2017 Fall: COMP-SCI 5590/490 - Special Topics

Python Programming

Lab Assignment 3

Assignment Overview

The following assignment focus on to make one familiar with python machine learning

Lab Assignment

1) Pick any dataset from the dataset sheet in class sheet and make one prediction model using your imagination with **Linear Discriminant Analysis***. Some examples are:
    a. In the report provide convincible explanations about the difference of logistic regression and Linear Discriminant Analysis.
    b. You can also pick dataset of your own.

*Logistic Regression is a classification algorithm traditionally limited **to only two-class classification problems**. If you have **more than two classes then the Linear Discriminant Analysis algorithm is the preferred** linear classification technique.

2) Implement Support Vector Machine classification,
    1) Choose one of the dataset using the datasets features in the scikit-learn
    2) Load the dataset
    3) According to your dataset, split the data to 20% testing data, 80% training data(you can also use any other number)
    4) Apply SVC with Linear kernel
    5) Apply SVC with RBF kernel
    6) Report the accuracy of the model on both models separately and report their differences if there is
    7) Report your view how can you increase the accuracy and which kernel is the best for your dataset and why

3) Write a program
    Take an Input file. Use the simple approach below to summarize a text file:

- Read the file
- Using Lemmatization, apply lemmatization on the words
- Apply the bigram on the text
- Calculate the word frequency (bi-gram frequency) of the words (bi-grams)
- Choose top five bi-grams that has been repeated most
- Go through the original text that you had in the file
- Find all the sentences with those most repeated bi-grams
- Extract those sentences and concatenate

- Enjoy the summarization

4) Report your views on the k nearest neighbor algorithm when we change the K how it will affect the accuracy. Provide a good justification about the changes of the accuracy when we change the amount of K.
   For example: compare the accuracy when K=1 and K is a big number like 50, why the accuracy will change

Submission Guidelines:

- Submit your code at Github and properly document it. Submit your screenshots as well.
- Properly document your code
- Submit only the code portion in text file to UMKC blackboard assignment.
- Remember code similarity to be less than 45%
- Use following link to submit your assignment:

  https://goo.gl/forms/cxvY8Kg1pvNNzrpw1