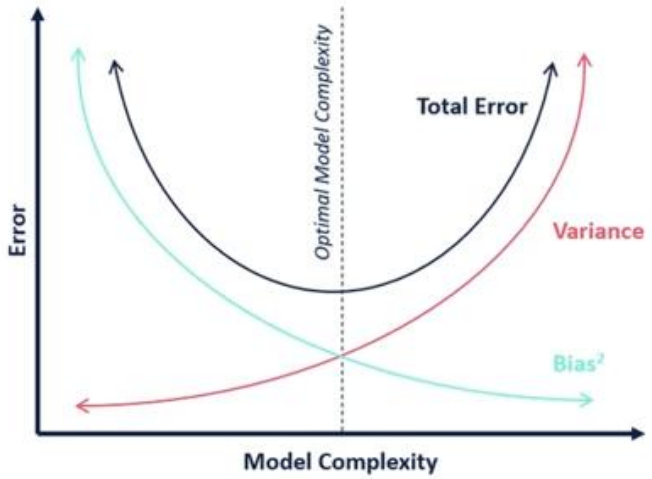
1. Choosing a wrong model
2. Less complexity of the model
3. Less variance but high bias

Prevent Underfitting by:

1. Choosing the 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

Siddhardhan

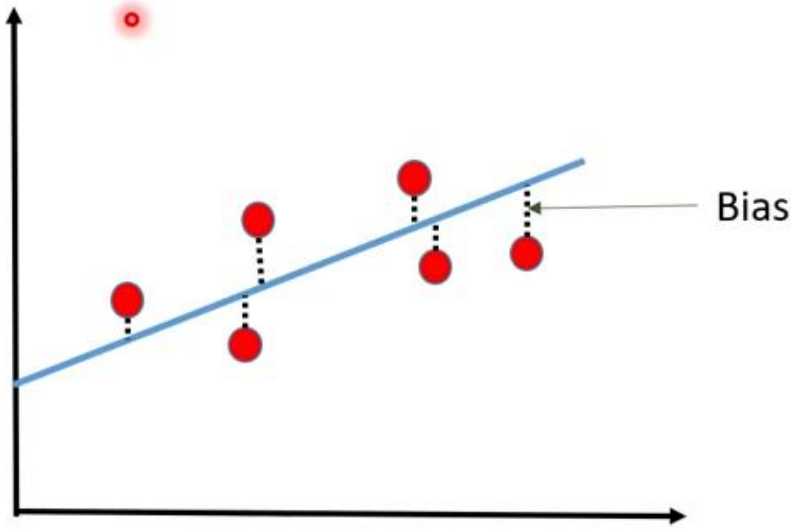
Bias – Variance Tradeoff in Machine Learning

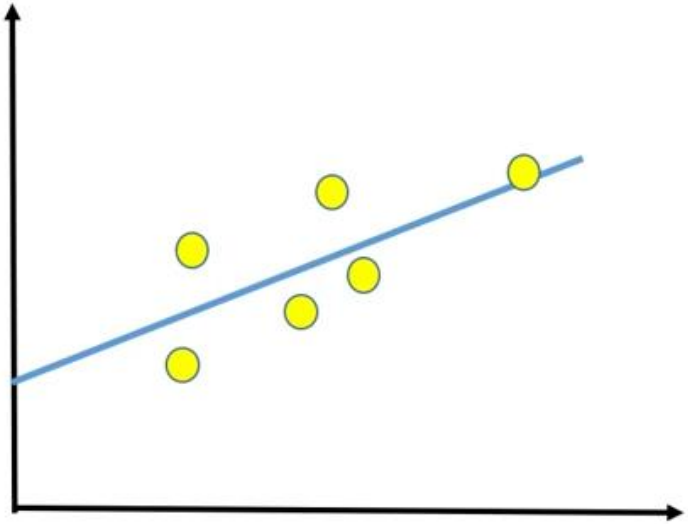


Bias – Variance Tradeoff

Bias :

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

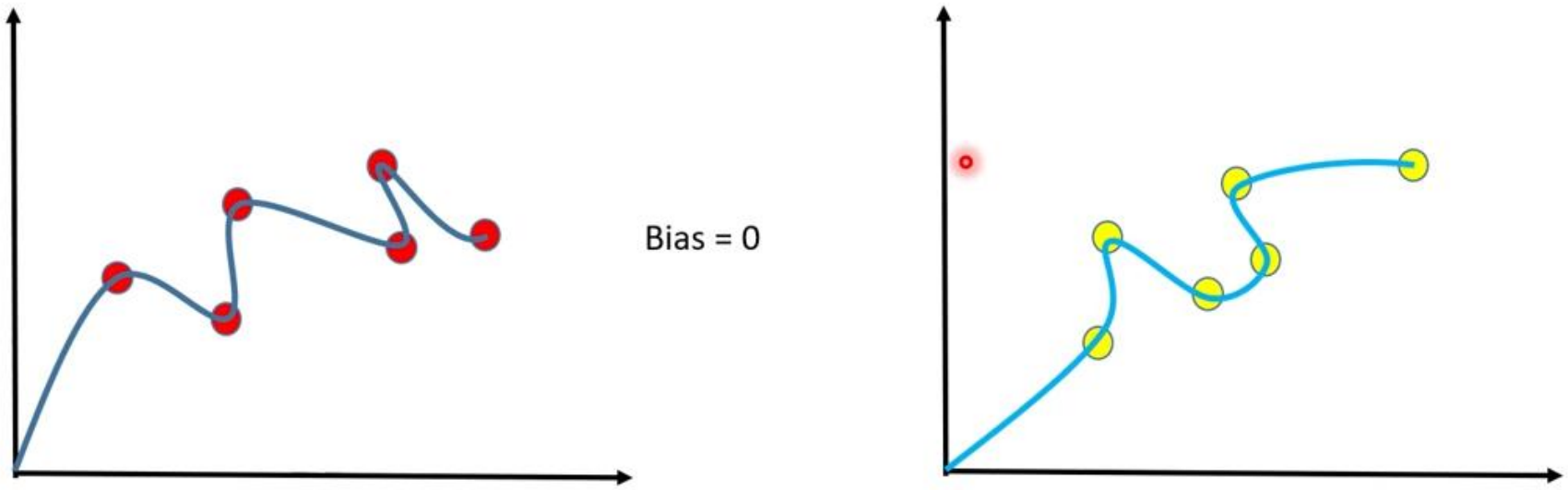




Bias – Variance Tradeoff

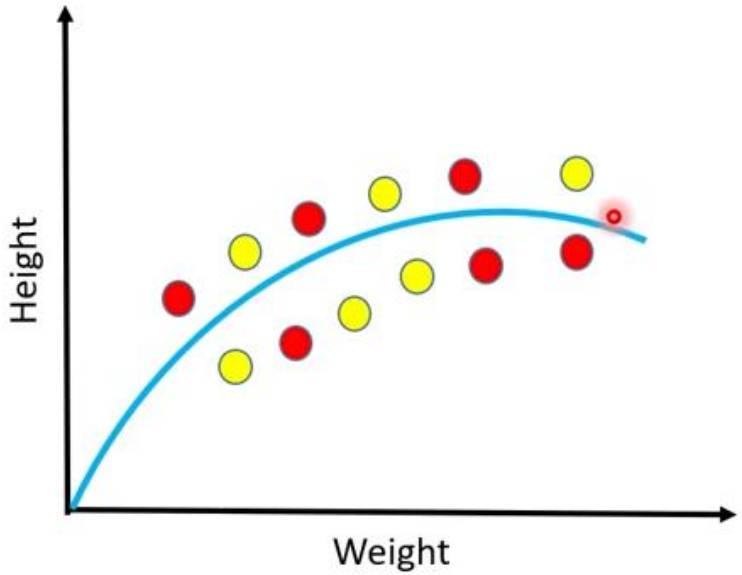
Variance :

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



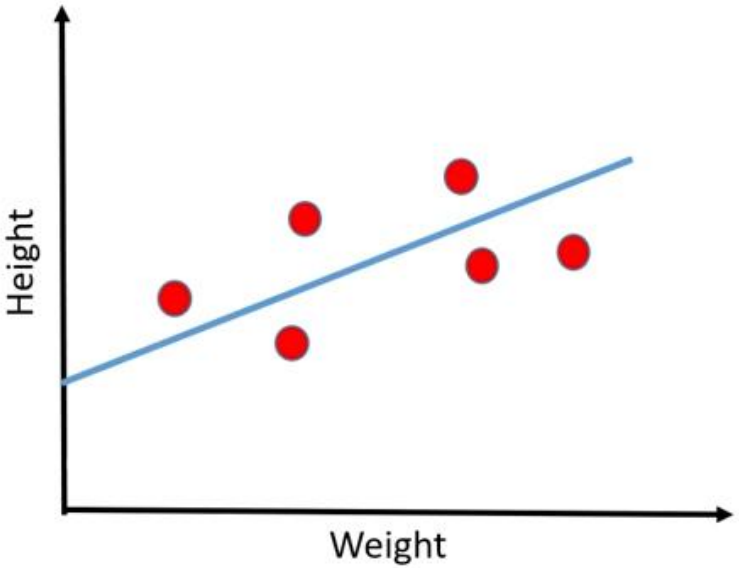
Bias – Variance Tradeoff

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

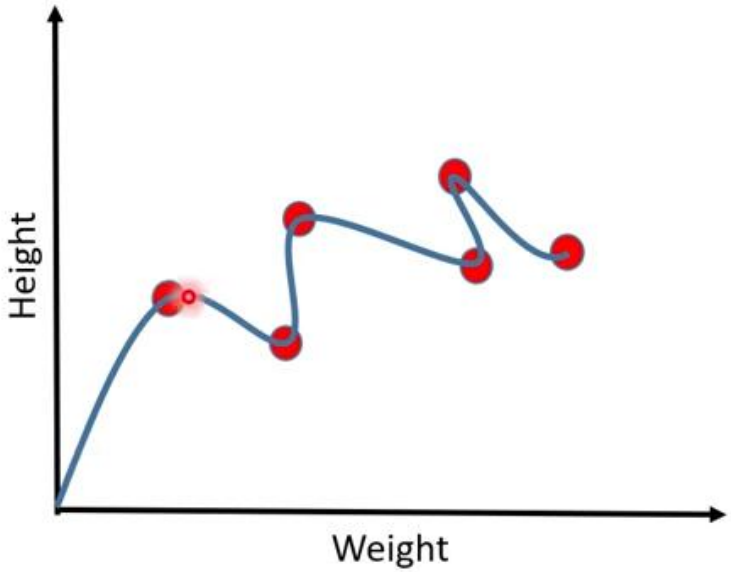


Underfitting & Overfitting

(Plot on training data)



