Chapter 2 Homework

2015011313 徐鉴劲 计54

To get all the figures in this assignments, run

```
python src/plot2.py
python src/plot4.py
python src/plot5.py
```

which will generator corresponding figure for problems.

Problem 1

(a)

Error probability $P(error|x) = min[P(\omega_1|x), P(\omega_2|x)]$, in which $P(\omega_1|x) = \frac{P(x|\omega_1)P((\omega_1)}{p(x)}$, and $P(\omega_2|x)$ is similar to this. Minimize error probability gives the following decision rule:

Select ω_1 if $P(x|\omega_1) > P(x|\omega_2)$. Select ω_2 otherwise.

(b)

Suppose $R(\omega|x)$ is the risk of selecting ω when x is observed. This error risk matrix gives the following risk expression:

$R(\omega_1|x) = P(\omega_2|x)$, $R(\omega_2|x) = 0.5P(\omega_1|x)$.

To minimize the risk, the following dicision rule can be reached:

If $\frac{P(\omega_1|x)}{P(\omega_2|x)} < \frac{1}{2}$, select ω_2. Otherwise, select ω_1.

Problem 2

Using a zero-one risk matrix, we obtain the risk selecting class i given feature x

$R(\omega_i|x) = \sum_j P(\omega_j|x)$

And minimizing this risk gives $R(x) = min_i R(\omega_i|x)$.

And the corresponding decision is $\omega_i = argmin_{\omega_i} R(\omega_i|x) = argmax_{\omega_i} P(\omega_i|x)$.

Transform the posterior to likelihood and prior

$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)}$

Notice that $P(x|\omega_i)$ is defined as $\left[\frac{\delta_i - |x-\mu_i|}{\delta_i^2}\right]^+$, so the decision function should be

$f(x) = argmax_i \left[\frac{\delta_i - |x-\mu_i|}{\delta_i^2}\right]^+$

(a)

Suppose that for ω_1 and ω_2, their delta are such that cause only one intersection, that is $\mu_2 - \mu_1 < \delta_2 + \delta_1$ and that $\mu_1 + \delta_1 < \mu_2 + \delta_2$.

And for ω_2 and ω_3, the constraint is similar.

In this case, there is only one decision point between each category.

For ω_1 and ω_2, the intersection is at x_1^*:

$\frac{\delta_1 - x + \mu_1}{\delta_1^2} = \frac{\delta_2 - \mu_2 + x}{\delta_2^2}$

this gives:

$x_1^* = \frac{\delta_2^2\delta_1 - \delta_1^2\delta_2 + \delta_1^2\mu_2 + \delta_2^2\mu_1}{\delta_1^2 + \delta_2^2}$

and x_2^* is the decision point between ω_2 and ω_3, which is similar:

$x_2^* = \frac{\delta_3^2(\delta_2+\mu_2) - \delta_2^2\delta_3 + \delta_2^2\mu_3)}{\delta_2^2 + \delta_3^2}$

(b)

In this case, one triangular is flat such that it intersects with another triangle at two points, x_1^* and x_2^*.

This place a constraint on δ that

$\delta_1 + \mu_1 > \delta_2 + \mu_2$

Also, $-\delta_1 + \mu_1 > -\delta_2 + \mu_2$ is also possible, which is similar to this.

this constraint indicates that line $y_0 = \frac{\delta_1 - x + \mu_1}{\delta_1^2}$ intersects with both $y_1 = \frac{\delta_2 + x - \mu_2}{\delta_2^2}$ and $y_2 = \frac{\delta_2 - x + \mu_2}{\delta_2^2}$, resulting in point x_1^* and x_2^*.

Solve the equation gives:

$x_1^* = \frac{\delta_2^2\delta_1 - \delta_1^2\delta_2 + \delta_1^2\mu_2 + \delta_2^2\mu_1}{\delta_1^2 + \delta_2^2}$

$x_2^* = \frac{\delta_2^2\delta_1 - \delta_1^2\delta_2 + \delta_1^2\mu_2 - \delta_2^2\mu_1}{\delta_2^2 - \delta_1^2}$

(c)

According to the formula above, we can get three decision point: $x_1^* = \frac{1}{3}$, $x_2^* = \frac{2}{3}$.

Select ω_1 if $-1 > x \leq \frac{1}{3}$.

