ML HW2 Report

姓名：陳品媛

學號：A063021

Problem 1 Bayesian Linear Regression

Problem 1-1

1. Compute the mean vector m_N and the covariance matrix S_N for the posterior distribution $p(w|t) = N(w|m_N, S_N)$
2. Given prior $p(w) = N(w|m_0, S_0^{-1} = 10^{-6}I)$.
3. The precision of likelihood function $p(t|w, \beta)$ or $p(t|x, w, \beta)$ is chosen to be $\beta = 1$.

My discussion

利用PRML 3.53, 3.54的公式：

$$m_N = \beta S_N \Phi^T t$$
$$S_N^{-1} = \alpha I + \beta \Phi^T \Phi$$

可求得posterior distribution中的m_N與S_N

```
data size 10
    mean [[   1.33354573]
 [  68.95680391]
 [-190.81664709]
 [ 131.24295547]
 [  -7.0662442 ]
 [  -3.00138623]
 [  -3.41399052]]
    convariance [[ 4.39583892e+00 -2.15200635e+02  6.22402798e+02 -4.13163243e+02
   1.56805189e+00 -2.91080487e-03  1.10583537e-04]
 [-2.15200635e+02  2.33087080e+04 -6.81888571e+04  4.52668408e+04
  -1.71798982e+02  3.18967156e-01 -1.21284397e-02]
 [ 6.22402798e+02 -6.81888571e+04  1.99509967e+05 -1.32445791e+05
   5.03217835e+02 -9.81110382e-01  4.66435848e-02]
 [-4.13163243e+02  4.52668408e+04 -1.32445791e+05  8.80294460e+04
  -4.45708111e+02  1.02810441e+01 -2.27640225e+00]
 [ 1.56805189e+00 -1.71798982e+02  5.03217835e+02 -4.45708111e+02
   1.21477827e+02 -1.09163376e+01  2.58319242e+00]
 [-2.91080487e-03  3.18967156e-01 -9.81110382e-01  1.02810441e+01
  -1.09163376e+01  2.30915051e+00 -1.22515939e+00]
 [ 1.10583537e-04 -1.21284398e-02  4.66435849e-02 -2.27640225e+00
   2.58319242e+00 -1.22515939e+00  2.04763344e+00]]

data size 15
    mean [[ 2.3063326 ]
```

```
 [ 3.69656535]
 [-0.29364143]
 [ 2.85192188]
 [-4.58075213]
 [-3.38209747]
 [-3.35406962]]
    convariance [[ 8.78681888e-01 -9.31284890e-01  7.89920410e-02 -4.81470532e-02
   2.31421650e-02 -1.75240698e-03  4.41129639e-04]
 [-9.31284890e-01  1.76056486e+00 -1.35287174e+00  9.57382756e-01
  -4.61387872e-01  3.49391471e-02 -8.79521959e-03]
 [ 7.89920410e-02 -1.35287174e+00  3.17165467e+00 -3.50700670e+00
   1.71162795e+00 -1.29640593e-01  3.26357785e-02]
 [-4.81470532e-02  9.57382756e-01 -3.50700670e+00  8.08640612e+00
  -5.84119684e+00  4.46673082e-01 -1.12740437e-01]
 [ 2.31421650e-02 -4.61387872e-01  1.71162795e+00 -5.84119684e+00
   5.20341852e+00 -8.49579977e-01  2.57109576e-01]
 [-1.75240698e-03  3.49391471e-02 -1.29640593e-01  4.46673082e-01
  -8.49579977e-01  1.32981585e+00 -1.01227910e+00]
 [ 4.41129639e-04 -8.79521959e-03  3.26357785e-02 -1.12740437e-01
   2.57109576e-01 -1.01227910e+00  1.99970790e+00]]

data size 30
    mean [[ 1.93108744]
 [ 4.06437784]
 [ 0.06453237]
 [ 2.16229663]
 [-4.20122193]
 [-3.36133565]
 [-3.20804399]]
    convariance [[ 3.12168567e-01 -3.45214136e-01  5.21914205e-02 -3.54372402e-02
