Assignment I

SHEN ZIJIN(沈子衿) 3160104734

1. Machine Learning Problems

(a)

1) BF 2) C 3) C 4) BG 5) AE 6) AD 7) BF 8) AE 9) BF

(b)

Answer: *FALSE*. On the one hand, the data at hand should be divided into training sets and test sets, not all for training. On the other hand, we need to properly control the number of training sets and test sets to maximize the performance of the test set. Too many training sets can lead to overfitting.

2. Bayes Decision Rule

(a)

(1) Since each box is selected as an independent event:

$$P(B_1 = 1) = \frac{1}{3}$$

(2) Since there is only one prize and that prize has been found in B1:

$$P(B_2 = 0 \mid B_1 = 1) = 1$$

(3) According to Bayes Theorem:

$$P(B_1 = 1|B_2 = 0) = \frac{P(B_2 = 0|B_1 = 1)P(B_1 = 1)}{P(B_2 = 0)} = \frac{1 \times \frac{1}{3}}{\frac{1}{1}} = \frac{1}{3}$$
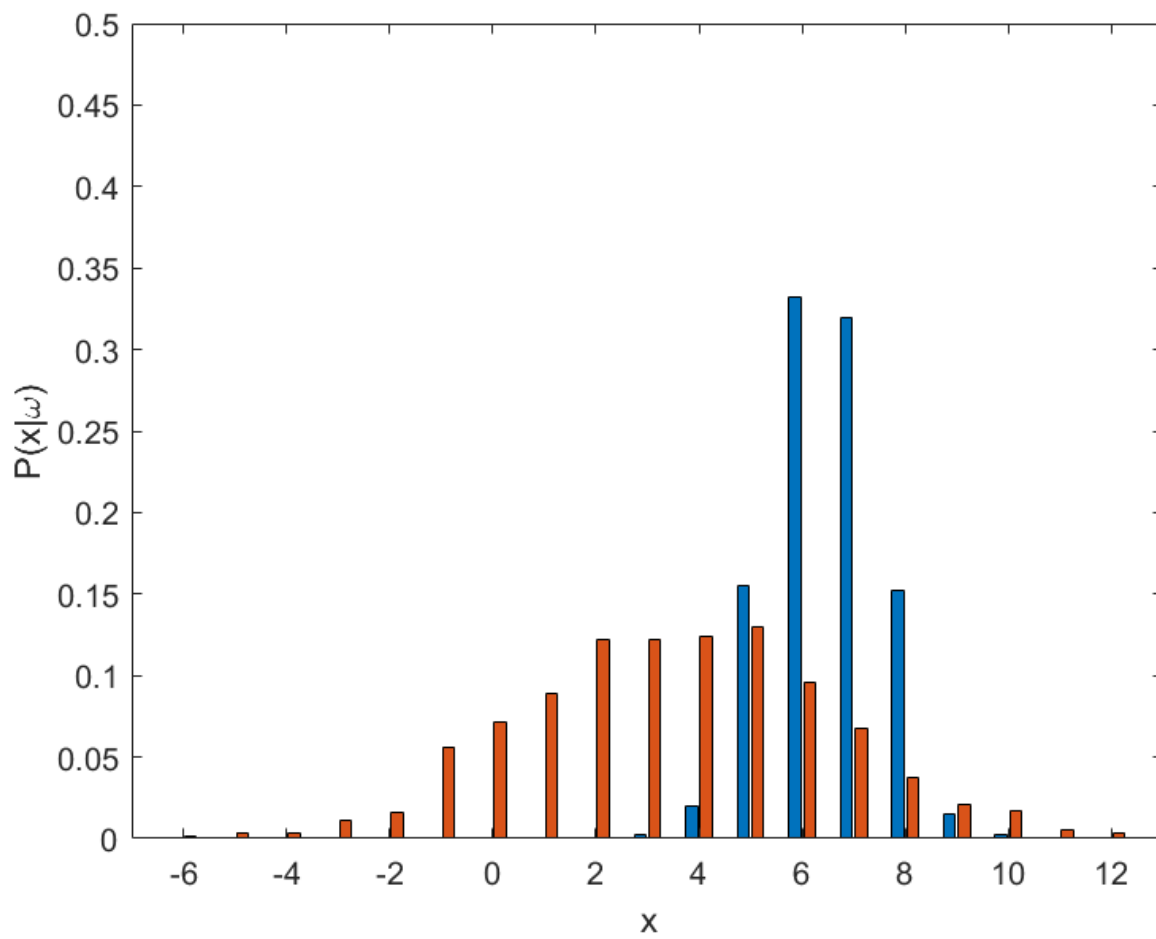
(4) Since