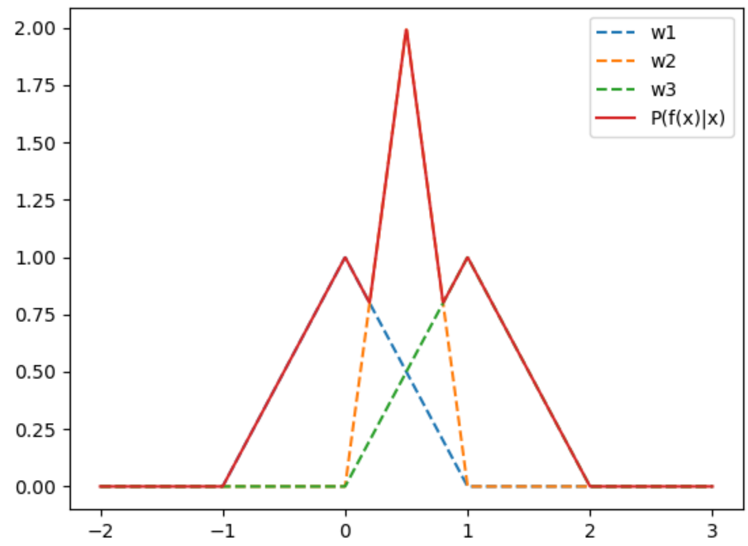
Select ω_2 if $\frac{1}{3} < x \leq \frac{2}{3}$.

Select ω_3 if $\frac{2}{3} < x \leq 2$

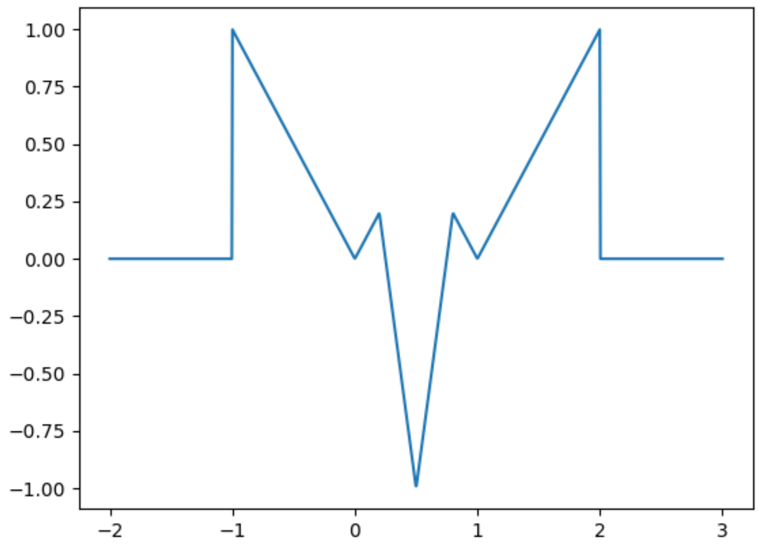
(d)

The risk is given by $R(x) = min_i R(\omega_i|x) = 1 - P(f(x)|x)$, which has shape:

The whole probability distribution, is shown in the figure below:



And the corresponding minimal risk function is:



Problem 3

(a) Show the minimal risk decision rule

Suppose the total risk is $R(\omega_i|x)$, we have

$R(\alpha_i|x) = \sum_j \lambda(\alpha_i|\omega_j)P(\omega_j|x) = \lambda_s \sum_{j\neq i} P(\omega_j|x)$

for i = 1,...,c and

$R(\alpha_{c+1} = \sum_j \lambda(\alpha_{c+1}|\omega_j)P(\omega_j|x) = \lambda_r \sum_j P(\omega_j|x) = \lambda_r$

To minimize the risk, we have $f(x) = argmax_{i=1,\ldots,c+1} R(\alpha_i|x)$.

If we want to decide ω_i, then $R(\alpha_i|x)$ should be smaller than all other terms, which indicates that $P(\omega_i|x)$ is the largest over other possibilities, $P(\omega_i|x) > P(\omega_j|x)$ for all other $j \neq i$.

Also it risk should be smaller than rejection

$\lambda_s \sum_{j\neq i} P(\omega_j|x) \leq \lambda_r \Rightarrow$

$\lambda_s \sum_j P(\omega_j|x) \leq \lambda_r + \lambda_s P(\omega_i|x) \Rightarrow$

$1 - \frac{\lambda_r}{\lambda_s} \leq P(\omega_i|x)$

(b) what happens if $\lambda_r = 0$

The condition above becomes

$1 \leq P(\omega_i)$

When $\lambda_r = 0$, it means that whatever the model predicts, choosing to reject resulting in zero risk. According to this decision rule, in most cases the algorithm will decide to reject, unless the predicted probability is one hundred percent.

