Michael Pollack
LING572 Hw3: NB and kNN
Due: 11pm on Jan 31, 2025

Q1 (35 points): Write a script, **build_NB.sh**, that implements the multinomial NB model (i.e., the model we discussed in class). It builds a NB model from the training data, classifies the training and test data, and calculates the accuracy.

Run build_NB.sh with **train.vectors.txt** as the training data, **test.vectors.txt** as the test data, and class_prior_delta set to 0:

- Fill out Table 1 with different values of cond_prob_delta when class_prior_delta = 0.

- Submit the model_file, sys_output and acc_file for the second row (when cond_prob_delta is 0.5) under q1/.

Table 1: Results of your Multinomial NB model when class_prior_delta = 0

cond_prob_delta	Training accuracy	Test accuracy
0.1	0.9596296296296296	0.91
0.5	0.9562962962962963	0.9033333333333333
1.0	0.952962962962963	0.8966666666666666

Q2 (40 points): Write a script, **build_kNN.sh**, that implements the kNN algorithm. It classifies a test instance x by letting the k nearest neighbors of x vote.

Run build_kNN.sh with **train.vectors.txt** as the training data and **test.vectors.txt** as the test data:

- Fill out Table 2 with different values of k and similarity function.

- Submit sys_output and acc_file with **k_val=5** and **similarity_function=2** under the directory q2/.

Table 2: Test accuracy using **real-valued** features

k	Euclidean distance	Cosine function
1	0.62	0.72
5	0.64	0.6966666666666667
10	0.6566666666666666	0.67

Submission: Submit the following to Canvas:

- Your note file *readme.(txt | pdf)* that includes Table 1 and 2, and any notes that you want the TA to read.

- hw.tar.gz that includes all the files specified in dropbox/24-25/572/hw3/submit-file-list, plus any source code (and binary code) used by the shell scripts.

- Make sure that you run **check_hw3.sh** before submitting your hw.tar.gz.