

Portfolio Component: Kernel and Ensemble Methods

Pair Assignment

You may work by yourself or with one other person on this assignment.

Objectives:

- Gain experience with machine learning using SVM linear, polynomial, and radial kernels
- Gain experience with ensemble techniques

Turn in:

- Print your Rmd notebooks to pdf, upload to your portfolio, and create links on your index page
- Upload your Rmd print-to-pdfs to eLearning

Instructions:

1. Notebook 1 Regression. Find a medium-size data set of at least 10K rows, suitable for regression. Provide a link to the data in markdown. You can use a data set you used before.
 - a. Divide the data into train/test
 - b. Explore the training data statistically and graphically
 - c. Perform SVM regression, trying linear, polynomial, and radial kernels with various C and gamma hyperparameters.
 - d. Provide some analysis on why the results were most likely achieved given how the kernels work.
2. Notebook 2 Classification. Repeat the steps above for a classification data set.
3. Notebook 3 Ensemble techniques. Using your classification data set, perform machine learning with decision tree as a baseline. Try Random Forest, XGBoost, and at least 1 other ensemble package. Write analysis comparing your results in terms of metrics and run times.
4. Upload the print-to-pdfs for all notebooks, and create links to them on your index page.
5. Narrative document. Write a 1-2 page narrative discussing:
 - a. how SVM works, and how SVM kernels work, your impression of the strengths and weaknesses of SVM
 - b. how Random Forest works, how the other 2 algorithms you used work compared to the simple decision tree, your impression of the strengths and weaknesses of these ensemble techniques

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.

Grading Rubric:

Element	Points
Step 1 Regression SVM notebook	50
Step 2 Classification SVM	50
Step 3 Ensemble notebook	50
Step 5 Narrative	40
Step 6 Create links to the document and code on the index page	10
Total	200

Grading Rubric:

- 90% and above for exceptional work
- 80-89% for good work
- 70-79% for average work
- below 70% for low quality work

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.