   1.73048924e-02 -1.14607354e-03  1.84124842e-04]
 [-3.45214136e-01  7.83987075e-01 -7.36094039e-01  5.50877094e-01
  -2.69330314e-01  1.78376703e-02 -2.86576105e-03]
 [ 5.21914205e-02 -7.36094039e-01  1.98755476e+00 -2.43224484e+00
   1.19881888e+00 -7.94124917e-02  1.27587687e-02]
 [-3.54372402e-02  5.50877094e-01 -2.43224484e+00  4.83028429e+00
  -3.09741517e+00  2.08089314e-01 -3.35519385e-02]
 [ 1.73048924e-02 -2.69330314e-01  1.19881888e+00 -3.09741517e+00
   2.61492787e+00 -5.38758187e-01  1.04225340e-01]
 [-1.14607354e-03  1.78376703e-02 -7.94124917e-02  2.08089314e-01
  -5.38758187e-01  6.42278420e-01 -3.61686068e-01]
 [ 1.84124842e-04 -2.86576105e-03  1.27587687e-02 -3.35519385e-02
   1.04225340e-01 -3.61686068e-01  8.58580921e-01]]

data size 50
    mean [[ 1.86868425]
 [ 3.91512444]
 [ 1.20853857]
 [-0.06959989]
 [-2.96267747]
 [-3.06244255]
 [-3.69748525]]
    convariance [[ 2.45574850e-01 -2.89642773e-01  5.62354601e-02 -2.04176294e-02
```

```
   9.03109464e-03 -9.12777690e-04  1.58772441e-04]
  [-2.89642773e-01  5.94480776e-01 -3.96015997e-01  1.53222995e-01
   -6.79189593e-02  6.86505853e-03 -1.19414697e-03]
  [ 5.62354601e-02 -3.96015997e-01  6.82640351e-01 -5.87233862e-01
    2.67532168e-01 -2.70660405e-02  4.70853975e-03]
  [-2.04176294e-02  1.53222995e-01 -5.87233862e-01  1.32406361e+00
   -9.54362000e-01  9.90809457e-02 -1.72957421e-02]
  [ 9.03109464e-03 -6.79189593e-02  2.67532168e-01 -9.54362000e-01
    1.07764604e+00 -3.95568127e-01  7.68098022e-02]
  [-9.12777690e-04  6.86505853e-03 -2.70660405e-02  9.90809457e-02
   -3.95568127e-01  5.24345328e-01 -2.52985537e-01]
  [ 1.58772441e-04 -1.19414697e-03  4.70853975e-03 -1.72957421e-02
    7.68098022e-02 -2.52985537e-01  3.67348372e-01]]

 data size 80
     mean [[ 2.20533623]
  [ 3.54732611]
  [ 1.17076039]
  [-0.21103937]
  [-2.56859254]
  [-3.20503753]
  [-3.68037447]]
     convariance [[ 1.36042477e-01 -1.72884093e-01  4.81504656e-02 -1.44087520e-02
   3.95097251e-03 -1.01664259e-03  1.91771151e-04]
  [-1.72884093e-01  3.95739750e-01 -2.94659650e-01  9.15191412e-02
   -2.51283867e-02  6.46637114e-03 -1.21976763e-03]
  [ 4.81504656e-02 -2.94659650e-01  4.86874907e-01 -3.08771515e-01
    8.72284853e-02 -2.24849525e-02  4.24173450e-03]
  [-1.44087520e-02  9.15191412e-02 -3.08771515e-01  4.31627247e-01
   -2.58473716e-01  6.99229523e-02 -1.32219335e-02]
  [ 3.95097251e-03 -2.51283867e-02  8.72284853e-02 -2.58473716e-01
    4.43916114e-01 -3.02973593e-01  5.96545476e-02]
  [-1.01664259e-03  6.46637114e-03 -2.24849525e-02  6.99229523e-02
   -3.02973593e-01  4.30141042e-01 -2.10692174e-01]
  [ 1.91771151e-04 -1.21976763e-03  4.24173450e-03 -1.32219335e-02
    5.96545476e-02 -2.10692174e-01  2.72494667e-01]]
```

