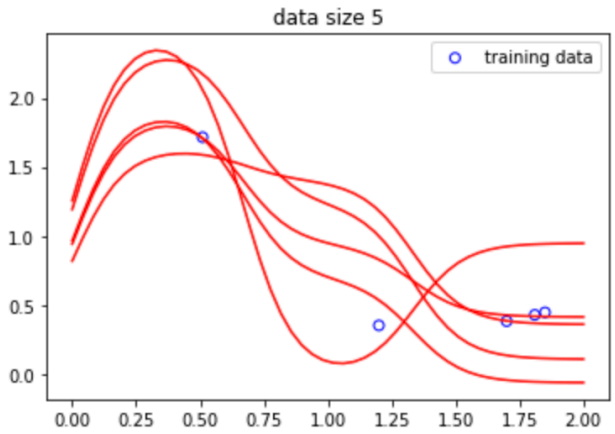
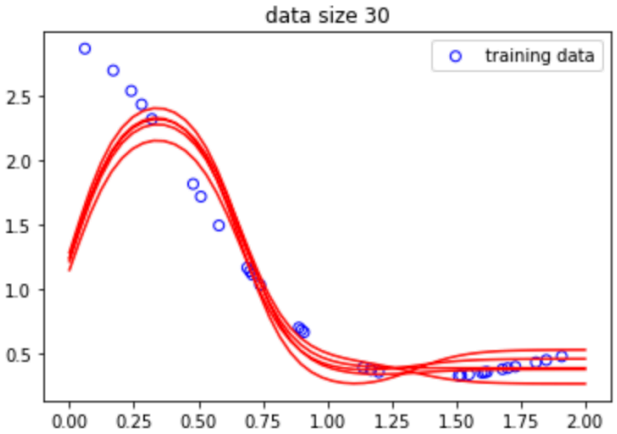
Machine Learning HW2 Report

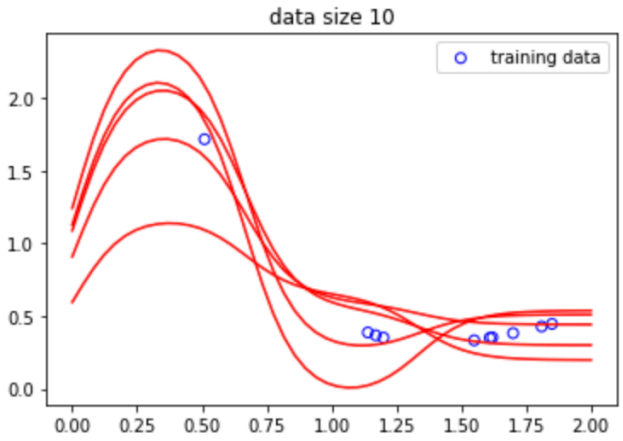
309505018 郭俊廷

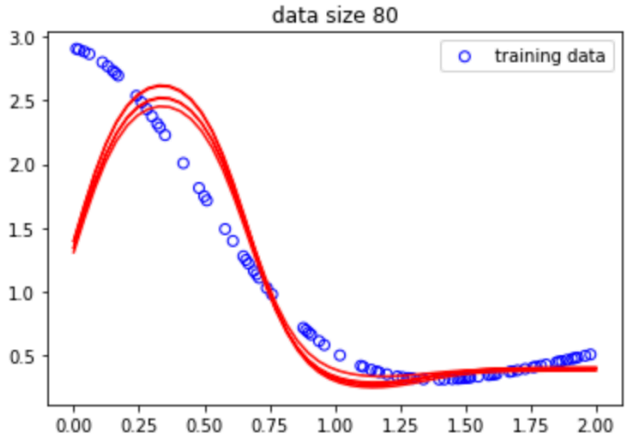
1. Sequential Bayesian Learning

(1) Plot five curves sampled from the parameter posterior distribution and N data points, e.g. (10%)

