Haolun Cheng
1882563827

(Written answers start at page (3))

EE 559 | **Homework 2 (Week 4)** | Posted: Tues., 2/1/2022
Jenkins | v2 | Due: Wed., 2/9/2022, 11:59 PM

Text in green is new in v2

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. rest method for multiclass classification. Let the discriminant functions be:

$$g_1(\underline{x}) = -x_1 - x_2 + 5$$
$$g_2(\underline{x}) = x_1 - 3$$
$$g_3(\underline{x}) = -x_1 + x_2 - 1$$

In this problem, use the OvR decision rule given in lecture.

Draw the decision boundaries and label decision regions Γ_i and any indeterminate regions. Classify the points $\underline{x} = (2,4), (4,3),$ and $(1,2)$. If there are indeterminate regions, show that a point in (one of) the region(s) doesn't get classified according to the OvR decision 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. one. Use the original unnormalized data given with Homework 1, and use the decision rule given in lecture. Note that the class means should always be defined by the training data. Run the 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 (S_k vs. S_j).
(c) A plot showing the final decision boundaries and regions ($\Gamma_1, \Gamma_2, \Gamma_3$, indeterminate if any). Please note that decision boundaries (which have area=0) don't count as indeterminate regions.

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 has a simple form. 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.

(e) Repeat part (a) except for a 3-class classifier, using the one vs. one (OvO) method: find the three discriminant function, $g_{1,2}(\underline{x})$, $g_{1,3}(\underline{x})$, $g_{2,3}(\underline{x})$ given three class means $\underline{\mu}_1$, $\underline{\mu}_2$, and $\underline{\mu}_3$. Express in simplest form. Is the classifier linear?

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 "expressions" in Discussion 2, 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 $\nabla_{\underline{w}} \left\| \underline{w} \right\|_2$. Express your answer in terms of $\left\| \underline{w} \right\|_2$ 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).

Written answers start from here:

