EE559 Homework 2 (week 3)

Jingquan Yan

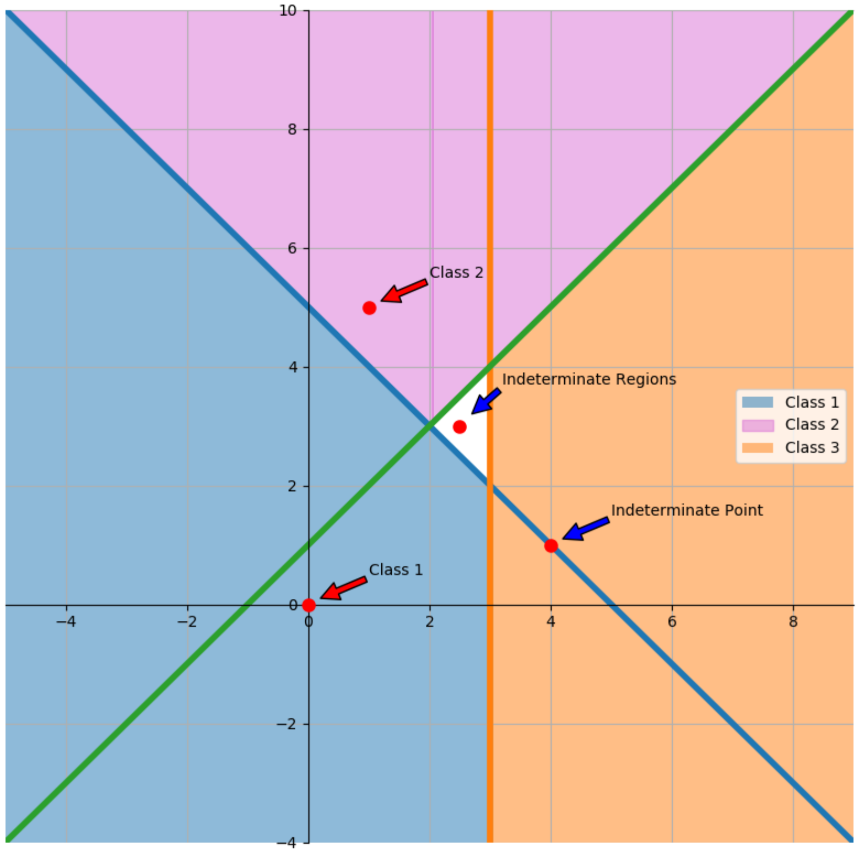
USC ID: 1071912676

Email: jingquan@usc.edu

EE559 repository: [Github]()

Problem 1:

Output:




```
C:\Users\Yan\AppData\Local\Programs\Python\Python37\python.exe "C:/Git/559/HW2/Problem 1.py"
Dot [4, 1] is in class: Indeterminate Point
Dot [1, 5] is in class: 2
Dot [0, 0] is in class: 1
Dot [2.5, 3] is in class: Indeterminate Point
```

This is an example dot in indeterminate regions

Problem 2:

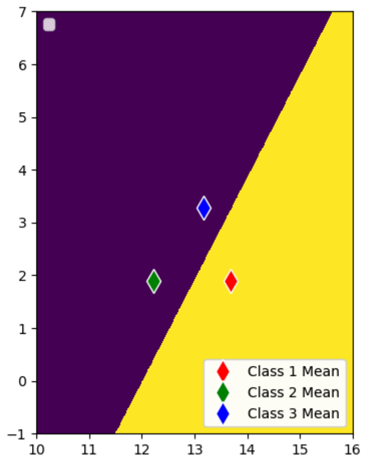
(a)

For training and testing dataset:

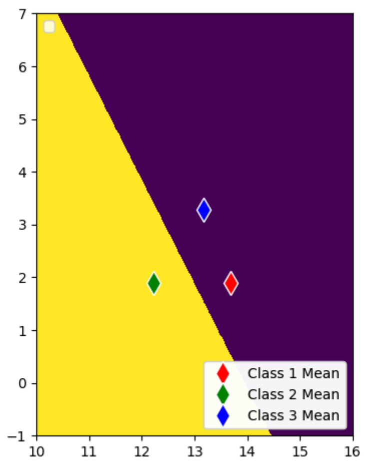
```
C:\Users\Yan\AppData\Local\Programs\Python\Python37\python.exe "C:/Git/559/HW2/Problem 2.py"
The accuracy of training dataset is 0.7415730337078652.
The accuracy of testing dataset is 0.7078651685393258.
```

(b)

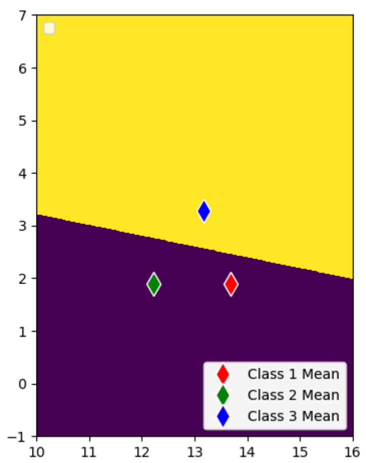
The 2-class decision for Class 1:



The 2-class decision for Class 2:

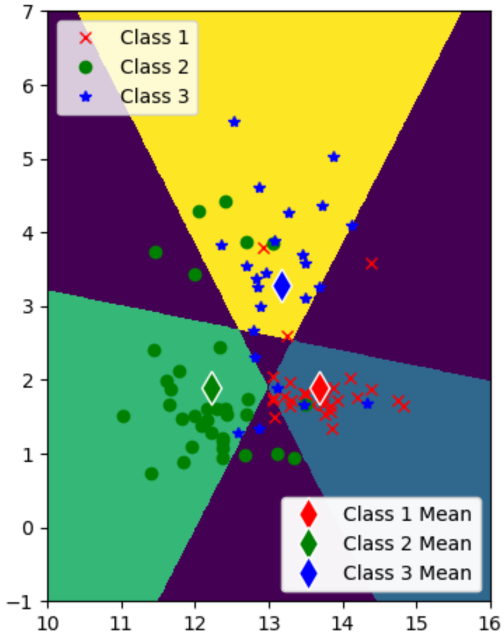


The 2-class decision for Class 3:

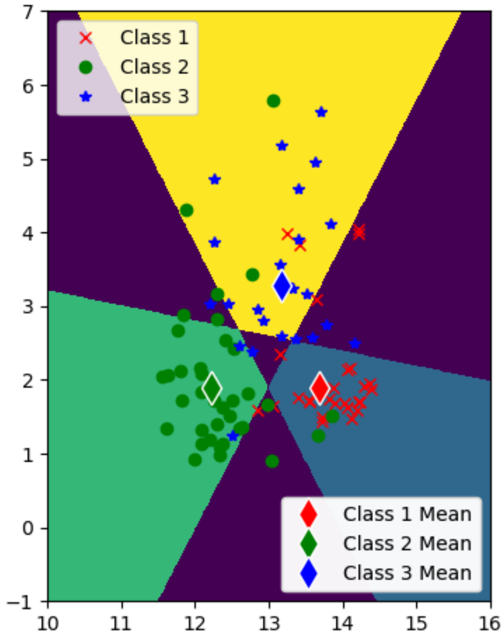


(c)

For training dataset, the purple regions are indeterminate regions, blue for class 1, green for class 2 and yellow for class 3.



For testing dataset, the purple regions are indeterminate regions, blue for class 1, green for class 2 and yellow for class 3.



Problem 3:

Solution:

(a) Proof:

We assume there're n features:

For 2-classes nearest-mean classifier, the hyperplane is the vertical bisector of the 2 means. So we have equations as follow:

$\because \vec{\mu}_1 = [\mu_{11}, \mu_{12}, \dots, \mu_{1n}]$, $\vec{\mu}_2 = [\mu_{21}, \mu_{22}, \dots, \mu_{2n}]$, $\vec{x} = [x_1, x_2, \dots, x_n]$

$\therefore g(x) = \|\vec{x} - \vec{\mu}_2\|^2 - \|\vec{x} - \vec{\mu}_1\|^2 \underset{2}{\overset{1}{\gtrless}} 0$

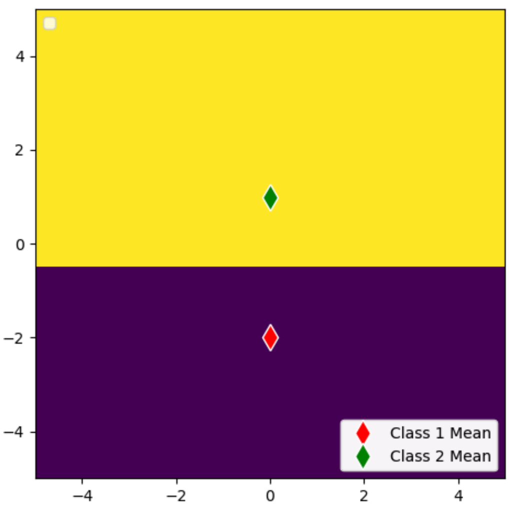
$\therefore (x_1-\mu_{21})^2 + (x_2-\mu_{22})^2 + \dots + (x_n-\mu_{2n})^2 - (x_1-\mu_{11})^2 - (x_2-\mu_{12})^2 - \dots - (x_n-\mu_{1n})^2 \underset{2}{\overset{1}{\gtrless}} 0$

$\iff 2(\vec{\mu}_1 - \vec{\mu}_2)^T \cdot \vec{x} + (\vec{\mu}_2^T \cdot \vec{\mu}_2 - \vec{\mu}_1^T \cdot \vec{\mu}_1) \gtrless 0$

Since $g(x) = 2(\vec{\mu}_1 - \vec{\mu}_2)^T \cdot \vec{x} + (\vec{\mu}_2^T \vec{\mu}_2 - \vec{\mu}_1^T \vec{\mu}_1)$ is the same form with $g(x) = \vec{w}\vec{x} + \vec{b}$

So this is a linear classifier.

(b)



(c) Proof:

Given a 3-class MvM classifier, it is still a nearest-mean classifier for each pair of classes.