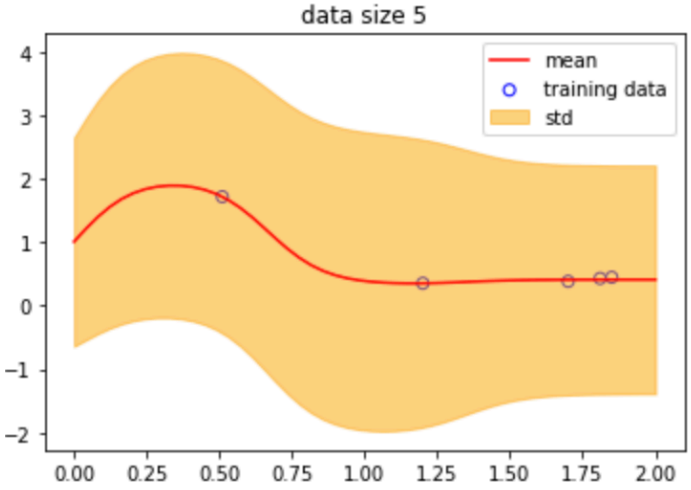


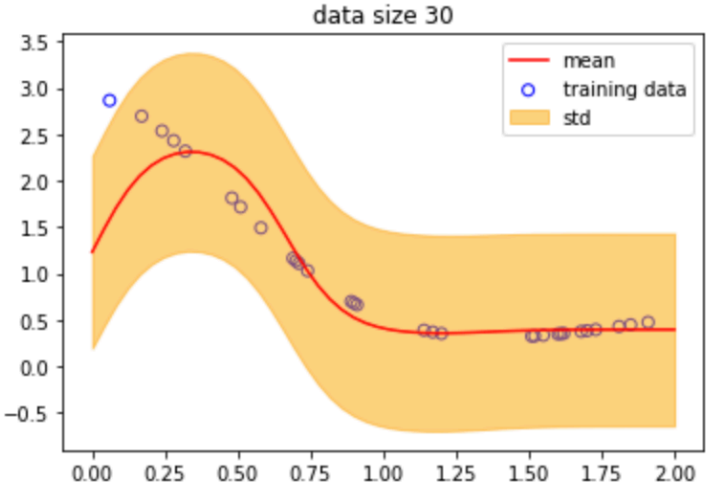


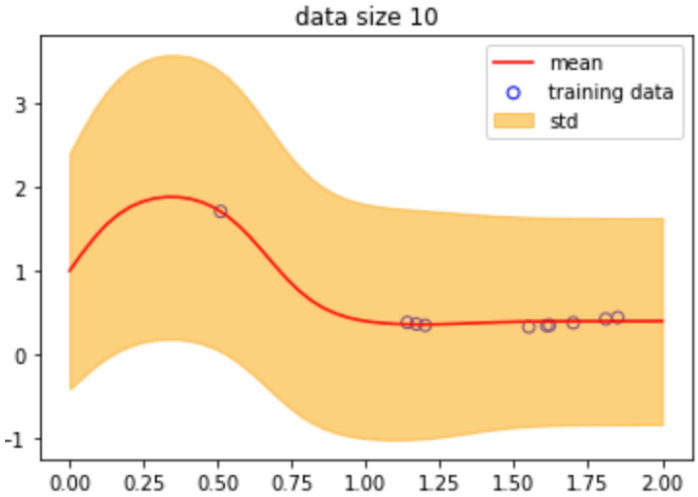


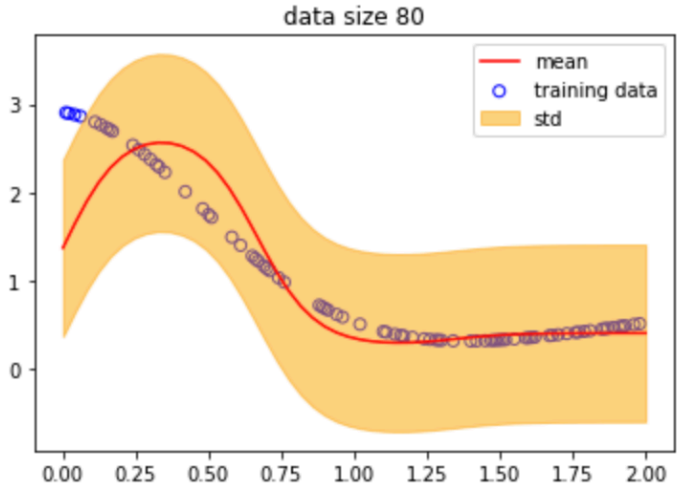


(2) Plot the predictive distribution of target value t by showing the mean curve, the region of variance with one standard deviation on both sides of the mean curve and N data points, (10%)

