1 Algorithms

1.1 Naive Bayes

Recall, in Naive Bayes we are interested in a hypothesis that gives maximum posterior probability \( P(h|D) \), or \( \arg\max_{h \in H} P(h|D) \) where \( h \) is a hypothesis, \( H \) is hypothesis space, and \( D \) is data. For classification, hypothesis space is all possible classes \( C \), and a single hypothesis is a class we are trying to predict \( \arg\max_{c_j \in C} P(c_j|D) \).

Using Bayes Rule, \( P(c_j|D) = \frac{P(D|c_j)P(c_j)}{P(D)} \). Since every data point is viewed as a tuple of attribute values and since Naive Bayes assumes conditional independence between the attributes, \( P(D|c_j) = P(v_1v_2...v_n|c_j) = \prod_{i=1}^{n} P(v_i|c_j) \) where \( v_i \) is a value of attribute \( a_i \). Finally, \( \arg\max_{c_j \in C} \prod_{i=1}^{n} P(v_i|c_j)P(c_j) \) is a rule used by Naive Bayes Classifier since \( P(D) \) is the same for every hypothesis.

1.2 Lee’s Classifier

1.2.1 Random walk:

In this case, we are interested in log posterior odds of a document belonging to a certain topic or not.

\[
\ln \frac{P(T|D)}{P(\neg T|D)} = \ln \frac{P(T)P(D|T)}{P(\neg T)P(D|\neg T)} = \ln \frac{P(T)\prod_{i=1}^{n} P(w_i|T)}{P(\neg T)\prod_{i=1}^{n} P(w_i|\neg T)}
\]

using independence assumption, this becomes \( \ln \frac{P(T)}{P(\neg T)} + \sum_{i=1}^{n} \ln \frac{P(w_i|T)}{P(w_i|\neg T)} \).

After the completion of the walk and calculation of the log posterior odds a decision is based on whether this value reached a threshold \( \epsilon \).

1.2.2 Accumulator Text Classifier

This method is similar to the one above, only in this case we compute the totals for a word being about the topic, or not being about the totals. That is, we keep track of \( A_T \) and \( A_{\neg T} \):

- if \( V_T(w_i) = \ln \frac{P(w_i|T)}{P(w_i|\neg T)} < 0 \), \( A_{\neg T} = V_T(w_i) \)
  - else \( A_T = V_T(w_i) \)

Then confidence measure is \( \frac{A_T - |A_{\neg T}|}{A_T + |A_{\neg T}|} \) and again, the decision is based on whether this value has reached a threshold \( \epsilon \).