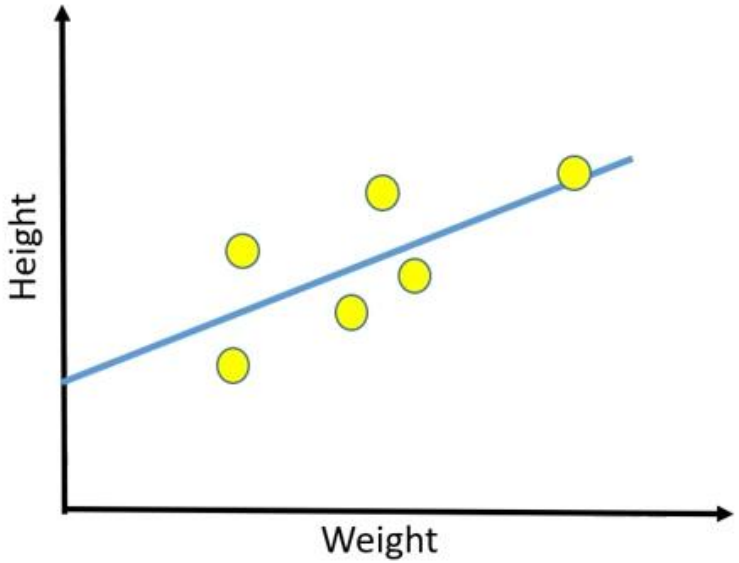
(i) Underfitting



(ii) Overfitting

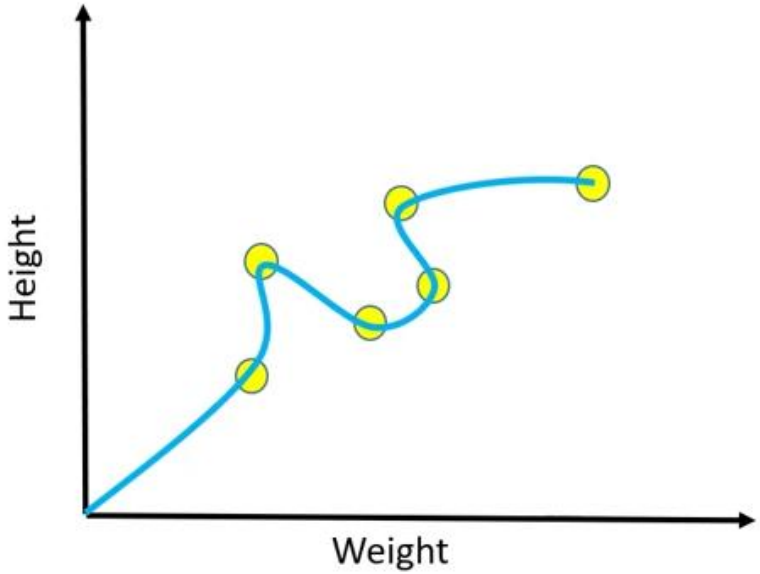
Bias – Variance Tradeoff

(Testing with different data)



(i) Underfitting

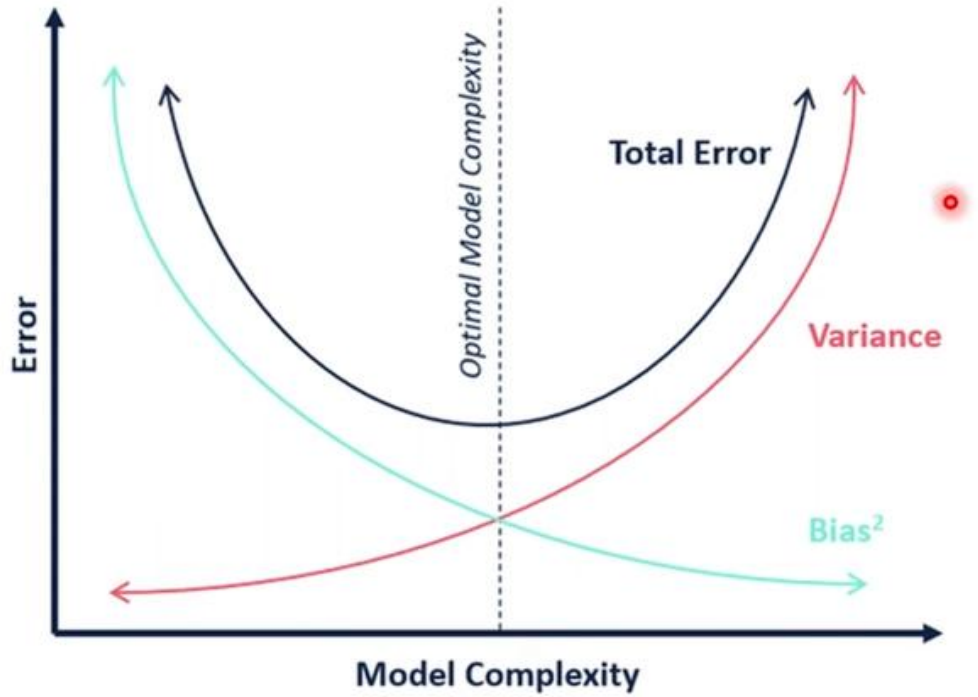
Inference:
a. *High Bias*
b. *Low Variance*