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

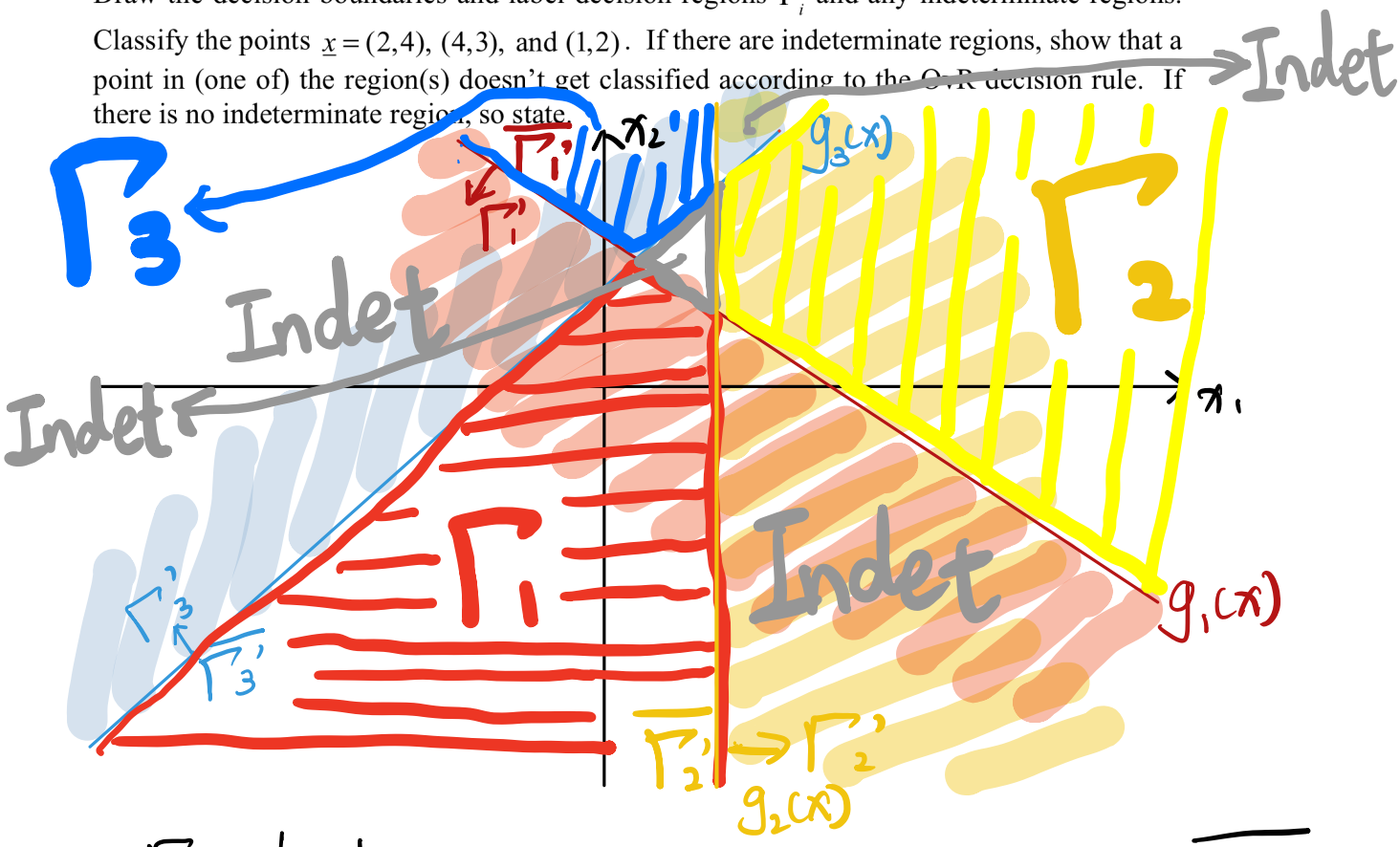
$$g_1(\underline{x}) = -x_1 - x_2 + 5$$

$$g_2(\underline{x}) = x_1 - 3$$

$$g_3(\underline{x}) = -x_1 + x_2 - 1$$

In this problem, use the OvR decision rule given in lecture.

Draw the decision boundaries and label decision regions Γ_i and any indeterminate regions. Classify the points $\underline{x} = (2,4)$, $(4,3)$, and $(1,2)$. If there are indeterminate regions, show that a point in (one of) the region(s) doesn't get classified according to the OvR decision rule. If there is no indeterminate region, so state.



Γ_1 – class 1

Γ_2 – class 2

Γ_3 – class 3

OvR: $\underline{x} \in \Gamma_k$ IFF $\underline{x} \in \Gamma_k'$ AND $\underline{x} \in \overline{\Gamma_j'}$ $\forall\, j \neq k$.

Indet: Indeterminate regions

Points: $\because \underline{x} = (2,4) \longrightarrow g_1(\underline{x}) = -2-4+5 = -1$

$g_2(\underline{x}) = 2-3 = -1$

$g_3(\underline{x}) = -2+4-1 = 1$

$\therefore \underline{x} = (2,4) \rightarrow$ class 3 (Γ_3)

$\because \underline{x} = (4,3) \rightarrow g_1(\underline{x}) = -4-3+5 = -2$

$g_2(\underline{x}) = 4-3 = 1$

$g_3(\underline{x}) = -4+3-1 = -2$

$\therefore \underline{x} = (4,3) \rightarrow$ class 2 (Γ_2)

$\because \underline{x} = (1,2) \longrightarrow g_1(\underline{x}) = -1-2+5 = 2$

$g_2(\underline{x}) = 1-3 = -2$

$g_3(\underline{x}) = -1+2-1 = 0$

$\therefore \underline{x} = (1,2) \longrightarrow$ is on the decision boundary of class 3 (Γ_3) and is also inside the class 1 (Γ_1) region. This point does not get classified according to the OvR decision rule.

2. For the wine dataset (from Homework 1 data files), code up a nearest-means classifier with the following multiclass approach: one vs. one. Use the original unnormalized data given with Homework 1, and use the decision rule given in lecture. Note that the class means should always be defined by the training data. Run the 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 (S_k vs. S_j).

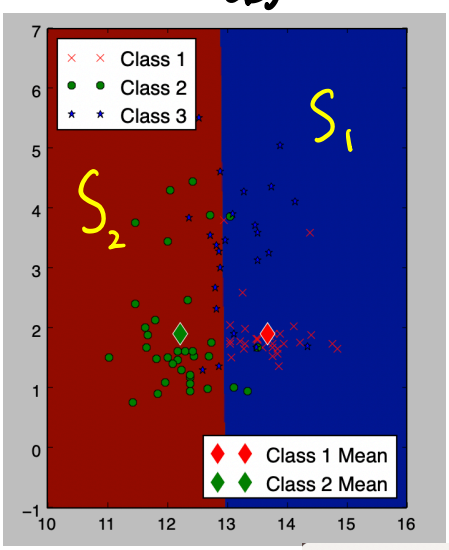
   (c) A plot showing the final decision boundaries and regions ($\Gamma_1, \Gamma_2, \Gamma_3$, indeterminate if any). Please note that decision boundaries (which have area=0) don't count as indeterminate regions.