$$P(B_3 = 1 \mid B_2 = 0) == \frac{2}{3}$$

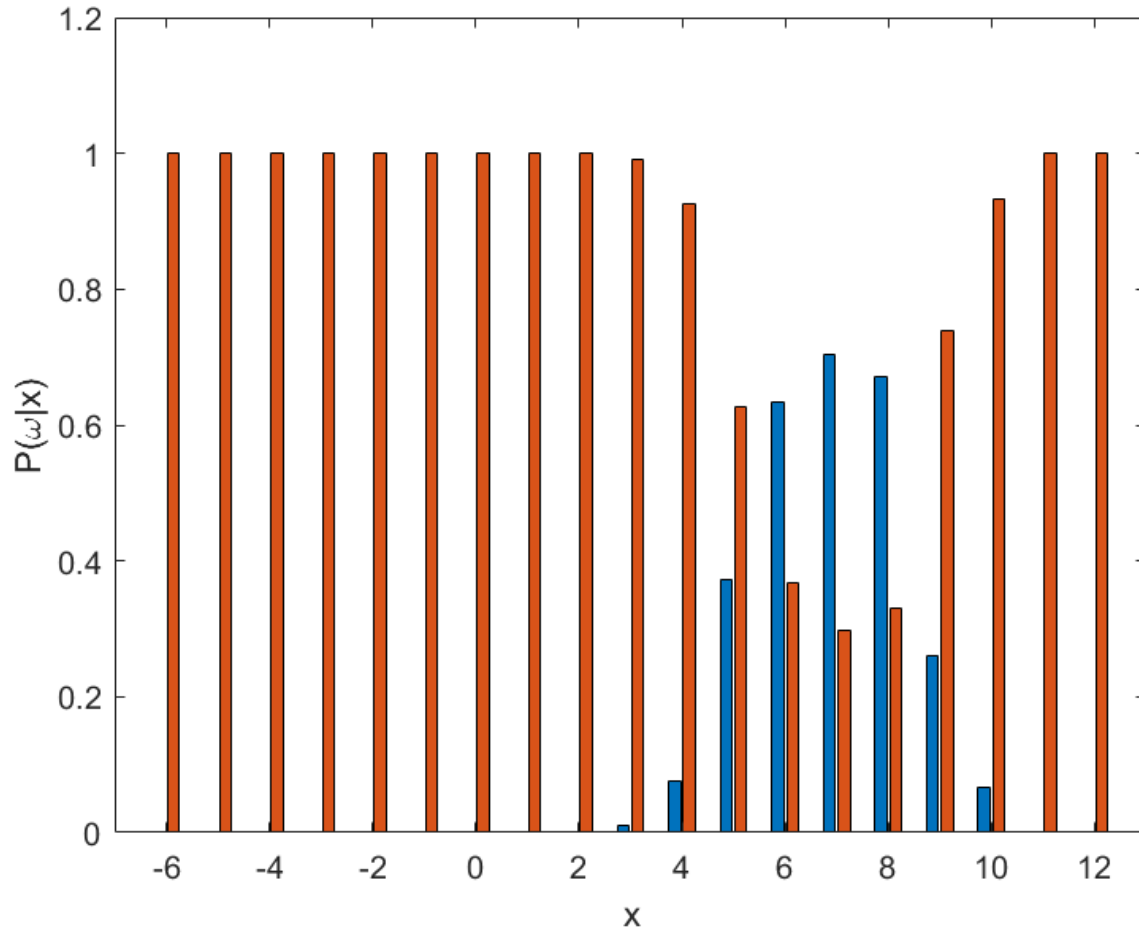We **should** change our decision.

(b)

(i) The distribution graph are as follows:

And the test error, according to the output, is 64.

(ii) The distribution graph are as follows:

And the test error, according to the output, is 47.

(iii) Minimum risk, according to the output, is 0.242667, (**TAKE ALL DATA AND P(X) INTO CONSIDERATION**).