(ii) Overfitting

Inference:
a. *Low Bias*
b. *High Variance*

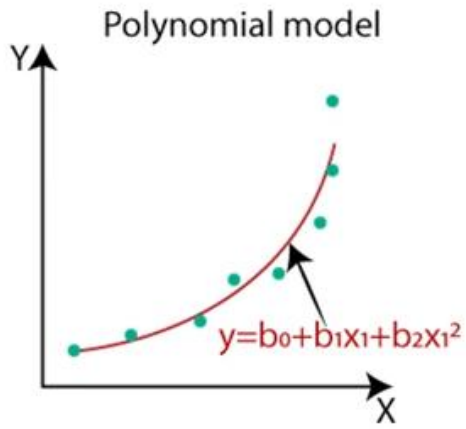
Bias – Variance Tradeoff



Bias – Variance Tradeoff

Techniques to have better Bias – Variance Tradeoff :

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



Siddhardhan

Loss Function in Machine Learning

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Loss Function

Loss Function

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

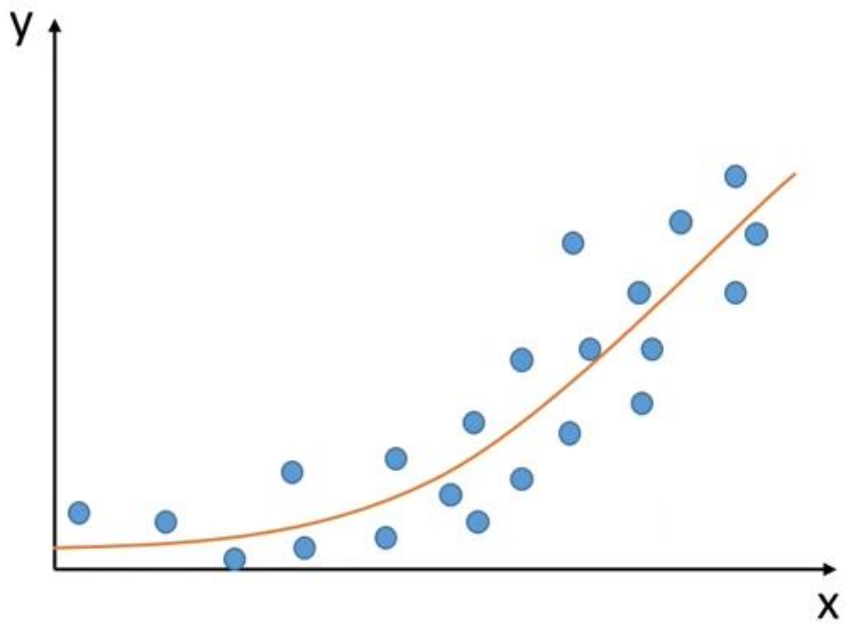
It is helpful to determine which model performs better & which parameters are better.

$$\textbf{\textit{Loss}} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Types of Loss Function:

- ❖ Cross Entropy Loss
- ❖ Squared Error Loss
- ❖ KL Divergence

Loss Function



$y = 0.0000003x^3 + 0.0002x^2 + 0.010x + 0.025$

Degree 3 Polynomial

Loss Function

$y_1 = 0.0000015x^3 + 0.0042x^2 + 0.020x + 0.035$

$y_2 = 0.0000023x^3 + 0.0001x^2 + 0.015x + 0.020$

$y = 0.0000003x^3 + 0.0002x^2 + 0.010x + 0.025$

$y_3 = 0.0000045x^3 + 0.0003x^2 + 0.040x + 0.028$

x	y	y_1	y_2	y_3
0.30	0.35	0.38	0.39	0.41
0.45	0.48	0.45	0.47	0.56
0.50	0.55	0.59	0.58	0.63
0.55	0.63	0.65	0.69	0.70
0.66	0.72	0.75	0.78	0.78

Loss Function

x	y	y_1	y_2	y_3
0.30	0.35	0.38	0.39	0.41
0.45	0.48	0.45	0.47	0.56
0.50	0.55	0.59	0.58	0.63
0.55	0.63	0.65	0.69	0.70
0.66	0.72	0.75	0.78	0.78

$$\textbf{Loss} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Loss_1 = [$(0.35-0.38)^2 + (0.48-0.45)^2 + (0.55-0.59)^2 + (0.63-0.65)^2 + (0.72-0.75)^2$] / 5

