# 1.

(a)

1. B F
2. C
3. A D
4. B G
5. A E
6. A D
7. B F
8. A E
9. C

(b)

False. We should focus on the result of model in testing data rather than training data. if we over concerned in training data, the model will overfit.

# 2.

(a)

(i)P(B1 = 1) = 1/3

(ii)P(B2 = 0 | B1 = 1) = P(B2 = 0, B1 = 1) / P(B1 = 1)

P(B2 = 0, B1 = 1) = 1 / 3

P(B1 = 1) = 1 / 3

So P(B2 = 0 | B1 = 1) = 1

(iii)P(B1 = 1 | B2 = 0) = P(B1 = 1, B2 = 0) / P(B2 = 0)

P(B1 = 1, B2 = 0) = 1/3

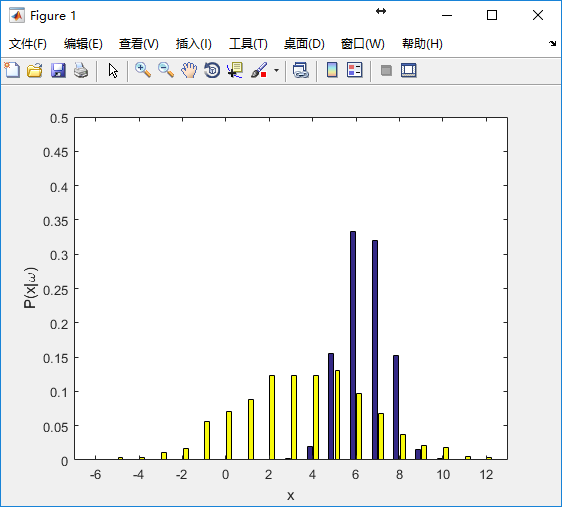
P(B2 = 0) = 1

So P(B1 = 1 | B2 = 0) = 1/3

(iv)According to the Bayes decision rule, the probability of changing choice is bigger, so I should change my choice.

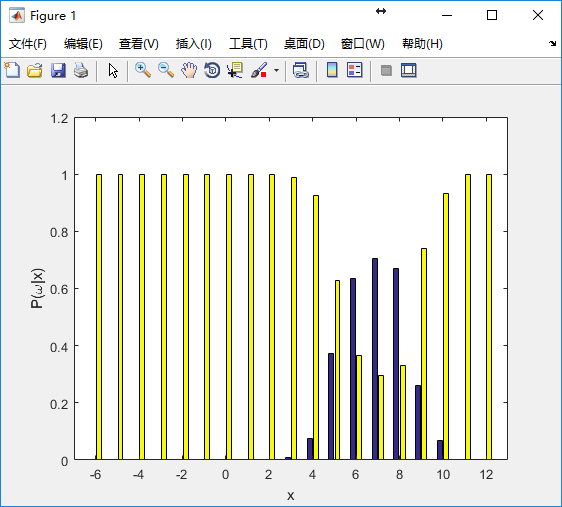
(b)

(i) The following drawing is the the distribution of P(x|ωi)



The number of misclassied test samples is 64.

(ii)The following drawing is the the distribution of P(x|ωi)



The number of misclassied test samples is 47.

(iii)The minimal total risk is 67.

3.

(a)

Make P(x| y = 0) \* P(y = 0) = P(x | y = 1) \* P(y = 1)

So P(x | y = 0) = P(x | y = 1)

