Ensemble techniques

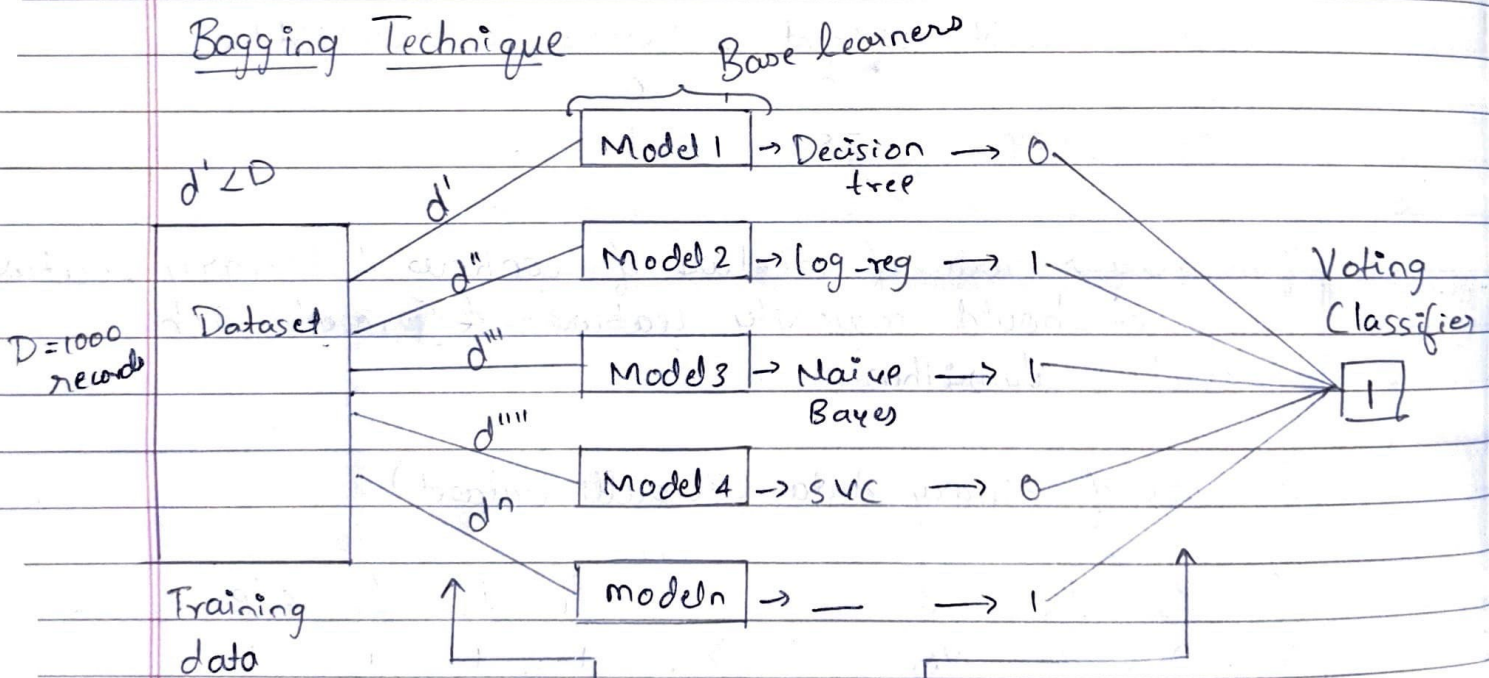
* Ensemble - Combining multiple models & predict using all models