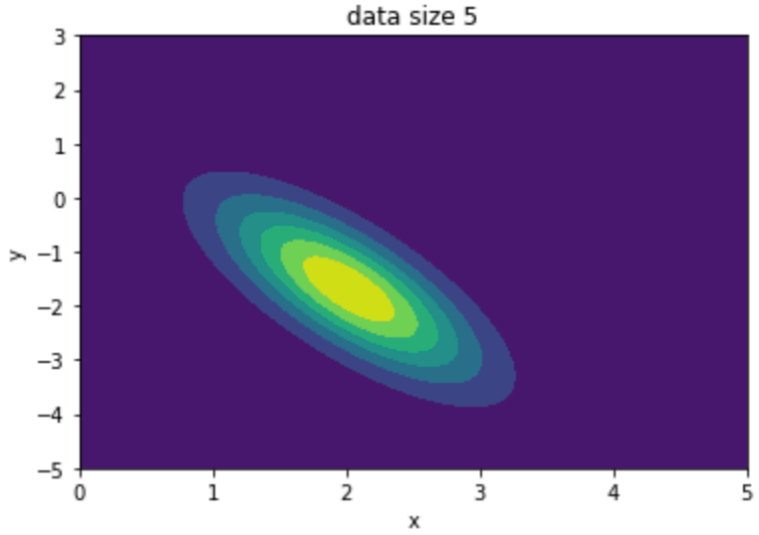


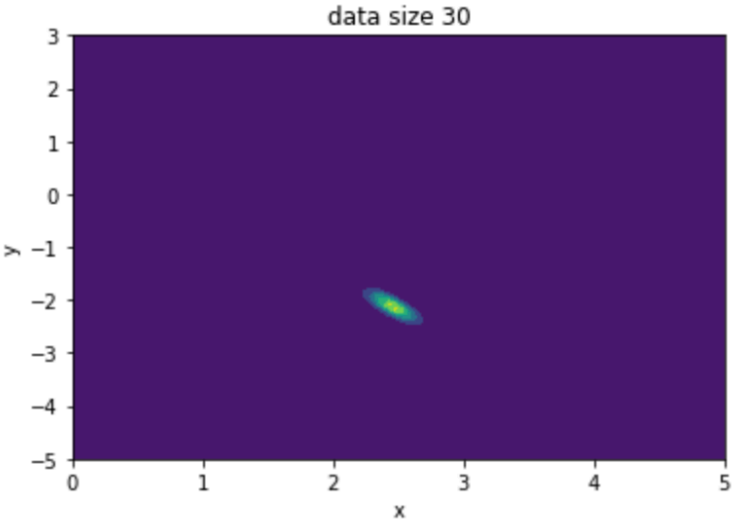


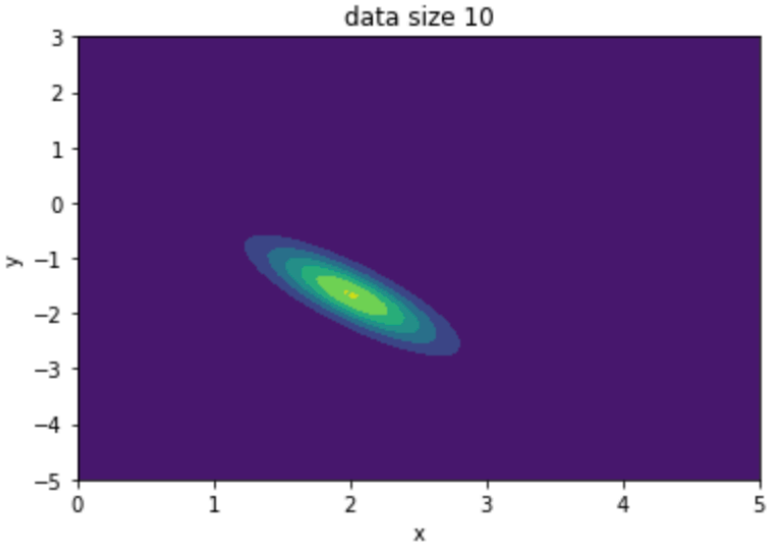


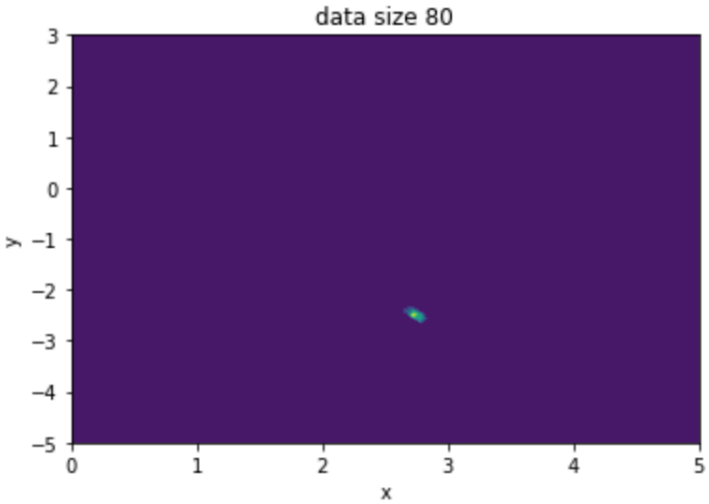


(3) Plot the prior distributions by arbitrarily selecting two weights, (10%)









(4) Make some discussion on the results of different N in 1, 2 and 3. (10%)

第一題是在針對不同的data size做 posterior distribution的sampled，可以看出data size越大，他的five curve sampled出來的曲線會越接近 每一個點所生成的曲線。

第二題是在針對不同的data size做 predictive distribution，並畫出mean curve和標準差，data size比較小時他的mean curve會幾乎通過每一個點，std範圍較寬，data size比較大時他有很多點不在mean curve上，但std範圍比較窄。

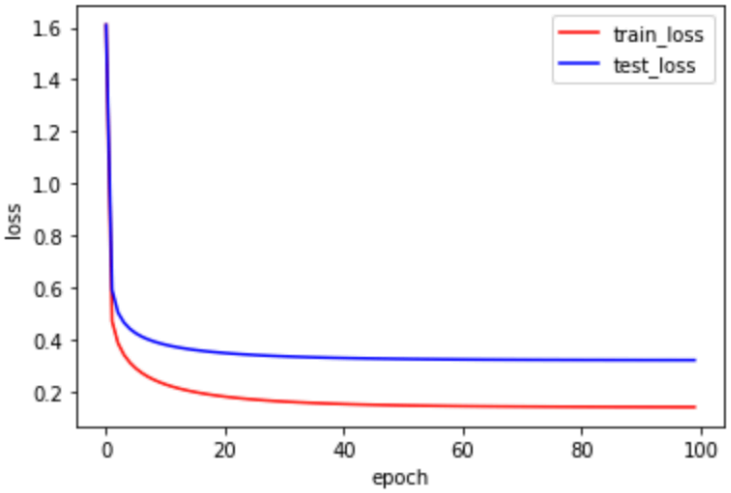
第三題是在針對不同的data size做prior distributions，選兩個weights印出，data size比較小時，兩個weights所能表示的prior distributions比較明顯，data size越大的話，可以看出他的橢圓越小。

