08 April 2024 12:06

1

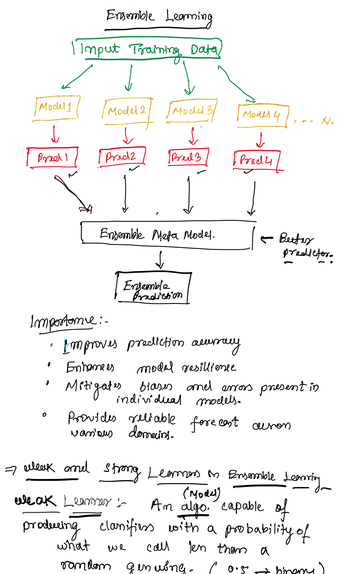
11th July. ML - Ensemble Learning.

⇒ Combines multiple models to enhance overall performance of ML algos.

→ Combine predictions from multiple models to produce more accurate and robust prediction.

Goals :-

- Enhance predictive accuracy by combining multiple models



Importance :-

- Improves prediction accuracy
- Enhances model resilience
- Mitigates biases and errors present in individual models.
- Provides reliable forecast across various domains.

⇒ weak and strong Learners in Ensemble Learning

weak Learner :- An algo. (Model) capable of producing classifiers with a probability of what we call less than a random guessing. (0.5 → binary)

weak Learner's prediction < 0.5

Error in prediction of weak Learner → High.

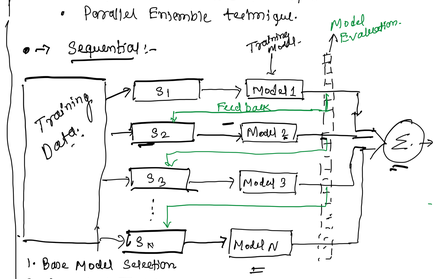
Strong Learner :- Small error probability in its prediction. It performs much better than random guessing.

2

Categories in Ensemble Learning

- Sequential ensemble technique.
- Parallel Ensemble technique.

• → Sequential :-



1. Base Model Selection
2. Sequential Training.
3. Error Correction → Each model learns from previous mistake.
4. Prediction Refinement → Iteratively refine predictions.
5. Combining Predictions, 6:- Evaluation - Assess ensemble prediction using metrics

20 1 → Browsing Data
20 2 → In house data.
20 3 → Cache.
20 4 → Firewall → xxx.
20 5 →

Test Training Sample.
60% →

10 →
10 →
10 →
10 →

Base Model Same Type → Homogeneous.

Base Model different type. → Heterogeneous

→ Insights gained from Model M_1 are passed to Model M_2 alongside sample S_2.

Model M_2 → adjusts the weights and biases based on outcome of Model M_1.

→ Process repeats → Model → 2, 3, 4.

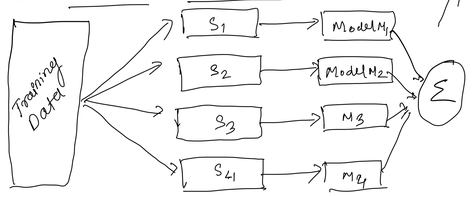
→ weighted avg. strategy used to combine predictions.

When to use Sequential Ensemble Technique. →

- Complex relationship between i/p → o/p variables
- Diverse data types → numerical, categorical, textual data.
- Imbalanced datasets.
- Incremental updates are needed to adapt models gradually to evolve data over time.

3

Parallel Ensemble Technique



Steps:

1. Data Partitioning
2. Model Training → Train models concurrently on subsets.
3. Prediction → Models make independent pred.
4. Combine Predictions :- Voting or averaging Freq. are used.
5. Evaluation.

Voting

Majority (Hard) Voting

M_1 → Yes.
M_2 → Yes
M_3 → No
M_4 → No
M_5 → No

→ No.

Weighted (Soft) Voting

M_1 → Yes (0.8) ✓
M_2 → Yes (1) ✓
M_3 → (No (0.5)
M_4 → (No) (0.5)
M_5 → (No (0.5)

⇒ 0.8Y + 1Y → (Yes)

⇒ 0.5N + 0.5N + 0.5N ⇒ No.

