605.744: Information Retrieval
Problem Set (Module 7)

Sabbir Ahmed

October 13, 2022

1. (40%) Briefly explain the following concepts in your own words and be sure to indicate how they are relevant to text classification:

    (a) bias-variance tradeoff

    Answer:

    (b) k-fold cross-validation

    Answer:

    (c) macro-averaging vs. micro-averaging

    Answer:

    (d) soft margin

    Answer:

2. (10%) Why is 3-Nearest-Neighbor (3-NN) almost always a better choice than 1-NN for a binary (i.e, two-class) text classification problem.

    Answer:

3. (50%) Naïve Bayes using the Binomial (also known as Bernoulli) model. First calculate estimates of P(c) and P(w|c) given the following training sentences. There are only three classes: Travel, Business, and Health. You should not use any smoothing. For features you should only use the following seven vocabulary terms: denver, employers, florida, hospital, jobs, nurses, vacation and you should ignore all other words and any punctuation. Next compute the probability of each class for the two test documents A & B below. Finally, indicate which is the predicted (i.e., the most likely) class for each test document. Please read these directions thoroughly, count carefully, and do show all of your work. Report probabilities using scientific notation (e.g., 1.563×10^{-5}) with three digits after the decimal point.

    Training data

    (a) Travel: denver hospital administrator takes vacation in florida
    (b) Travel: nurses plan a trip to florida
    (c) Travel: employers offering more jobs with vacation benefits
    (d) Business: employers see growth in information science
    (e) Business: hospital nurses in denver say high paying jobs are vanishing

(f) Business: employers say florida is nice vacation spot and there are good jobs

(g) Health: study: more hospital nurses need to take a vacation

(h) Health: local doctors attend florida conference on diabetes

(i) Health: hospital trains maternity ward nurses

(j) Health: denver hospital says local employers have jobs for nurses

Test documents:

(a) florida nurses take skiing vacation in denver

(b) jobs available for experienced nurses at florida hospital

Answer: