CS771: Machine Learning: Tools, Techniques, Applications
Assignment 1 : Naive Bayes, NN, LDF

Divyanshu Shende
Roll No.: 13264

February 4, 2016

1 Spam Detection : Naive Bayes

Here are the result obtained on using Naive Bayes under various situations. The Score, Precision and Recall are in range $[0, 1]$.

$$\text{Score} = \frac{\textit{correct predictions}}{\textit{total predictions}}$$

$$\text{Precision} = \frac{\textit{true positives}}{\textit{true positives + false positives}}$$

$$\text{Recall} = \frac{\textit{true positives}}{\textit{true positives + false negatives}}$$

Type	Validation	Score	Precision	Recall
Simple	Ten Fold	0.9948168230300999	0.9958880497408638	0.9979253112033195
Stop Words	Ten Fold	0.9951628437912999	0.9967093662774239	0.9975103734439834
Lemmatizers	Ten Fold	0.9948168230300999	0.995881162688522	0.9979253112033195

2 Spam Detection : Linear Discriminant Functions

Here are the result obtained on using Naive Bayes under various situations. The Score, Precision and Recall are in range $[0, 1]$. Cross-Validation was 10-fold. A few details about the experiments:
Learning Algorithm : Perceptron
Number of Iterations : 100
Learning Rate : 0.002

$$\text{Score} = \frac{\textit{correct predictions}}{\textit{total predictions}}$$

$$\text{Precision} = \frac{\textit{true positives}}{\textit{true positives + false positives}}$$

$$\text{Recall} = \frac{true\ positives}{true\ positives\ +\ false\ negatives}$$

Type	Score	Precision	Recall
Bag of Words (BoW)	0.9934339413606998	0.9946432917051455	0.9975120880628235
Term Frequency (TF)	0.9906717165444	0.9938483385179893	0.9950207468879668
TF-IDF	0.9923970640515	0.9958930672826186	0.995022461506807

3 Handwritten Digit Recognition on MNIST Dataset

Learning Algorithm : k-Nearest Neighbours

$$\text{Score} = \frac{correct\ predictions}{total\ predictions}$$

Metric Used	Definition of metric	Score
Euclidean	$\sqrt{\sum(x-y)^2}$	0.9705
Manhattan	$\sum(\mid x-y \mid)$	0.9705
Chebyshev	$\sum(max(\mid x-y \mid))$	0.7487