(*Tips.*) The screenshot of the final output of MATLAB script(run.m):

```
>> run
Error = 64.
Error = 47.
Min Risk = 0.242667.
```

3. Gaussian Discriminant Analysis and MLE

(a)

According to Bayes' theorem:

$$P(y = 1 \mid \mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = \frac{P(\mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1 \mid y = 1)P(y = 1)}{P(\mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1)}$$

According to the definition of conditional probability:

$$P(\mathbf{x}) = P(\mathbf{x} \mid y = 0)P(y = 0) + P(\mathbf{x} \mid y = 1)P(y = 1)$$

According to the conditions given, we can also know that:

$$\Sigma_0 = \Sigma_1 = \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I, \phi = \frac{1}{2}, \mu_0 = (0,0), \mu_1 = (1,1)^T$$

Combine the above expressions to get:

$$P(y = 1 \mid \mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = \frac{P(\mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1 \mid y = 1)P(y = 1)}{P(\mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1)}$$

$$= \frac{P(\mathbf{x} \mid y = 1)P(y = 1)}{P(\mathbf{x} \mid y = 0)P(y = 0) + P(\mathbf{x} \mid y = 1)P(y = 1)}$$

$$= \frac{\frac{1}{4\pi}e^{-\frac{1}{2}[(x_1-1)^2+(x_2-1)^2]}}{\frac{1}{4\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)} + \frac{1}{4\pi}e^{-\frac{1}{2}[(x_1-1)^2+(x_2-1)^2]}}$$

$$= \frac{e^{-\frac{1}{2}[(x_1-1)^2+(x_2-1)^2]}}{e^{-\frac{1}{2}(x_1^2+x_2^2)} + e^{-\frac{1}{2}[(x_1-1)^2+(x_2-1)^2]}}$$

$$= \frac{\frac{1}{4\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)}e^{x_1+x_2-1}}{\frac{1}{4\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)} + \frac{1}{4\pi}e^{-\frac{1}{2}(x_1^2+x_2^2)}e^{x_1+x_2-1}}$$

$$= \frac{e^{x_1+x_2-1}}{1 + e^{x_1+x_2-1}}$$

$$= \frac{1}{1 + e^{1-x_1-x_2}}$$

Since $P(y = 0 \mid \mathbf{x}; \dots) = 1 - P(y = 1 \mid \mathbf{x}; \dots)$, we can get:

$$P(y = 0 \mid \mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = \frac{1}{1 + e^{x_1+x_2-1}}$$

Let $P(y = 1 \mid \mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1) = P(y = 0 \mid \mathbf{x}; \phi, \mu_0, \mu_1, \Sigma_0, \Sigma_1)$, We can get a plane:

$$x_1 + x_2 = 1$$

Since these two functions are monotonous with respect to $u = x1 + x2$, it is easy to know:

✳ if $x_1 + x_2 > 1, P(y = 1 \mid \mathbf{x}) > P(y = 0 \mid \mathbf{x})$;

