Jerry Cheng

Machine Learning – Sentiment Analysis Question

1/11/2018

1. What is nltk and what purpose does it serve in Python?
   1. “Natural language toolkit”. Frequency distribution of words, stemming.
2. What is a stemmer?
   1. Stemmers remove affixes from words, leaving only the word stem
3. What are stop words?
   1. A stop word is a commonly used word (and, an, I) that a search engine, or in this case nltk is programmed to ignore.
4. In the line “for I, row in enumerate(plots):”, what is stored in I and what is stored in row?
   1. I contains the line number for the file, and row stores the actual data.
5. What is a naïve bayes classifier? What are some advantages of it versus other ML algorithms? What are some disadvantages?
   1. A naïve bayes classifier is a family of probabilistic classifiers based on applying Bayes’s theorem. The naïve bayes classifier is extremely simple, fast, and easy to implement. However, it makes very strong assumptions about your data, sometimes resulting in errors.



1. Experiment:
   1. Vary the amount of training Tweets that we use to train the classifier. Plot the accuracy as a function of the runtime using matplotlib, with the legend indicating the number of training Tweets used.