CAP 6638 **Pattern Recognition** **Spring 2015**
MW 05:00 PM – 06:15 PM in NES 013

Project #3

This assignment invites you to experiment with the Nearest Neighbor method, a simple and effective classification procedure.

It is generally used with much larger sample sizes than what we have.

With each of the following methods, use Data Set A of 100 samples as the *reference set*, and classify the remaining three sets. Print a **summary table of the error rates**, with one row for each method, and one column for each data set. (On Data Set A you should have no errors; because Nearest Neighbors classification yields zero error on the reference set.) Print out all three confusion tables for Methods 1 and 2 only.

1. **NN on the bitmaps of the samples**, with Euclidian metric. Note that you may replace the squared distance with an equivalent linear expression. Decide ties by lexical precedence.

2. **5-NN on the bitmaps of the samples** (vote among the five nearest neighbors of each unknown sample, break ties lexically).

3. **NN on the eight moment features**, with Euclidian metric.

4. **5-NN on the eight moment features**, with Euclidian metric.

5 (Optional – extra credit). Prune the reference set (using moments) *to no more than 25 samples* using any method documented in the literature, then classify the unknown samples using the pruned reference set. (Synonyms for pruning are "editing" and "condensing".) **Indicate the exact number of patterns left in the pruned reference set** and cite the source of the method. Attach a **brief explanation** of your method (~half page).

6 (Optional – extra credit). Reduce the number of distance calculations for classification using any method documented in the literature. If there were no reduction, you would have to perform 100 distance calculations for each unknown character (one for each reference character). **Report the average number of distance calculations per character.** Use the unpruned reference set here. Note that the speed-up should not change your classification. **Explain your method** briefly (~half page) and cite source.

A Very Incomplete Bibliography on Nearest Neighbors:

Dasarathy, B., *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*, IEEE Computer Society Press, Silver Spring MD, Library of Congress Number 90-45250, ISBN 0-8186-8930-7, 1991.

Cover, T.M., Hart, P.E., Nearest Neighbor Pattern Classification, *IEEE-IT 13*, 21-27, January 1967.

Hart, P.E., The Condensed N-N Rule, *IEEE-IT 14*, 515-6, May 1968.

Gates, G.W., The Reduced N-N Rule, *IEEE-IT*, 431-433, May 1972.

Ritter, G.L. et al, An Algorithm for Selective N-N Decision Rule, *IEEE-IT* 665-669, Nov. 1975.

Tomek, I., Two modifications of CNN, *IEEE-SMC*, 769-772, Nov. 1976.

Vidal, E., An algorithm for finding nearest neighbors in (approximately) average constant time, *Pattern Recognition Letters 4*, 3, 145-157, 1986.

Dwyer, R.A., Higher Dimensional Voronoi Diagrams in Linear Expected Time, *Procs. Fifth Symp. on Computational Geometry,* 326-332, 1989.

J. McNames, A Fast Nearest-Neighbor Algorithm Based on a Principal Axis Search Tree, PAMI-23, 9, September 2001, 964-976.

Ramasubramanian, V., Paliwal, K.K., An efficient approximation-elimination algorithm for fast nearest-neighbor search based on a spherical distance coordinate formulation, *Pattern Recognition Letters 13*, 7, 471-480, 1992.

Jaromczyk, J.W. & Toussaint, G.T., Relative Neighborhood Graphs and their Relatives, *Proceedings of the IEEE 80*, 9, 1502-17, Sept. 1992.

Mico., L., Oncino, J., Vidal, E, A new version of the Nearest-Neighbor Approximating and Eliminating Search Algorithm (ASEA) with linear preprocessing time and memory requirements, *Pattern Recognition Letters 15,* 1, 9-17, January 1994.

Wu, Y., Ianakiev, K., Govindaraju, V., Improved k-nearest neighbor classification, *Pattern Recognition 35,* 10, 2311-2318, October 2002.