So (x1 - 1)^2 + (x2 - 1)^2 = x1^2 + x2^2

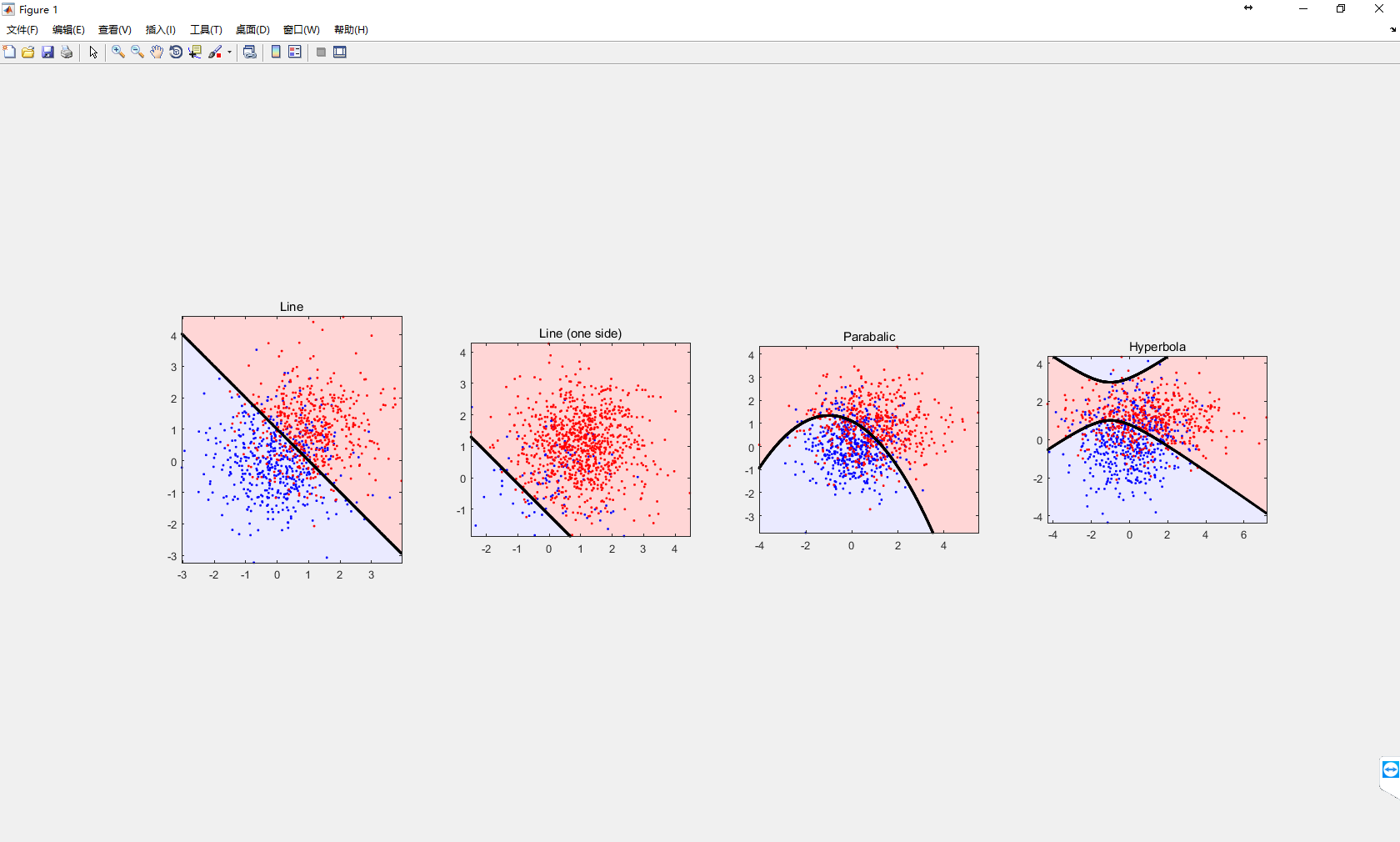
So the boundary is x1 + x2 = 1

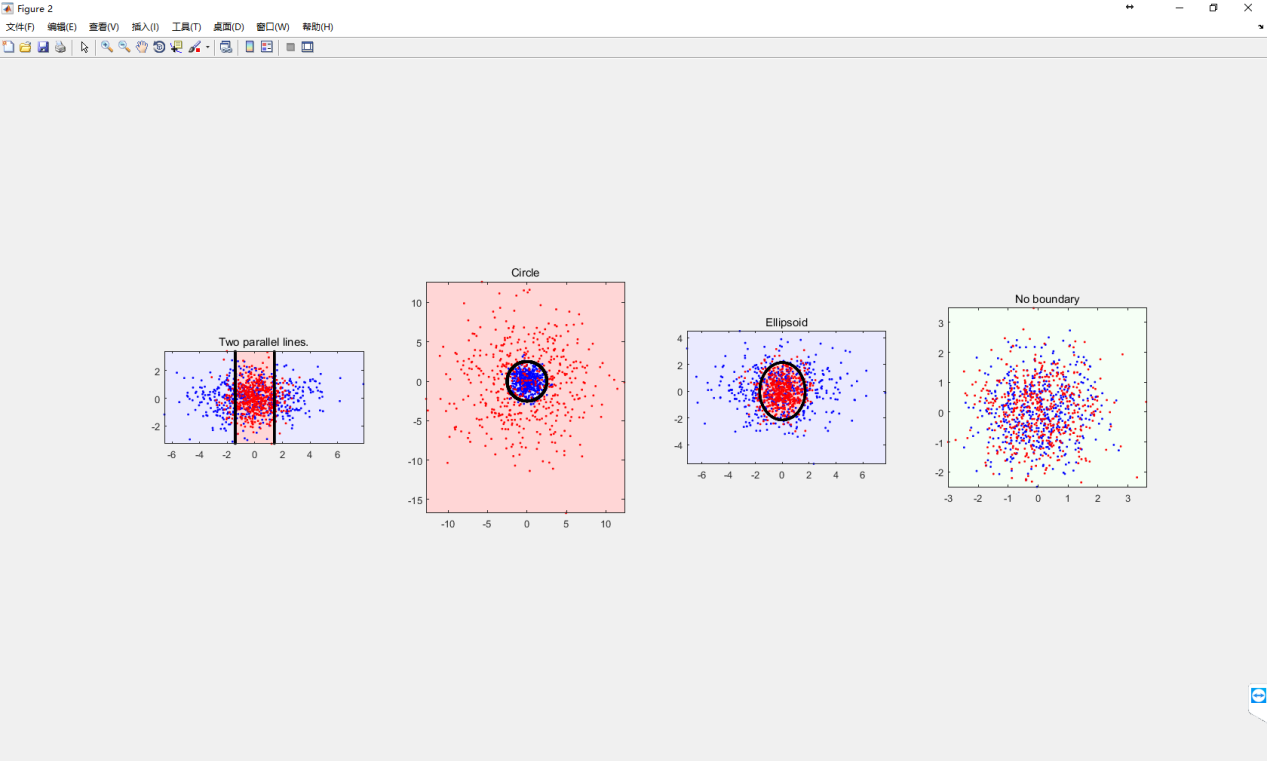
(b)

I have finished gaussian\_pos\_prob.m to calculate posterior probability.

(c)

My result:





(d)

Φ = sum(y=1) / (sum(y=0) + sum(y=1))

μ0 = mean(xi,y=0)

μ1 = mean(xi,y=1)

# 4.

1. The top 10 words list as follow:

gauze id=19957

opportune id=56930

recess id=13613

mexzx id=37568

superbowl id=65398

swam id=9494

noaa id=45153

ejwah id=38176

allyannis id=75526

phenotype id=30033

(b)

The accuracy of my spam filter on the testing set is 98.57%

(c)

False, if the ratio of positive(negative) sample is lower than 1%, a model with 99% accuracy is not always a good model.

(d)

|  |  |  |
| --- | --- | --- |
|  | Spam(label) | Ham(label) |
| Spam(predict) | TP = 1093 | FP = 28 |
| Ham(predict) | FN = 31 | TN = 2983 |

Precision = 97.50%

Recall = 97.24%

(e)

For a spam filter, precision is more important, because classify ham to spam is terrible.

For a classifier to identify drugs and bombs at airport, recall is more important, if a class’ label is drug or bomb, but our prediction is normal, it is dangerous.