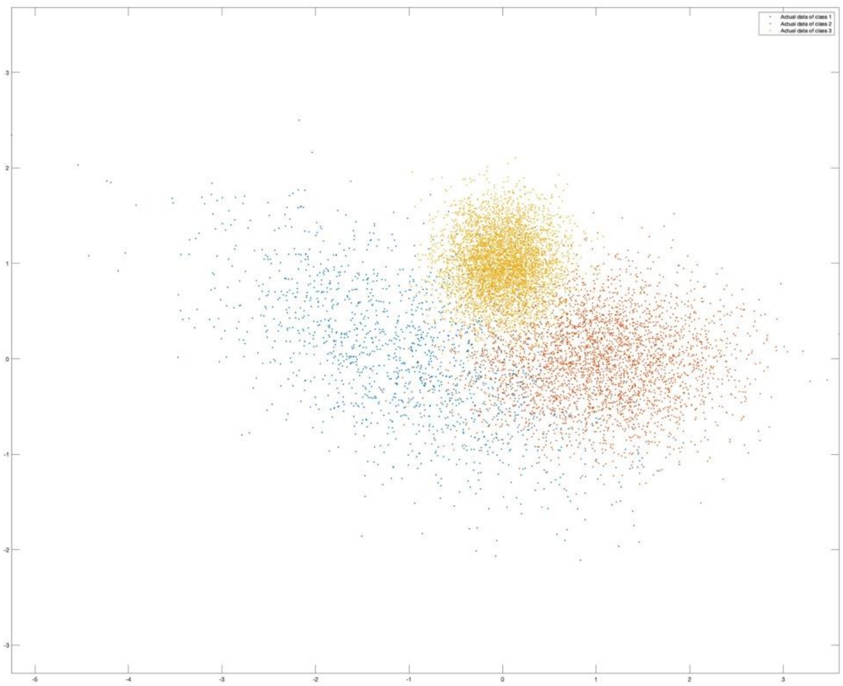
EECE5644
Yu Shun Lin
ID 001459022 Exam 1

1. Three class with priors are respectively P(L=1)=0.15,P(L=2)=0.35,P(L=3)=0.5.
   The data distribution are as follows

$$\mathcal{N}\left(\begin{bmatrix}-1\\0\end{bmatrix},\begin{bmatrix}1 & -0.4\\-0.4 & 0.5\end{bmatrix}\right),\mathcal{N}\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0.5 & 0\\0 & 0.2\end{bmatrix}\right),\mathcal{N}\left(\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}0.1 & 0\\0 & 0.1\end{bmatrix}\right).$$



The plot shows the true data of each three classes
Receive the discriminant score for the function of the evalGaussian. Which indicate the decision as P(wi|x) is the largest of three class. Then make the decision for the data by choosing the largest discriminant score.

The confusion matrix result is as follow

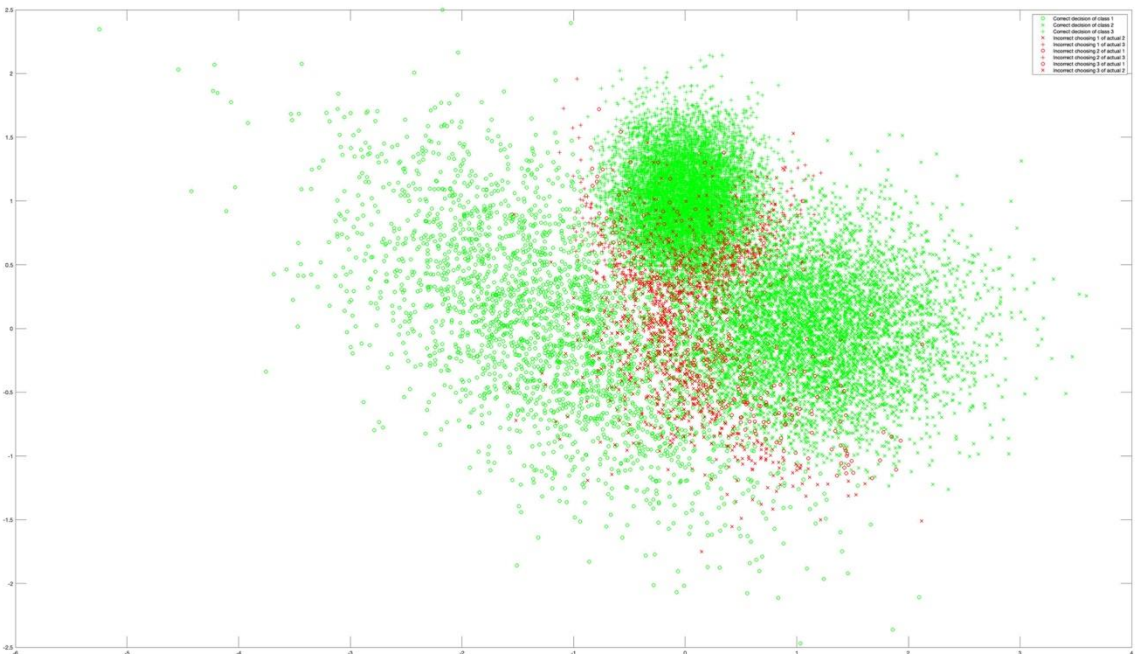
```
confusion_matrix =

    1289         135          56
     293        3023         186
      81         147        4790
```

