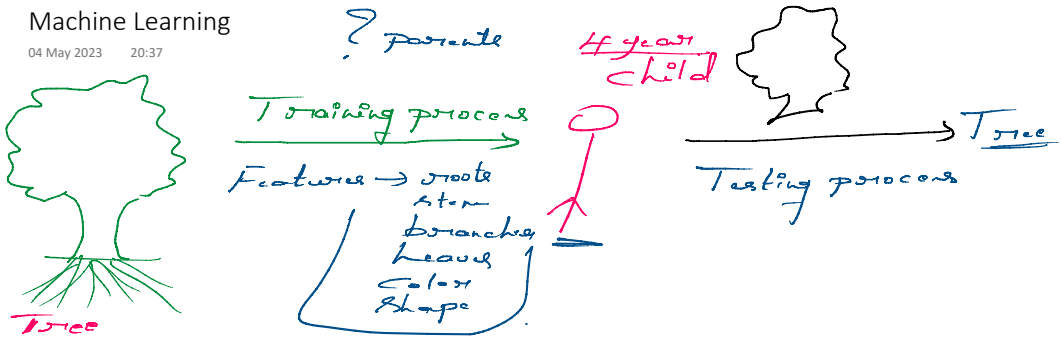
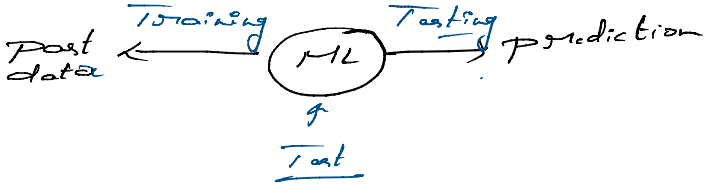
Machine Learning

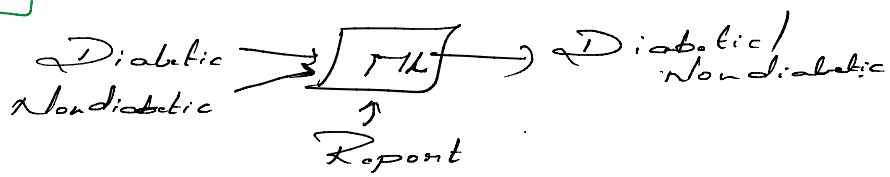
04 May 2023 20:37



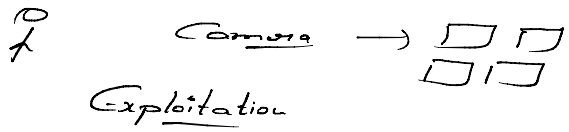


Def of Machine Learning

Machine Learning is branch of Computer science which has several machine learning algorithm which learns from past data and make prediction.



Amazon Return Policy



1. Customer history
   Machine Learning → It learns Customer history
2. Recommendation
   Google, Youtube, Netflix, Facebook, LinkedIn
3. Google Assistant, Siri
4. Chat GPT
5. Self driven
6. Yolo

Types of Machine Learning

4 = x

10/10

Types of Machine Learning

* Supervised
* Unsupervised
* Reinforcement

$y = x$

x = input / features

y = output / target / label

$x \rightarrow$ independent variable

$y \rightarrow$ dependent variable.

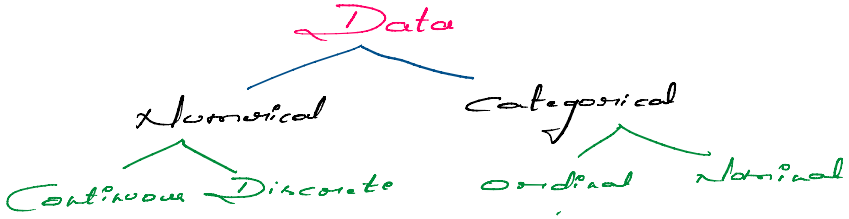
Supervised Learning

If we train machine learning model with both input and output variable.

Types of Supervised Learning

It is classified based on datatype of target variable

- Regression
- Classification



Regression

Regression algorithms are used when target variable is continuous.

Eg: House price data

x_1 BHK	x_2 Sqft	x_3 No of floors	x_4 Location	y Price
				2.8
				3.4
				5.2
				1.2

Continuous

Types of Regression

1. Linear Regression ✓
2. Time series forecasting..
3. KNN Regressor
4. SVM Regressor
5. Decision Tree Regressor
6. XG Boost Regressor
7. Gradient descent Regressor
8. Random Forest Regressor.

Classification

Classification algorithms are used when target variable is categorical or discrete.

Eg: Diabetic Dataset

x_1 RMT	x_2 BP	x_3 Insulin	x_4 Glucose	y Result

Eg:- Diabetic Dataset

x_1 BMI	x_2 BP	x_3 Insulin	x_4 Glucose	y Result
				Diabetic / NonDiabetic / Diabetic

Result: Diabetic, NonDiabetic, Diabetic — Categorical

Classification Algorithms

1. Logistic
2. KNN
3. SVM
4. NaiveBayes
5. Decision Tree
6. Random Forest
7. XGBoost
8. Gradient Boosting
9. Bagging

Classify given data into Regression or classification

1. Flight price prediction
   Target → price → Regression.
2. Covid19 prediction.
   Target < positive / Negative → Classification
3. Loan amount prediction.
   Amount → Regression.
4. Cancer dataset → Classification
5. Heart attack prediction → Classification
6. Liver disease prediction → Classification
7. Sales prediction → Regression
8. Iris data. → Classification
9. Email classification → Classification

Linear Regression

* It comes under supervised learning [both x and y].
* Linear regression is regression algorithm which is used to predict the "continuous target variable".
* Linear Regression is used to find linear relationship between dependent(y) and independent variable(x) by fitting straight line.

