

Advances in Machine Learning

Aim: Implementation of ensemble and performance measurement

Title: Ensemble learning methods like random forest, bagging, and boosting to improve accuracy.

Frequently Asked Question

Question 1: What is an ensemble model?

⇒ An ensemble model is one where multiple basic models are combined to improve the accuracy of the regression / classification task.

Advantages:

- Achieve better performance than any single contributing model
- Reduce the spread or dispersion of the predictions

Question 2: What is bagging, boosting and stacking?

- 1) Bagging: In bagging, we combine the output of multiple algorithms that have been trained on smaller segments of the data and all of them operate independently.
- 2) Boosting: Here, the training of every subsequent model depends on the performance of the previous model.
- 3) Stacking: Stacking uses different types of ML algorithms to create an ensemble. One model is trained to combine the predictions.

Question 3: Can we ensemble multiple models of the same algorithm?

- ⇒ Yes, in bagging and boosting, models of the same algorithm are combined to create an ensemble.

Question 4: How can we identify the weight of different models for ensemble?

Then are ~~collinearity~~ various methods to find the optimal methods:

- Find the collinearity between bon learners and based on this table, then identify the bon to ensemble
- Find the algorithm to return the optimal weights for bon learners.
- Using methods like forward selection, selection with replacement and bagging.

Question 5 What are the benefits of an ensemble model?

=> Some benefits are:

- Better accuracy
- More stability
- Low variance & chances of overfitting.