BEFORE ML

(PREREQUISTES)

① LEARNING ALGO

SUPERVISED LEARNING
↓
Labelled Data (Target ✓)
↓
model trained on label data
- → Classification
- → Regression

UNSUPERVISED LEARNING
↓
Unlabelled Data (Target ×)
↓
Model ~~learn~~ Train-ed on unlabell-ed Data
↓
Algo try to detect Pattern, relationship in data w/o explicit supervision
- ↳ Clustering
- ↳ Dimensionality Reduction (PCA)

REINFORCEMENT LEARNING.
↓
Model learns from its Environment
↓
Agent learn to make decision by interacting with env. & Recieves Rewards or Penalities Based on its action & Try to maximize Reward
- ↳ Q-Learning
- ↳ Policy Based
- ↳ DQN.

* DeepSeek is trained on Deep Q-Network (DQN) specifically Q-learning which is value based reinforcement learning

(11) ML Algo Purely For CLASSIFICATION

(I) LOGISTIC REGRESSION
(II) NAIVE BAYES

ML Algo Purely for REGRESSION

(I) Linear Regression
(II) LASSO REGRESSION
(III) RIDGE REGRESSION
(IV) ELASTIC NET (L1 + L2)
LASSO + RIDGE

ML Algo For BOTH CLASSIFICATION & REGRESSION

(I) KNN
(II) Decision Tree
(III) Random Forest.
(IV) SVM
(V) ADABOOST
(VI) XGBOOST

PAGE NO.:

(iii) Classification Problem ⇒ Output is out of fixed categorical label

Regression Problem ⇒ output is contineous (Real) value

(iv) LOSS FXN FOR EACH ML ALGO

Loss function is also known as error function and it measures the value b/w model prediction & actual value.

Goal of ML algo is to Minimize loss function

i) LINEAR REGRESSION → MSE

$$MSE = \frac{1}{n} \sum (\underbrace{y}_{\text{actual}} - \underbrace{\bar{y}}_{\text{predicted}})^2$$

ii) LOGISTIC REGRESSION ⇒ BINARY CROSS ENTROPY

(BCE)

~~log loss = $-\frac{1}{n}\sum_{i=1}^{n}$ [pred log(pred) +~~

$$\text{log loss} = -\frac{1}{n}\sum_{i=1}^{n}\left[\text{actual}\log(\text{pred}) + (1-\text{actual})\log(1-\text{pred})\right]$$

III) SVM → Hinge Loss = $\sum_{i=1}^{n} \max(0, 1-y_i$

$$\text{Hinge Loss} = \sum_{i=1}^{n} \max(0, 1-\text{actual} \times \text{pred})$$

IV) Decision Tree Classification

↳ Gini Impurity or Entropy (Cross Entropy)

$$\text{Gini Entropy (default)} = 1-\sum_{i=1}^{K} p_i^2$$

$$\text{Entropy} = -\sum_{i=1}^{K} p_i \log(p_i)$$

V) D.T for Regression: MSE

vi) KNN ⇒ No Loss function

vii) K-Mean Clustering ⇒ Square Eucledian Distance

viii) PCA ⇒ Reconstruction Error.

ix) Naive Bayes ⇒ Maximum Likelihood Estimation

(V) MODEL TO USE WHEN

(I) Linearly Seperable Data

i) Logistic Regression
ii) SVM with Linear Kernel
iii) KNN.
--> Can also use other
v) Linear Regression

(II) Non-Linear Seperable Data

i) D.T
ii) Random Forest
iii) KNN
iv) SVM with Non-linear Kernel (RBF)
v) Neural Network

VI HOW TO KNOW DATA IS LINEARLY SEPERABLE OR NOT. (EDA or Data Visualisation)

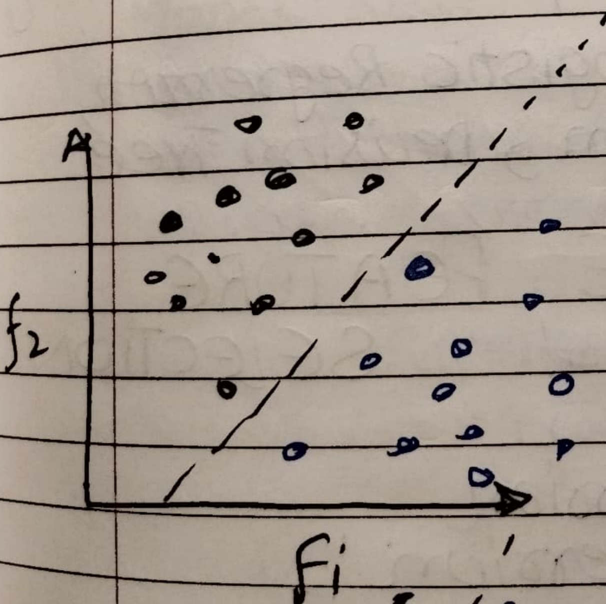
↳ 1. Use Pairplot.