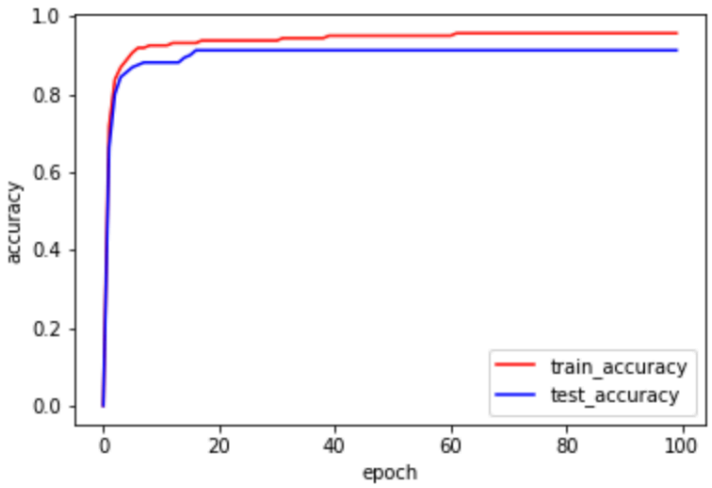
2. Logistic Regression

(1) Implement batch GD, SGD, mini-batch SGD (batch size = 32) and Newton-Raphson algorithms (15%)

(a) Batch GD

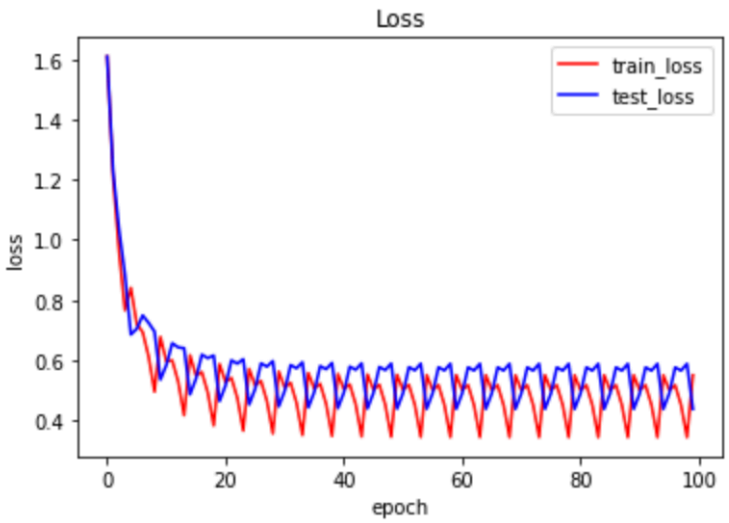
```
train_accuracy:0.95625
test_accuracy:0.9125
```

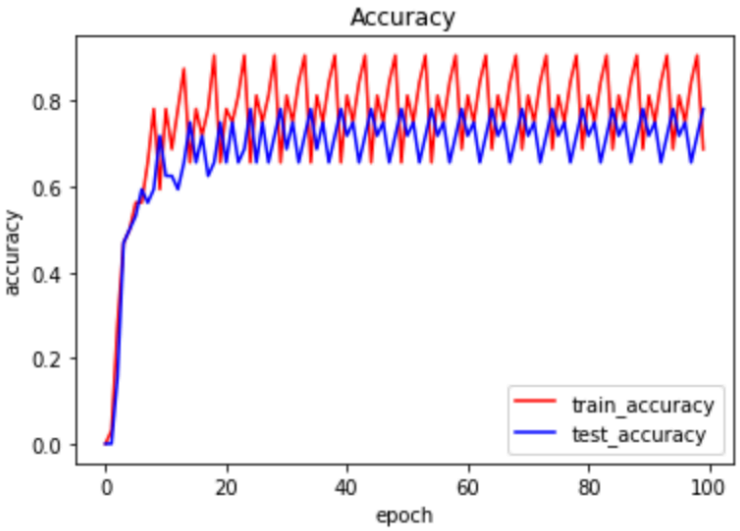




(b) Minibatch GD

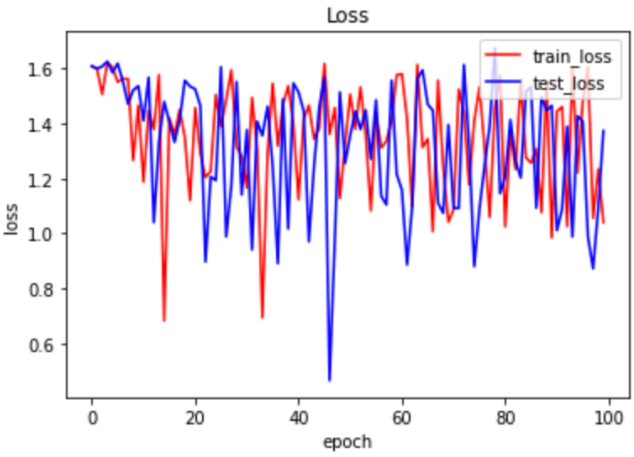
```
train_accuracy:0.90625
test_accuracy:0.78125
```