(c) what happens if $\lambda_r > \lambda_s$

The condition above becomes

$0 \leq P(\omega_i)$

This means that whatever the model predicts, the risk of reject is always larger than accept a category. So in all cases, the algorithm will not choose to reject.

Problem 4

(a)

Problem 3 use $P(\omega_i|x)$ to decide the category and use $1 - \frac{\lambda_r}{\lambda_s} \le P(\omega_i|x)$ to decide rejection.

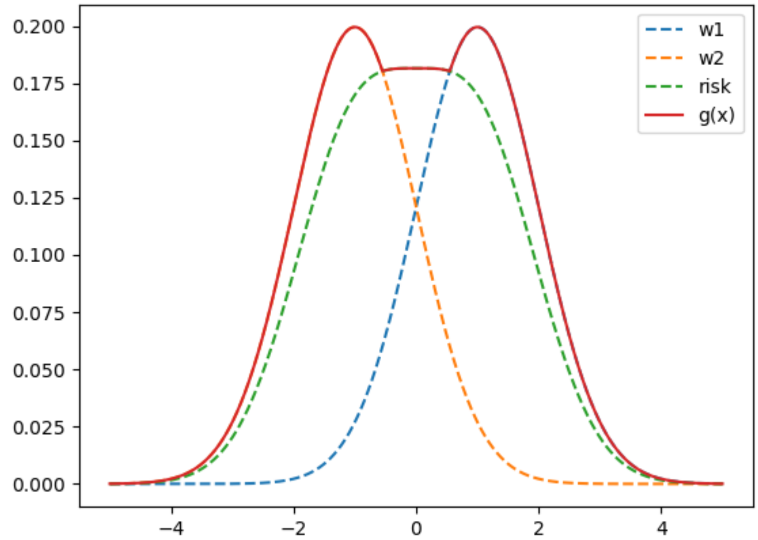
In this problem, we have $g_i(x) = P(x|\omega_i)P(\omega_i) = P(x, \omega_i) = P(\omega_i|x)P(x)$.

and $g_r(x) = (1 - \frac{\lambda_r}{\lambda_s})P(x)$.

Divide all the functions with $P(x)$, we get $G_i(x) = \frac{g_i(x)}{P(x)} = P(\omega_i|x)$ and $G_c(x) = 1 - \frac{\lambda_r}{\lambda_s}$.

This is consistent with the result of problem 3.

(b)



On left and right, the decision can be easily made, but in the middle region, the classifier refuse to make a decision.

(c)

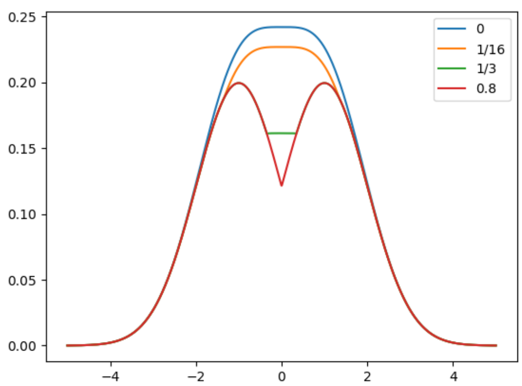
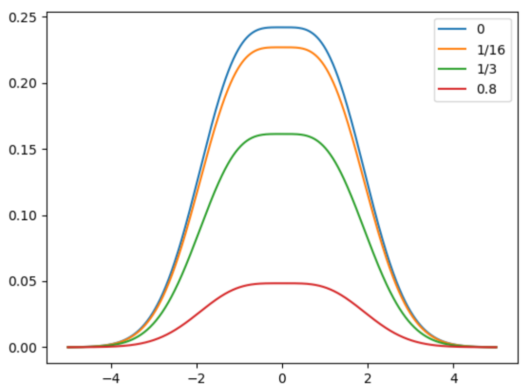
As $r = \frac{\lambda_r}{\lambda_s}$ increase from 0 to 1, the rejection line raises.

When r is small, rejection does not appear, because the risk of rejection is too high.