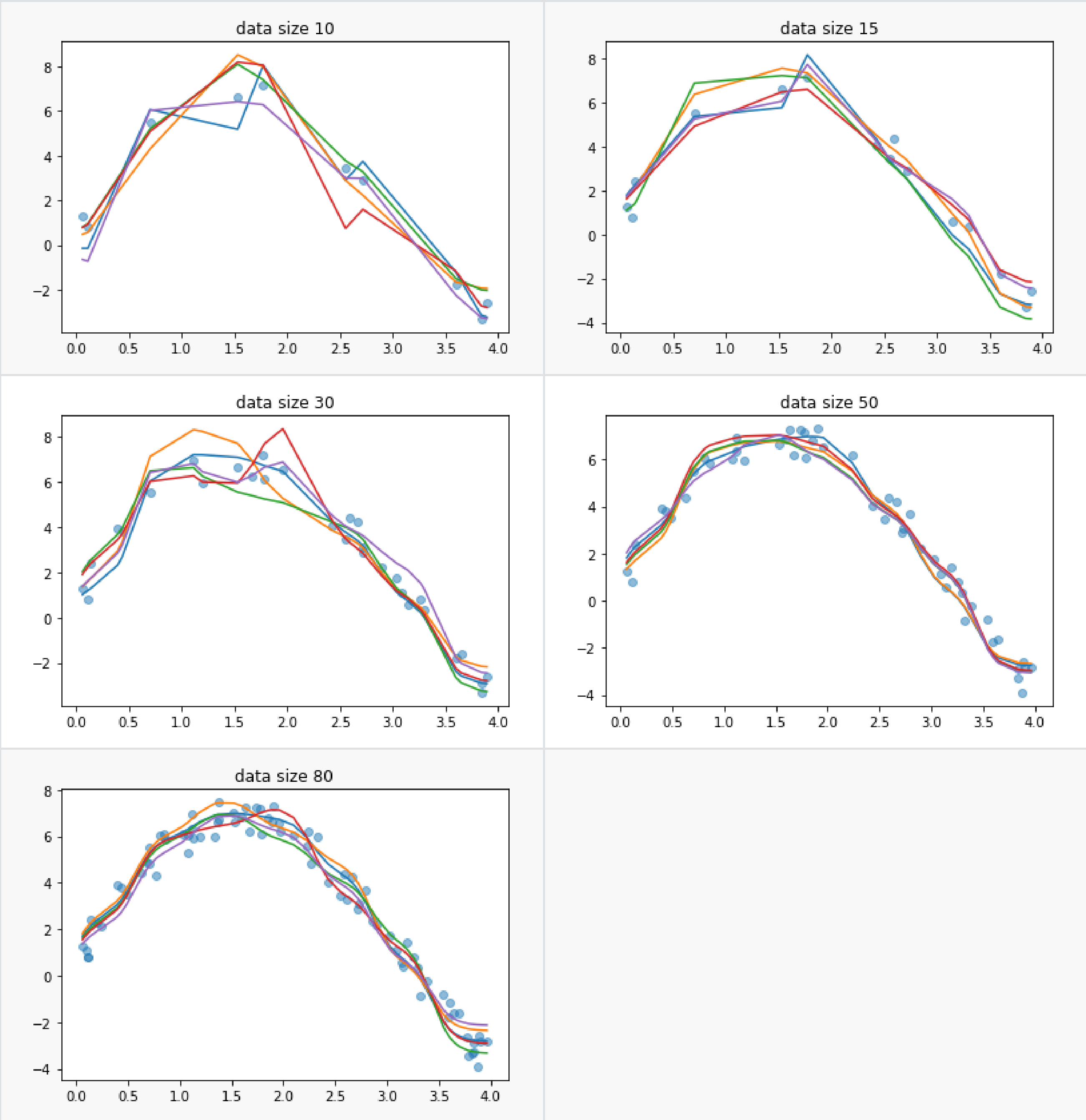
Problem 1-2

Similar to Fig. 3.9, please generate five curve samples from the parameter posterior distribution.

