CS 545

Machine Learning

Programming #2

Submitted by Andrew Huffman

March 1, 2020

Portland State
UNIVERSITY

Analysis:

For this assignment I applied Gaussian Naïve Bayes classification to the Spambase data from UCI ML repository (https://archive.ics.uci.edu/ml/datasets/spambase) to classify each example as spam or not spam. I created my training and test data sets by splitting the spam examples in half and splitting the non-spam examples in half and combining the first halves to from the training set and the second halves to for the test set. I calculated the mean and standard deviation for each feature for each class. Using these feature parameters and assuming normal distributions, I calculated the probability of each feature value given spam and not spam classes for each example. Then, I used MAP to assign each example to a class as follows:

$$class(\boldsymbol{x})_{NB} = argmax_{class}(\boldsymbol{x})[logP(class) + logP(x_1|class) + \cdots + logP(x_n|class)]$$

Results:

accuracy	78.3%
precision	66.1%
recall	92.1%

Table 1: Summary of model prediction performance

		Predicted Class	
		0	**1**
Actual Class	**0**	967	427
	1	72	834

Table 2: Confusion matrix (0—not spam; 1—spam)

The accuracy I got was in accordance with expectations. The model incorrectly predicted spam class much more often than it incorrectly predicted not spam. There 427 predicted 1's when the actual class was 0, but only 72 predicted 0's when the actual class was 1. I did not expect this since the prior for the spam class (≈0.4) was smaller than the prior for the non-spam class (≈0.6). I do not think that the features in the data are all truly independent. Despite this poor assumption, the classifier still performs better than the baseline 60% accuracy that could be achieved by merely guessing not spam for every example. This model does not provide high reliability and would incorrectly classify a significant amount of messages as spam when they are not, which would not be good if it were applied as a filter. Partially the classifier's error is likely to feature dependence. I suspect that another impediment is that the chosen features are not well suited to distinguish between spam and non-spam messages using a Naïve Bayes classifier.