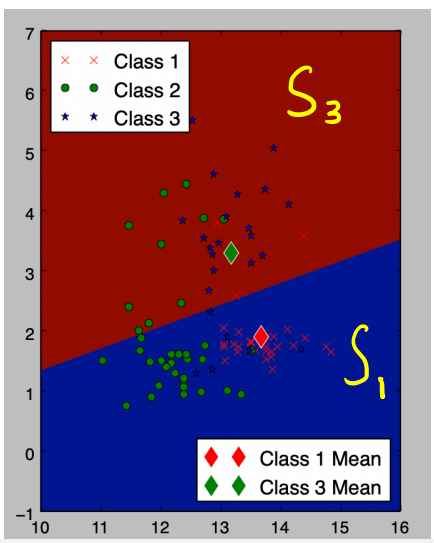
   Hint 1: For (b) and (c), you can use PlotDecBoundaries(). Modify it if necessary.

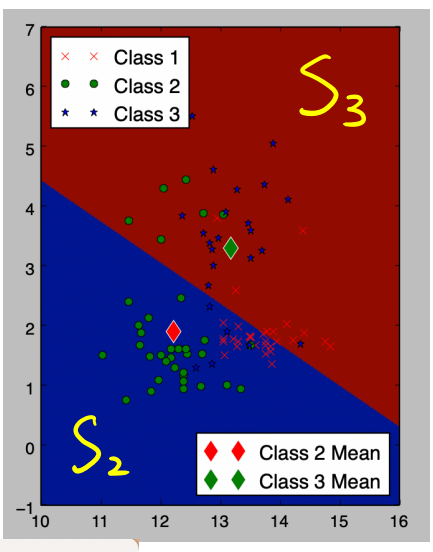
   Hint 2: *If using Python, you may optionally use scipy.spatial.distance.cdist in calculating Euclidean distance between matrix elements.

(a) Please check the other PDF for code.

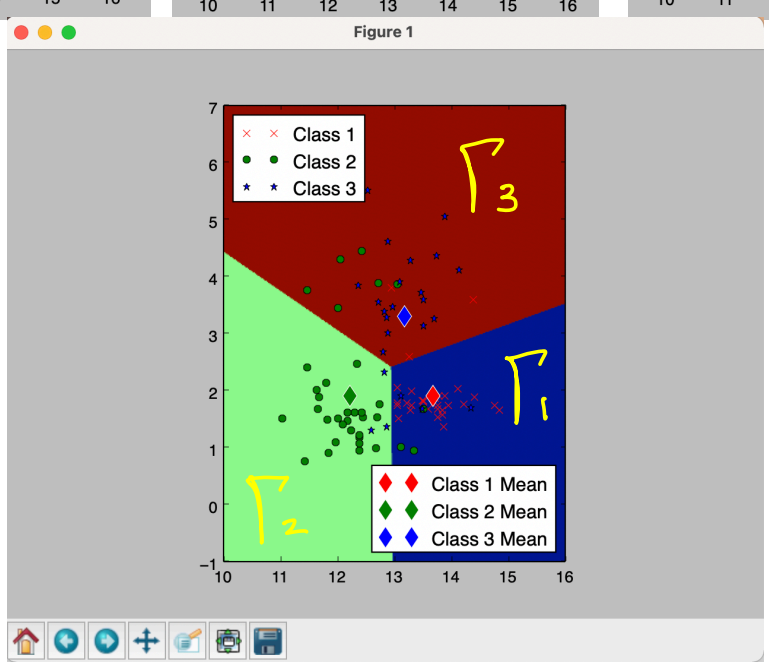
(b)







(c)



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 has a simple form. What matters is how $g(x)$ compares to 0.
[2]You can check your answer by comparing with a plot of the decision boundary.