Loss Function

x	y	y_1	y_2	y_3
0.30	0.35	0.38	0.39	0.41
0.45	0.48	0.45	0.47	0.56
0.50	0.55	0.59	0.58	0.63
0.55	0.63	0.65	0.69	0.70
0.66	0.72	0.75	0.78	0.78

$$\textbf{Loss} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$Loss_1 = [\ (0.35-0.38)^2 + (0.48-0.45)^2 + (0.55-0.59)^2 + (0.63-0.65)^2 + (0.72-0.75)^2\ ]\ /\ 5$

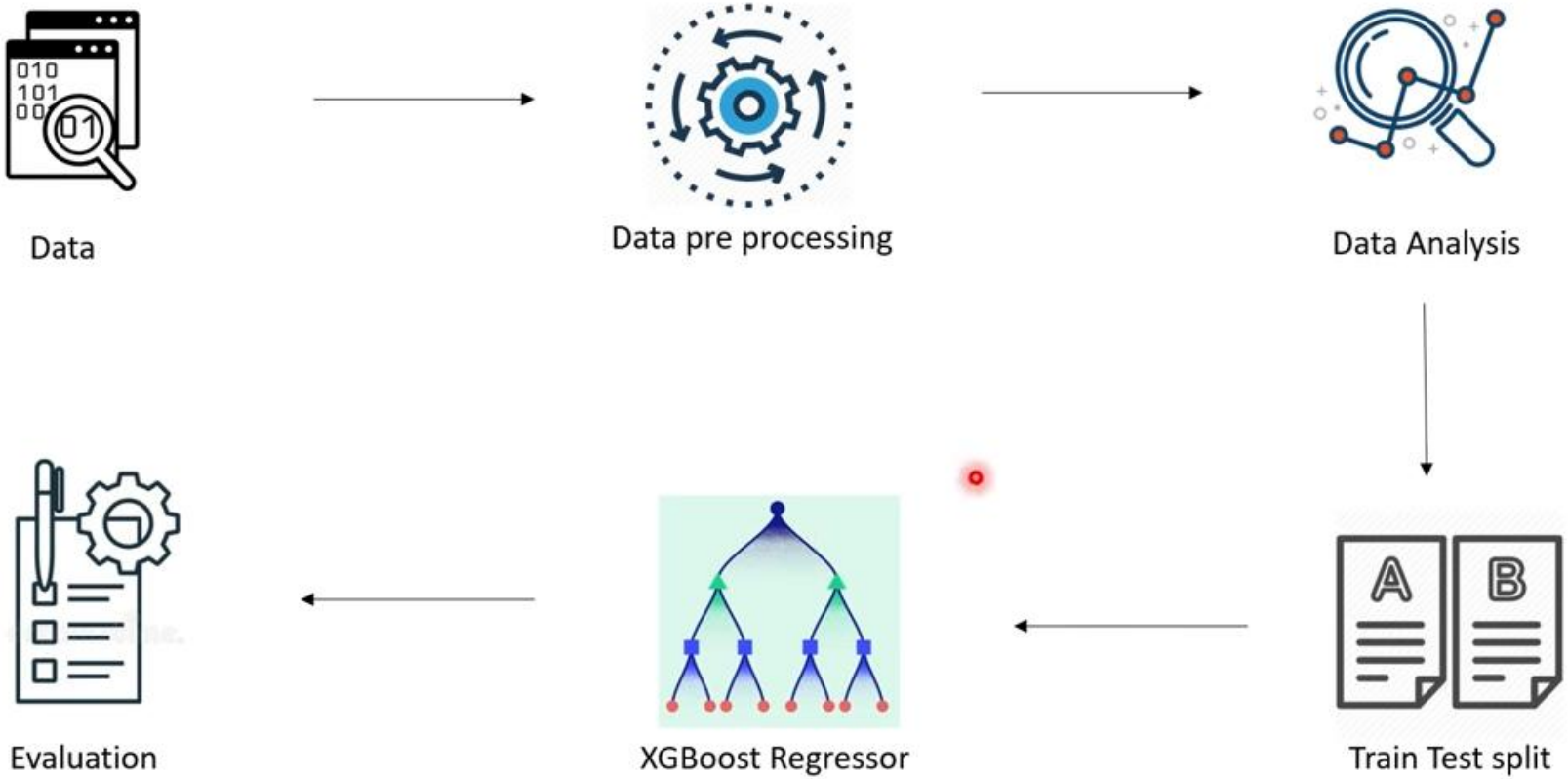
$Loss_1 = 0.173$

Low Loss value → High Accuracy

Siddhardhan

Model Evaluation in Machine Learning

Work Flow of a ML Project



Types of Supervised Learning

Classification

Classification is about predicting a class or discrete values
Eg: Male or Female; True or False

Evaluation metric for Classification: **Accuracy score**

Regression

Regression is about predicting a quantity or continuous values
Eg: Salary; age; Price.

