

Experiment 2: Ensemble Methods in ML

Instruction: Create a new .Rmd file and write code for each section in separate R code block

Ensemble methods are techniques that create multiple models (often called *base learners* or *weak learners*) and then combine their predictions to produce a more accurate and robust overall result. Bagging is a specific type of ensemble method in which we train multiple models on different random samples (with replacement) from the training data.

Perform the following steps and validate the efficiency of bagging methods:

1. Create 10 Bernoulli Random Variables (RV). Each has a success probability of 0.8. Evaluate the majority output. If 6 RVs output a 1 then majority output is 1, otherwise majority output is 0.
2. Implement step 1 10000 times and evaluate the probability of 1 as majority output in 10000 trials.
3. Train a Decision Tree model on IRIS Dataset using “rpart” library in R. Use 120 samples randomly for train dataset and 30 samples for test set. Evaluate train and test set classification accuracy.
4. **Bagging:** Develop a Random Forest Model using the Decision Tree in Step 3. Implement bagging by training the model (in step 3) on $n=10$ different random subsets of data (with replacement). Find training and test classification accuracy of each model. Combine all 10 models using majority voting to get Random Forest Model. Compare results obtained in 4 and 2. You may have to change success probability in 2 for comparison.
5. Exploratory: Bagging is not just training similar classifiers on random subsets of data. It is also about combining different classifiers. In step 4, use different feature combinations for the trees. Create a Random Forest with 6 trees, each having only two features. Print training and test accuracy for the individual trees and the forest.
6. Answer the questions in the shared report.