

Machine Learning (Day-10) :-

Agenda :-

1) Ensemble Technique

- a) Boosting
- b) Bagging
- c) Stacking

2) Different between Bagging and Boosting.

3) Random Forest

4) Ada Boost

5) Random Forest Classifier

6) Random Forest Regressor.

★) Ensemble Technique :-

⇒ Ensemble Techniques are a powerful tools for improving the accuracy and robustness of Machine-Learning model. These techniques involve combining the predictions of a multiple models to produce a final prediction that is more accurate than any of the individual models alone.

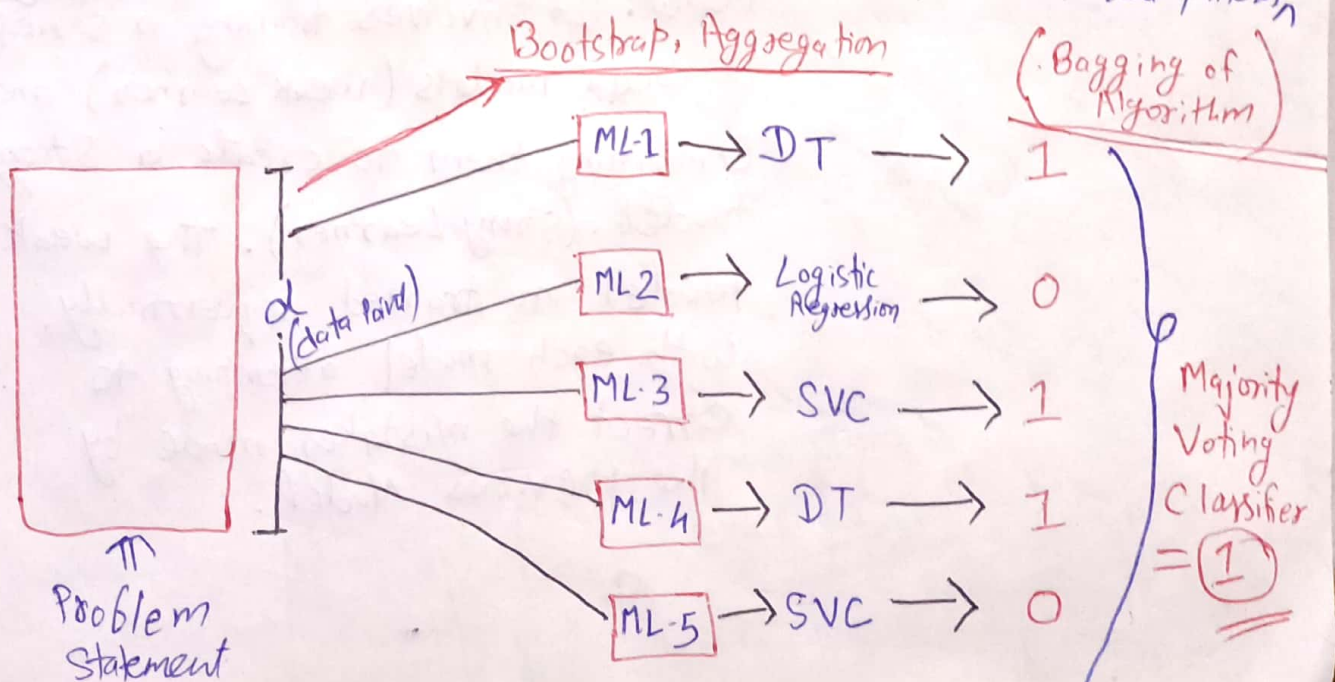
These Techniques can be applied to a wide variety of Machine Learning Models, including Decision Tree, Neural Networks and Linear Regression model.

There are several Different Types of Ensemble Techniques, including Bagging, Boosting and Stacking. (124)