Evaluation metric for Regression: **Mean Absolute Error**

Accuracy Score

In Classification, **Accuracy Score** is the ratio of **number of correct predictions** to the **total number of input data points**.

$$\text{Accuracy Score} = \frac{\text{Number of correct predictions}}{\text{Total Number of data points}} \times 100\ \%$$

Number of correct predictions = 128

Total Number of data points = 150

Accuracy Score = 85.3 %

```
from sklearn.metrics import accuracy_score
```

Mean Squared Error

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

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

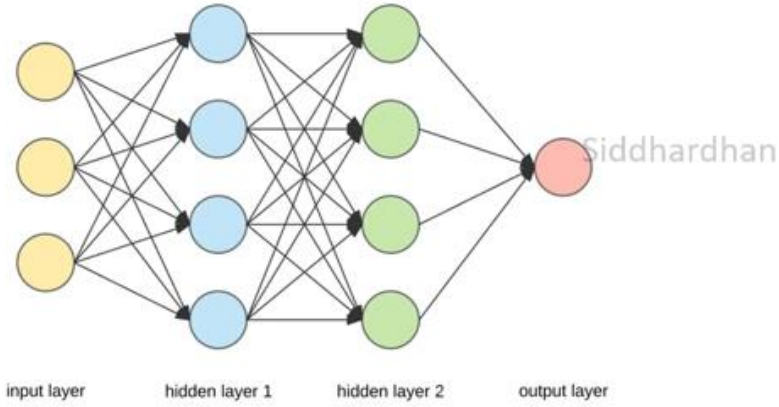
Actual Value (Y_i = 140 mg/dL)

Predicted Value (Y_i = 160 mg/dL)

```
from sklearn.metrics import mean_squared_error
```

Siddhardhan

Model Parameters & Hyperparameters



Types of Parameters

Model Parameters

These are the parameters of the model that can be determined by training with training data. These can be considered as internal Parameters.

- **Weights**
- **Bias**

Y = w*X + b

Hyperparameters

Hyperparameters are parameters whose values control the learning process. These are adjustable parameters used to obtain an optimal model. External Parameters.

- **Learning rate**
- **Number of Epochs**

Model Parameters

Weights: Weight decides how much influence the input will have on the output.

Applicant's Details

Name	Degree	College	C	C++	Python	Height	Weight	No. of Backlogs
A	B.E	ABC college	✓	✗	✓	165	72	1
B	M.E	XYZ College	✓	✓	✗	168	80	0
C	M.C.A	State College	✓	✗	✗	175	67	0
D	B.E	ZYX College	✓	✓	✓	168	70	2

✗ ✓ ✓ ✓ ✓ ✓ ✗ ✗ ✓

Model Parameters

Weights:

Weight decides how much influence the input will have on the output.

$$Y = w*X + b$$

$$Y = w_1*X_1 + w_2*X_2 + w_3*X_3 + b$$

X – feature or input variable
Y – Target or output variable
w – weight
b – bias

Bias:

Bias is the offset value given to the model. Bias is used to shift the model in a particular direction. It is similar to a Y-intercept. 'b' is equal to 'Y' when all the feature values are zero.

Model Parameters

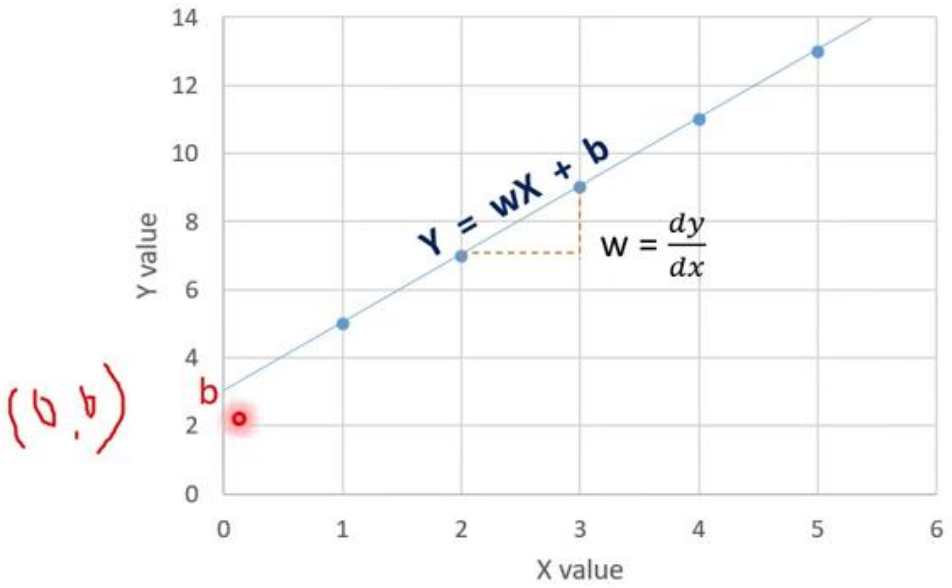
Weights: Weight decides how much influence the input will have on the output.