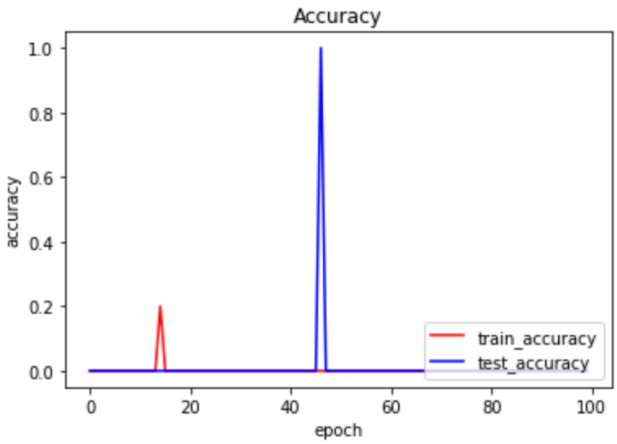




(c) SGD

```
train_accuracy:0.2
test_accuracy:1.0
```





(d) Newton-Raphson

Singular matrix

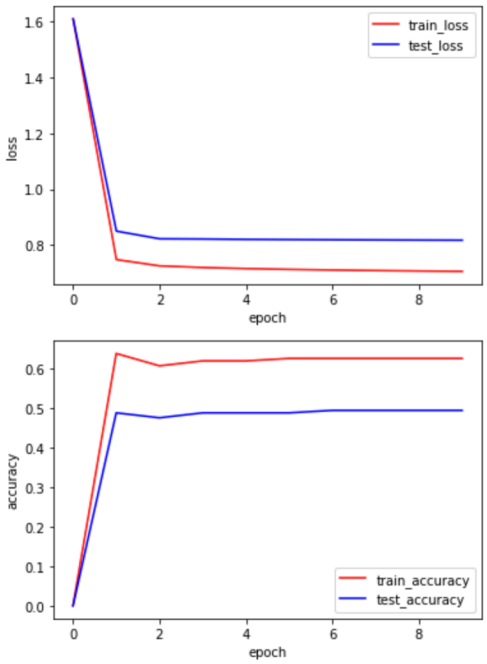
(2) Use principal component analysis (PCA) to reduce the dimension of images to d = 2, 5, 10 (15%)

(a) Repeat 1 by using PCA to reduce the dimension of images to d.

1.Batch GD

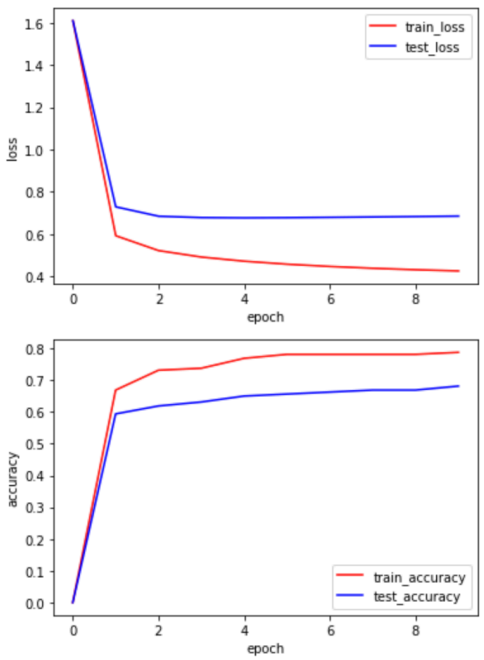
d=2.

```
train_accuracy:0.625
test_accuracy:0.49375
```



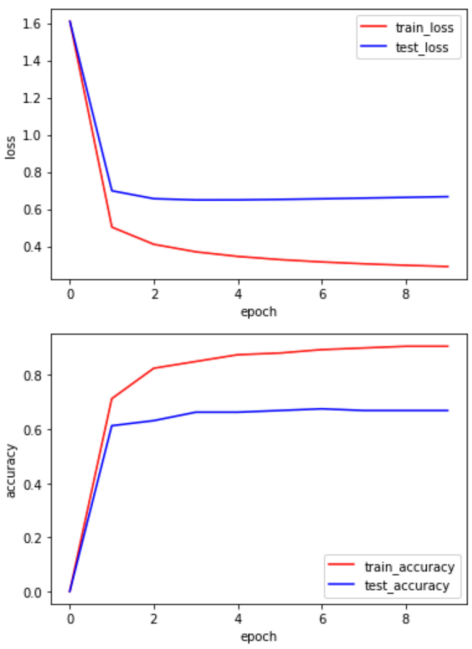
d=5

```
train_accuracy:0.7875
test_accuracy:0.68125
```



d=10

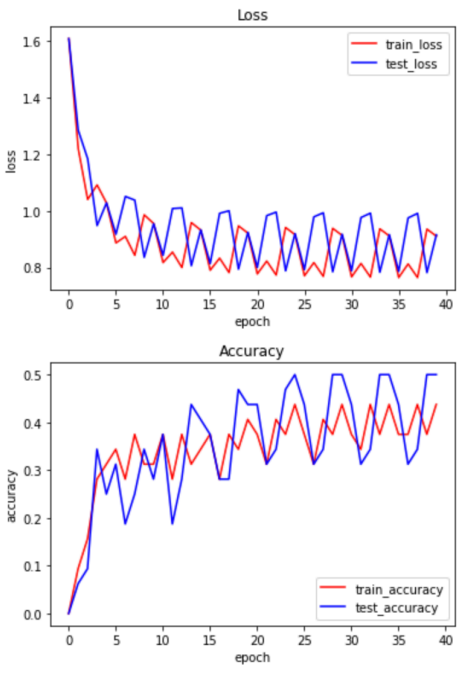
```
train_accuracy:0.90625
test_accuracy:0.66875
```