Now,

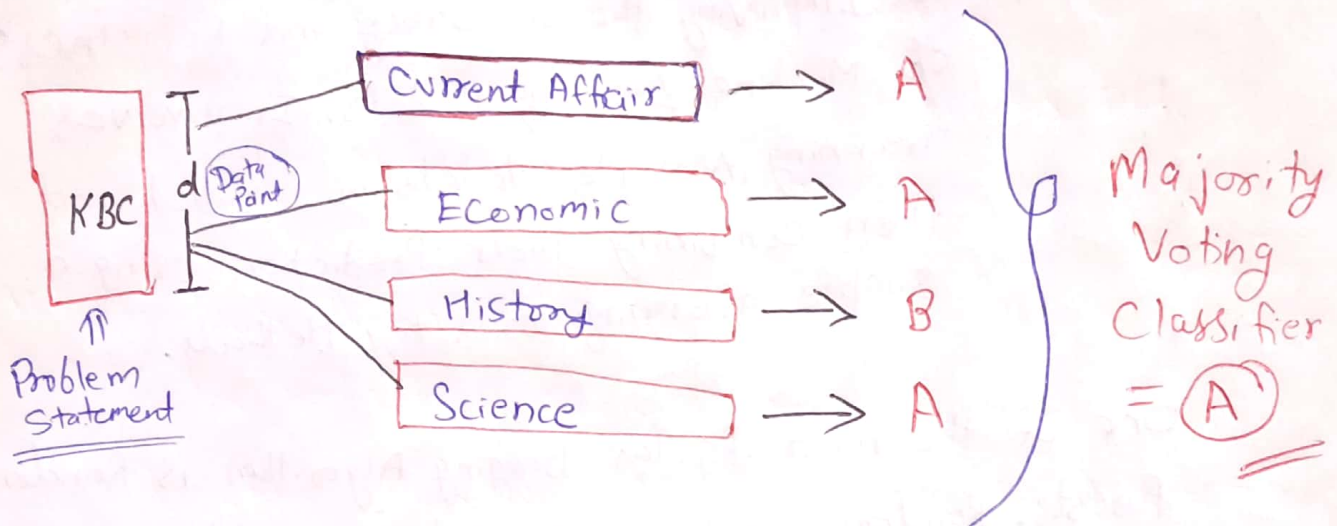
★) Bagging :- Bagging is a Powerful Ensemble Technique for Improving the accuracy and Robustness of Machine Learning Models. It involves Training Multiple Models in Parallel and then Combining their Prediction using a Simple averaging or Voting Methods.

One of the most Popular Bagging Algorithm is Random-Forests. Random Forests work by training a large Number of Decision Trees on Random Subsets of the Training data. The Final Prediction is made by averaging the Predictions of all the Decision Trees. Random Forests are Particularly Effective at reducing overfitting, as the Multiple Decision Trees are able to capture different patterns in the Data.



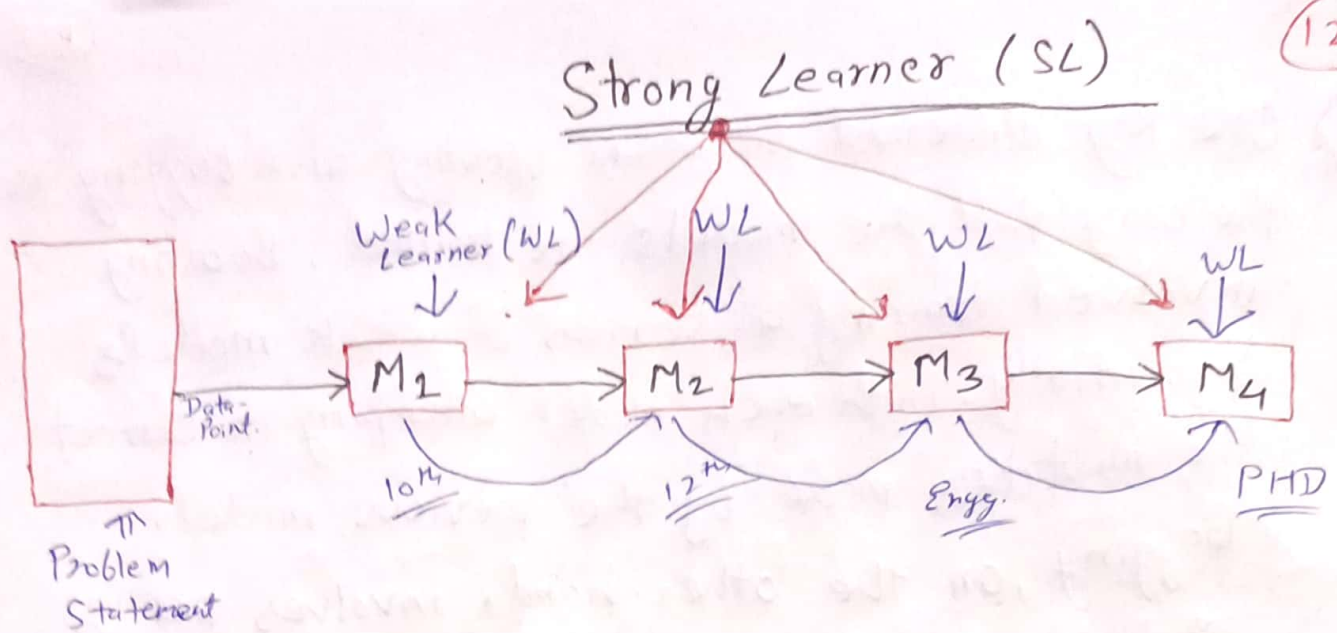
Example of Bagging Algorithm:-

⇒ KBC :- Kaun Banega Crorepati (Amitabh Bachchan)



Now,

#) Boosting :- Boosting is a Powerful Ensemble Technique for improving the accuracy and robustness of Machine Learning Model. It involves training a series of weak models (weak Learner) and Combining them to create a Strong Model. (Strong Learner). The weak model are trained Sequentially, with each model attempting to Correct the mistakes made by the previous Model.



