

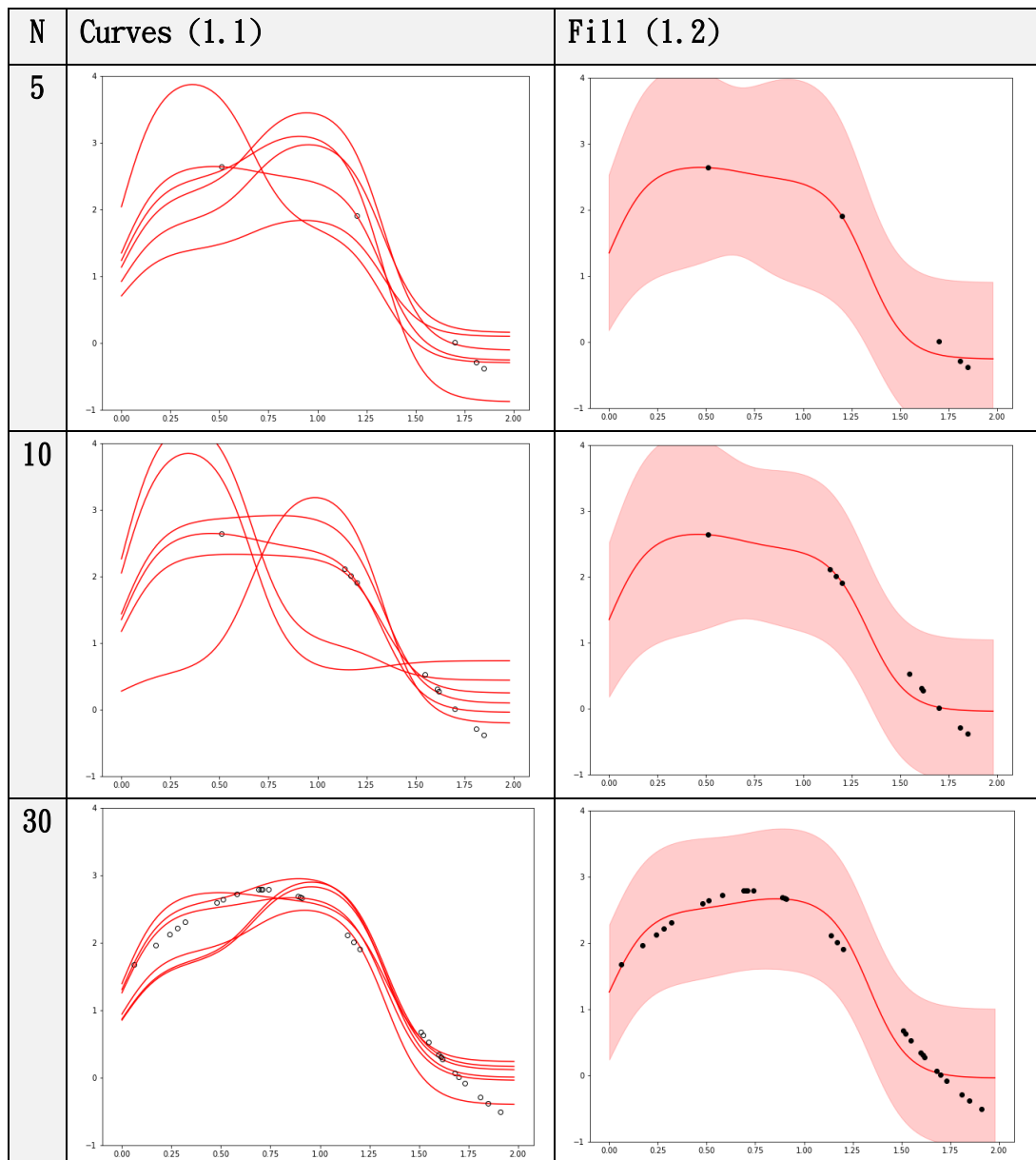
Machine Learning HW2

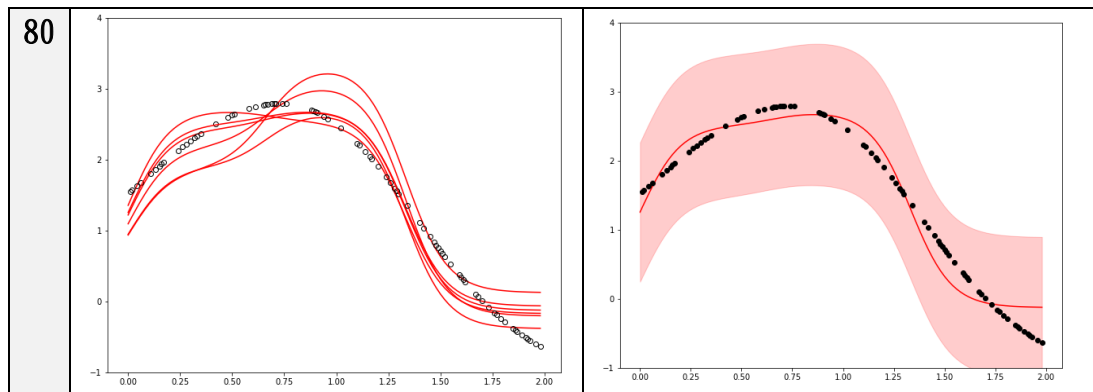
資管碩一 0853412 吳宛儒

1. Sequential Bayesian Learning

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