✳ if $x_1 + x_2 < 1, P(y = 1 \mid \mathbf{x}) < P(y = 0 \mid \mathbf{x})$;
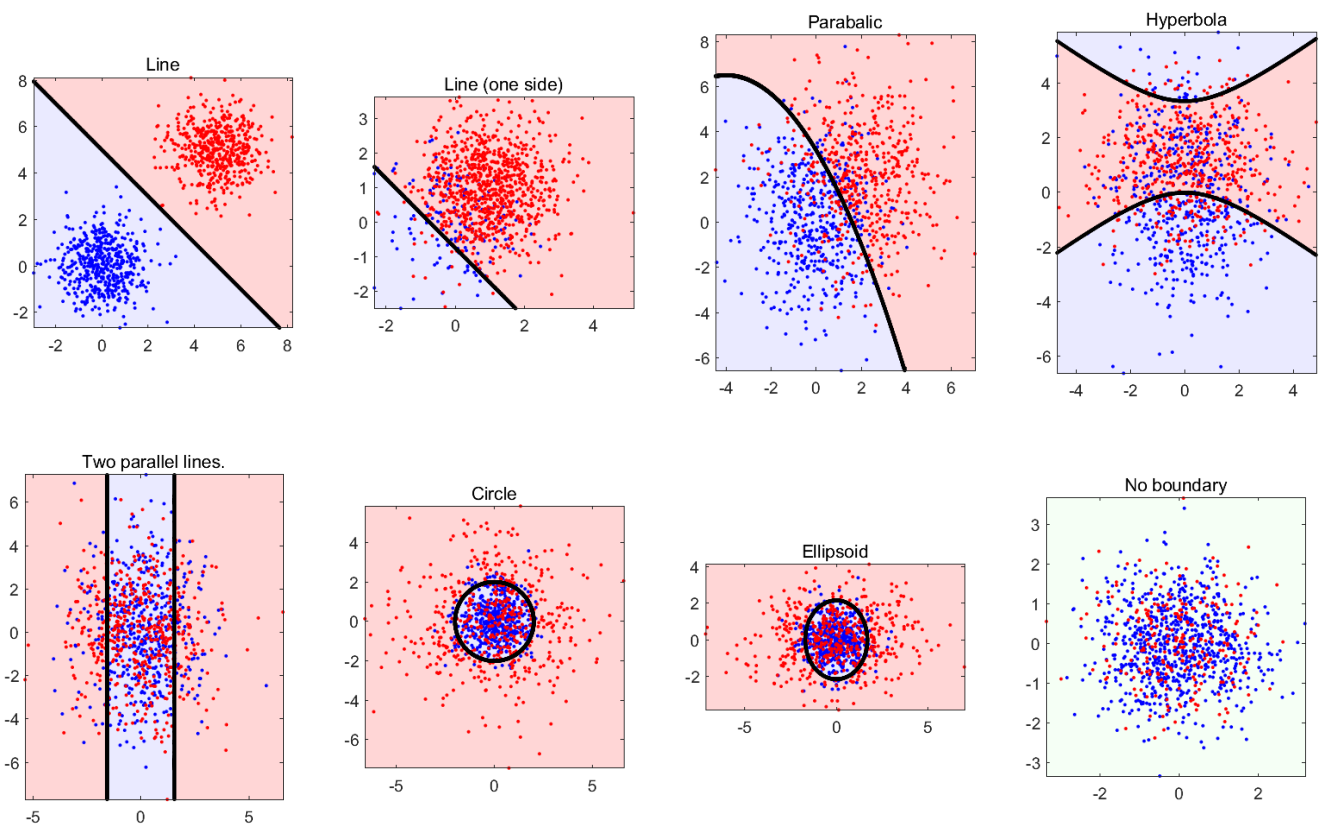
So that is the distribution plane.

(b)

Code Added (View $gaussian\_pos\_prob.m$ for details):

```
M = size(X, 1);
for i = 1:N
    for j = 1:K
        x = X(:,i);
        mu = Mu(:,j);
        sigma = Sigma(:,:,j);
        P(i, j) = Phi(j) * exp((x-mu)' * sigma^-1 * (x-mu)/-2);
        P(i, j) = P(i,j)./ ((2 * pi)^(M/2) * sqrt(det(sigma)));
    end
end
P = P ./ sum(P,2);
p = P;
```

(c)

$N = 1000, Step = 0.01$

(d)

[k-Classes Cases]

Promote the case described in the problem to k classes, $\forall k$:

$$data_k = \left\{ (\mathbf{x}^{(i)}, y^{(i)}) \mid y^{(i)} = k, i = 1, \ldots, m \right\}$$

$$\mathbf{x}_k = \left\{ \mathbf{x}^{(i)} \mid y^{(i)} = k, i = 1, \ldots, m \right\}, N_k = |\mathbf{x}_k|$$

According to the concept of MLE, We need to now:

$$argmax(P\left(\mathbf{x}_k \mid \mu_k, \boldsymbol{\Sigma}_k\right))$$

According to the formula in the slides:

$$p(D \mid \theta) = \prod_{k=1} p(\mathbf{x}_k \mid \theta)$$

we get:

$$P\left(\mathbf{x}_k \mid \mu_k, \boldsymbol{\Sigma}_k\right) = \prod_{\mathbf{x} \in \mathbf{x}_k} \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{|\boldsymbol{\Sigma}_k|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\mu_k)}$$