Suppose point A has a coordinate (x_1, x_2)
and we have two class means u_1 and u_2.

$$g(x) = \|u_1 - x\|_2^2 - \|u_2 - x\|_2^2$$

$$= u_1^T u_1 + x^T x - 2u_1^T x - (u_2^T u_2 + x^T x - 2u_2^T x)$$

$$= u_1^T u_1 + x^T x - 2u_1^T x - u_2^T u_2 - x^T x + 2u_2^T x$$

$$= 2x^T(u_2 - u_1) + u_1^T u_1 - u_2^T u_2$$

$\because$ $g(x) = w_0 + w^T x$

$\therefore$ $w = 2(u_2 - u_1)$

$w_0 = u_1^T u_1 - u_2^T u_2$

Thus, if $g(x) > 0$, x belongs to class 2,
if $g(x) < 0$, x belongs to class 1,
if $g(x) = 0$, x is on the decision boundary.

The classifier is 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.

$\because$ from part (a), we have $\begin{cases} w_0 = u_1^T u_1 - u_2^T u_2 \\ w = 2(u_2 - u_1) \end{cases}$

$\therefore w_0 = u_1^T u_1 - u_2^T u_2 = [0 \ -2]\begin{bmatrix} 0 \\ -2 \end{bmatrix} - [0 \ 1]\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3$

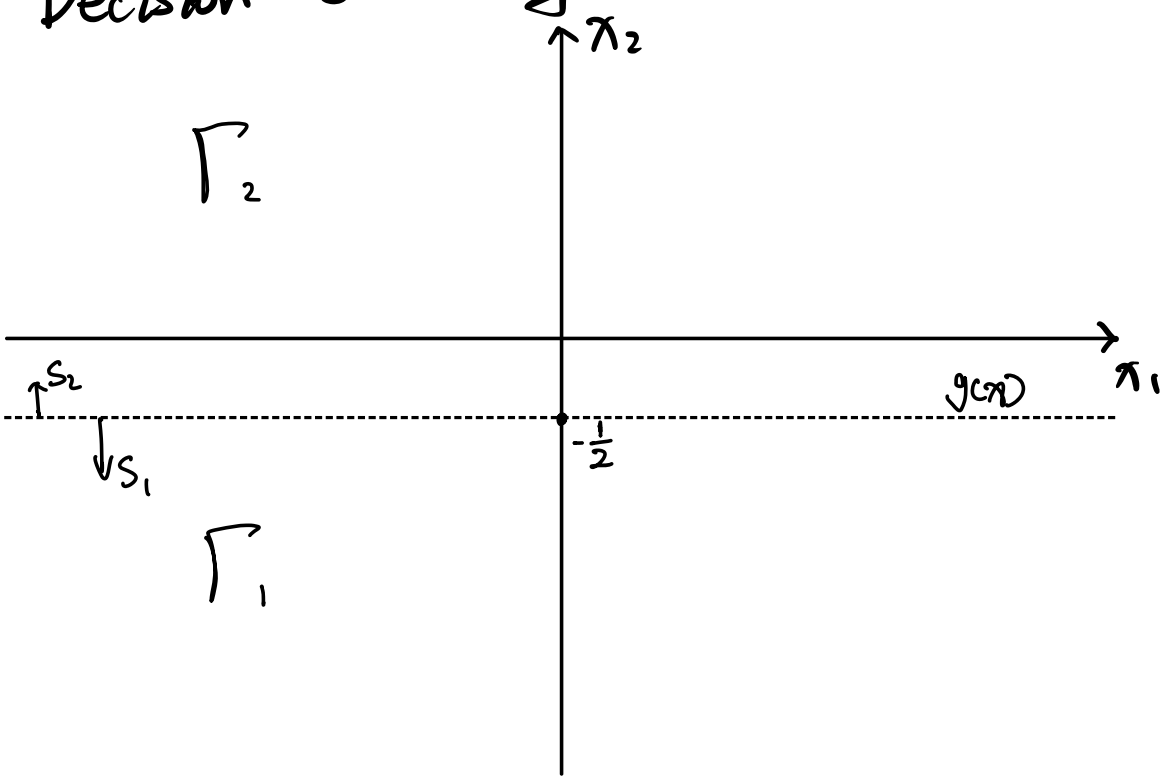
$w^T = 2[u_2 - u_1] = 2\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

$\therefore g(x) = 3 + [x_1 \ x_2]\begin{bmatrix} 0 \\ 6 \end{bmatrix}$

$= 3 + 6x_2$

$\therefore x_2 = -\frac{1}{2}$

Decision Boundary:



(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?