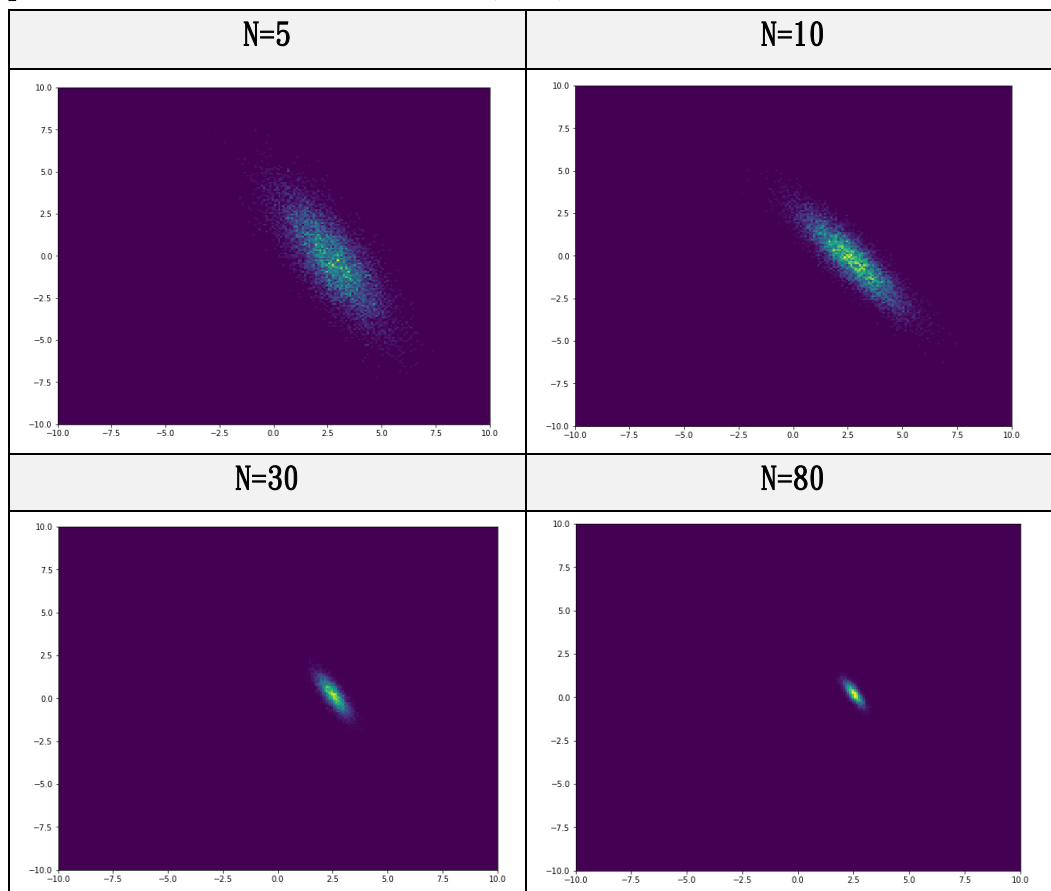
1.2 Similar to Fig. 3.8, please plot the predictive distribution of target value t and show the mean curve and the region of variance with one standard deviation on either side of the mean curve.





