Challenge Assignment

You may work with a partner for this assignment.

The purpose of this Assignment is to familiarize you with approaching a problem and trying to analyze what method(s) might be the best suited for a particular domain/data. You are free to choose your validation framework, techniques, etc. You will use a validation stage to select your favorite methods. Once you select your best method from the validation set, you will generate predictions on the testing set.

Two datasets are provided – one for classification and one for regression. Please read below for the Assignment FAQ, which outlines the details of the data and your work. The highest scorer for each task (classification and regression will be announced at the Class Workshop. Good Luck & Have Fun!

FAQ

1. *Is it a group project?* Yes, you are to work in pairs..

2. *What will be provided?*
   - A training set
   - A testing set with dummy class values values, just so that you can make predictions but not any optimizations. Please don't use the testing set for any validation or evaluation as the class labels mean nothing. I have just attached the dummy class labels so your data format is not getting messed up. I have the actual labels with me that I will use for evaluation.

3. *What is the classification dataset?* It is a dataset with 48 features (both binary and continuous) and 4147 examples. Although the number of features is moderate, feature selection may be helpful here; it is up to you to investigate. The testing set has 415 examples.

4. What is the regression dataset? It is a dataset with 21 continuous features and 5460 examples. The testing set has 2730 examples.

5. What classifier can I use? You can use any classifier. You will have to utilize a validation stage to select the best method. You are free to use any methods for validation and for learning a classifier. You must try at least two classifiers and select the best from validation.

6. What regression function can I use? You may use any regressor – linear regression, regression trees, or various other choices. You must utilize a validation method. You must try at least two different regression methods and choose the one that gives you the best performance on the validation for making predictions on the testing set.

7. How do I evaluate which is a good classifier/regressor choice? You should evaluate your methods by employing a validation framework. You are free to choose how you want to construct your framework. You can use 10-fold cross-validation, boostrapping, or even use hold-out sets. You can use accuracy as a performance benchmark for classification and root mean squared error for regression.

8. What am I required to submit?

a. Your predictions on the testing set. You have to submit one classification result and one regression result on the testing set examples. For classification, you will provide a one-column flat file. The column will be the predicted class. For regression, you will provide a single-column file, which will contain the predicted value for each example.

c. The performance of two classifiers/regressors during the validation stage. A one-two paragraph write-up that losts the classification and regression methods you tried. The validation framework you chose. If you did any feature selection, please note that. And then finally what classifier did you choose to apply on the testing set. You can only make one (final) testing set submission for each of the classification and regression tasks.

9. When will this assignment be due? In your AFS (webfile) dropbox by 11:59 PM, December 13th, 2014.

10. When will the highest scoring team/method be announced? At the class workshop.

11. How will performance be measured?

a. For the classification method, I will utilize accuracy.

b. For the regression method, I will utilize root mean squared error.

12. How will I be graded? A complete submission will suffice for grading purposes. Essentially, you use at least two methodologies for each classification and regression, deploy a validation framework, and follow the requirements for submission, irrespective of the final outcome of your predictions, you get full points on this assignment. I encourage all of you to give it your best try and win the challenge!