Tianchu Li

001826214

10.19.2019

Exam 1

Question 1

1. Actual number of samples: Class 1 = 1544; Class 2 = 3551; Class 3 = 4905

```
ans =

      1544


ans =

      3551


ans =

      4905
```

2. Confusion matrix for your classifier consisting of number of samples decided

```
ConfusionMatrix =

        1261         142          25
         185        3190          97
          98         219        4783
```

	True Class 1	True Class 2	True Class 3
Decision Class 1	1261	142	25
Decision Class 2	185	3190	97
Decision Class 3	98	219	4783

3. The total number of samples misclassified = 766
4. Estimate of the probability of error = 7.66%
5. From figure 1, we can see that when true class is 1, the most error of decision is class 2. While for true class is 1, the most error of decision is class 3. Besides, most of misclassified

decision class 3 concentrate on the area of true class 3 (yellow area) and misclassified decision class 2 are distributed almost all of the plot.

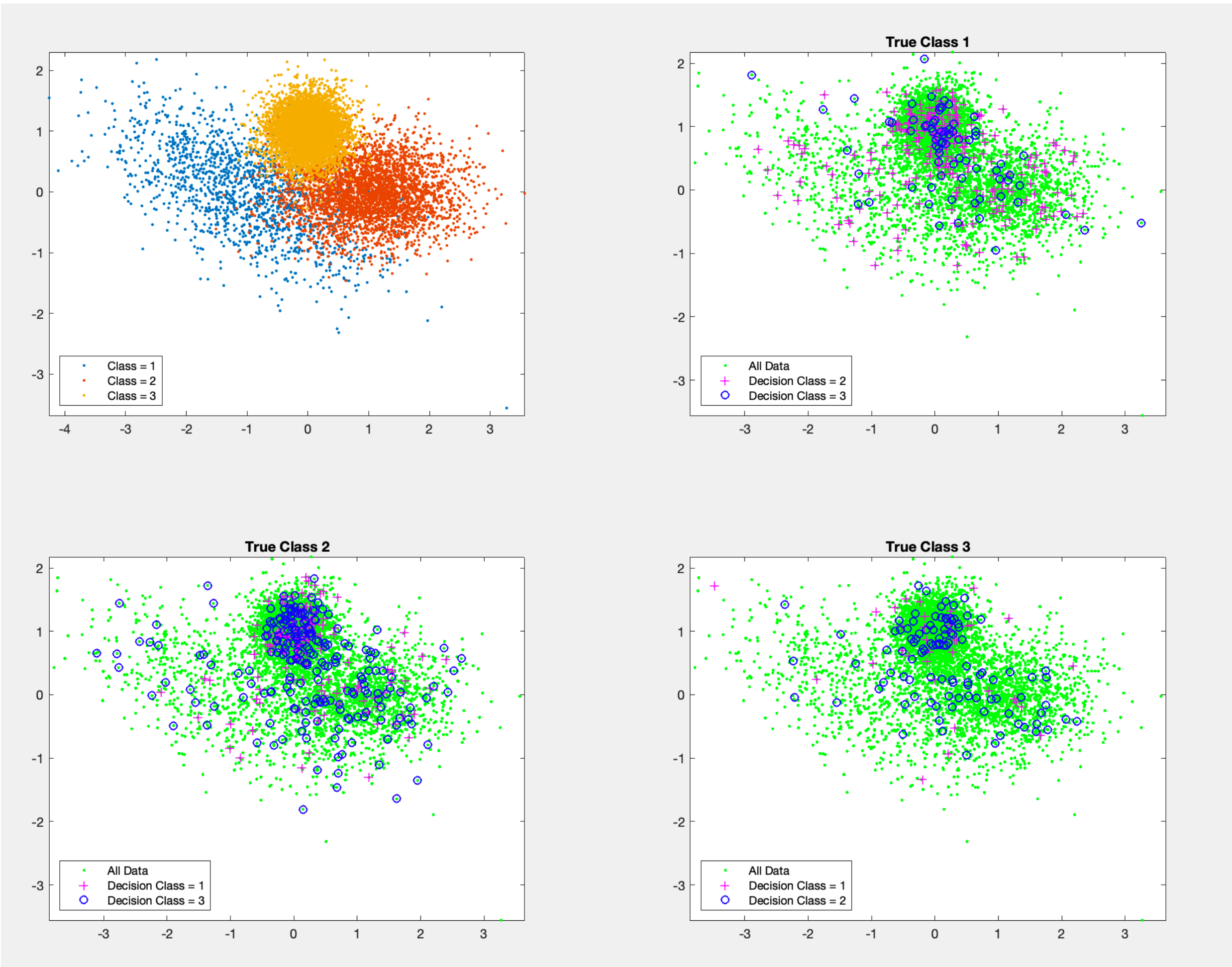

Figure 1

Question 2

1.

