UNIT-4.

Ensemble Effect:

⇒ means combining predictions of multiple models to improve overall performance.

⇒ Reduces noise in predictions, improves model stability.

Model Ensembles- Motivation:

1) Accuracy Improvement: combining models ⇑ prediction correctness.

2) Error Reduction: Individual makes models different errors; ensemble cancels out those errors.

3) Diversity in Models: Different models look at data differently, leading to improved performance.

4) Reduction of overfitting: Ensemble smooths out extreme predictions of single models.

5) Better Decision- making: Out performs most single models in competitions.

Wisdom of Crowds:

⇒ Means that a group of diverse individuals can make better decisions or predictions than a single expert.

⇒ When we combine all their opinions, the errors cancel out, and the correct information strengthens.

Working:

1) Diversity of Opinions:

⇒ Each person/ model thinks differently.

⇒ Different errors cancel out.

2) Independence:

⇒ Each model makes its own prediction without getting influenced by others.

3) Decentralization:

⇒ Each model looks at different parts of data or uses different parts of data or uses different methods.

4) Aggregration:

⇒ All predictions are combined using voting or averaging.

* Many models = crowd, combined result = wisdom.

Adv: 1) Better accuracy

2) Reduced errors

3) Less overfitting

Conditions:

* Models must be independent

* Models must be diverse (not same errors)

* There must be enough models.

* Proper aggregration method must be used.

Bagging:

Bagging is an ensemble technique where multiple models are trained on different random samples of the data, predictions are combined to improve accuracy.

Step-by-step working:

1) Bootstrap sampling:
⇒ From original dataset, several new datasets are created using sampling with replacement

2) Train Multiple Models:
⇒ separate model is trained on each bootstrap sample.
⇒ Models are trained independently & parallely.

3) Combine predictions:
⇒ ~~All~~ model predictions are combined
Regression → average
classification → majority voting

Adv
* Reduces overfitting
* Reduces variance
* Improves acurracy
* Parallel training

DisAdv
* Requires more accuration
* Does not reduce bias, only variance

Uses:
* when model easily overfits
* when dataset is noisy
* prediction variance is high

→ Final Prediction = majority vote / averaging.

Boosting:

→ Boosting is an ensemble learning technique where multiple weak models are trained sequentially, each new model focuses on correcting the errors made by previous one.

Step by step working:

1) **Initialize:**
Train the first weak model on dataset

2) **Identify Errors:**
check which data points were misclassified or predicted wrongly

3) **Increase weights of wrong Predictions:**
More importance given to incorrectly predicted samples.

4) **Train Next Model:**
Focuses on correcting the mistakes of previous model

5) **Combine All Models:**
Final o/p = weighted vote / weighted average of all models.

Characteristics:

* Sequential training (not parallel)
* Very high accuracy on structured / tabular data
* Reduces bias significantly.

Adv

* Very high accuracy
* Reduce bias
* Handles complex patterns in data

DisAdv:

* Sensitive to noise
* Harder to interpret
* Training is slow (models train one after another)

Random Forests:

Forest → collection of many decision trees

Random → randomness in

- selecting data samples (bagging)
- selecting features

It is a method that builds many decision trees using bagging + random feature selection & combines their predictions using majority voting.

Working:

1) Create Bootstrap Samples:

From original dataset, many random samples are created

2) Build Multiple Decision Trees:

Each tree is trained on one bootstrap sample

3) Make Predictions:

Each tree gives its own prediction

4) Aggregate the results:

classification → majority vote regression → average

→ Combined result is final o/p.

Adv: * High accuracy

* can handle large datasets.
* Handles missing values

DisAdv: * Requires more memory

* Slower prediction bcz of many trees.

Applications: * Fraud detection

* Medical diagnosis
* Feature selection
* Stock market prediction

Stochastic Gradient Boosting:

→ It is an advanced boosting technique where models are built sequentially and each new tree is trained to correct the errors of previous trees.

→ It is called stochastic bcz uses random sampling during training.

→ Gradient descent used to minimize the errors.

Working:

1) Train the first model of small decision tree is trained on data.

2) Calculate the error:

Error = Actual − predicted, these errors called residuals.

3) Random sampling:

Used to train next tree

4) Train next tree on residuals.

5) Repeat the process & many trees added sequentially.

6) Final Prediction: All tree's predictions are combined

Concepts a) Learning Rate b) Sub sampling c) weak learners.

Adv: * Less overfitting bcz of randomness

* works well for large datasets.

Dis Adv: * Complex and requires tuning

* Sensitive to noise if parameters are not set correctly.

Ex: 1) XGBoost

2) LightBGM

3) CatBoost

Heterogeneous Ensembles:

→ Method where different types of ML models are combined to make a final prediction.

→ Each model has different strengths & weakness.

Uses: Higher Diversity, Better accuracy, Reduces Overfitting.

Working: 1) Train Multiple Different models:

Decision Tree, Logistic Regression, Neutral n/w.

2) Combine predictions:

using voting (for classification), averaging (for regression), stacking, blending.

Techniques: a) Voting classifier

→ Hard voting → majority vote

→ Soft voting → avg of probabilities.

b) Stacking

→ Very powerful & widely used

→ Base models make predictions

c) Blending

→ Similar to stacking but uses a validation set instead of cross-validation.

Adv: * Better model accuracy

* Less sensitive noise

DisAdv: * Complex to build and interpret

* Requires more time for training & testing

Applications: Credict scoring

Fraud detection

Real life Ex: Imagine u ask 3 experts to predict a student's mark

1) A teacher (knows performance) 2) A parent (knows habits) 3) A friend (knows daily study)

Each gives different insights

when you combine their predictions, you get the most accurate result

This is a heterogeneous ensemble.