

Machine learning homework 4

Nan Hu

This program follow the exact same step as the homework requirement describe. I created a training set and a testing set. Each of them has about 40% spam example, and 60% not-spam example. Then I created probability model, which is calculating the means and standard deviation for each feature as the condition of same class. Finally, use Naïve Bayes Algorithm to classify the instance. Here is the result I got.

The accuracy is: 0.784441547153

Confusion matrix:

	predicted_spam	predicted_not_spam
actual_spam	842	65
actual_not_spam	431	963

Precision is 0.661429693637

Recall is 0.928335170893