My discussion

由上一小題所得到的m_N和S_N所形成的高斯分佈，隨機抽樣五個w，並利用五個不一樣的w對dataset中的抽樣點算出對應的y，並畫出相應的curve。

另外，從以下五張圖，可見隨著dataset的大小增加，curve可以fit的越好，curve越圓滑。



Problem 1-3

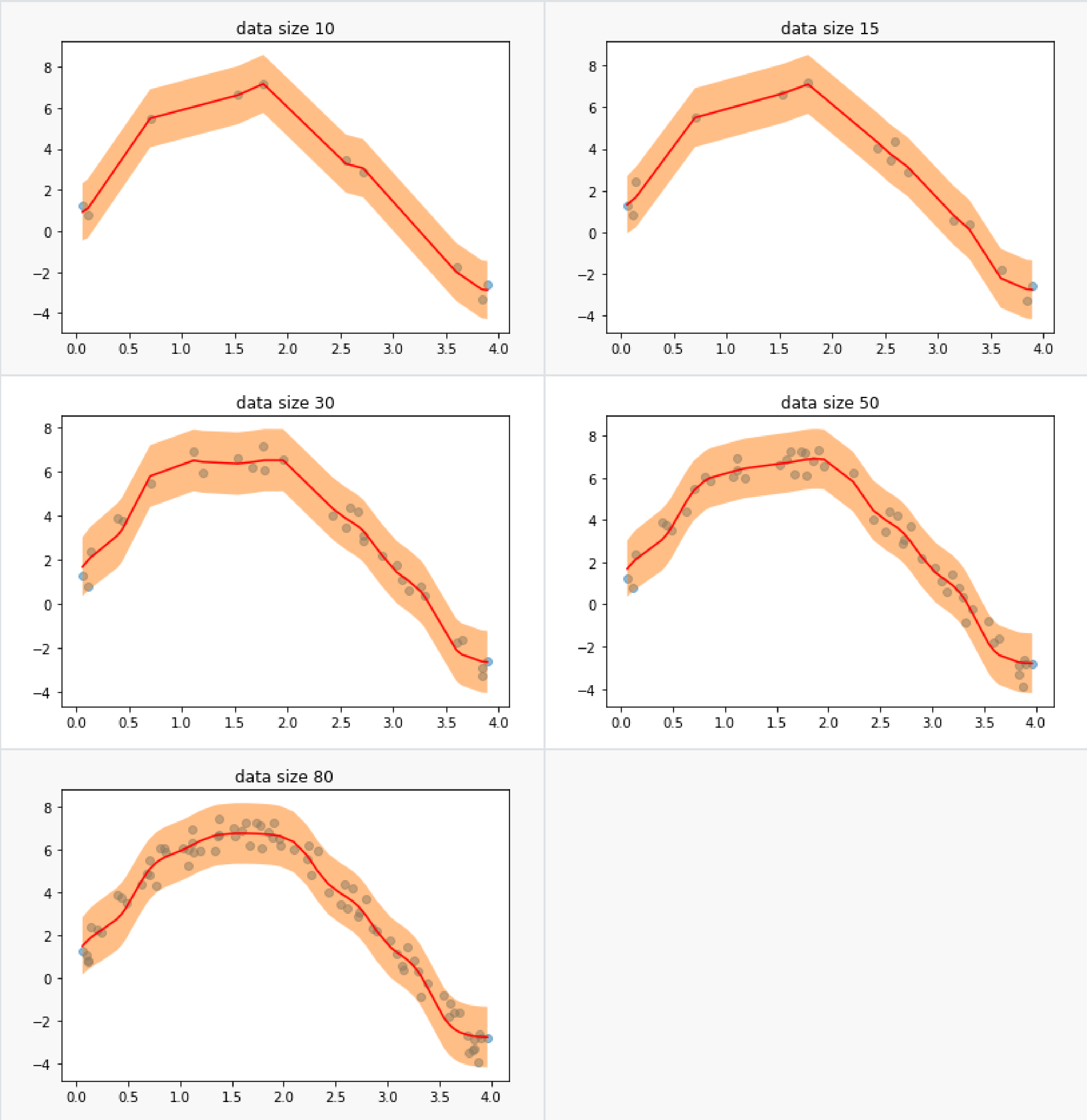
Similar to Fig. 3.8, please plot the predictive distribution of target value t and show the mean curve and the region of variance with one standard deviation on either side of the mean curve.