⇒ Yes → 1.8.
No → 1.5.

⇒ final prediction ⇒ Yes.

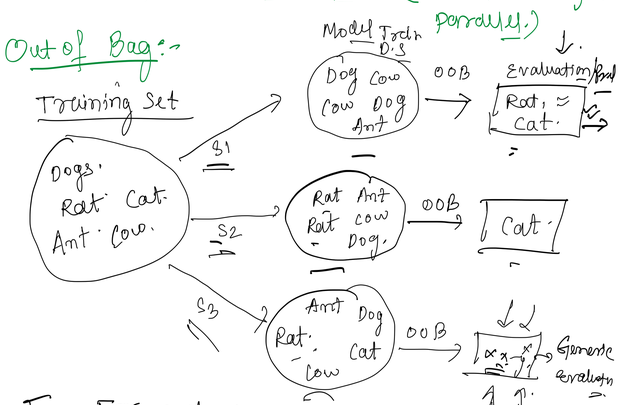
4 Advanced Techniques in Ensemble Learning

* Bagging :- Bootstrap Aggregating.

→ Parallel Ensemble Learning.

Eg:- Random Forest (Multiple DT put in parallel.)

Out of Bag :-



- For each sample - Model combination once trained.
  → evaluation is performed on the OOB data.

→ OOB concept minimizes the cross-val-tering.

5 Sequential Ensemble Adv. Technique.

- Boosting

✓ Seq. ensemble technique, each model corrects its predecessor's mistake.

→ Finally, → The final prediction,
→ weighted sum of all model's prediction.

Popular boosting algos → AdaBoost, Grad. Boosting, XG Boosting.

ADABoost :- Adaptive Boosting (Classification Prob)

⇒ It increases weights of misclassified instances, directing subsequent classifiers to pay more attention to error prediction done by previous classifier.

Steps:

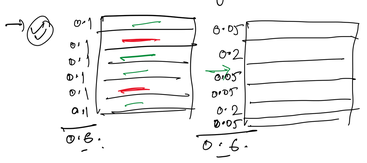
→ Initially, equal weights to all observation in dataset.

→ Model is constructed using the subset data.

→ Prediction is made on the entire dataset using this model.

→ Algorithm computes errors by comparing the predictions to actual value.

→ In subsequent model creation ⇒ Algorithm assigns higher weights to incorrectly predicted data.



→ How weight is being determined???
Ans:- based on error values, higher errors result in heavier observation weights.

→ Iteration is continued → till error function stabilizes or the algorithm reaches max. limit of estimators.

6 Gradient Boosting

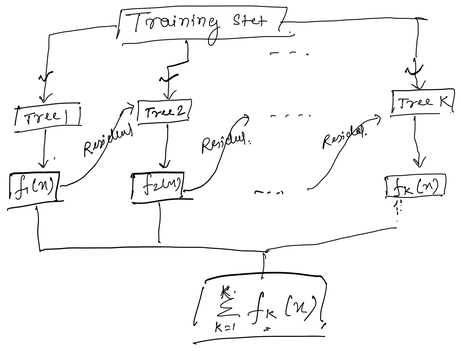
Both classification or regression predictive modeling problems.

⇒ weak Learners are :- DT with depth = 1 (stubs)

Main components of Grad. Boosting :-

- Loss Function :- Error in prediction of each record.
  Goal :- Minimize Loss function.
- Weak Learners :- DT are used as default weak Learners.
- Additive Model :- Instead of adjusting the weights of data points (like AdaBoost), G.B fits new models to residual errors made by previous models.

∴ Each new model is built on errors of the whole ensemble.



7 XG Boost :-

Supervised Learning Tasks,
→ Classification, Regression, Ranking.

Reasons to use :-

High Performance :- XG Boost is optimized for speed and efficiency.
↳ Large data sets.
→ real time application.

Regularization Methods.

L1 (Lasso) and L2 (Ridge) regularization terms are included in XG Boost to avoid overfitting and increase generalization.

Handles missing data :- XG Boost handles missing data automatically.

CAT Boost :- Variant of Gradient Boosting.
→ handle both numerical and categorical data.
→ Does not require any feature encoding techniques. (OHEncoder, L. Encoder.)