From part (A), we have a point (x_1, x_2)

$\therefore g_1(x) = -2x^T u_1 + u_1^T u_1$

$g_2(x) = -2x^T u_2 + u_2^T u_2$

$g_3(x) = -2x^T u_3 + u_3^T u_3$

Yes, the classifier is 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.

$g_1(x) = -2[x_1 \ x_2]\begin{bmatrix} 0 \\ -2 \end{bmatrix} + [0 \ -2]\begin{bmatrix} 0 \\ -2 \end{bmatrix}$

$= 4x_2 + 4$

$g_2(x) = -2[x_1 \ x_2]\begin{bmatrix} 0 \\ 1 \end{bmatrix} + [0 \ 1]\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$= -2x_2 + 1$

$g_3(x) = -2[x_1 \ x_2]\begin{bmatrix} 2 \\ 0 \end{bmatrix} + [2 \ 0]\begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$= -4x_1 + 4$

$\because$ According to MVM:

① $g_1(x) = g_2(x) \Rightarrow \quad 4x_2 + 4 = -2x_2 + 1$

$6x_2 = -3$

$x_2 = -\frac{1}{2}$

② $g_1(x) = g_3(x) \Rightarrow \quad 4x_2 + 4 = -4x_1 + 4$

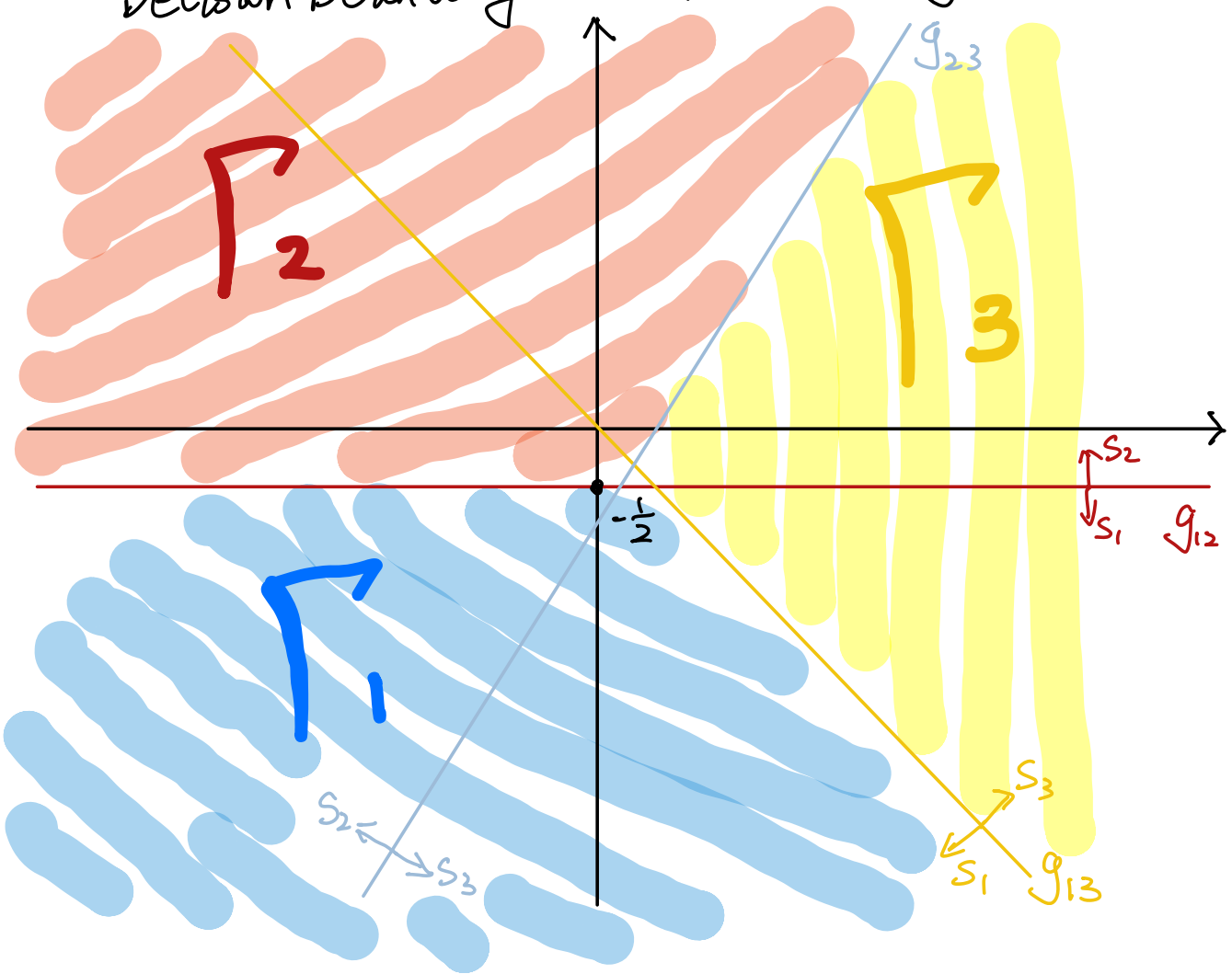
$x_1 + x_2 = 0$

③ $g_2(x) = g_3(x) \Rightarrow \quad -2x_2 + 1 = -4x_1 + 4$

$$4x_1 - 2x_2 = 3$$

$$2x_1 - x_2 = \frac{3}{2}$$

Decision Boundary and Decision Regions:



(e) Repeat part (a) except for a 3-class classifier, using the one vs. one (OvO) method: find the three discriminant function, $g_{1,2}(\underline{x})$, $g_{1,3}(\underline{x})$, $g_{2,3}(\underline{x})$ given three class means $\underline{\mu}_1$, $\underline{\mu}_2$, and $\underline{\mu}_3$. Express in simplest form. Is the classifier linear?

∴ From part (a), we have: $\begin{cases} w = -2u_i \\ w_0 = u_i^T u_i \end{cases}$ for $i = 1, 2, 3$.

And from part (c), we have

$$g_1(x) = -2x^T u_1 + u_1^T u_1$$
$$g_2(x) = -2x^T u_2 + u_2^T u_2$$
$$g_3(x) = -2x^T u_3 + u_3^T u_3$$