My discussion

利用PRML 3.58, 3.59的公式：

$$p(t|x, t, \alpha, \beta) = N(t|m_N^T \phi(x), \sigma_N^2(x))$$
$$\sigma_N^2(x) = \frac{1}{\beta} + \phi(x)^T S_N \phi(x) \beta$$

可求得predictive distribution中的mean與standard deviation，即可利用數值畫出圖形。



Problem 2 Logistic Regression

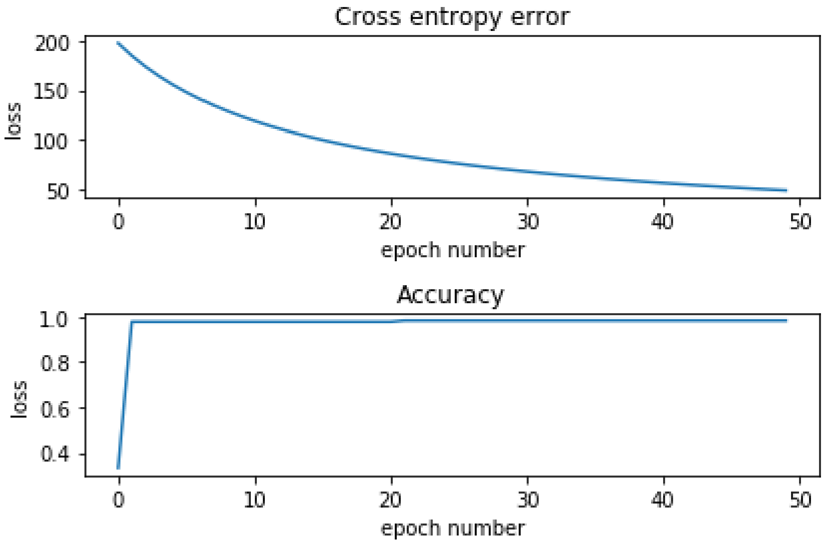
Problem 2-1

Set the initial w to be zero, and show the learning curve of E(w) and the accuracy of classification versus the number of epochs until convergence of training data.

My discussion

首先，我參照PRML section 4.4.4，實作Newton-Raphson algorithm，並額外為update項增加learning rate，讓model在訓練時，可以慢慢校正方向。

我除了設stopping criterion $E(w) < \epsilon$，我另外設了stop epoch=50，以防止始終由於達不到ϵ而無止盡的訓練下去。



Problem 2-2

Show the classification result of test data.