可以看出當 sample 越多點的時候，資料的分布就會越確定，寬帶就會較窄；也可以進一步查看右邊的寬帶圖上凡是有點經過的地方，寬帶較窄。寬帶之定義為加減一個標準差的寬度。

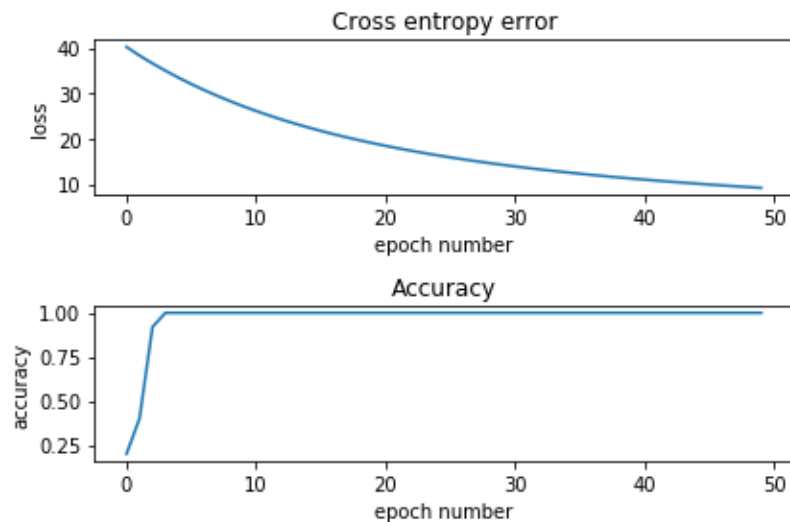
Addition similar to the middle column of Fig. 3.7, please arbitrarily select two weights by yourself and carefully plot the corresponding prior distributions when $N = 5; 10; 30$ and 80



可以看出同 1.2 所討論的結果，當 N 較大的時候，資料分布較為確定。

2. Logistic Regression

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. Gradient descent algorithm is applied.



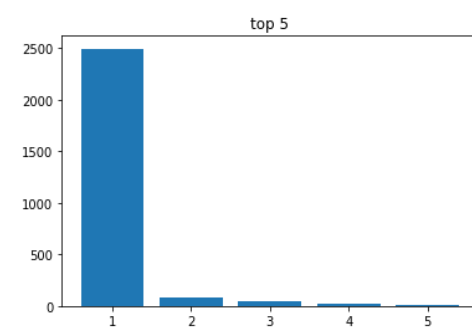
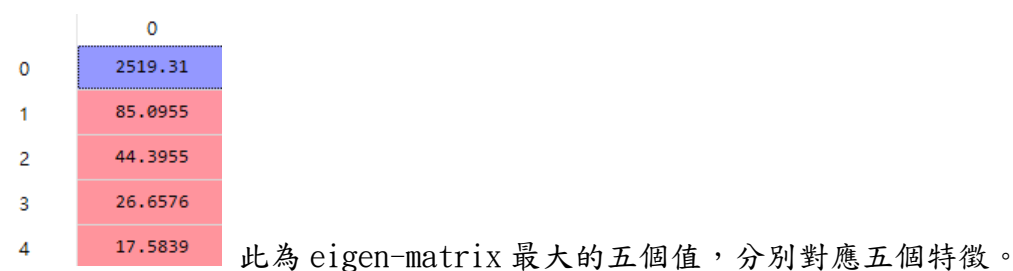
可以看到 error 在 50 個 epoch 後逐漸收斂，accuracy 則在 10epoch 以前就有明顯的提升。最終在 test data 上面也有全對的效果。

2.2 Show the classification result of test data.

```
In [6]: print(LM.predict(test_x))  
[0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4]
```

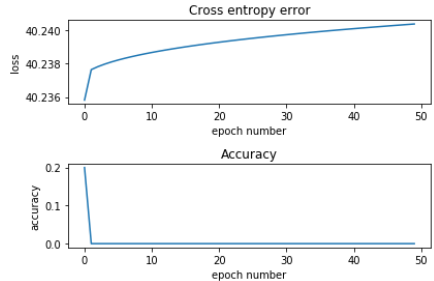
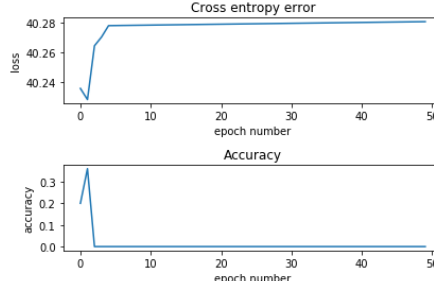
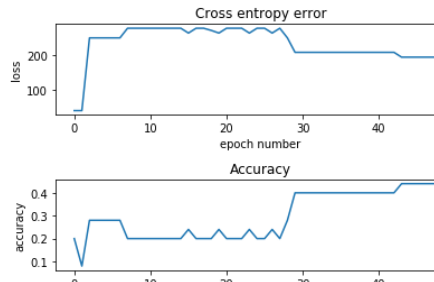
此為預測 test 之結果。

2.3 Use the principal component analysis (PCA) to reduce the dimension of data and plot five eigenvectors corresponding to top five eigenvalues.