$$\therefore g_{1,2}(x) = g_1(x) - g_2(x) = -2x^T u_1 + u_1^T u_1 - 2x^T u_2 - u_2^T u_2$$
$$= -2x^T(u_1 + u_2) + u_1^T u_1 - u_2^T u_2$$

$$g_{1,3}(x) = g_1(x) - g_3(x) = -2x^T u_1 + u_1^T u_1 - 2x^T u_3 - u_3^T u_3$$
$$= -2x^T(u_1 + u_3) + u_1^T u_1 - u_3^T u_3$$

$$g_{2,3}(x) = g_2(x) - g_3(x) = -2x^T u_2 + u_2^T u_2 - 2x^T u_3 - u_3^T u_3$$
$$= -2x^T(u_2 + u_3) + u_2^T u_2 - u_3^T u_3$$

The classifier is linear.

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.]

$\because \nabla_{\underline{x}} f[p(\underline{x})] = \left(\frac{\partial f[p(\underline{x})]}{\partial \underline{x}_1}, \ldots, \frac{\partial f[p(\underline{x})]}{\partial \underline{x}_D}\right)$

$\nabla_{\underline{x}} p(\underline{x}) = \left(\frac{\partial p(\underline{x})}{\partial \underline{x}_1}, \ldots, \frac{\partial p(\underline{x})}{\partial \underline{x}_D}\right)$

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

$\frac{\partial f[p(\underline{x})]}{\partial \underline{x}_i} = \left[\frac{d}{dp} f(p)\right]\left[\frac{\partial p(\underline{x})}{\partial \underline{x}_i}\right]$, $i = 1, 2, 3, \ldots, D$

Since we know $\frac{\partial f(p)}{\partial p} = \frac{d}{dp} f(p)$.

$\therefore \frac{\partial f[p(\underline{x})]}{\partial \underline{x}_i} = \left[\frac{\partial f(p)}{\partial p}\right]\left[\frac{\partial p(\underline{x})}{\partial \underline{x}_i}\right] = \left[\frac{d}{dp} f(p)\right]\left[\frac{\partial p(\underline{x})}{\partial x_i}\right]$

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

(b) Use relation (4) of "expressions" in Discussion 2, to find $\nabla_{\underline{x}}\left(\underline{x}^T \underline{x}\right)$.