Case, decision = 1, L = 1 has 1289 samples
Case, decision = 2, L = 2 has 3023 samples
Case, decision = 3, L = 3 has 4790 samples
These are the correct decisions.

Case, decision = 2, L = 1 has 135 samples. Case decision = 3, L = 1 has 56 samples.
Case, decision = 1, L = 2 has 293 samples. Case decision = 3, L = 2 has 186 samples.
Case, decision = 1, L = 3 has 81 samples. Case decision = 2, L = 3 has 147 samples.

Confusion matrix shows the number of samples correct and incorrect. Where column indicate to decision class and the row indicate to actual class.

EECE5644
Yu Shun Lin
ID 001459022 Exam 1
Total number of miss equal to (10000-total correct) = 10000-9102 = 898.
Thus, the error can be calculated by 898/10000 = 0.0898 = 8.98%



The plot shows every samples with shape and color indicate for the correct decision and the wrong decision.

EECE5644
Yu Shun Lin
ID 001459022 Exam 1

2. An true point is locate by land marks by measurement r1 = dT + n. Where n is the noise, and each noise is independent to each other in each K measurement. The objective function to determine the estimation is as follow.

$$r_i = d_{T_i} + n_i \qquad P\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}[x \ y]\begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix}}$$

we determine the MAP estimation by $P(r_i | \begin{bmatrix} X \\ Y \end{bmatrix}) P(\begin{bmatrix} X \\ Y \end{bmatrix})$

$$\Rightarrow \arg\max (P(r_i | \begin{bmatrix} X \\ Y \end{bmatrix}) P(\begin{bmatrix} X \\ Y \end{bmatrix}) \qquad r_i \sim N(d_T, \sigma_n)$$

$$= \arg\max \left(\frac{1}{\sqrt{2\pi}\,\sigma_n^2} e^{-\frac{(r_i - d_{T_i})^2}{2\sigma_n^2}} \cdot \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_y^2}\right)} \right)$$

$$= \arg\max \ln \left(\frac{1}{\sqrt{2\pi}\,\sigma_n^2} \cdot \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\frac{(r_i - d_{T_i})^2}{\sigma_n^2}} \cdot e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_y^2}\right)} \right)$$

$$= \arg\max \left(-\frac{1}{2} \cdot \frac{(r_i - d_{T_i})^2}{\sigma_n^2} + \left(-\frac{1}{2}\right)\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_y^2}\right) \right)$$

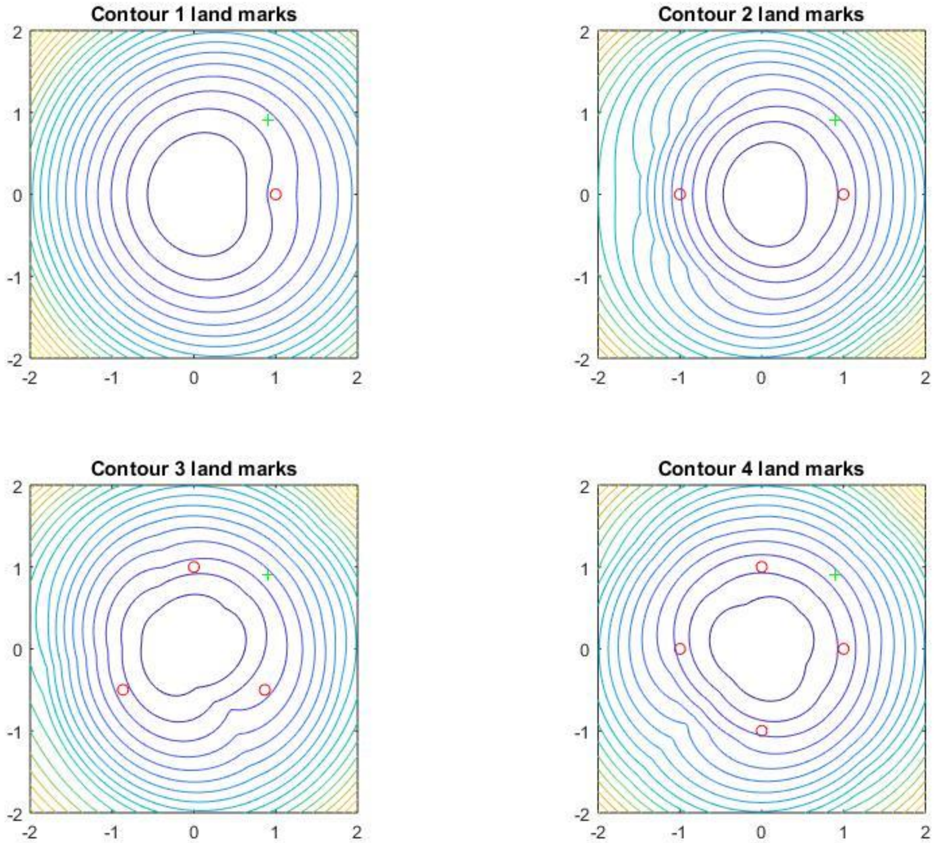
$$= \arg\min \left(\frac{(r_i - d_{T_i})^2}{\sigma_n^2} + \frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_y^2} \right)$$

