Assignment 1.

Name: 沈锴
ID: 21921071

1. Machine Learning Problems

1.1 Choose proper word(s) from

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

1.2

False. Because if we use all data as train set, the model is totally fit for this dataset but has no idea of the performance on unknown data. So it can lead to bad result on the unknown data, which is called overfitting. We should use part of the dataset as training data and use the rest data as test data for evaluation.

2. Bayes Decision Rule

a) Suppose you are given a chance to win bonus grade points

1 1/3

2 1

3 $P(B_1 = 1 \mid B_2 = 0) = \dfrac{P(B_2 = 0 \mid B_1 = 1) * P(B_1 = 0)}{P(B_2 = 0)} = \dfrac{1 * 1/3}{1} = \dfrac{1}{3}$

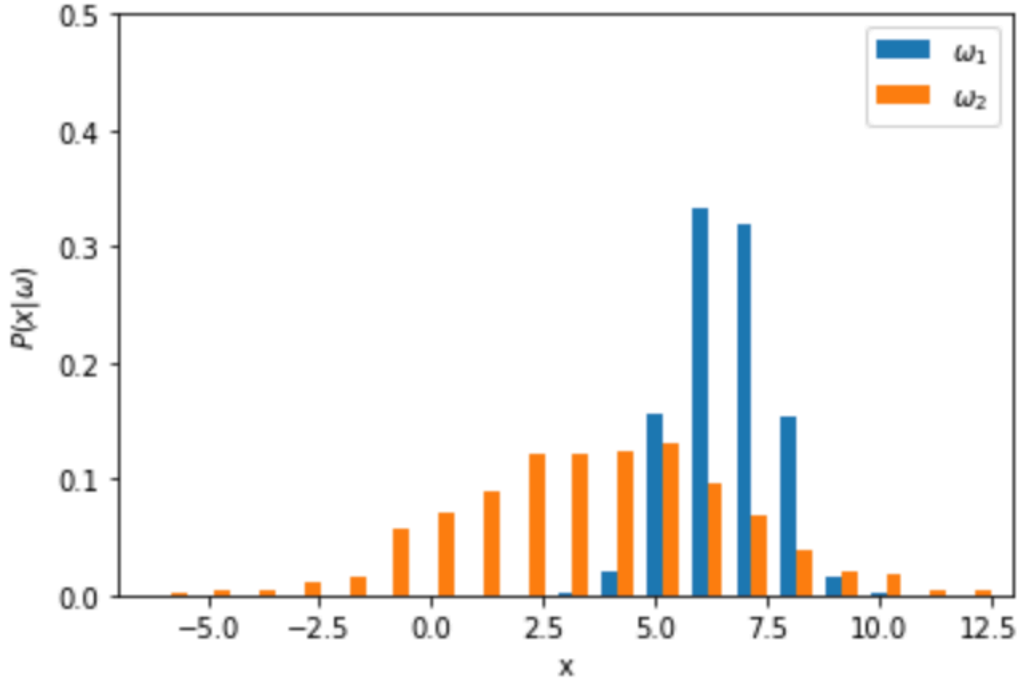
4 Yes. $P(B_3 = 1 \mid B_2 = 0) = 1 - P(B_1 = 1 \mid B_2 = 0) = \dfrac{2}{3} > P(B_1 = 1 \mid B_2 = 0)$

b)

1)