Straight line eqn
$y = mx + c$
m → slope
c → y-intercept

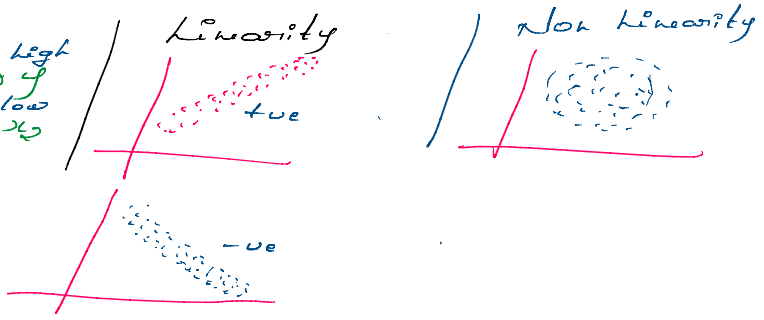
Slope / Gradient
$m = \frac{dy}{dx}$ → unit change in x results in change in y

Assumption of Linear Regression

* Linearity → Linear relationship b/w x and y / High correlation between x and y
* Normal distribution → Data should be normal.
* Little or no multicollinearity.
  → There should not be high correlation between independent variables
  x_1 x_2
  Corr should be less

Eg: Area / sqft

x Note: Corr high $x \rightarrow y$; Corr low $x_1 \rightarrow x_2$



How Linear Regression works?

Data

x: No of Hours Studied	y: Marks obtained
4	50
7	80
10	92
6	75
8	85
9	?

Qs: what will be the marks of student studies for 9 hours?

Linear regression learns data by fitting / drawing straight line.

$y = mx + c$

$m \rightarrow$ slope

$c \rightarrow$ intercept.

→ Linear regression will find slope and intercept.

$m = ?, c = ?$

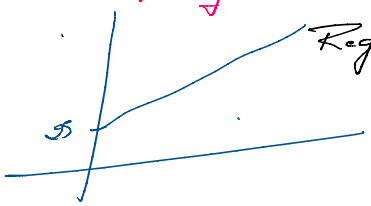
$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$c = \bar{y} - m\bar{x}$$

$\bar{x}, \bar{y}$ are mean of x and y

→ $m = 7, c = 25$

→ $\hat{y} = 7x + 25$ → This equation is called as Best fitline / Best model / Regression model / Regression line



Prediction

$\hat{y} = 7x + 25$

$x = 9$

$\hat{y} = 7(9) + 25$

$\hat{y} = 88$ → prediction made by Regression model

Find marks of x = 4

$\hat{y} = 7x + 25$

$\hat{y} = 7(4) + 25$

$\hat{y} = 53$

How do you say model is Best or not?

→ If error is less then it's best model otherwise it's bad model.

x	actual y	predicted $\hat{y}$	error
4	60	58	2
2	30	31	1
10	92	90	2
8	85	80	5

Imp

Error → Difference between actual value and predicted value

Error = how: $y - \hat{y}$

How to calculate average error / Total error?

How to calculate average error / Total error?

① Mean Squared error (MSE)

$$MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

MSE is very important error function.

② Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum|y_i - \hat{y}_i|$$

③ Root mean squared error (RMSE)

$$RMSE = \sqrt{MSE}$$

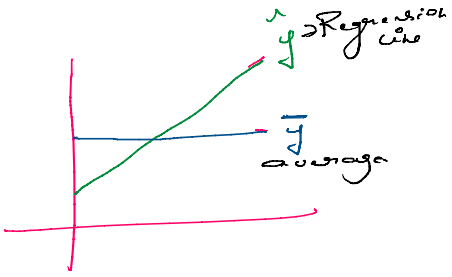
$y_i \rightarrow$ actual value
$\hat{y}_i \rightarrow$ predicted value
$n \rightarrow$ no of observations

Imp
* MSE, MAE, RMSE are the error functions which tells about total error made by the model.
* Model is Best only if total error is less.
* Error = actual − predicted.

How to evaluate model?
How to measure performance of model.

→ R-Squared / R2-Score / R^2-Score

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$



RSS = Residual Sum of square

$$RSS = \sum(y_i - \hat{y}_i)^2$$

TSS = Total sum of square

$$TSS = \sum(y_i - \bar{y})^2$$

R^2 Compares Regression line with average line.

$R^2 = 90\% \rightarrow$ 90% better model

Range of R^2: [−1, 1]

$R^2 = 1 \rightarrow$ Best model

$R^2 = 0 \rightarrow$ Bad model

Types of Regression

① Simple Linear Regression
If data has single input variable, then straight line is $y = mx + c$

② Multi Linear Regression
If data has multiple input variables, then straight

(3) Multi Linear Regression

If data has multiple input variables, then straight line is given by

$$y = m_1 x + m_2 x_2 + m_3 x_3 + C \ldots$$