My prediction

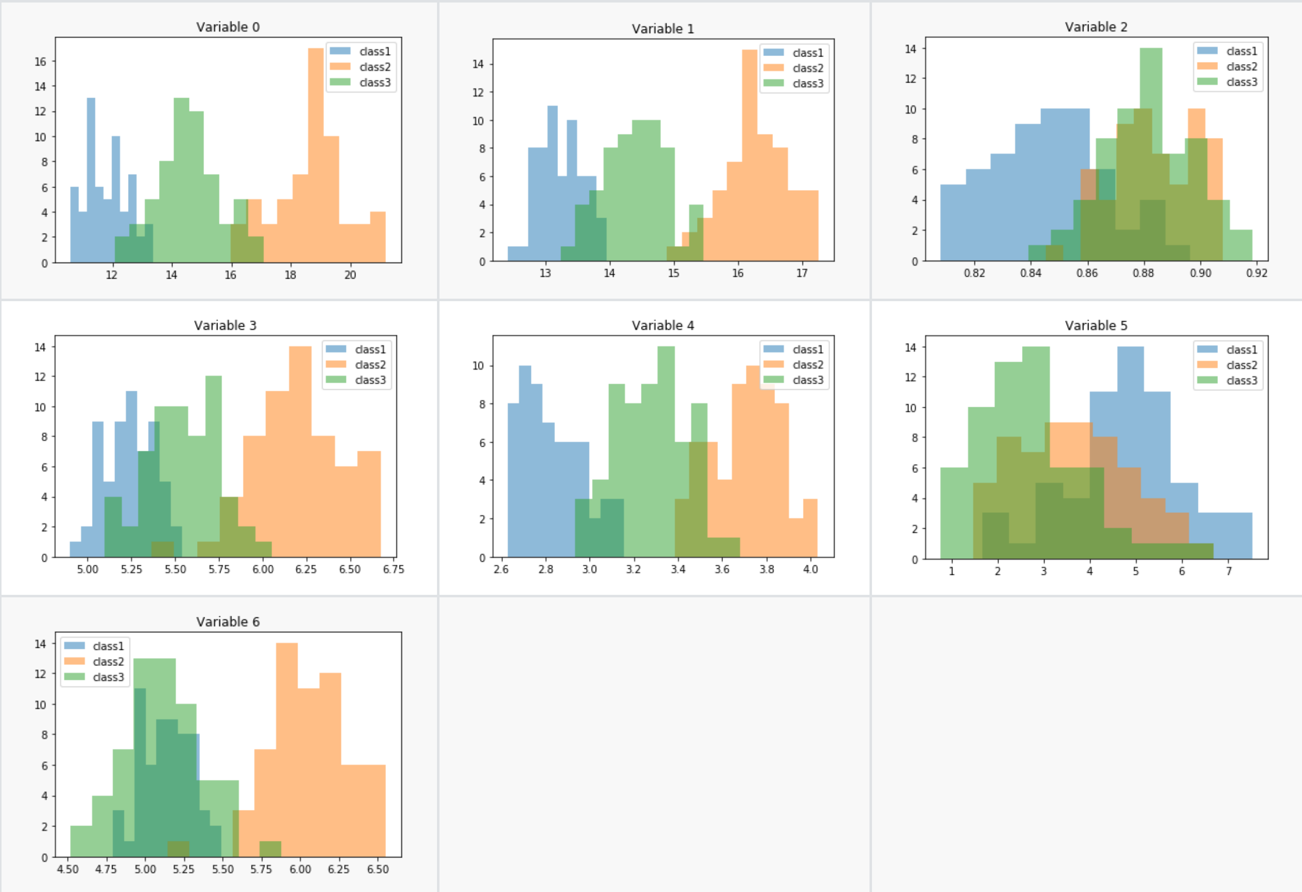
```
2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 2, 0, 0, 0, 0, 0, 0, 0,
0
```

Problem 2-3

Please plot the distribution (or histogram) of the variable in each dimension of training data and map different colors to each class.

My discussion

從training data畫出的 distribution (or histogram) of the variable，以我的直覺上，三個類別重疊（overlap）面積最小的分佈，應該是最有能力可以分類data的variable。所以，我認為variable 0和variable 1最有可能是很有貢獻的變數。



Problem 2-4

Explain that how do you know the model you trained is on the way to global minimum.

My discussion

從E(w)的learning curve可以看出，error持續下降，另外由於error function是convex，所以只要learning rate夠小，可以持續往global minimum converge。一開始，我只有設定stop_critirion的參數以及learning rate=0.01，但一旦訓練時超過global minimum，error就會由原本漸漸變小，但是達某次epoch後，error會往上升。所以可以看出我的model有往global minimum訓練。

