

# Exercises #7 - Ensemble Learning

Thales Menezes de Oliveira

Brazilian Center for Physics Research (CBPF)

April 1, 2022

1. If you have trained five different models on the exact same training data, and they all achieve 95% precisions, is there any chance that you can combine these models to get better results? If so, how? If not, why?

A: It is possible to combine these five models in a voting ensemble. This voting ensemble is capable of achieve a better performance than the best individual model. It can be even better if the models in the ensemble are very different, like SVM classifiers, a Decision Tree, Logistic Regression and so on.

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

A: The hard voting classifier just counts the votes of each classifier in the ensemble and picks the class with the majority of the votes. On the other hand, a soft voting classifier is capable to compute the average estimated class probability for each class and pucks the class with the highest probability. For different classifiers, it is necessary to all of them are in position do estimate probabilities.

3. Is it possible to speed up training of a bagging ensemble by distributing it across multiple servers? What about pasting ensembles, boosting ensembles, random forests, or stacking ensembles?

A: .

4. What is the benefit of out-of-bag evaluation ?

A:

6. If your AdaBoost ensemble underfits the training data, what hyperparameters should you tweak and how?

A: For the underfit of the training data, it is possible to increase the number of estimators or reducing the regularization hyperparameters. It is also possible to decrease the learning rate.

7. If your Gradient Boosting ensemble overfits the training set, should you increase or decrease the learning rate?

A: The better strategy to avoid the overfitting is to decrease the learning rate. It is also possible to evaluate other hyperparameters, as the number of estimators, since the learning can be good but it have too many estimators. The early stopping strategy can help with this task.