Applicant's Details

Name	Degree	College	C	C++	Python	Height	Weight	No. of Backlogs
A	B.E	ABC college	✓	✗	✓	165	72	1
B	M.E	XYZ College	✓	✓	✗	168	80	0
C	M.C.A	State College	✓	✗	✗	175	67	0
D	B.E	ZYX College	✓	✓	✓	168	70	2

✗ ✓ ✓ ✓ ✓ ✓ ✗ ✗ ✓

Linear Regression



Y = wX + b

X --> X value
Y --> Y value
w --> weight
b --> bias

Bias:

Bias is the offset value given to the model. Bias is used to shift the model in a particular direction. It is similar to a Y-intercept. 'b' is equal to 'Y' when all the feature values are zero.

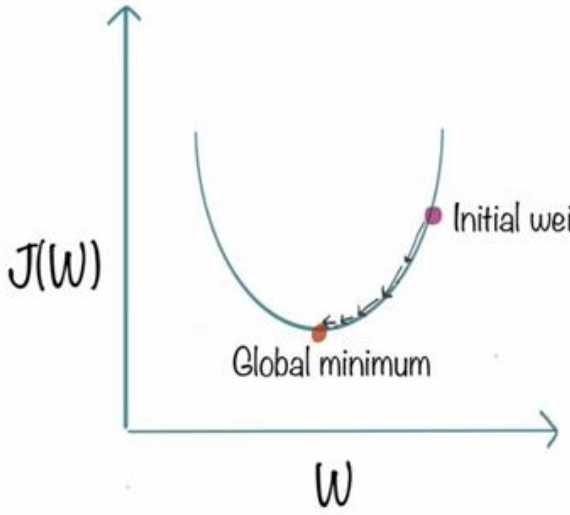
Hyperparameters

Learning Rate:

The **Learning Rate** is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function.

Number of Epochs:

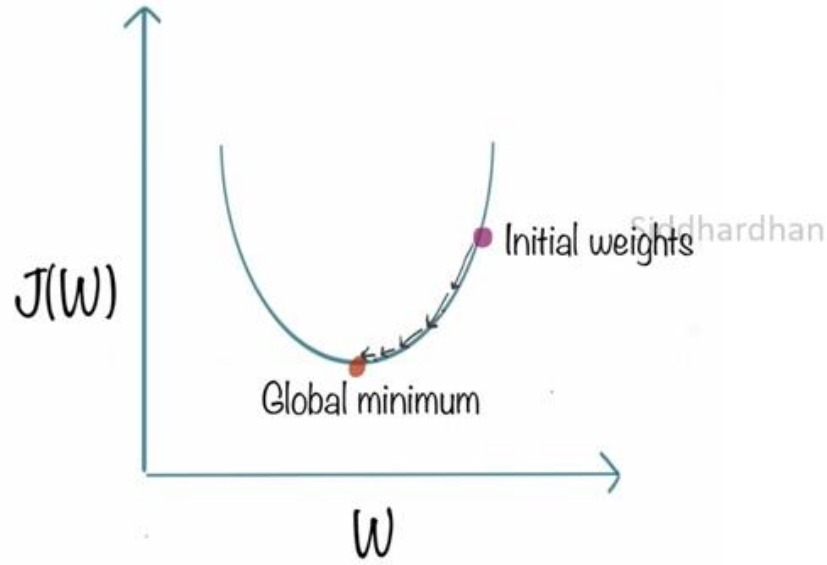
Number of Epochs represents the number of times the model iterates over the entire dataset.



Gradient Descent

Siddhardhan

Gradient Descent in Machine Learning



Loss Function

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

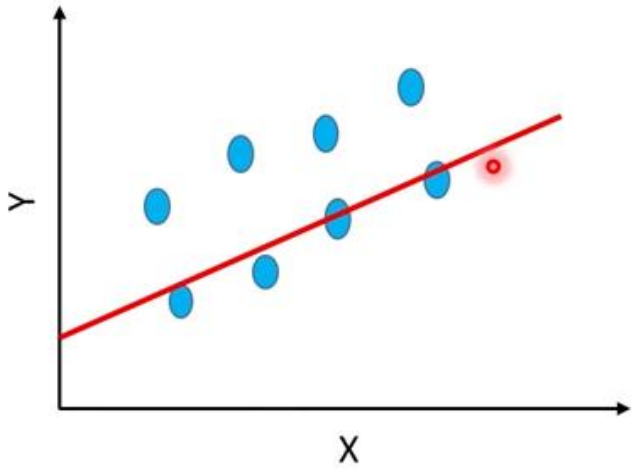
It is helpful to determine which model performs better & which parameters are better.

