EE 559 **Homework 2 (Week 3)** Posted: Fri., 2/5/2021
Jenkins Due: Fri., 2/12/2021, 11:59 PM

Note: for a dataset, **classification accuracy** is defined as number of correctly classified data points divided by total number of data points.

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:

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

Draw the decision boundaries and label decision regions Γ_i and any indeterminate regions. Classify the points $\underline{x} = (2,4),\ (4,3),\ \text{and } (1,2)$. 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 (from Homework 1 data files), code up a nearest-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 Homework 1 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 test set.

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

(c) A plot showing the final decision boundaries and regions (Γ_1, Γ_2, Γ_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 nearest-means classifier, based on Euclidean distance, for class means $\underline{\mu}_1$ and $\underline{\mu}_2$. Keep the number of dimensions variable. Express in simplest form.[1] Is the classifier linear[2]?

Hints: [1]Remember that the expression for $g(x)$ is not unique; choose an expression that yields the simplest result. What matters is how $g(x)$ compares to 0.
[2]You can check your answer by comparing with a plot of the decision boundary.

(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.

(c) Repeat part (a) except for a 3-class classifier, using the maximal value method (MVM): 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) using MVM, 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. (a) Let $p(\underline{x})$ be a scalar function of a D-dimensional vector $\underline{x}$, and $f(p)$ be a scalar function of p. Prove that:

$$\nabla_{\underline{x}} f\left[p(\underline{x})\right] = \left[\frac{d}{dp} f(p)\right] \nabla_{\underline{x}}\, p(\underline{x})$$

i.e., prove that the chain rule applies in this way. [**Hint:** you can show it for the i^{th} component of the gradient vector, for any i. It can be done in a couple lines.]

(b) Use relation (4) of "Import expressions" in Discussion 1, to find $\nabla_{\underline{x}}\left(\underline{x}^T \underline{x}\right)$.

(c) Prove your result of $\nabla_{\underline{x}}\left(\underline{x}^T \underline{x}\right)$ in part (b) by, instead, writing out the components.

(d) Use (a) and (b) to find $\nabla_{\underline{x}}\left[\left(\underline{x}^T \underline{x}\right)^3\right]$ in terms of $\underline{x}$.

5. (a) Use relations above to find $\nabla_{\underline{w}} \|\underline{w}\|_2$. Express your answer in terms of $\|\underline{w}\|_2$ where possible. **Hint:** let $p = \underline{w}^T \underline{w}$; what is f ?

(b) Find: $\nabla_{\underline{w}} \left\| \underline{\underline{M}}\,\underline{w} - \underline{b} \right\|_2$. Express your result in simplest form. **Hint:** first choose p (remember it must be a scalar).

6. **[Extra credit]** For $C > 2$, show that total linear separability implies linear separability, and show that linear separability doesn't necessarily imply total linear separability. For the latter, a counterexample will suffice.