As an analogy from (a), we have:

$H_{12}: \|\vec{x} - \vec{\mu}_2\|^2 - \|\vec{x} - \vec{\mu}_1\|^2 \gtrless 0$; $H_{13}: \|\vec{x} - \vec{\mu}_3\|^2 - \|\vec{x} - \vec{\mu}_1\|^2 \gtrless 0$; $H_{23}: \|\vec{x} - \vec{\mu}_3\|^2 - \|\vec{x} - \vec{\mu}_2\|^2 \gtrless 0$

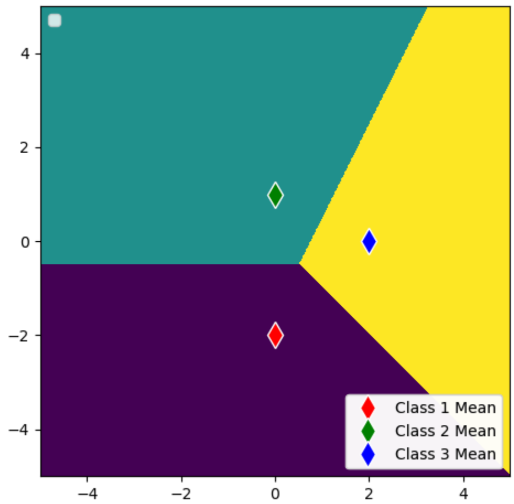
For each hyperplane we have:
$$\begin{cases} 2(\vec{\mu}_1 - \vec{\mu}_2)^T \vec{x} + (\vec{\mu}_2^T \vec{\mu}_2 - \vec{\mu}_1^T \vec{\mu}_1) \gtrless 0 \\ 2(\vec{\mu}_1 - \vec{\mu}_3)^T \vec{x} + (\vec{\mu}_3^T \vec{\mu}_3 - \vec{\mu}_1^T \vec{\mu}_1) \gtrless 0 \\ 2(\vec{\mu}_2 - \vec{\mu}_3)^T \vec{x} + (\vec{\mu}_3^T \vec{\mu}_3 - \vec{\mu}_2^T \vec{\mu}_2) \gtrless 0 \end{cases}$$

Since $g(\underline{x})$ is not unique, one possible $g(\underline{x})$ set is as follows:

$$g_1(x) = 2\vec{\mu}_1 \vec{x} - \vec{\mu}_1^T \vec{\mu}_1$$
$$g_2(x) = 2\vec{\mu}_2 \vec{x} - \vec{\mu}_2^T \vec{\mu}_2$$
$$g_3(x) = 2\vec{\mu}_3 \vec{x} - \vec{\mu}_3^T \vec{\mu}_3$$

the equations above are in "$g(x) = \vec{w}\vec{x} + \vec{b}$" form, So this is a linear classifier.

(d)



(d)

Proof:

Assume two convex hulls are linearly separable and intersect with each other as well. Now we'll find the contradiction:

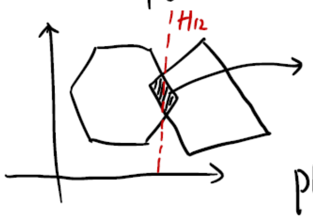
The discriminant function $g(\underline{x}) = w^T x + w_0 \underset{2}{\overset{1}{\gtrless}} 0$, denote the two classes data with $x^{(1)}$ and $x^{(2)}$ respectively.

Now we have:

$$\begin{cases} g(\underline{x}^{(1)}) = w^T x^{(1)} + w_0 \\ g(\underline{x}^{(2)}) = w^T x^{(2)} + w_0 \end{cases}$$

Plug in the given formula: $x^{(k)} = \sum_{i=1}^{n} \alpha_i^{(k)} x_i^{(k)}$ we will have:

$$\begin{cases} g(\underline{x})^{(1)} = w^T \cdot \sum \alpha_i^{(1)} \cdot x_i^{(1)} + w_0 \\ g(\underline{x})^{(2)} = w^T \sum \alpha_i^{(2)} \cdot x_i^{(2)} + w_0 \end{cases} \xLeftrightarrow{\sum \alpha_i^{(k)} = 1} \begin{cases} g(\underline{x})^{(1)} = \sum \alpha_i^{(1)} (w^T \cdot x_i^{(1)} + w_0) \\ g(\underline{x})^{(2)} = \sum \alpha_i^{(2)} (w^T x_i^{(2)} + w_0) \end{cases}$$



For the intersection area, we denote the data as $x^{(12)}$.

plug $x^{(12)}$ into $g(\underline{x})$ as a class 1 and class 2 respectively and their equations should be equal:

Contradiction

① $g(\underline{x}^{(12)}) = g(\underline{x}^{(12)})^{(1)} = \sum \alpha_i^{(1)} (w^T x_i^{(12)} + w_0) = g(\underline{x}^{(12)})^{(2)} = \sum \alpha_i^{(2)} (w^T x_i^{(12)} + w_0)$

But for each class we should have:

② $g(\underline{x}^{(12)})^{(1)} > 0$ and $g(x^{(12)})^{(2)} < 0$

So the dataset could be either linearly seperable or their convex hull intersect with each other.