將其畫圖呈現。

2.4 Repeat 1 and 2 by applying Netwon-Raphson algorithm. PCA should be used to reduce the dimension of face images to 2, 5 and 10. Make some discussion.

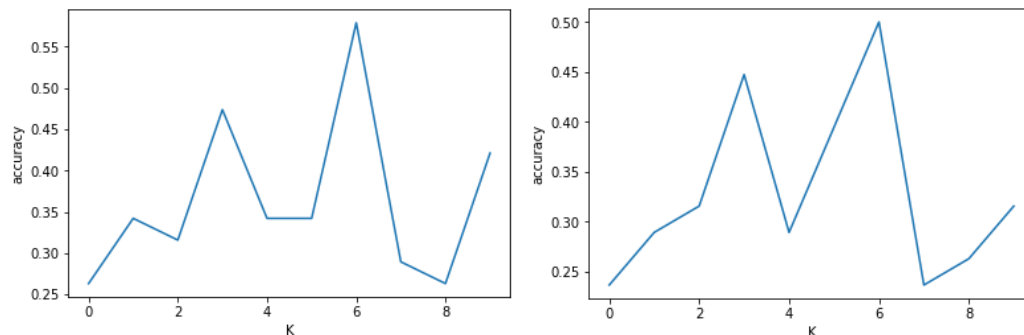
dim	figure	explanation
2		最後 error 直接飆升，accuracy 也在前幾個 epoch 就降到幾乎零的地方。效果非常差。
5		略比 PCA 取 2 的時候好，但是整體上來看效果一樣沒有明顯的起色。
10		Accuracy 比前面兩者好許多，也沒有下降到 0 的趨勢。Cross entropy error 不太確定為何反而飆升到 100 至 200 間，有待進一步討論。

2.5 Make some discussion on the results of Netwon-Raphson and gradient descent algorithms.

討論兩者差異 Netwon-Raphson and gradient descent algorithms，由實驗結果來看，經過 PCA 降維以後的數據在經過 Netwon-Raphson 方法以後，效果明顯比 gradient descent 方法差。推估是因為 PCA 降維過程將數據維度刪減，讓數據能考量的 features 變少，資訊同時變少，或甚至破壞原有數據架構，使得效果變差。

3. Nonparametric Methods

3.1 K-nearest-neighbor classifier

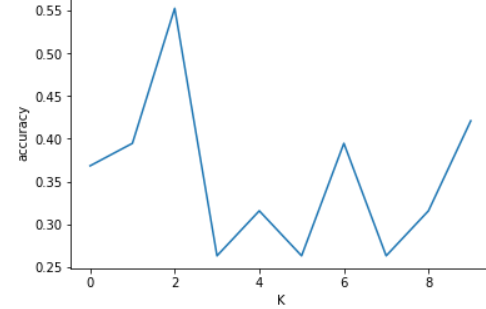
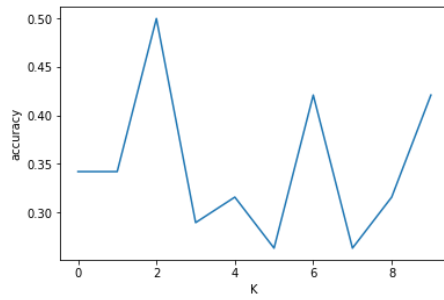


可以看出 K 越大不見得準確度會較好。左圖：當三個類別有兩個同票數時，取具有最小距離的那個類別(也就是最近的類別)；右圖：當三個類別有兩個同票數時，找第一個類別，其中排序為 Normal, Psychic, Water。原本預期左圖會比右圖好許多(因為較為合理)，但發現其實只有好一點點。

3.2 Please implement the principal component analysis (PCA) for training samples and reduce the dimension of training and test data to 7, 6, and 5 by using the bases obtained from PCA. Repeat the above procedure.

