What's for Dinner - part 2

Submit Assignment

Due Dec 5 by 4:59pm **Points** 100 **Submitting** a text entry box or a file upload
File Types doc, docx, py, pdf, txt, zip, ipynb, and html

In this assignment, you will be tuning two models - a Naive Bayes Classifier and a Support Vector Machine classifier. Template code for using these models can be found in the Kaggle dicussion forum.

For the Naive Bayes (NB) classifier, you will want to rerun the cross-validation to choose the best preprocessing steps, like you did for the kNN classifier. In addition, the one parameter for the NB model to tune is called alpha, the Laplace correction. For more information on the model, see **http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html (http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html#sklearn.naive_bayes.MultinomialNB)** .

For the Support Vector Machine (SVM) classifier, you will also need to use cross-validation for choosing the best prepocessing steps and parameters. The main parameters to try to change are:

- penalty, which how we calculate the length of the weight vector, either based on Manhattan distance ("l1") or Euclidean distance ("l2"). Note that this relates to how we define the width of the optimal margin.

- C, which is the coefficient that specifies the relative importance of correct classification and the margin size. Larger values for C lead to fewer misclassified examples in the training set, but a smaller margin. Smaller values for C lead to more misclassified examples, but a larger margin. To equally weight these two competing objectives, use a value of 1 for this parameter.

Target Accuracies

For full credit on the assignment, you should aim for at least 0.86 accuracy on the Private Leaderboard for your NB model, and 0.90 accuracy on the Private Leaderboard for your SVM model.

Deliverables and Target Code

Submit the following:

1) The name of your group, and the names of all members in your group.

2) The code used to build the NB submission.

3) The NB submission.

4) Your estimate of the NB submission's accuracy based on crossvalidation, and the score for the submission based on the public leaderboard.

5) The code used to build the SVM submission.

6) The SVM submission.

7) Your estimate of the SVM submission's accuracy based on crossvalidation, and the score for the submission based on the public leaderboard.