Lab 2: Sentiment Classification with Support Vector Machine

Aim:

- Understanding Support Vector Machine Classification technique
- Get familiar with data analysis process
- Learning to employ Python to do data analysis

Materials:

We provide you with three data datasets we got from a social media site:

1) training dataset

2) test dataset

3) evaluation dataset

Requirements:

1. Using Python to create your own SVM classifier and train the network to classify data into two sentiment classes: ”positive” and ”negative”. You decide how to build your own SVM classifier, for example, using linear or nonlinear SVM, the type of kernels you would like to use.

2. Validate your classifier with the test dataset.

3. Applying your classifier to the evaluation dataset to measure the performance. You decide which criteria you would like to use. In the best case you can try to measure the performance with all combinations of indicators (accuracy/error rate, sensitivity/specificity, precision/recall, or ROC).

4. Write a lab report containing 1) title 2) names and team 3) systematic diagram of your sentiment analysis process, including training and testing; 4) feature extraction; 5) evaluation; 6) results; 7) comparison with the results from your Naïve Bayesian classifier; 8) the link to your code (or upload your code).