Question 2

1. From question, we know that,

$$r_i = d_{Ti} + n_i \sim N(d_{Ti}, \sigma_i^2) \quad \text{for } i \in \{1, \cdots K\}$$

So

$$[X\ Y]_{MAP} = \underset{[X\ Y]}{\arg\max} \prod P(r_i \mid [X, Y]^T)\, P([X\ Y]^T)$$

$$= \underset{[X\ Y]}{\arg\max} \prod_{i=1}^{K} P(r_i \mid [X, Y]^T)\, P([X\ Y]^T)$$

$$= \underset{[X\ Y]}{\arg\max} \sum_{i=1}^{K} \ln P(r_i \mid [X\ Y]^T) + \ln P([X\ Y]^T)$$

$$= \underset{[X\ Y]}{\arg\max} \sum_{i=1}^{K} \ln \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(r_i - d_{Ti})^2}{2\sigma_i^2}} + \ln \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}[X\ Y]\begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}X\\ Y\end{bmatrix}}$$

$$= \underset{[X\ Y]}{\arg\max} -\frac{1}{2}\sum_{i=1}^{K} \frac{(r_i - d_{Ti})^2}{\sigma_i^2} - \frac{1}{2}[X\ Y]\begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}^{-1}\begin{bmatrix}X\\ Y\end{bmatrix}$$

$$= \underset{[X\ Y]}{\arg\min} \sum_{i=1}^{K} \frac{(r_i - d_{Ti})^2}{\sigma_i^2} + \frac{X^2}{\sigma_x^2} + \frac{Y^2}{\sigma_y^2}$$

$$\left(d_{Ti} = \sqrt{(X - X_i)^2 + (Y - Y_i)^2}\right)$$

2.

In my program, I choose 0.25 for σx and σy, and 0.3 for the noise variance values σi. And I also set true position at [0,0] for easy observation.

As the increase of K, the MAP estimate gets closer to the true position [0,0]. From figure 2 we can see that, the center of circles of contour is become closer to origin. Besides, the gap between contour lines become smaller, also suggesting that the MAP estimate is more accurate than other K smaller than current K value and get less affected by noise.