↳ If every two feature based on target.

Pairplot -> Multiple Scatter plot.

sns.pairplot(df, hue='Target')

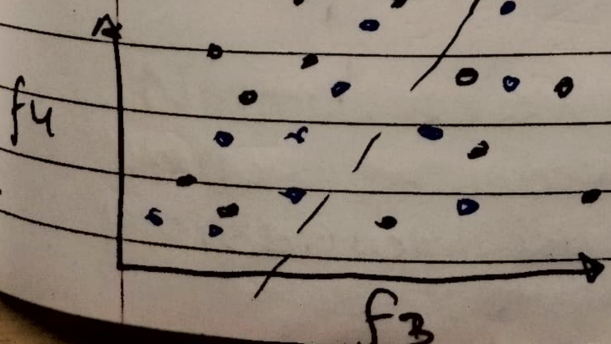
↳ x_train



=> Linearly seperable

• → Target => 0

• => Target => 1



=> Non-Linearly Separable

• => Target => 0

• => Target => 1

PAGE NO.:

★ If you found some linearly seperable feature and you want to apply Linear Model. Do feature selection only feed that feature to the model that are linearly seperable

(VII) MODEL ACC. TO THEIR T.C

PCA, GC NN, CNN, RNN, Transformers, SVM (RBF) > Random Forset, KNN, SVM (Linear Kernel), K-Mean Clustering > Naive Bayes, Logistic Regression, Linear Regression, Decision Tree.

(VIII) MODEL REQUIRE FEATURE SELECTION

i) Linear Regression
ii) Logistic Regression
iii) SVM
iv) KNN.

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

PAGE NO.:

(ix) NORMALIZATION (MIN-MAX SCALING)

O/P ⇒ (0 to 1)

→ Normalization means scaling the data in the range of 0 and 1.

When to Use Normalization?

↳ If your model is sensitive to feature magnitude and uses Distance Based Calculation

MODEL REQUIRE DATA NORMALIZATION?

↳
(I) Neural N/ws
(II) K-Mean Clustering
(III) KNN
(IV) SOM -> self organizing map
(V) Linear Regression or logistic Regression
(VI) SVM

Normalization Benefits

↳ (I) Improve Model performance
(II) Faster Convergence.

Important

PAGE NO.:

---→ Before Applying Normalization you need to remove Outlier as they can skew the Data else you can use Robust Scaler

Remove Outlier
↓
Normalization (Min-Max) (0→1)

ⓧ STANDARDIZATION

(Z-SCORE NORMALIZATION)

↳ Process of scaling your data to have mean 0 & standard deviation 1.

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

X → datapoint

μ → mean

σ → standard deviation
↳ spread of feature values

PAGE NO.:

$$\text{variance}(\sigma^2) = \frac{1}{n} \sum_{i=1}^{n} (x-\mu)^2$$

$$\text{standard Deviation } (\sigma) = \sqrt{\sigma^2}$$

(V) When to use Standardization?
↳ If your model is sensitive to the scale of data & uses Distance base Calculation

MODEL REQUIRE STANDARDIZATION?
↳
(I) Linear, Logistic Regression
(II) KNN
(III) K-Mean Clustering
(IV) SVM
(V) Neural Networks (NN)

Important
---→ Before Applying Standardization you need to remove outliers as they can skew the Data else you can use Robust scaler.

★ You either need to Perform Normali-zation or Standardization or

PAGE NO.:

Robust Scaling not all or both.

--→ <u>Outlier Sensitivity on Scaler</u>

Normalization > Standardization
> Robust Scaler.

★ For CNN, we perform Normalization
(min-max scaling)

★ On Sirs experience, on most of the ML algo Standardization perform well

★ ~~we will see features that need Normalization or Standardization & will only apply on that features~~

★ Algo uses Gradient Descent needs Normalization for faster convergence

XI OUTLIERS

↳ Outliers are the datapoint that deviate significantly from rest of the data in Dataset & can affect the model Training

For Example => Salary => $30,000 to $70,000
--> $500,000 is outlier

HOW TO IDENTIFY OUTLIERS

↳ (i) Box Plot
(ii) Scatter Plot

Imp
★ It is very important to note that whether we need to remove outlier or not cuz In some problem outlier are very important
For e.g => Fraud Detection System
=> Breast Cancer Detection
(use Robust Scaling)
=> Earth Quake Prediction
=> Stock Price Prediction

Outlier Not Important
↳ (i) House Price Prediction
(ii) Sensor Data in IOT
(iii) Retail Sales Data

Important
★★ Analyze yourself whether outliers in that particular problem are important or not & also ask chatgpt about that

