(DS&ML)

DOMS | Page No.
Date / /

U-2 classification Methods

Support Vector Machine :

- supervised ML algo
- binary classification → goal is to separate datapoints into classes
- Hyperplane → finds it that best separates.
- Support vectors → datapoints closest to hyperplane
- Kernel Trick — SVM → handles nonlinear data by using kernel func.
- Multi Class Classifier.
  - sensitive to scal
  - strengths → effective in high dimensional spaces.
    - ↳ no. of features > no. of samples
    - robust against → overfitting → when tuned.
  - Demerits → expensive for large datasets
    - model interpretation → challenging.

Appln → image classifn, text categorization, etc

Hyperplane :

decision boundary divides input space into two or more regions.

- directly proportional to no. of features

DOMS | Page No.
Date / /

-optimal separating hyperplanes-

- correctly classifies all data which being farthest away from datapoint

Kernal functions -

- used to take data as input & transfer into required form of processing data.
- set of mathematical functions used in Support Vector Machine. providing window to manipulate data
- returns inner product b/w two points in standard feature dimension
- symmetric • smoothness • complexity • positive definite • non-negativity

Kernel Selection -

choosing appropriate kernel kernal fun. using kernel method.

1) Purpose.
- ↓
transform data in higher-dimensional space, to find complex, non-linear relationship

2) common Kernals
- ↳ Linear
- ↳ Polynomial
- ↳ Radial Basic Fun
- ↳ Sigmoid Kernel.

3) Kernel selection

4) Hyperparameter tuning.

5) Domain Knowledge

6) Experimentation

7) Performance Evaluation

DOMS | Page No.
Date / /

Ensemble methods

→ techniques → improve accuracy of results by combining multiple models instead of using single model.

→ Ideal for regression & classification reduce bias & variance to boost accuracy of models.

	Bagging	Boosting	Stacking
	Reduce Variance	Reduce Bias.	Improve accuracy
Base Learner Types	Homogeneous	Homogenous	Heterogeneous
Base Learner Training	Parallel	Sequential	Refer more.
Aggregation	Max Voting averaging	weighted Avg	Weighted avg
~~Decision Tree.~~			
Eg Algo	Decision Tree	XGBoost	

DOMS | Page No.
Date / /

Random forest

- consists of multiple random decision tree.
- Two types of randomness are built into trees.
  - 1) each tree is built on randdom sample from orgral data.
  - 2) each tree node → subset of feature are randomly selected to generate best split.

- greater no. of trees leads to higher accuray & prevents overfitting

- takes less training time when capared to others
- predicts output → high accuracy → even for large datasets.
- maintains accuracy → when large portion of data is missing
- Performs both classification & Regression task.
- can handle large dataset with high dimensionality
- used for both classif^n & Regression isn't suitable for Regression task

Demerits →
- Random Feature selection

Hyperparameter tuning → optimize model performace

Growing of Random forest

↳ process of creating multiple decision trees with ensemble.

- combines decision trees → improve accuray & reduce overfitting

Steps → 1) Data preparation 2) Bootstapping 3) Random Feature selection
4) Decision Tree Building 5) Parallelization
6) Ensemble Creation 7) Prediction Aggregation
8) Out of Bag Error estimation 9) Feature Importance.
10) Hyperparameter tuning

DOMS | Page No.
Date / /

* Random Feature Selection -

Technique → Random Forest
- → Prevents Overfitting
- → Improves robustness
- → Increases accuracy

steps → 1) Bootstrap Samples.

2) Random Subset of Fetures.

3) Diversity & Decorrelation.

4) Preventing Dominance

5) Hyperparameta Control

6) Balacing Bias & Variance

7) Feature Importance