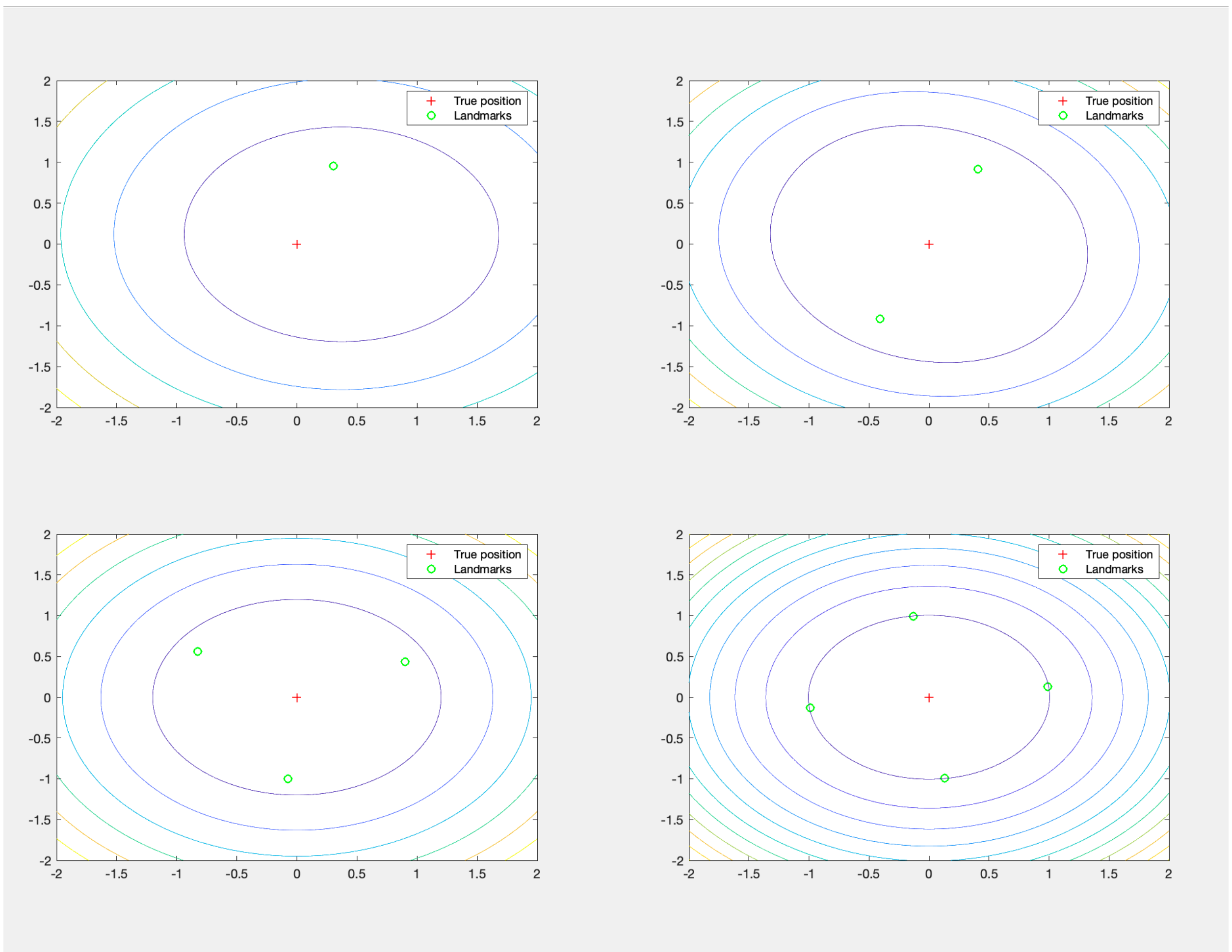

Figure 2. From up to down, left to right are contours of K=1,2,3,4

Question 3

1.

Question 3

1. From question, make $Y=\begin{bmatrix}Y_1\\ \vdots\\ Y_N\end{bmatrix}$, then we have

$$\begin{bmatrix}Y_1\\ \vdots \\ \vdots \\ Y_N\end{bmatrix} = \begin{bmatrix} x_1^3 & x_1^2 & x_1 & 1\\ & \vdots & & \\ & \vdots & & \\ x_N^3 & x_N^2 & x_N & 1\end{bmatrix}\begin{bmatrix}a\\b\\c\\d\end{bmatrix} + \begin{bmatrix}v_1\\ \vdots\\ \vdots \\ v_N\end{bmatrix}$$

$$Y = Xw + v \sim N\Big(Xw, \underbrace{\begin{bmatrix}\sigma^2 & & 0\\ & \ddots & \\ 0 & & \sigma^2\end{bmatrix}}_{\sigma^2 I}\Big)$$

So, for MAP, we have

$$w_{MAP} = \arg\max_w \; p(Y|w)\, p(w)$$

$$= \arg\max_w \; (2\pi)^{-\frac{N}{2}} |\sigma^2 I|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y-Xw)^T(\sigma^2 I)^{-1}(Y-Xw)} \cdot (2\pi)^{-2} |\gamma^2 I|^{-\frac{1}{2}} e^{-\frac{1}{2} w^T (\gamma^2 I)^{-1} w}$$

$$= \arg\max_w \; \ln(\cdots)$$

$$= \arg\max_w \; -\frac{1}{2}(Y-Xw)^T(\sigma^2 I)^{-1}(Y-Xw) - \frac{1}{2} w^T(\gamma^2 I)^{-1} w$$

make $\sigma^2 I = \Sigma$, $\gamma^2 I = \Sigma_0$, then

$$w_{MAP} = \arg\min_w \; (Y-Xw)^T\Sigma^{-1}(Y-Xw) + w^T\Sigma_0^{-1}w$$

$$= \arg\min_w \; -w^TX^T\Sigma^{-1}Y - Y^T\Sigma^{-1}Xw + w^TX^T\Sigma^{-1}Xw + w^T\Sigma_0^{-1}w$$

Through derivation, we have,

$$-2Y^T\Sigma^{-1}X + 2X^T\Sigma^{-1}Xw_{map} + 2\Sigma_0^{-1}w_{map} = 0$$

So

$$w_{MAP} = \frac{Y^T\Sigma^{-1}X}{X^T\Sigma^{-1}X + \Sigma_0^{-1}}, \quad \Sigma = \sigma^2 I, \quad \Sigma_0 = \gamma^2 I$$

2.