d	Figure(找第一個類別，上右圖)	Figure(最近的類別，上左圖)																																												
7	<table><caption>Data for d=7, First Class</caption><thead><tr><th>K</th><th>accuracy</th></tr></thead><tbody><tr><td>0</td><td>0.425</td></tr><tr><td>1</td><td>0.34</td></tr><tr><td>2</td><td>0.45</td></tr><tr><td>3</td><td>0.45</td></tr><tr><td>4</td><td>0.34</td></tr><tr><td>5</td><td>0.29</td></tr><tr><td>6</td><td>0.29</td></tr><tr><td>7</td><td>0.26</td></tr><tr><td>8</td><td>0.42</td></tr><tr><td>9</td><td>0.32</td></tr></tbody></table>	K	accuracy	0	0.425	1	0.34	2	0.45	3	0.45	4	0.34	5	0.29	6	0.29	7	0.26	8	0.42	9	0.32	<table><caption>Data for d=7, Nearest Class</caption><thead><tr><th>K</th><th>accuracy</th></tr></thead><tbody><tr><td>0</td><td>0.37</td></tr><tr><td>1</td><td>0.47</td></tr><tr><td>2</td><td>0.39</td></tr><tr><td>3</td><td>0.39</td></tr><tr><td>4</td><td>0.29</td></tr><tr><td>5</td><td>0.29</td></tr><tr><td>6</td><td>0.29</td></tr><tr><td>7</td><td>0.29</td></tr><tr><td>8</td><td>0.42</td></tr><tr><td>9</td><td>0.34</td></tr></tbody></table>	K	accuracy	0	0.37	1	0.47	2	0.39	3	0.39	4	0.29	5	0.29	6	0.29	7	0.29	8	0.42	9	0.34
K	accuracy																																													
0	0.425																																													
1	0.34																																													
2	0.45																																													
3	0.45																																													
4	0.34																																													
5	0.29																																													
6	0.29																																													
7	0.26																																													
8	0.42																																													
9	0.32																																													
K	accuracy																																													
0	0.37																																													
1	0.47																																													
2	0.39																																													
3	0.39																																													
4	0.29																																													
5	0.29																																													
6	0.29																																													
7	0.29																																													
8	0.42																																													
9	0.34																																													
6	<table><caption>Data for d=6, First Class</caption><thead><tr><th>K</th><th>accuracy</th></tr></thead><tbody><tr><td>0</td><td>0.39</td></tr><tr><td>1</td><td>0.29</td></tr><tr><td>2</td><td>0.41</td></tr><tr><td>3</td><td>0.37</td></tr><tr><td>4</td><td>0.26</td></tr><tr><td>5</td><td>0.26</td></tr><tr><td>6</td><td>0.39</td></tr><tr><td>7</td><td>0.29</td></tr><tr><td>8</td><td>0.39</td></tr><tr><td>9</td><td>0.34</td></tr></tbody></table>	K	accuracy	0	0.39	1	0.29	2	0.41	3	0.37	4	0.26	5	0.26	6	0.39	7	0.29	8	0.39	9	0.34	<table><caption>Data for d=6, Nearest Class</caption><thead><tr><th>K</th><th>accuracy</th></tr></thead><tbody><tr><td>0</td><td>0.37</td></tr><tr><td>1</td><td>0.34</td></tr><tr><td>2</td><td>0.39</td></tr><tr><td>3</td><td>0.37</td></tr><tr><td>4</td><td>0.32</td></tr><tr><td>5</td><td>0.26</td></tr><tr><td>6</td><td>0.39</td></tr><tr><td>7</td><td>0.26</td></tr><tr><td>8</td><td>0.41</td></tr><tr><td>9</td><td>0.37</td></tr></tbody></table>	K	accuracy	0	0.37	1	0.34	2	0.39	3	0.37	4	0.32	5	0.26	6	0.39	7	0.26	8	0.41	9	0.37
K	accuracy																																													
0	0.39																																													
1	0.29																																													
2	0.41																																													
3	0.37																																													
4	0.26																																													
5	0.26																																													
6	0.39																																													
7	0.29																																													
8	0.39																																													
9	0.34																																													
K	accuracy																																													
0	0.37																																													
1	0.34																																													
2	0.39																																													
3	0.37																																													
4	0.32																																													
5	0.26																																													
6	0.39																																													
7	0.26																																													
8	0.41																																													
9	0.37																																													

5



原本以為 pca 取 7 的時候可能會因為資訊較為充足，便一定會有平均較高的準確度，但其實發現取 5 的時候效果最好，推論是當取超過 5 的時候，因為有相關性極低的特徵被納入，才會引發這樣的結果。