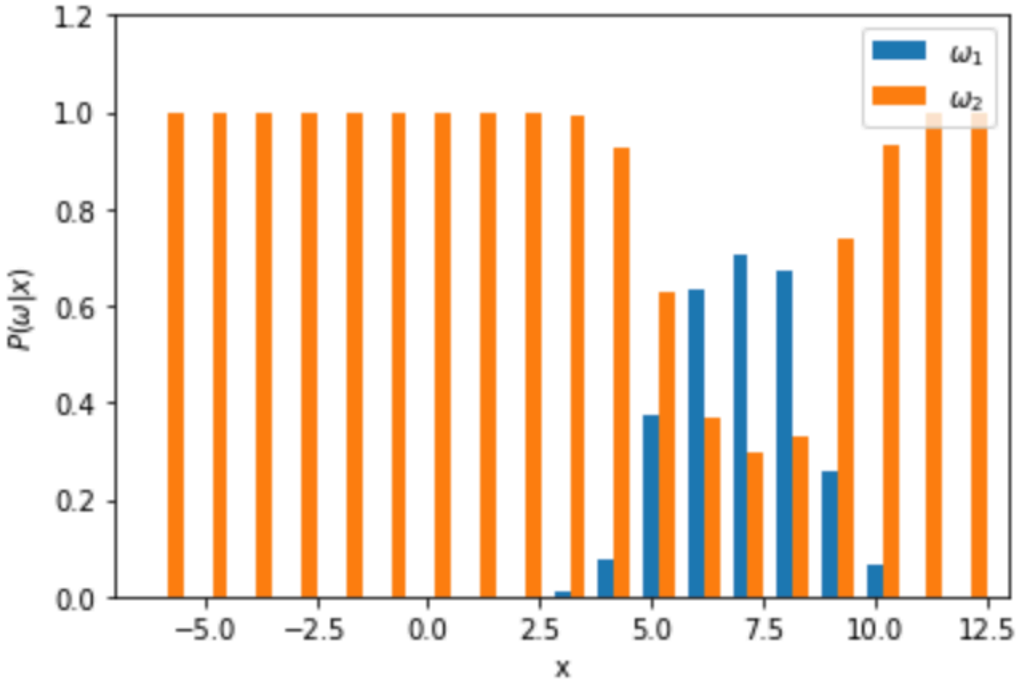
misclassified = 64

rate = 64/300 = 21.33%



2)

misclassified = 47
rate = 47/300 = 15.67%



3)
minrisk = 0.2475

3.Gaussian Discriminant Analysis and MLE

a)

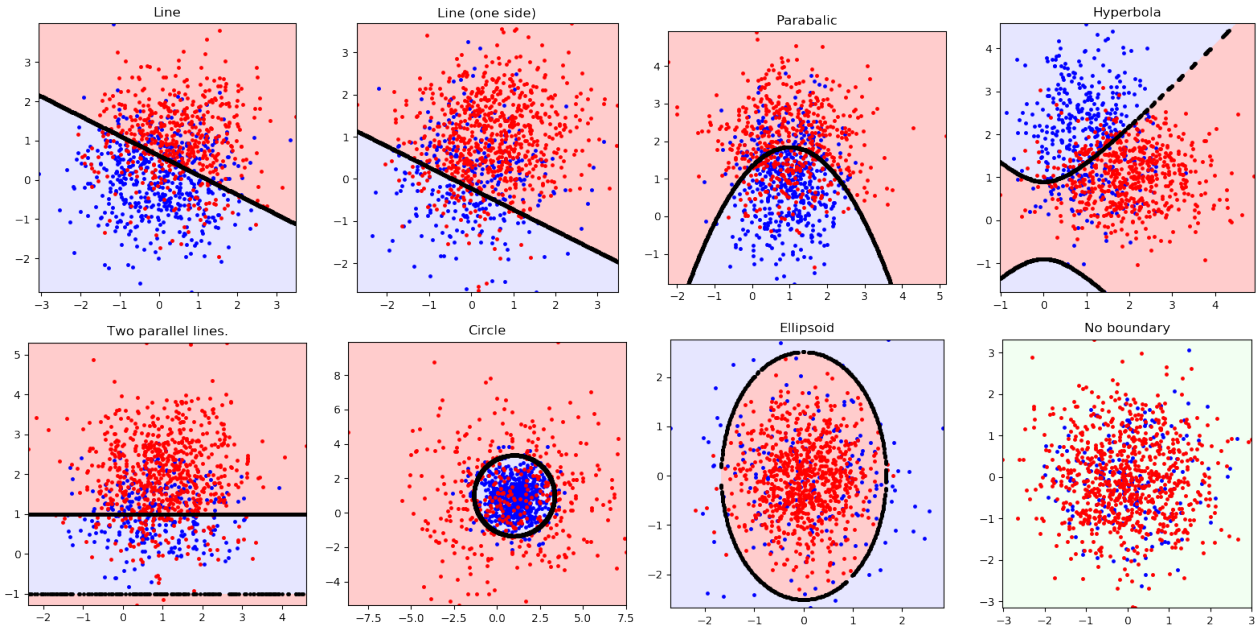
$$P(y=1\,|\,x)=\frac{P(x\,|\,y=1)*P(y=1)}{P(x)}=\frac{P(x\,|\,y=1)*P(y=1)}{P(x\,|\,y=0)*P(y=0)+P(x\,|\,y=1)*P(y=1)}=\frac{N(\mu_1,\Sigma_1)*\phi}{N(\mu_1,\Sigma_1)*\phi+N(\mu_0,\Sigma_0)*(1-\phi)}$$