$\therefore \nabla_x x^T M x = (M + M^T) x$

$x^T x \Rightarrow M = I$

$\therefore \nabla_x (x^T x) = (I + I^T) x = 2Ix = 2x$

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

$x^T = [x_1, x_2, \cdots, x_n]$

$\therefore \nabla_x (x^T x) = \frac{d}{dx} [x_1^2, x_2^2, \cdots, x_n^2]$

$= [2x_1, 2x_2, \cdots, 2x_n]$

$\therefore \nabla_x (x^T x) = 2x$.

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

Suppose $f(x)=x^Tx$, $\nabla_x f(x)=2x$

we have $\nabla_x[(x^Tx)^3]=\nabla_x(f^3(x))$

$$=\frac{d}{df}g(f_x^3)\nabla_x g(x)$$

$$=3(x^T\cdot x)^2\cdot 2x$$

$$=6(x^Tx)^2x$$

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

(a) $\nabla_{\underline{w}}\|\underline{w}\|_2=\nabla_{\underline{w}}(\underline{w}^T\underline{w})^{\frac{1}{2}}=\frac{1}{2}(\underline{w}^T\underline{w})^{-\frac{1}{2}}\cdot 2\underline{w}=(\underline{w}^T\underline{w})^{-\frac{1}{2}}\underline{w}$

(b) Let

$$\begin{cases} f(p)=p^{\frac{1}{2}} \\ p(\underline{w})=(\underline{\underline{M}}\,\underline{w}-\underline{b})^T(\underline{\underline{M}}\,\underline{w}-\underline{b}) \end{cases}$$

we have $f[p(\underline{w})]=[(\underline{\underline{M}}\,\underline{w}-\underline{b})^T(\underline{\underline{M}}\,\underline{w}-\underline{b})]^{\frac{1}{2}}=\|(\underline{\underline{M}}\,\underline{w}-\underline{b})\|_2$

Suppose $\underline{b}$ is a B-dimensional vector and $\underline{\underline{M}}$ is a BxD-dimensional matrix, we have

$$p(\underline{w})=(\underline{\underline{M}}\,\underline{w}-\underline{b})^T(\underline{\underline{M}}\,\underline{w}-\underline{b})$$

$$=\sum_{i=1}^{B}\left(\sum_{j=1}^{D}\underline{\underline{M}}_{ij}\underline{w}_j-\underline{b}_i\right)^2$$

$\because \nabla_{\underline{w}}p(\underline{w})=\left(\frac{\partial p(\underline{w})}{\partial \underline{w}_1},\ldots\frac{\partial p(\underline{w})}{\partial \underline{w}_k},\ldots,\frac{\partial p(\underline{w})}{\partial \underline{w}_D}\right)$

$\therefore$ ⇓

$$\frac{\partial p(\underline{w})}{\partial \underline{w}_k} = \frac{\partial \sum_{i=1}^{B} (\sum_{j=1}^{D} \underline{\underline{M}}_{ij} \underline{w}_j - \underline{b}_i)^2}{\partial \underline{w}_k}$$

$$= \sum_{i=1}^{B} \left[\frac{\partial (\sum_{j=1}^{D} \underline{\underline{M}}_{ij} \underline{w}_j)^2}{\partial \underline{w}_k} - \frac{\partial (2\underline{b}_i \sum_{j=1}^{D} \underline{\underline{M}}_{ij} \underline{w}_j)}{\partial \underline{w}_k} \right]$$

$$= 2 \sum_{i=1}^{B} \left[\underline{\underline{M}}_{ik} (\sum_{j=1}^{D} \underline{\underline{M}}_{ij} \underline{w}_j - \underline{b}_i) \right]$$

$$= 2 (\underline{\underline{M}} \underline{w}_k - \underline{b})^T \underline{\underline{M}}$$

$\therefore$ $\nabla_{\underline{w}} p(\underline{w}) = 2 (\underline{\underline{M}} \underline{w} - b)^T \underline{\underline{M}}$

From question 4 part (a), we have

$$\nabla_{\underline{w}} \| (\underline{\underline{M}} \underline{w} - \underline{b}) \|_2 = \nabla_{\underline{w}} f[p(\underline{w})]$$

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

$$= [(\underline{\underline{M}} \underline{w} - \underline{b})^T (\underline{\underline{M}} \underline{w} - \underline{b})]^{-\frac{1}{2}} (\underline{\underline{M}} \underline{w} - b)^T \underline{\underline{M}}$$