Siddhardhan

$$\textbf{Loss} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Model Optimization

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

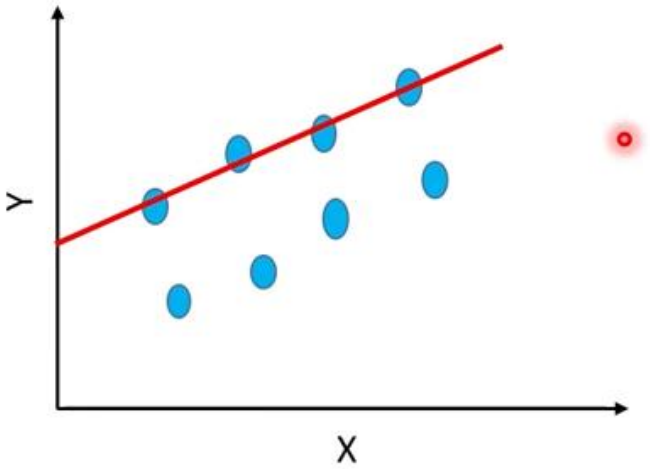


$Y = w_1X + b_1$

(w_1 & b_1 are the parameters of the line)

Model Optimization

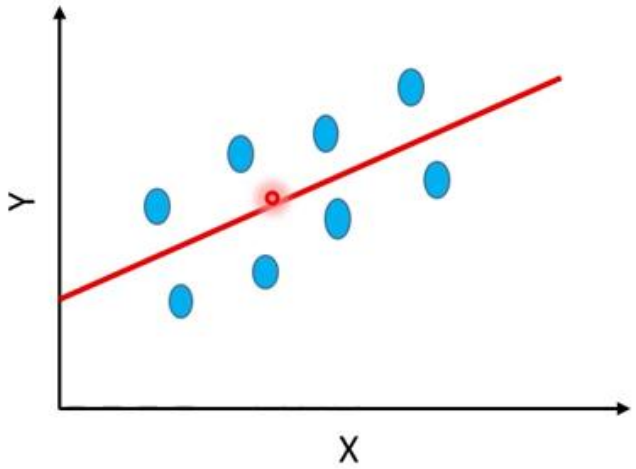
Optimization refers to determining best parameters for a model, such that the loss function of the model decreases, as a result of which the model can predict more accurately.



$$Y = w_2X + b_2$$

Model Optimization

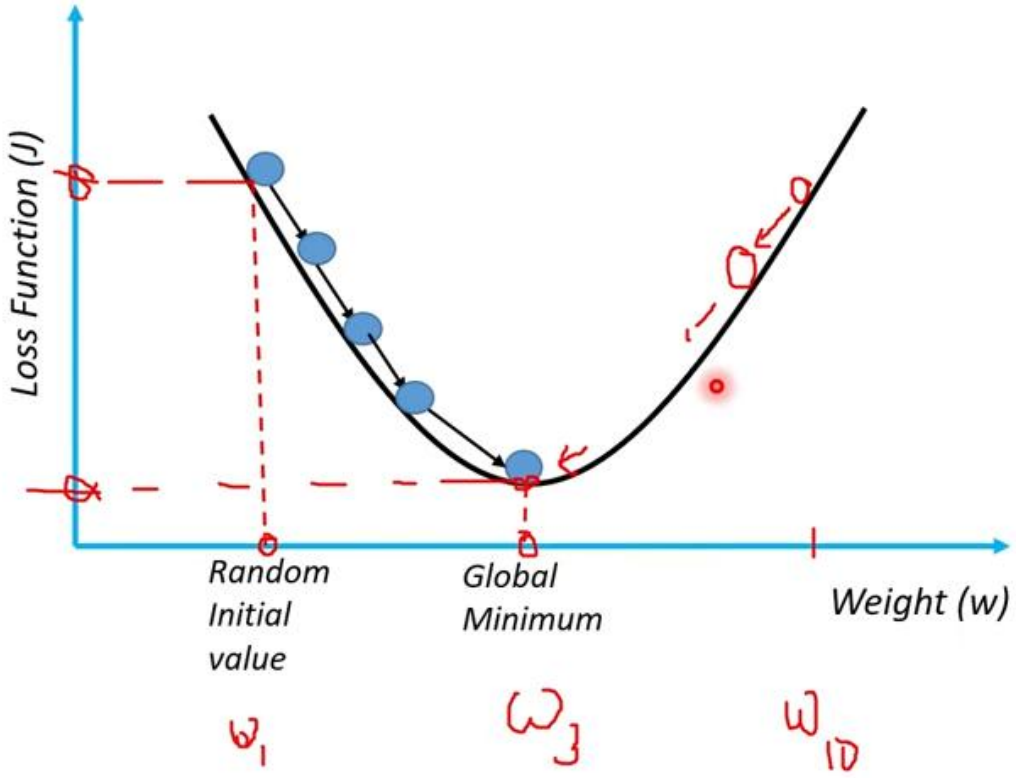
Optimization refers to determining best parameters for a model, such that the loss function of the model decreases, as a result of which the model can predict more accurately.



$Y = w_3X + b_3$

Hence, w_3 & b_3 are the best parameters

Gradient Descent



Gradient Descent

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

$$w = w - L*dw$$

$$b = b - L*db$$

w --> weight
b --> bias
L --> Learning Rate

dw --> Partial Derivative of loss function with respect to m ~~m~~ w

db --> Partial Derivative of loss function with respect to c ~~c~~ b

Gradient Descent in 3 Dimension