To simplify the problem, we take the logarithm transformation:

$$\ln P\left(\mathbf{x}_k \mid \mu_k, \boldsymbol{\Sigma}_k\right) = \ln(\prod_{\mathbf{x} \in \mathbf{x}_k} \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{|\boldsymbol{\Sigma}_k|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\mu_k)})$$

$$= -\frac{1}{2} N_k D \ln 2\pi - \frac{N_k}{2} \ln |\boldsymbol{\Sigma}_k| - \frac{1}{2} \sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mu_k)$$

$$= -\frac{N_k}{2} \left(D \ln 2\pi + \ln |\boldsymbol{\Sigma}_k|\right) - \frac{1}{2} \sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mu_k)$$

Partial derivative of μ_k (partly refer to slides):

$$\frac{\partial}{\partial \mu_k} \ln P\left(\mathbf{x}_k \mid \mu_k, \boldsymbol{\Sigma}_k\right) = \sum_{\mathbf{x} \in \mathbf{x}_k} \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \mu_k) = 0$$

$$\boxed{\mu_{max} = \frac{\sum_{\mathbf{x} \in \mathbf{x}_k} \mathbf{x}}{|\mathbf{x}_k|}}$$

Partial derivative of $\boldsymbol{\Sigma}_k$ (partly refer to slides):

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_k} \ln P\left(\mathbf{x}_k \mid \mu_k, \boldsymbol{\Sigma}_k\right)$$

$$= -\frac{N_k}{2} \boldsymbol{\Sigma}_k^{-1} - \frac{1}{2} \boldsymbol{\Sigma}_k^{-1} \left(\sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)^T\right) \boldsymbol{\Sigma}_k^{-1} = 0$$

Sort the equation above, get:

$$N_k \Sigma_k^{-1} = \Sigma_k^{-1} \left(\sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)^T \right) \Sigma_k^{-1}$$

$$\boxed{\Sigma_{max} = \frac{\sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_{max})(\mathbf{x} - \mu_{max})^T}{|N_k|}}$$

Get:

$$\mu_k = \frac{\sum_{\mathbf{x} \in \mathbf{x}_k} \mathbf{x}}{|\mathbf{x}_k|}$$

$$\phi_k = \frac{|\mathbf{x}_k|}{\sum_{i=1}^{K} |\mathbf{x}_i|}$$

$$\Sigma_k = \frac{\sum_{\mathbf{x} \in \mathbf{x}_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)^T}{|\mathbf{x}_k| - 1}$$

4. Text Classification with Naïve Bayes

(a)

Answer : By running corresponding code, we can know the order (viewed in MATLAB):

1	2	3	4	5	6	7	8	9	10
30033	75526	38176	45153	9494	65398	37568	13613	56930	9453

Look up the file, the top 10 word with the highest ratio is:

1	2	3	4	5	6	7	8	9	10
nbsp	viagra	pills	cialis	voip	php	meds	computron	sex	ooking

(b)

Answer :The accuracy is 0.9857.(for calculation process see **(d)**)

(c)

Answer : This statement is $FALSE$. According to Bayes theorem, if $p(Spam) = 0.01, p(Ham) = 0.99, S'$ or H' means **considered** as spam & ham (but may not BE spam & ham). then:

$$P(S \mid S') = \frac{P(S' \mid S)P(S)}{P(S' \mid S)P(S) + P(H' \mid H)P(H)} = 0.5$$

The probability of precision is 0.5. It's not a good ratio actually. So this statement is $FALSE$.

(d)

According to the output of the code(View $run.m$ for details):

```
accuracy = (tp + tn) / (num_ham_test + num_spam_test);
precision = tp / (tp + fp);
recall = tp / (tp + fn);
fprintf("accuracy= %f.\n",accuracy);
fprintf("precision= %f.\n",precision);
fprintf("recall= %f.\n",recall);
```

```
>> run
accuracy= 0.985732.
precision= 0.975022.
recall= 0.972420.
```

The result is:

$$\begin{cases} accurancy = 0.985732 \approx 0.9857 \\ precision = 0.975022 \approx 0.9750 \\ recall = 0.972420 \approx 0.9750 \end{cases}$$

(e)

$Answer$: **Precision** should be the most important. Because it is basically acceptable that some spam is unshielded, but the result of missing important emails is devastating because of the high level of shielding.

For airport security, however, **Recall** becomes especially important. There may be some passengers be misjudged, but for safety, it is always not too much to be careful.