2.minibatch GD

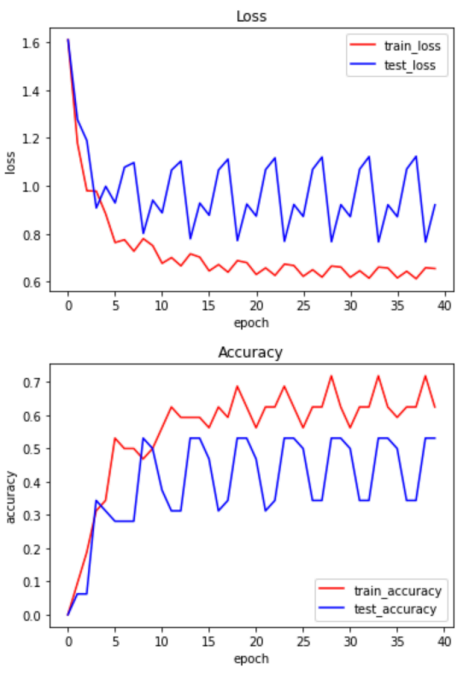
d=2.

```
train_accuracy:0.4375
test_accuracy:0.5
```



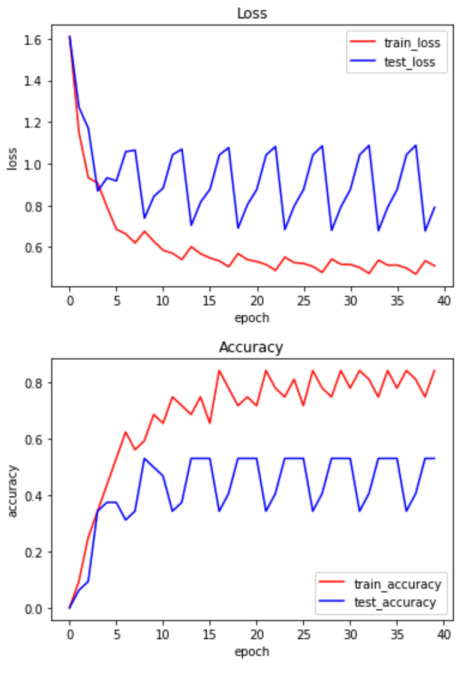
d=5

```
train_accuracy:0.71875
test_accuracy:0.53125
```



d=10

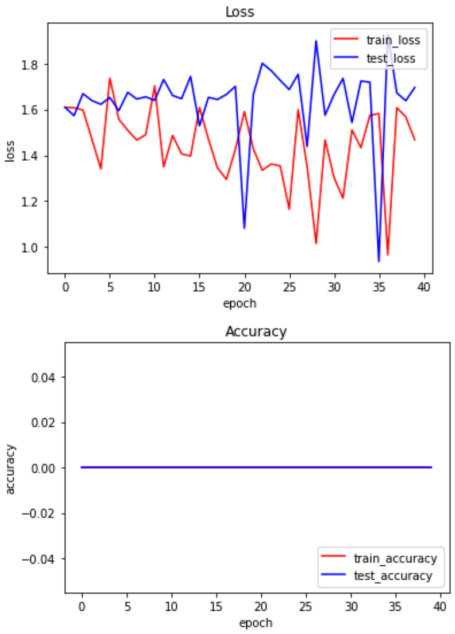
```
train_accuracy:0.84375
test_accuracy:0.53125
```



3. SGD

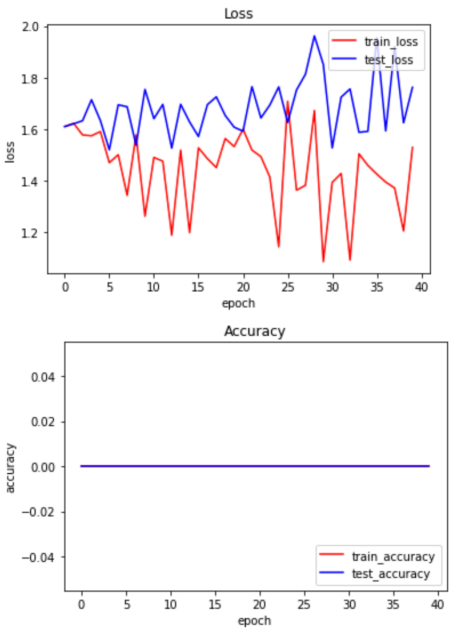
d=2.

```
train_accuracy:0
test_accuracy:0
```



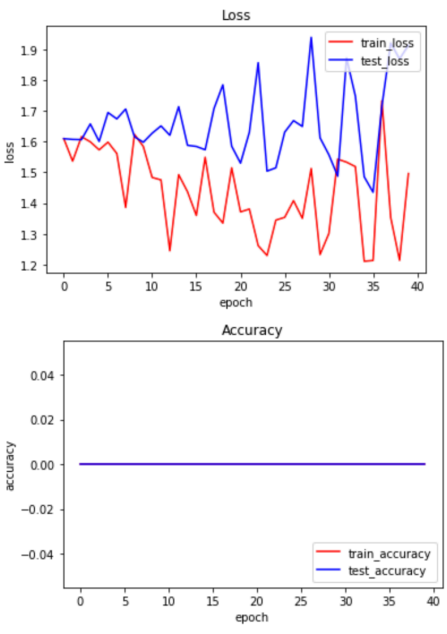
d=5

```
train_accuracy:0
test_accuracy:0
```

