

COMP90051 Statistical Machine Learning

Semester 2, 2015

Workshop Week 4: Ensemble Learning; Project 1 working

August 17, 2015

In this week's workshop you'll have an opportunity to ask your tutor about

- Material from class you're unsure about (such as regularisation or ensemble methods from week #3, or probability which forms the basis of the material in lectures for weeks #4-5)
- Project 1

You can also spend time starting on the following exercises to get hands-on experience with ensemble methods. The exercises ask you to explore some of the ideas discussed in class, such as the effect of hyperparameters of common ensemble methods. When you're done, think up other experiments to do, such as exploring similar questions with AdaBoost.

1 Ensemble Approaches

Recall the Ripley data from Week #2 workshop Question 2.3.

1. Fit a random forest to the Ripley 2D synthetic dataset. Repeat the same plots you did for the linear approaches. How does it compare? What is this classifier's accuracy?
2. Now try this again, but with the wine dataset in the UC Irvine machine learning dataset repository: <http://archive.ics.uci.edu/ml/datasets/Wine> . Download the csv file and load it into R, note the first column is the class label.
3. What is the effect of the number of trees and number of features per tree? E.g. Vary one of these, and plot the resulting accuracy. Try varying the other.