EE 559 **Homework - Week 3** Posted: Thur., 1/24/2019
Jenkins Due: Fri., 2/1/2019, 2:00 PM

Note: for a dataset, **classification accuracy** is defined as number of correctly classified data points divided by total number of data points.
Reminder: **include a copy of your code** as part of your homework submission, as one separate computer-readable pdf file, for all assignments that include computer problems in this class.

1. In this 3-class problem, you will use the one vs. one method for multiclass classification. Let the discriminant functions be:

$$g_{12}(\underline{x}) = -x_1 - x_2 + 5$$
$$g_{13}(\underline{x}) = -x_1 + 3$$
$$g_{23}(\underline{x}) = -x_1 + x_2 - 1$$
and $g_{ji}(\underline{x}) = -g_{ij}(\underline{x})$.

The decision rule is:

$$\underline{x} \in S_k \text{ iff } g_{kj}(\underline{x}) > 0 \text{ for all } j \neq k.$$

Draw the decision boundaries and label classified regions and any indeterminate regions. Classify the points $\underline{x} = (4,1),\ (1,5),\text{ and } (0,0)$. If there is an indeterminate region prove it by finding a point that doesn't get classified according to the above rule. If there is no indeterminate region, so state.

2. For the wine dataset, code up a minimum-distance-to-class-means classifier with the following multiclass approach: one vs. rest. Use the original unnormalized data. Note that the class means should always be defined by the training data. Run the one vs. rest classifier using only the following two features: 1 and 2.

Note that the same guidelines as HW 2 apply on coding the classifier(s) yourself vs. using available packages or routines , with one possible exception*.

Give the following:

(a) Classification accuracy on training set and on testing set.

(b) Plots showing each resulting 2-class decision boundary and regions ($S_k{}'$ $vs.$ $\overline{S_k{}'}$)

(c) A plot showing the final decision boundaries and regions ($\Gamma_1, \Gamma_2, \Gamma_3,$ indeterminate).

Hint 1: For (b) and (c), you can use `plotDecBoundaries()`. Modify it if necessary.

Hint 2: *If using Python, you may optionally use scipy.spatial.distance.cdist in calculating Euclidean distance between matrix elements.

3. (a) Derive an expression for the discriminant function $g(x)$ for a 2-class minimum-distance-to-class-means classifier, based on Euclidean distance, for class means $\underline{\mu}_1$ and $\underline{\mu}_2$. Is the classifier linear?

(b) Continuing from part (a), for the following class means:

$$\underline{\mu}_1 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \quad \underline{\mu}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Plot the decision boundaries and label the decision regions.

Note: *Parts (c) and (d) below use the maximal value (or "linear machine") method for multiclass classification. This method is described briefly in the reading (DHS Sec. 5.2.2) and in Discussion 3. It will also be covered in more detail in Lecture 6 (Week 4) on Monday.*

(c) Repeat part (a) except for a 3-class classifier, using the maximal value method: find the three discriminant functions $g_1(\underline{x}),\ g_2(\underline{x}),\ g_3(\underline{x})$, given three class means $\underline{\mu}_1$, $\underline{\mu}_2$, and $\underline{\mu}_3$. Express in simplest form. Is the classifier linear?

(d) Continuing from part (c), for the following class means:

$$\underline{\mu}_1 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \quad \underline{\mu}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \underline{\mu}_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Plot the decision boundaries and label the decision regions.

4. **Extra credit.** DHS Problem 5.9. (Note that DHS has a set of "Problems", and a set of "Computer Exercises", both at the end of each chapter. This is "Problem" 9 of Chapter 5.) The problem statement starts "The *convex hull* of a set of vectors…". Some versions of the DHS text may have a slightly different numbering of problems, so it's best to check every time that you are going to solve an assigned problem.

Additional hint: Classify the point $\underline{x}$ twice, once based on $\underline{x}$ in the convex hull of S_1 data points, and a second time based on $\underline{x}$ in the convex hull of S_2 data points.