You are given data associated with a speech classification task, where there are 8 classes. You have four sets of data: a general training set from 25 speakers, a small set of matched data for each of the test speakers, an independent test set with known labels for development, and a second unlabeled test set. Pick one of the following two options, and use the data to design and test three different classifiers: i) train on the general set only, ii) train on the matched set only, and iii) use the matched set together with the general set in an adaptation mode. The two options are:

1. Gaussian class-conditional distributions: Choose between full and diagonal covariance and justify your choice. For condition (iii), use MAP estimation of the mean, where the prior is the model that you found from general training, and use either the general variance or Bayesian learning to find the adapted model variance.

2. Nearest-neighbor: For condition (i), use a computation reduction technique that you explored in HW4. (No code sharing with others from the class.) For condition (ii), there is much less data, so you may not need to reduce computation. For condition (iii), you can get an adapted model by first using the matched data to edit the general data (i.e. find the nearest neighbors in the general set to the matched samples, and throw out all general samples that misclassify the matched data). Combine the remaining data and do further computation reduction as needed.

Use the three models to predict classes on the unlabeled test. Create a separate ascii file for each of the 3 conditions, with the predicted class label on a separate line for each test sample. Use the following naming convention for the files:

<yournamename>_<Gauss/KNN>_C<condition number>.txt

The classifications and code should be emailed to the grader. The written discussion of your approach is due in hard copy with a signed version of this cover page. The discussion should include details of your approach, results on the development set, and discussion, written in the style of a technical report.

To constrain the scope of the problem and facilitate comparison of results, please do not experiment with any type of feature transformation, and only implement one of the two approaches (Gaussian vs. KNN). You will be graded on your description of the method you chose and how well you do for that particular method, but grades will not be based on performance relative to those who chose a different method.

Honor Code:
This exam represents only my own work. I did not give or receive help on this exam.