From the figure we can find that, when γ is under 10^-3, all different percent of squared-error values are very similar. And they will decrease as γ increase from 10^-5 to 10^-2.

The different of minimum, 25%, median, 75%, and maximum values of these squared-error values shows and becomes larger with the increase of γ starting from 10^-2.

After γ=0, minimum, 25%, median, 75% of squared-error values tend to be 10^-2, 10^-1.5, 10^-1, and 10^-0.5. However, maximum values of squared-error keep going and fluctuate around 10^1.

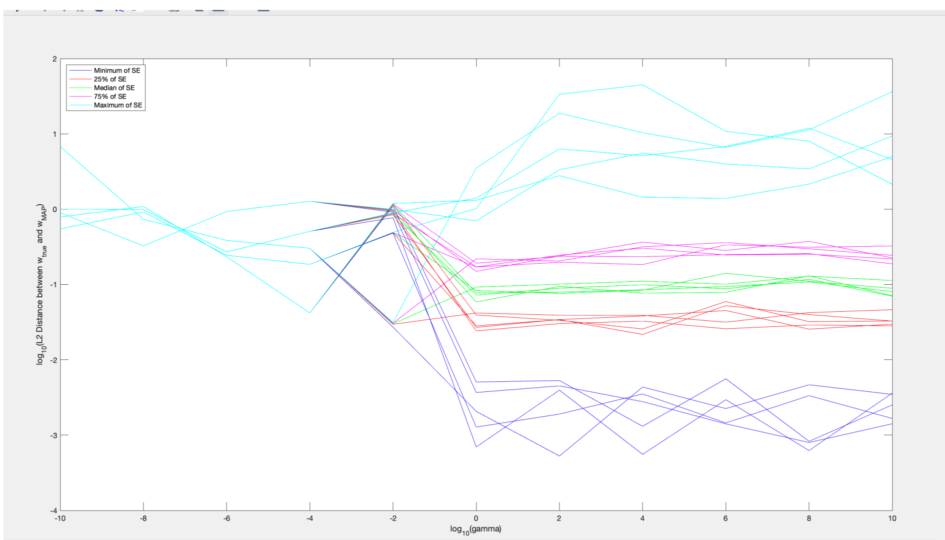

Figure 3. Five results of gamma from 10^-10 to 10^10

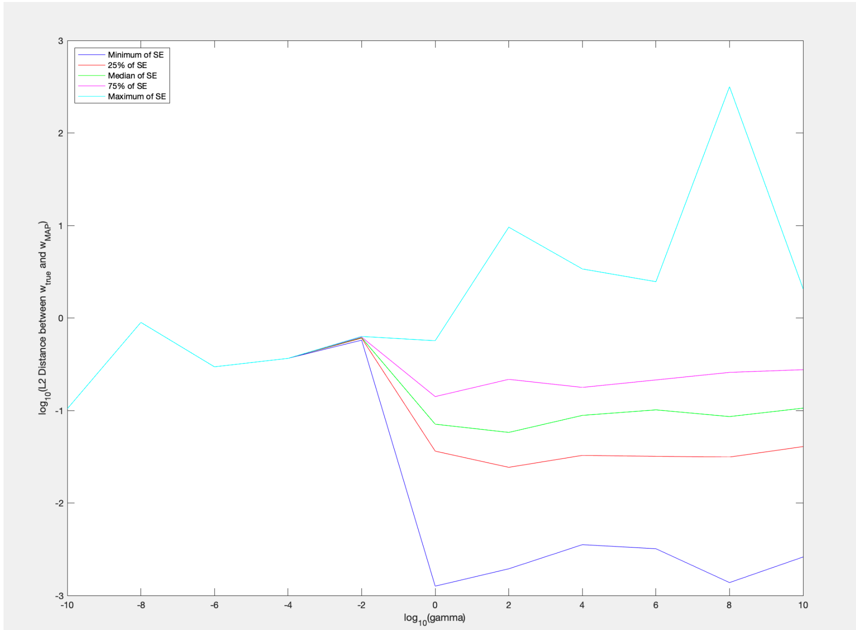

Figure 4. One result of gamma from 10^-10 to 10^10