Assignment 22 Solutions

1. Is there any way to combine five different models that have all been trained on the same training data and have all achieved 95 percent precision? If so, how can you go about doing it? If not, what is the reason ?

Ans: Hybrid Model: A technique that combines two or more different machine learning models in some way. But we can't get 95 percent precision as all other models give different precision rate accuracy is differed.

2. What's the difference between hard voting classifiers and soft voting classifiers ?

Ans: In hard voting (also known as majority voting), every individual classifier votes for a class, and the majority wins. In statistical terms, the predicted target label of the ensemble is the mode of the distribution of individually predicted labels.

In soft voting, every individual classifier provides a probability value that a specific data point belongs to a particular target class. The predictions are weighted by the classifier's importance and summed up. Then the target label with the greatest sum of weighted probabilities wins the vote.

3. Is it possible to distribute a bagging ensemble's training through several servers to speed up the process? Pasting ensembles, boosting ensembles, Random Forests, and stacking ensembles are all options ?

Ans: When sampling is performed without replacement, it is called pasting. In other words, both approaches are similar.In both cases you are sampling the training data to build multiple instances of a classifier.

Boosting is a general ensemble method that creates a strong classifier from a number of weak classifiers. This is done by building a model from the training data, then creating a second model that attempts to correct the errors from the first model. It is the best starting point for understanding boosting.

4. What is the advantage of evaluating out of the bag ?

Ans: The advantage of the OOB method is that it requires less computation and allows one to test the model as it is being trained.

5. What distinguishes Extra-Trees from ordinary Random Forests? What good would this extra randomness do? Is it true that Extra-Tree Random Forests are slower or faster than normal Random Forests ?

Ans: Random forest uses bootstrap replicas, that is to say, it subsamples the input data with replacement, whereas Extra Trees use the whole original sample. This may increase variance because bootstrapping makes it more diversified.

Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Extra Trees is much faster.

This is because instead of looking for the optimal split at each node it does it randomly.

6. Which hyperparameters and how do you tweak if your AdaBoost ensemble underfits the training data ?

Ans: If your AdaBoost ensemble underfits the training data, you can try increasing the number of estimators or reducing the regularization hyperparameters of the base estimator. You may also try slightly increasing the learning rate.

7. Should you raise or decrease the learning rate if your Gradient Boosting ensemble overfits the training set ?

Ans: If your Gradient Boosting ensemble overfits the training set, you should try decreasing the learning rate. You could also use early stopping to find the right number of predictors (you probably have too many)