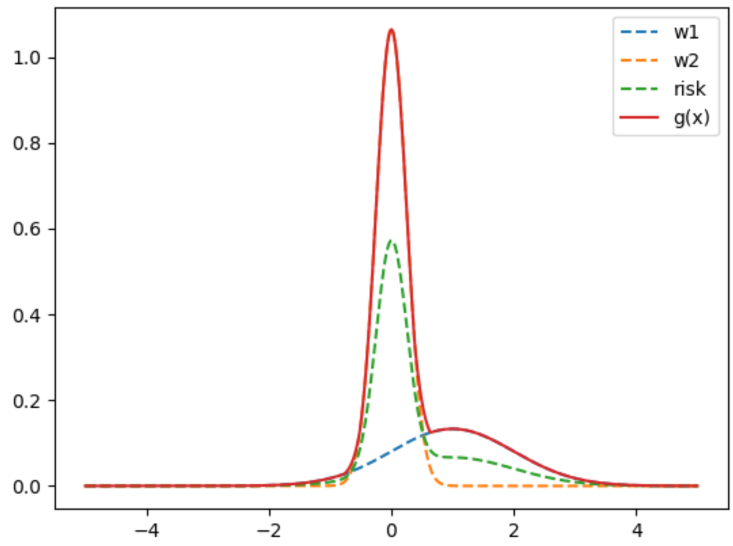
When r increases, then bewteen two classes there appears a rejection area.

When r continues to increase, the risk of rejection is low enough to force all decision to become a rejection.

In the following figure, the process is seen more clearly.

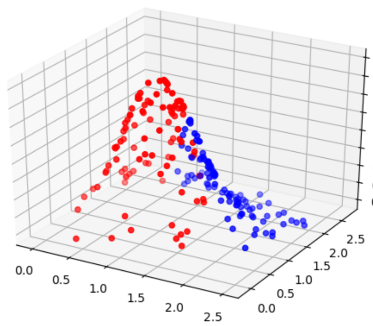
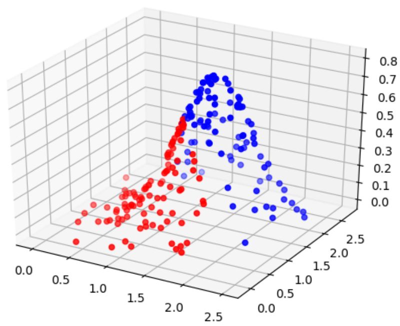
Total Decision Curve	Risk Curve
	

(d) repeat expr with different parameter

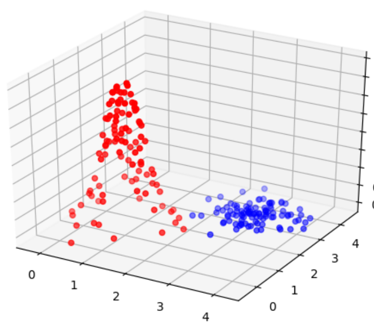
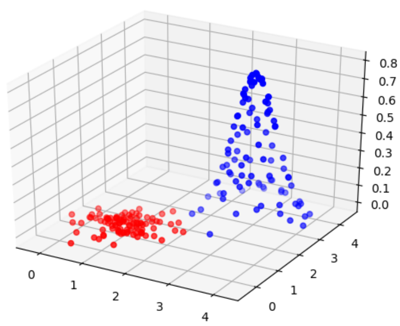


Problem 5

Set μ to be 1.5, the accuracy is about 70% ~ 80%, which varies greatly. To be specific, the accuracy of ω_1 and ω_2 are close to each other. In addition, the figure below shows the feature probability $P(x|\omega_1)$. Red and blue represent ω_1, ω_2 respectively.

Set μ to be 3, the accuracy is 100%. The feature probability is also shown below.

(a)

As the covariance matrix Σ can be divided into 2 blocks,

$P(x_1, x_2, x_3) = P(x_1)P(x_2, x_3) = \mathcal{N}(x_1; 1, 1)\mathcal{N}(x_1, x_2; \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix})$.

The formulae for two dimensional normal distribution is $\frac{1}{2\pi\sqrt{|\Sigma|}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$. $|\Sigma| = 21$, $\Sigma^{-1} = \frac{1}{21}\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$.

$P(x_0|\omega) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{8}} \sim 0.35206532676.$

$P(x_1, x_2|\omega) = \frac{1}{42\pi}e^{-\frac{1}{42}[2,1]\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \frac{1}{42\pi}e^{-\frac{17}{42}} = 5.056092087 \times 10^{-3}$

So $P(\mathbf{x}_0) = 1.78 \times 10^{-3}$.