Types of Ensemble techniques
- Bagging
- Boosting

i) Bagging
* Random forest classifier & regressor

ii) Boosting
* Adaboost
* Gradient boost
* Xg boost

Bagging Technique

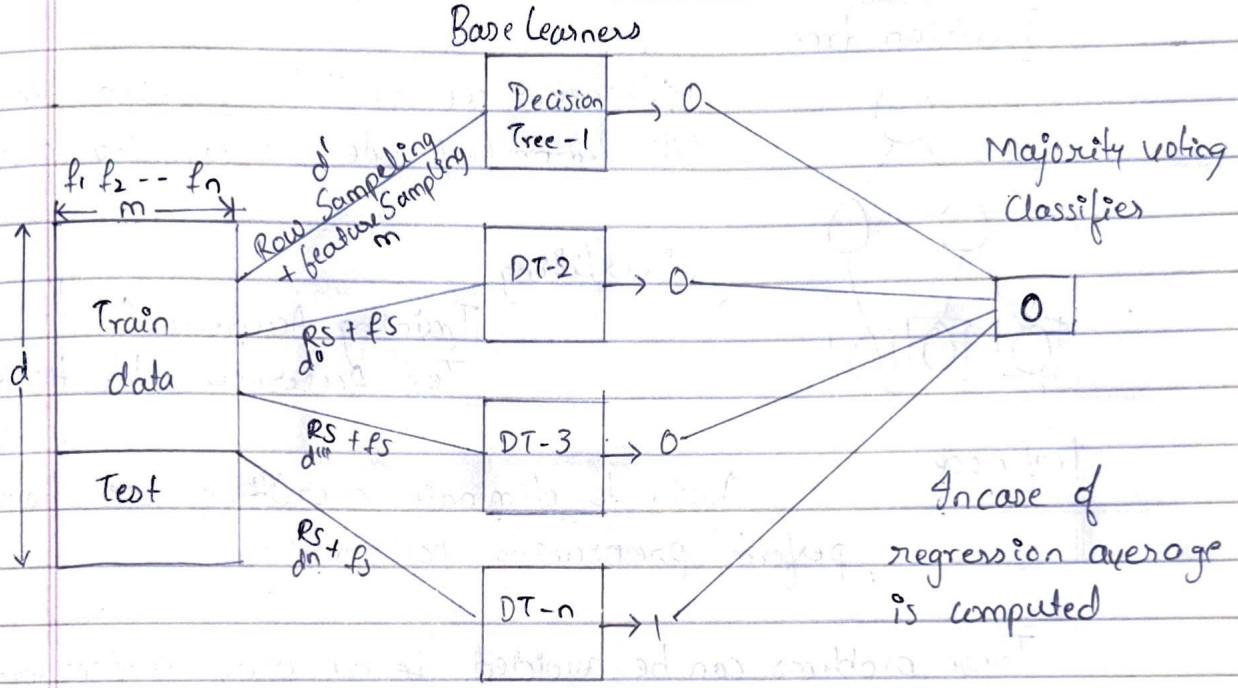


Classification
↳ Custom bagging or custom ensemble technique

In case of regression we take average of all outputs

Random forest Classifier & Regressor

Similar to bagging technique but all the base learners are decision trees



Row Sampeling → Selecting sample of records

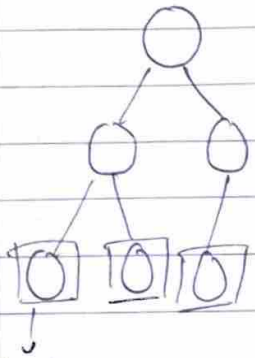
Feature sampeling → Selecting n features (where n<m)

Note

* Classification - Majority voting classifier
* Regression - Average of models

Why should we use Random forest insted of Decision tree?

Decision tree



leaf node

Whenever we split the decision tree till ~~left~~ no leaf node → overfitting occurs

Overfitting

Training Accuracy ↑↑ low Bias
Test Accuracy ↓↓ High variance

Again to eliminate overfitting we should perform prepruning post pruning

These problems can be avoided ~~in~~ by using Random forest as we have trained many decision trees with sample features thus increases test data accuracy

Out of Bag Data

Those rows which are missed during row sampeling of random forest

Parameter → oob_score = True

↳ Out of bag data = validation data = validation accuracy

↓

oob_score