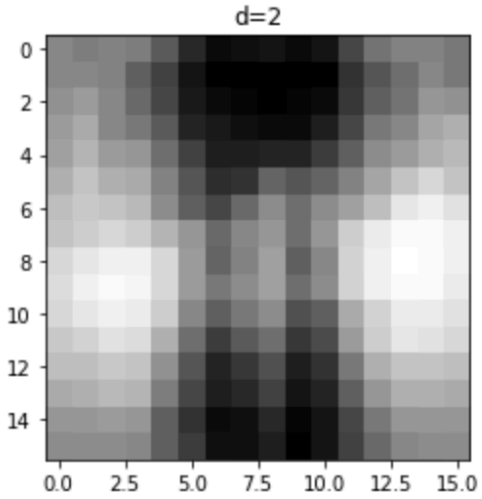


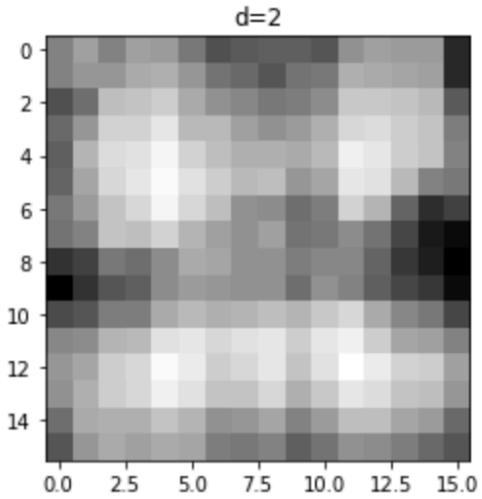
d=10

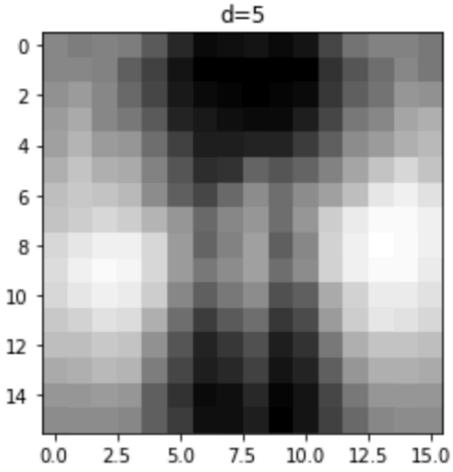
```
train_accuracy:0
test_accuracy:0
```

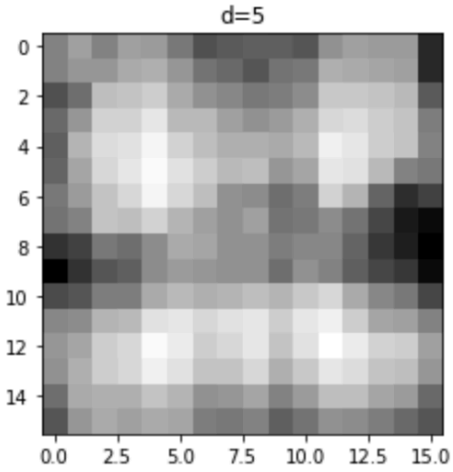


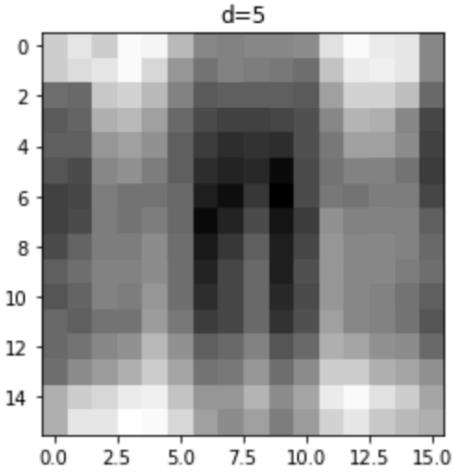
(b) Plot d eigenvectors corresponding to top d eigenvalues, e.g.