the decision boundary:

$$P(y=0\,|\,x)=\frac{P(x\,|\,y=0)*P(y=0)}{P(x)}=\frac{P(x\,|\,y=0)*P(y=0)}{P(x\,|\,y=0)*P(y=0)+P(x\,|\,y=1)*P(y=1)}=\frac{N(\mu_0,\Sigma_0)*(1-\phi)}{N(\mu_1,\Sigma_1)*\phi+N(\mu_0,\Sigma_0)*(1-\phi)}=P(y=1\,|\,x)$$

and $x_1+x_2=1$

c)



d)

$$L(\phi,\mu_0,\mu_1)=\prod_{y_i=0}(1-\phi)*N(\mu_0,\Sigma_0,\mathrm{x}_i)*\prod_{y_i=1}\phi*N(\mu_1,\Sigma_1,\mathrm{x}_i)$$

$$l(\phi,\mu_0,\mu_1)=log(L)=\Sigma_{y_i=0}ln(1-\phi)+\Sigma_{y_i=1}ln\phi+\Sigma_{y_i=0}ln(N(\mu_0,\Sigma_0,\mathrm{x}_i))+\Sigma_{y_i=1}ln(N(\mu1,\Sigma_1,\mathrm{x}_i))$$

to make it easier to calculate, we denote C_0 as the number of examples which $y_i = 0$, and C_1 as $y_i = 1$

For ϕ:

$$\frac{dl}{d\phi} = -\frac{C_0}{1-\phi} + \frac{C_1}{\phi} = 0$$

we get: $\phi = \frac{C_1}{C_0 + C_1}$

For μ_0:

$$\frac{dl}{d\mu_0} = \Sigma_{y_i=0}\Sigma_0^{-1}(\mathrm{x}_i - \mu_0) = 0$$

so we get: $\mu_0 = \frac{\Sigma_{y_i=0}\mathrm{x}_i}{C_0}$

and similarly, for μ_1:

$$\mu_1 = \frac{\Sigma_{y_i=1}\mathrm{x}_i}{C_1}$$

4 Text Classification with Naive Bayes

a)

the top 10 will be:

```
nbsp, 1325.1002358991152
viagra, 1249.5763882571969
pills, 1101.9615951389017
cialis, 847.9268348888121
voip, 837.6281283921868
php, 768.9700850813518
meds, 672.8488244461829
computron, 652.2514114529324
sex, 614.4894876319731
width, 518.3682269968041
```

b)

accuracy: 0.9857315598548972

c) False. Because if we consider the situation of a spam filter when the ratio of spam and ham is 1:99, and we label all email to ham, so the accuracy is 99%. But this model actually doesn't work.

d)
precision: 0.9750223015078054
recall: 0.9724199288169715

e)
I think for a spam filer email, precision is more important. Because we care more about the correctness of the labeled email.
But for a classifier of drugs and bombs, recall is more important. Because we care more about labeling all possible things.