Here,

Weak Learner = WL

Strong Learner = SL

Note :- We Combined Multiple Weak Learners to perform Strong Learner.

★ Different Between Boosting and Bagging?

⇒ Boosting and Bagging are both Ensemble techniques that are used to improve the accuracy and robustness of Machine Learning Models.

However, they work in different ways and have some key Differences :-

P.T.O

i) One Key Different between boosting and bagging is the way that the models are trained. Boosting involved training a series of weak models sequentially, with each model attempting to correct the mistakes made by the previous model.

Bagging, on the other hand, involves training multiple models in parallel.

ii) Another Different is the way that the predictions of the individual models are combined.

Boosting Algorithm typically combine the predictions of the weak models using a weighted average, with the weights determined by the performance of the individual models. Bagging Algorithm, on the other hand, typically combine the predictions of the model using a simple average or voting method.

iii) A third Difference is the way that the models are evaluated. Boosting Algorithm are typically evaluated using Loss Function, which measure the difference between the predicted and Actual values.

Bagging Algorithm, on the other hand, are typically evaluated using ~~loss~~ accuracy metrics, which measures the percentage of correct predictions.

★) Stacking :- Stacking is an Ensemble Technique that involves training a series of models and using their predictions as input to a higher-Level model, Which makes the final Prediction. Stacking is often used to combine the Strengths of Multiple Model and Produce a more accurate final Prediction.

★) Random Forest :-

⇒ Random Forest is a Machine Learning Algorithm that is Used for Classification and Regression Tasks. It is an ensemble method, which means that it combines the predictions of Multiple Decision Trees to improve the overall prediction accuracy.

The basic idea behind Random Forests is to train a large Number of Decision Trees using a random Subset of the data, and then have each tree make a Prediction. The final Prediction is made by Averaging the Prediction of all the Trees. This process helps to reduce Overfitting, Since each tree only sees a small Subset of the data, and it also helps to improve the overall accuracy of the model.

One of the main advantages of random Forests is that they can handle large and complex datasets, and they are also relatively easy to train and tune. They are widely used in a variety of applications, including image Classification, Natural - Language Processing (NLP) and Predictive Modelling.

*> Two Type of Random Forests are there :-

- ① Random ~~Classifier~~ Forest Classifier.
- ② Random Forest Regressor.

*> Random Forest Classifier :-

⇒ It's used For Predicting a Categorical outcome, Such as whether an email is Spam or Not Spam. They work by Training Multiple Decision Trees on a Training Dataset, and then using the Trees to Classify new data points based on the features of the data. The final Prediction is made by Combining the predictions of all the trees.

P.T.O

* Random Forest Regressor :-

⇒ It's used for predicting a Continuous outcome, such as the price of a House. They work in a similar way to ~~Classification Problem~~ Random Forest Classifier, but the predictions made by each tree are Average instead of Combined through Voting.

* ADA-BOOST :-

⇒ Ada Boost, short for Adaptive Boosting, is a Machine Learning Algorithm that can be used for Classification or Regression Tasks. It is a Types of Ensemble Learning method, which means that it combines the predictions of Multiple base models to ~~re~~produce a more accuracy overall Prediction.

P.T.O

AdaBoost works by building a series of base Models each of which is trained to make prediction on the data. The base models are typically Decision Trees, but they can be any type of model. After each model is trained, AdaBoost assigns a weight to each training example based on how well the model predicted the label of the example. Examples that are misclassified are given higher weights, while examples that are correctly classified are given lower weights. This process is repeated for multiple rounds, with each round producing a new base model that is trained on the weighted examples.

At the end of the process, AdaBoost combines the predictions of all of the base models to produce a final prediction. The weight of each base model's prediction is based on how well it performed during training. This means that models that perform better are given more influence in the final prediction, while models that perform poorly are given less influence.

AdaBoost is a popular algorithm because it is relatively simple to implement and has shown good results in a wide range of applications. It is also relatively robust to overfitting, which means that it is likely to produce models that perform well on the training data but poorly on new, unseen data.