Assignment 2

BITS F464 Machine Learning

Submission Date:15/04/2019

1. ***Logistic Regression*** (Probabilistic Discriminative Model)-as described in Bishop ch 4.

   - Visualize the dataset labels

   - Use the gradient descent algorithm to implement logistic regression

   - Try using different learning rates and initialization

   - Add regularization and contrast the learnt weight values and model performance. (Elucidate these observations in your report)

   - Libraries may be used only for data handling and visualisation. The algorithm must be coded from scratch.

   Dataset Information
   1. variance of Wavelet Transformed image (continuous)
   2. skewness of Wavelet Transformed image (continuous)
   3. curtosis of Wavelet Transformed image (continuous)
   4. entropy of image (continuous)
   5. class (integer)

2. ***Sentiment analysis with Naive Bayes***

   - The attached dataset has category (ignore this), sentiment (neg/pos), document identifier, and the document respectively

   - Make a 80-20 data split for validation. Note that there is no need to perform tokenisation or normalization on this data.

   - Learn the probabilities using the naive bayes model on the train set and use them to classify the unseen examples as negative or positive

- As always, libraries except numpy/pandas and their equivalents are not permitted

Java, C++, C or Python may be used for this assignment. Submitted codes will be tested on standard plagiarism detection tools and those with plagiarism above a threshold will not be evaluated.

Report

- Prepare a formal report detailing your observations and results obtained.

- Present the results of the naive bayes technique as a confusion matrix along with precision, recall and F1 measure values

- Note that the code must be properly indented and commented.

For any queries reach out to Kshitij Grovor at f20150070@hyderabad.bits-pilani.ac.in