另外，以對training data的分類正確率也能看出正確率達98%，但也有可能會產生overfit的現象。

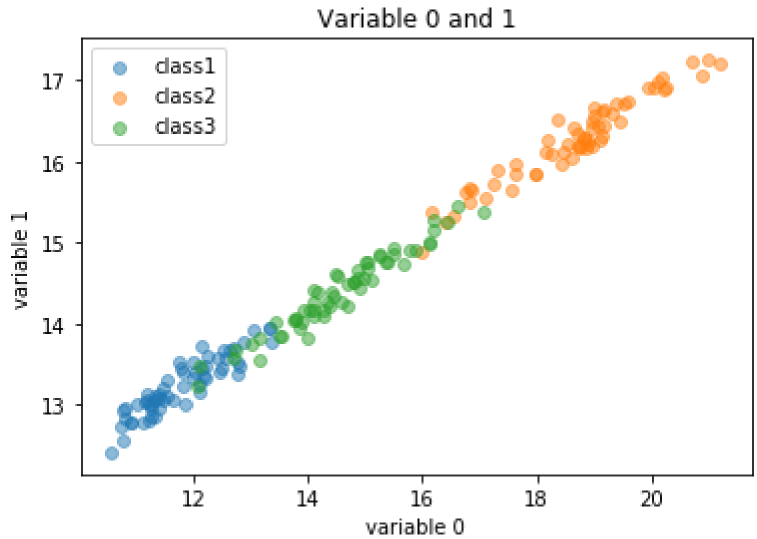
Problem 2-5

Please choose a pair of the most contributive variables and plot the samples in training data via 2D graph.

My discussion

我還是跑完所有的pair來看每一組的error，找出error最小的pair，結果也有與我在problem 2-3的猜想一樣是variable 0與variable 1。

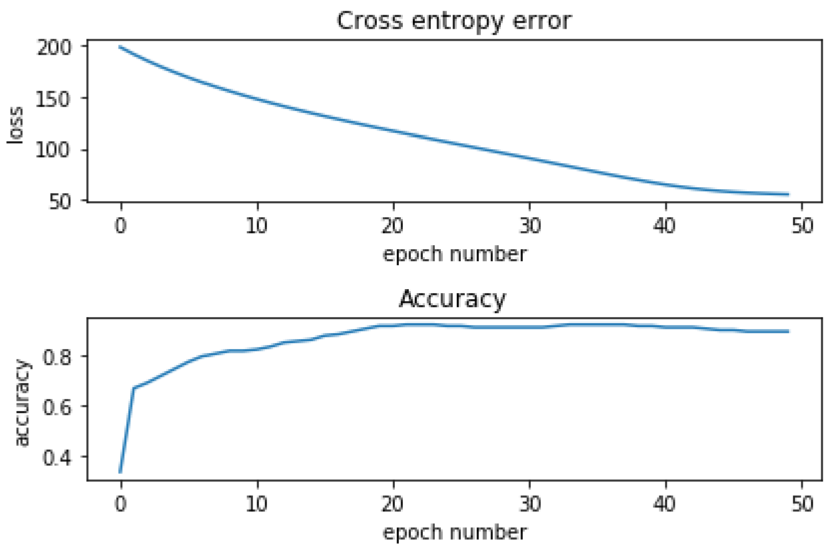
從下方圖，分別看variable 0與variable 1，皆可看出兩者對每個class皆可以明顯的分別出各個類別。



Problem 2-6

Use the variables you choose in (5) and redo (1) and (2).

My result



My prediction

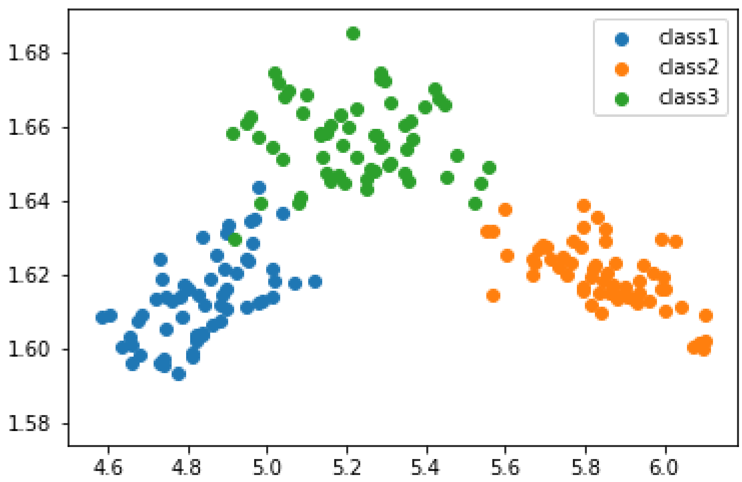
```
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1
```

Problem 2-7

Use the Fisher’s linear discriminant (or the linear discriminant analysis) in Section 4.1 to project the data on a two-dimensional (2D) space and plot the training samples in a 2D graph.