(b)

To transform the matrix into identity matrix, first we do eigen value decomposition:

$B\begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}B^T = \begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$, in which $B = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$.

Let the original random variables to be $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, and $\tilde{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ the tranformation would be

$\tilde{X} = diag(1, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{7}})\tilde{B}(X - \mu)$

(c)

Apply the transformation to $\mathbf{x}_0$, the result is $\tilde{\mathbf{x}}_0 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{6}} \\ \frac{3}{\sqrt{14}} \end{bmatrix}$

(d)

Mahalanobis distance is $B_M(x) = \sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}$.

For original distribution, $d_1 = \sqrt{\frac{1}{21}[\frac{1}{2}, 2, 1]\begin{bmatrix} 21 & 0 & 0 \\ 0 & 5 & -2 \\ 0 & -2 & 5 \end{bmatrix}\begin{bmatrix} \frac{1}{2} \\ 2 \\ 1 \end{bmatrix}} = \frac{1}{2}\sqrt{\frac{89}{21}}$.

For transformed distribution, $d_2 = \sqrt{||[\frac{1}{2}, \frac{1}{\sqrt{6}}, \frac{3}{\sqrt{14}}]||^2} = \sqrt{\frac{1}{4} + \frac{1}{6} + \frac{9}{14}} = \frac{1}{2}\sqrt{\frac{89}{21}}$.

$d_1 = d_2$.

(e)

Original probability density is $P(\mathbf{x}_0) = Ce^{-\frac{1}{2}(x_0-\mu)^T\Sigma^{-1}(x_0-\mu)}$.

The tranformed probability density is

$P(\tilde{\mathbf{x}}_0) = Ce^{-\frac{1}{2}(\tilde{x}_0 - T^t\mu)^T(T^t\Sigma T)^{-1}(\tilde{x}_0 - T^t\mu)}$

in which $\tilde{\mathbf{x}}_0 = T^t\mathbf{x}_0$.

Thus we have

$P(\tilde{\mathbf{x}}_0) = Ce^{-\frac{1}{2}(x_0-\mu)^tT(T^t\Sigma T)^{-1}T^t(x_0-\mu)}$

As T is a linear tranformation, it is not singular, we have

$(T^t\Sigma T)^{-1} = T^{-1}\Sigma^{-1}(T^{-1})^t$

so all the T can be canceled out:

$P(\tilde{\mathbf{x}}_0) = Ce^{-\frac{1}{2}(x_0-\mu)^tTT^{-1}\Sigma^{-1}(T^{-1})^tT^t(x_0-\mu)} = Ce^{-\frac{1}{2}(x_0-\mu)^T\Sigma^{-1}(x_0-\mu)} = P(\mathbf{x}_0)$.

(f)

Let the gaussian random vector to be X, and the original parameter to be μ and Σ. Apply the whitening transformation $\tilde{X} = \Phi\Lambda^{-\frac{1}{2}}X$, then $\mathcal{N}(\mu, \Sigma)$ is tranformed into $\mathcal{N}(\Phi\Lambda^{-\frac{1}{2}}\mu, \Phi\Lambda^{-\frac{1}{2}}\Sigma(\Phi\Lambda^{-\frac{1}{2}})^T)$.

As $\Phi\Sigma\Phi^T = \Lambda$, so $\Phi\Lambda^{-\frac{1}{2}}\Sigma(\Lambda^{-\frac{1}{2}})^T\Phi^T = \Lambda^{-\frac{1}{2}}\Phi\Sigma\Phi^T(\Lambda^{-\frac{1}{2}})^T = I$.

Problem 6

(a)

$$P(x_0, x_1, x_2, x_3|\omega_1, \omega_3, \omega_3, \omega_2) = P(0.6|\omega_1)P(0.1|\omega_3)P(0.9|\omega_3)P(1.1|\omega_2) = \frac{1}{4\pi^2}e^{-\frac{0.4^2}{2}}e^{-\frac{0.9^2}{2}}e^{-\frac{0.1^2}{2}}e^{-\frac{0.6^2}{2}} = \frac{1}{4\pi^2}e^{-\frac{0.16+0.81+0.01+0.36}{2}} = 0$$

(b)

$$P(0.6, 0.1, 0.9, 1.1|\omega_1, \omega_2, \omega_2, \omega_3) = \frac{1}{4\pi^2}e^{-\frac{0.36+0.16+0.16+0.01}{2}} = 0.018$$

(c)

The sequence is $\omega_2, \omega_1, \omega_3, \omega_3$.