

Problem Set 2

Remember that for qualitative questions, you must answer *in your own words*. When submitting this assignment, provide a hardcopy of your answers (including code and plots). Submit your code online using [Blackboard](#), see the [instructions](#) on how to do this.

1. (30 points) Write a function called *distortion* that computes the Euclidean distortion¹ of a vector against a set of column vectors contained in a matrix. The inputs should be a column vector and the matrix. Two outputs should be produced, the minimum distortion of the vector when compared to all of the vectors in the matrix, and the index of the matrix column that produced the minimum distortion.

Test your function using the mean of the vectors contained in file `cluster.mfc` which can be obtained from rohan in directory `~mroch/lib/cs682/` along with the file function `spReadFeatureDataHTK` which can be used to read the features. The features represent an individual saying the digit sequence: 3 3 0 3 1.

2. (30 points) Write a function called *partition* that selects five column vectors at random from a $2 \times N$ matrix. Plot these vectors using a different symbol and color for each. For each of the vectors, determine which of the five column vectors it is closest to and plot it with a point of the same color. The function should return the average of the minimum distortions. As an example, a linespec of 'g.' would be a green dot, see doc `linespec` for more details on colors and symbols. Test your algorithm by passing in the first two rows from the cluster matrix that you read in the previous problem.
3. (20 points) Why it is important to have both development and evaluation sets whenever possible?
4. (20 points) What are feature vectors?

¹ Euclidean distance squared