My discussion

參考PRML section 4.1.6以及http://goelhardik.github.io/2016/10/04/fishers-lda/之公式與解釋實作，將training data投影至2D空間。



Problem 3 Nonparametric Methods (Bonus Question)

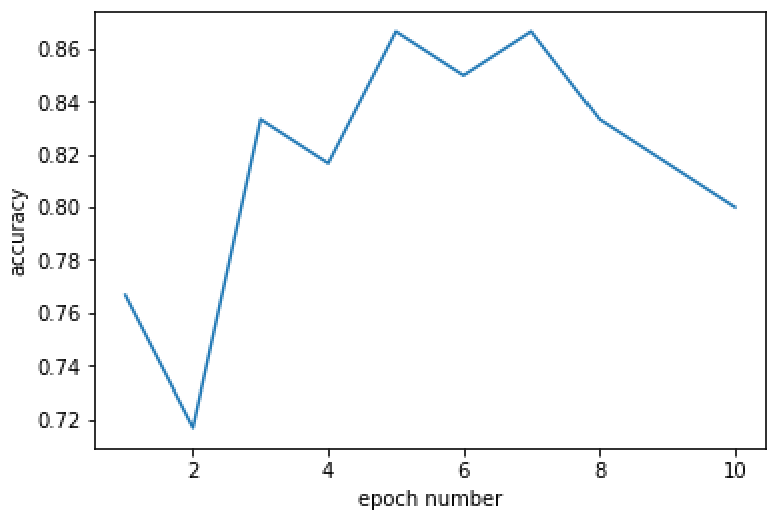
Problem 3-1

K-Nearest-Neighbor Classifier

My discussion

找最近的K個點，做類別的統計。

從下圖可看出K需要做適當的選擇才能有較好的分類結果，當K過大時，可能會涵蓋過多的其他類別進來，所以效果並不會好。



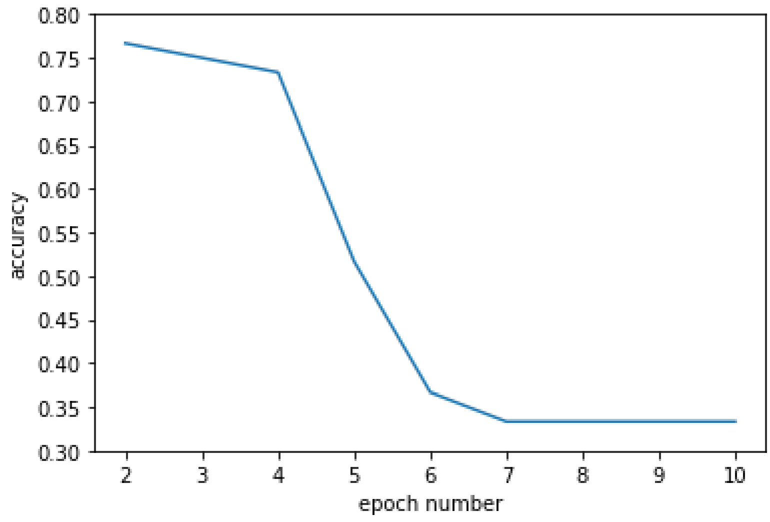
Problem 3-2

Fixing the distance and determining the K from training data

My discussion

找小於V距離的點，做類別統計。

從下圖可以看出當距離大超過一個程度，會沒有判斷類別的能力，因為可能將其他類別的point也包進來了。


0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
accuracy
2
3
4
5
6
7
8
9
10
epoch number