Here the minimum of the function determine the most possible of the point
By the function create the estimate point from -2 to 2 both on the plot and generate the contour

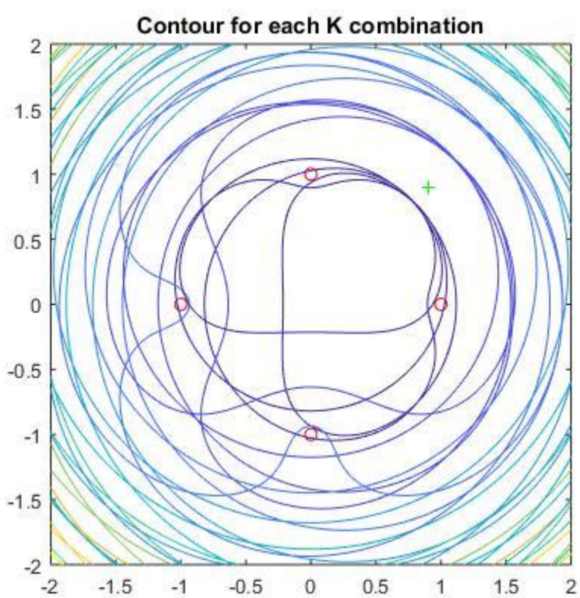
Take the sigma_noise as 0.3, sigma_x and sigma_y = 0.25; true point (0.9,0.9)

EECE5644
Yu Shun Lin
ID 001459022

Exam 1



The plot show that the contour for different land marks. Since the objective function of the prior domain the determination of the result. However, different land marks leads to the different result. I consider the more land marks will estimate more correctly because the contour center seems moving to the true point when the land mark number increase.



This plot shows the different contour combination of different land marks by each estimation.

Since each noise is independent to each other. Same color contour have the rage on it. This range can be seen as normal distribution mean value at dT and Sigma_noise = 0.3.

EECE5644
Yu Shun Lin
ID 001459022 Exam 1

3. The estimation function could be find as the following processing

The polynomial $y = ax^3 + bx^2 + cx + d + v$ where $v \sim \mathcal{N}(0, \sigma^2)$

we could write it as $y = w \cdot X + v$ where $w = [a, b, c, d]^T$

$X = [x^3\ x^2\ x^1\ x^0]$

$\Rightarrow$ that y becomes $y \sim \mathcal{N}(wX, \sigma^2)$

$$\Rightarrow P(y|x.w) = \mathcal{N} \sim (y|wX, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-wX)^2}{2\sigma^2}}$$

then estimate w by $D \cdot \{(x_1, y_1) \ldots (x_N, y_N)\}$.