HOW TO REMOVE OUTLIERS

① Use Z-Score

$$Z = \frac{X-u}{\sigma}$$

if $Z > 3$ or $Z < -3$
↳ It is an Outlier

(XII) ALGO IMPACTED BY OUTLIER

↳ (I) Linear Regression
↳ (II) Logistic Regression
↳ (III) K-Nearest Neighbour (KNN)
↳ (IV) SVM
↳ (V) PCA

(XIII) BENEFITS & PROBLEM OF EACH ML ALGO

i> Linear Regression
↳ Benefit
↳ (I) Less Time Complexity
↳ (II) Works well with linearly seperable data

PAGE NO.:

Problem
↳ (I) Sensitive to Outlier
↳ (II) Can't solve non-linearly seperable problem

ii) Logistic Regression

Benefit
(I) Great for Binary classification
(II) works well with linearly seperable data

Disadvantage
↳ (I) Struggle with non-linear Data
↳ (II) Sensitive to Outlier.
↳ (III) Sensitive to Correlated features

Q How to identify correlated features & what to do with them in Logistic Regression?

i) Identify correlated feature
↳ i) x = df.corr()
ii) sns.heatmap(x)
iii) Correlation = 1 or -1
↳ Highly Correlated

ii) Handling correlated feature

↳ (I) Remove One correlated feature

or

(II) Apply PCA.

⇢ Before applying PCA data is standardized

III) Decision Tree

Benefits

↳ (I) Don't Require Normaliz-ation or Standardization of Data

(II) Not much need of Removing outlier

(III) Can capture non-linear Relationship

(IV) Less T.C.

(V) Can handle missing value.

(VI) No Need of feature selection

Problem

↳ (I) Prone to overfitting

(II) Poor Performance when imbalanced data

PAGE NO.:

IV) Random Forest

Benefits

can still overfit if not properly tuned

↳ I Reduces Overfitting
II No Need of feature Selection
III No Need of feature Scaling.
IV Can capture non-linear relation
V Can handle missing value (null value)

Problem

↳ I Slower litlle bit (T.C)
↳ II Harder to interpret

V) KNN

Benefit

↳ I Work well with Multiclass Classification
↳ II Can also be used for non-linear Data.

Problem

↳ I Computionally expensive for Large Dataset.
↳ II Need Proper feature selection.

PAGE NO.:

↳ (II) Need Standardization or Normalization

Q How to Do Feature selection for KNN?

↳ (I) Check which according to you are important feature for prediction

(II) Remove Highly Correlated feature

↳ i) PCA

ii) remove one of out of them

↳ if corr > 0.9

VI) NAIVE BAYES

Benefits

↳ (I) Fast & Efficient.

↳ (II) Works well with Text Data (Spam Classification)

Problem

↳ (I) Assume feature indepe-ndence

(II) Struggle with smaller Datasets

VII) SVM

Benefits

↳ (I) Works well with for many features

(II) No Need of feature Selection

↳ Perform only if

i) Features > 1000

ii) Feature Highly Correlated

iii) Model overfitting.

(III) Can handle non-linear Data.

Problem

↳ (I) Slow for Large Dataset (T.C)

(II) Need Standardization and Normalization.

↳ (III) Harder to Tune Hyper Parameter.

VIII) PCA

Benefits

↳ (I) Reduce Dimensionality while keeping max. variance

↳ (II) Removes feature correlation.

Problem
↳ (I) T.C
(II) Harder to interpret transformed feature
(III) Sensitive Outliers

(XIV)

HYPERPARAMETER FOR DIFFERENT MODEL

(I) Linear Regression → No hyperparameter

(II) Logistic Regression → No hyperparameter

(III) Decision Tree

~~(I) Max Depth:~~

(I) max_depth: Max. depth of Tree (To prevent overfitting)

(II) criterion: gini or entropy (loss fxn)

(III) min_samples_split: min. samples needed to split a node into child node.

(IV)

PAGE NO.:

(IV) Random Forest

(I) n_estimators : No. of Trees in forest
(II) max_depth :
(III) criterion

(V) KNN

(I) n_neighbors : No. of Neighbors to consider (K)
(II) metric : euclidean or manhattan (Default)

(VI) SVM

(I) kernel : linear, poly, rbf, sigmoid
(II) C :
(III) gamma
(IV) degree

(VII) Naive Bayes : No hyperparameters

(I) var_smoothing

(VIII) PCA : (I) n_components : No. of P.C to keep

(IX) K-Mean Clustering

(I) n_clusters : No. of clusters
(II) init → Initialization method (K-Means ++, random)

(III) max_iter : Max. no. of iterations for convergence

(XV) PERFORMANCE METRICS

(I) CLASSIFICATION

↳ i) Accuracy
ii) Precision
iii) Recall
iv) F1-Score
v) ROC Curve

(II) REGRESSION

↳ (I) MSE
(II) MAE
(III) R^2
(IV) Adjusted R^2