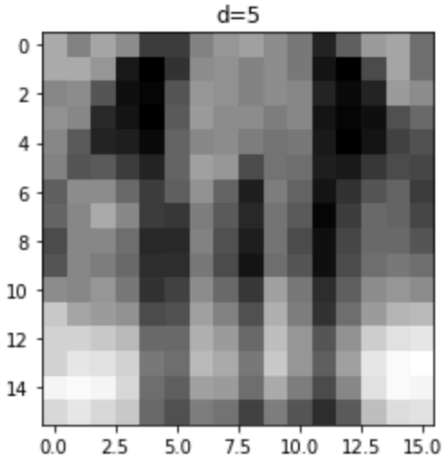


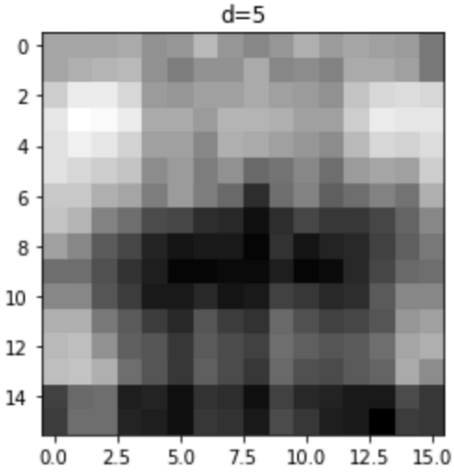




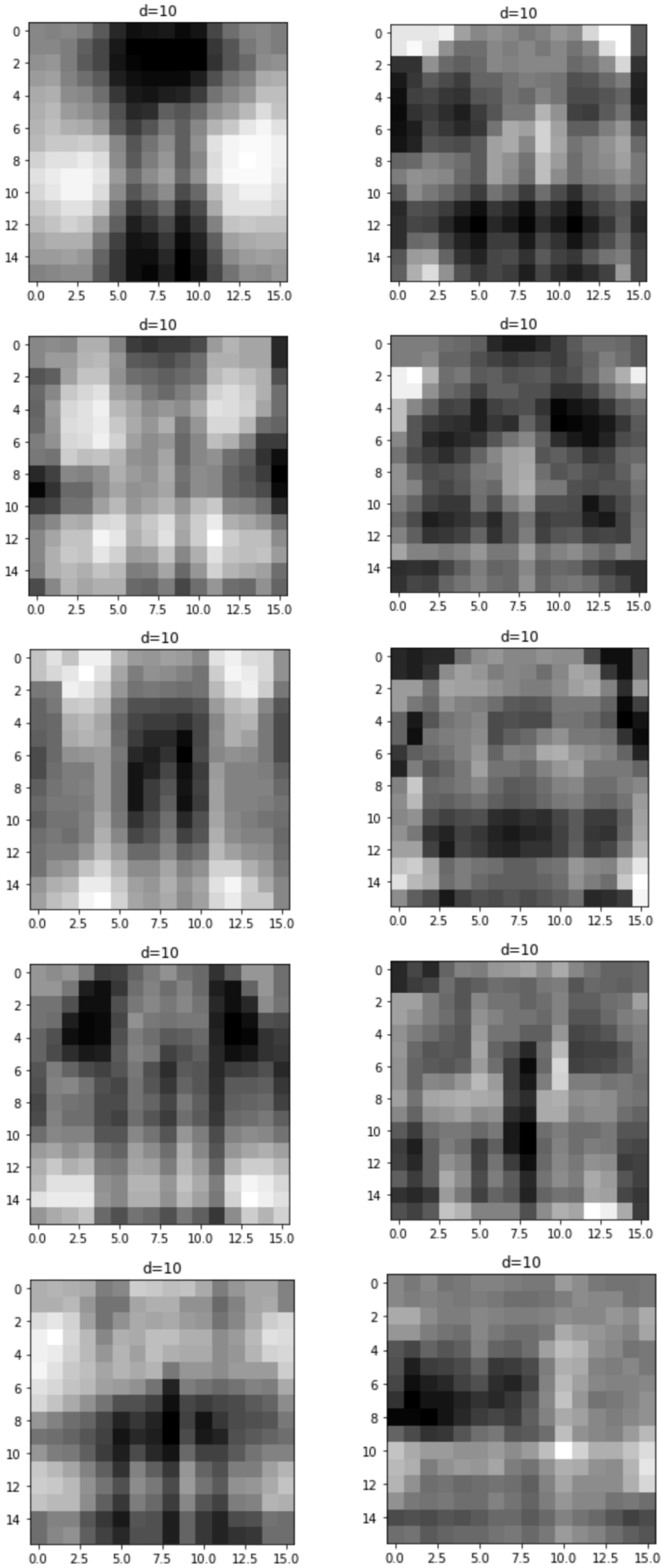






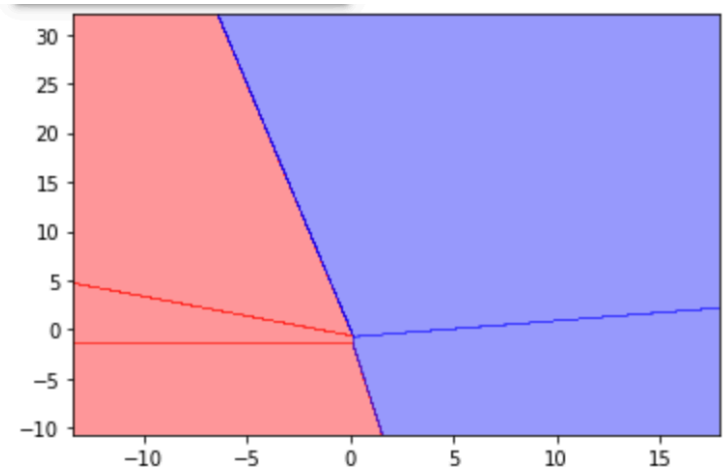


d = 10



(3) What do the decision regions and data points look like on the vector space? (15%)

我使用的是PCA(d=2) 降至二維，再利用 Batch Gradient Descent 預測結果



(4) Make some discussion on the results of 1, 2 and 3. (15%)

第一題是用batch GD, minibatch GD, SGD, Newton-Raphson 四種方法計算 logistic regression，batch GD的loss圖比較平滑下降，accuracy是最高的，minibatch GD每次讀32筆data，loss圖雖會上下擺動，但整體還是有向下，accuracy也不錯高，SGD因為每次都隨機抽取一筆資料，所以loss圖會大幅的上下擺動

第二題是用PCA降維降至2,5,10，用PCA降維完之後可以看出資料量變少，batch GD, minibatch GD, SGD, Newton-Raphson所預測的正確率也跟著降低

第三題是用PCA降維降至2維後，再畫出decision regions，所以可以很清楚看到點在哪個區域