$$\Rightarrow P(w) \sim \mathcal{N}(0, \gamma^{-1} I) = \frac{1}{\sqrt{2\pi}^{D}} e^{\frac{\gamma}{2} ww^T}$$

$$\Rightarrow \log P(w|D) = \log P(w) + \log P(D|w) - \log P(D)$$

$$\text{MAP} \Rightarrow \underset{w}{\text{argmax}} (\log P(w) + \log P(D|w) - \log P(D))$$

$$= \underset{w}{\text{argmax}} (\log P(w) + \log P(D|w))$$

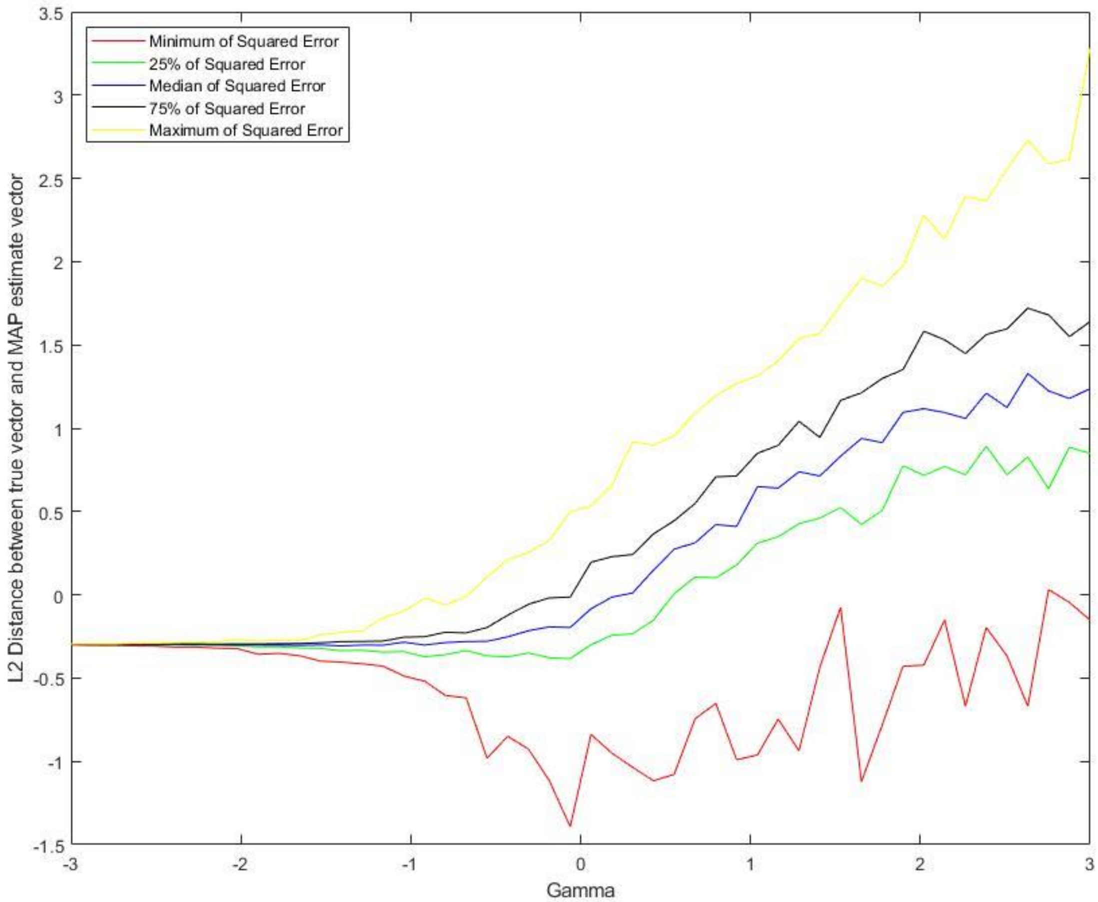
$$= \underset{w}{\text{argmax}} \left\{ -\frac{D}{2} \log(2\pi) - \frac{\gamma}{2} ww^T + \sum_{n=1}^{N} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{(y_n - wX_n)^2}{2\sigma^2} \right. \right.$$

$$= \underset{}{\text{argmin}} \frac{1}{2\sigma^2} \sum_{n=1}^{N} (y_n - wX_n)^2 + \frac{\gamma}{2} ww^T$$

$\leftarrow$ determine

Since we have the estimation function, we could estimate the w on matlab

EECE5644
Yu Shun Lin
ID 001459022 Exam 1



The plot shows the gamma value from -3 to 3, as 50 gamma samples, and the vertical axis is the distance of the true vector and the estimated vector.

Here we observe that as the gamma being larger, the estimation will have more error because Gamma value determine the prior. Thus, the gamma values should not be to large for the MAP estimation.

EECE5644
Yu Shun Lin
ID 001459022 Exam 1

Reference:

1. Richard O. Duda, Peter E. Heart, David G. Stork, Pattern Classification (2006) Ch2.4, Ch3.3.2
2. Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press.
3. https://wiseodd.github.io/techblog/2017/01/01/mle-vs-map/
4. https://www.mathworks.com/help/index.html

Code Resource on Github:
https://github.com/MakiseYuki/EECE5644-Machine-Learning/tree/master/Exam%201