Diploma in Computer Science and Engineering

5^{TH} Semester

Artificial Intelligence and Machine Learning

Ensemble Learning - WEEK 8 – Session 5

- Introduction

  Ensemble learning is primarily used to improve a model's performance (classification, prediction, function approximation, etc.) or lessen the chance that a poor model may be chosen accidentally.

  Multiple machine learning models are used in ensemble learning to increase the predictability and accuracy of outcomes.

  Example - **Problem –** to develop a machine learning model for your business that forecasts inventory stock orders based on historical data from prior years.

- The techniques linear regression, support vector machine, regression decision tree, and fundamental artificial neural network are used to train four machine learning models. However, none of them manage the requisite 95 percent forecast accuracy even after thorough tuning and configuration. As a result of their inability to converge to the target level, these machine learning models are referred to as "weak learners“ But being helpless is not the same as being weak. They can be combined to create a group. You run your input data through all four models for each new prediction, then average the outcomes.
- When you look at the updated result, you'll see that the overall accuracy of 96 percent is more than acceptable. Ensemble learning is effective because your machine learning models operate differently. Each model might work well on some data sets, but not on others. Their combined flaws make up for each other's strengths.

- Basic Ensemble Techniques
  - Max Voting

    Many models produce predictions, or "votes," where each prediction represents one vote. Decisions are frequently taken in favour of the majority of votes, as is frequently the case with voting.

    It is best suited for classification problems.

      As an example, we attempt to forecast car prices. The yield prices (in Rupees) for the models that were used were 8000, 7000, 8000, 8000, 4000, 8000, 7000, 6000, and 8000. The majority of the models predict 8000, as can be seen. The ultimate forecast with the most votes is 8000.

  - Averaging

The predictions from all the models are averaged, making this the easiest ensemble technique to define. The average of the predictions is used to make the final result.

Problems involving classification and regression can be solved using this technique.

- Consider a situation where we wish to predict car pricing. The rates of it are 6000, 7000, 5000, and 6000 Rupees are predicted by our models.
  - The final projection, since we're averaging, is:
  - 5000+7000+8000+4000 / 4=6000

- Weighted Average

A weighted average provides a model with better prediction ability and more weight. Weights serve as a representation of this importance.

- **Weighted Average – Example**

Imagining you have a collection of five classifiers You discover that, if the classifiers are all independent, each of them may result in an error of 0.2. And when the ensemble classifier fails on an instance, it does so because at least three of your five classifiers collectively made a mistake (majority voting). As a result, the likelihood of a false forecast is determined as follows:

When 3 of 5 things go wrong, the result is 10,

When 4 of 5 things go wrong, the result is 5,

And When all 5 go wrong, the result is 1.

If you can see, the overall error has fallen from 0.2 to 0.058, but as more classifiers are added to the combination, a bottleneck is to be expected, and the trend of errors decreasing will plateau.

- Advanced Ensemble Techniques
  - Stacking

This method functions by allowing a training algorithm to aggregate the predictions of multiple learning algorithms similar to it.

Regression, density estimation, distance learning, and classification have all succeeded in stacking.

Additionally, the error rate during bagging can be determined using it.

Creating a Meta-Model is the primary goal of the stacking generalization technique. Using the k-fold cross-validation technique, such a Meta-Model is constructed from the predictions of several ML base models (i.e. weak learners). Finally, a second ML model—

often referred to as the "final estimator" or "final learner"—is used to train the meta-model.

- Blending

The **Stacking Generalization** method comprises training stages, better known as "*level 0*" and "*level 1*".

A method derived from stacking generalization is blending.

The only distinction is that in Blending, the training data for the meta-model is not created using the k-fold cross-cross-validation procedure.

Blending uses a "one-holdout set," or a small fraction of the validation data, to produce predictions that will be "stacked" to create the meta-training model's data. Predictions are also formed from the test data to create the meta-model test data.

- Bagging

Classification and regression are the main applications for bagging, which is short for bootstrap aggregation.

Decision trees are used to reduce variance, which increases model accuracy drastically. By lowering variance, removing overfitting, which is a problem for many predictive models, improves accuracy.

Bagging comes in two flavours: bootstrapping and aggregation.

Bootstrapping is a sampling method that gathers samples from the entire population using the replacement mechanism (set).

Aggregation is used in bagging to include all potential outcomes of the prediction and randomize the result. As a result, the aggregate is based either on all predictive model results or probability bootstrapping processes.

- Boosting

Boosting is an ensemble method that improves future predictions by learning from past predictor errors.

The method greatly increases model predictability by combining numerous weak base learners into one strong learner.

Boosting techniques include gradient boosting, adaptive boosting (AdaBoost), and XGBoost (Extreme Gradient Boosting).