



Machine Learning (Homework 2)



Due date : 11/30

1 Bayesian Linear Regression (40%)

- Please compute the mean vector \mathbf{m}_N and the covariance matrix \mathbf{S}_N for the posterior distribution $p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$ with the given prior $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0 = 0, \mathbf{S}_0^{-1} = 10^{-6}\mathbf{I})$. The precision of likelihood function $p(\mathbf{t}|\mathbf{w}, \beta)$ or $p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)$ is chosen to be $\beta = 1$.

SN - NumPy array							mN - NumPy array	
	0	1	2	3	4	5	6	0
0	0.135212	-0.166162	0.0394506	-0.0108146	0.00288083	-0.000700566	0.000154005	2.19027
1	-0.166162	0.337971	-0.221884	0.0637437	-0.0170121	0.00413745	-0.000909539	3.58776
2	0.0394506	-0.221884	0.358293	-0.225888	0.0622915	-0.0151776	0.00333683	1.04987
3	-0.0108146	0.0637437	-0.225888	0.326348	-0.193189	0.0492857	-0.0108627	-0.0429693
4	0.00288083	-0.0170121	0.0622915	-0.193189	0.304233	-0.199326	0.0459618	-2.9204
5	-0.000700566	0.00413745	-0.0151776	0.0492857	-0.199326	0.308321	-0.169245	-2.76258
6	0.000154005	-0.000909539	0.00333683	-0.0108627	0.0459618	-0.169245	0.220465	-3.97906

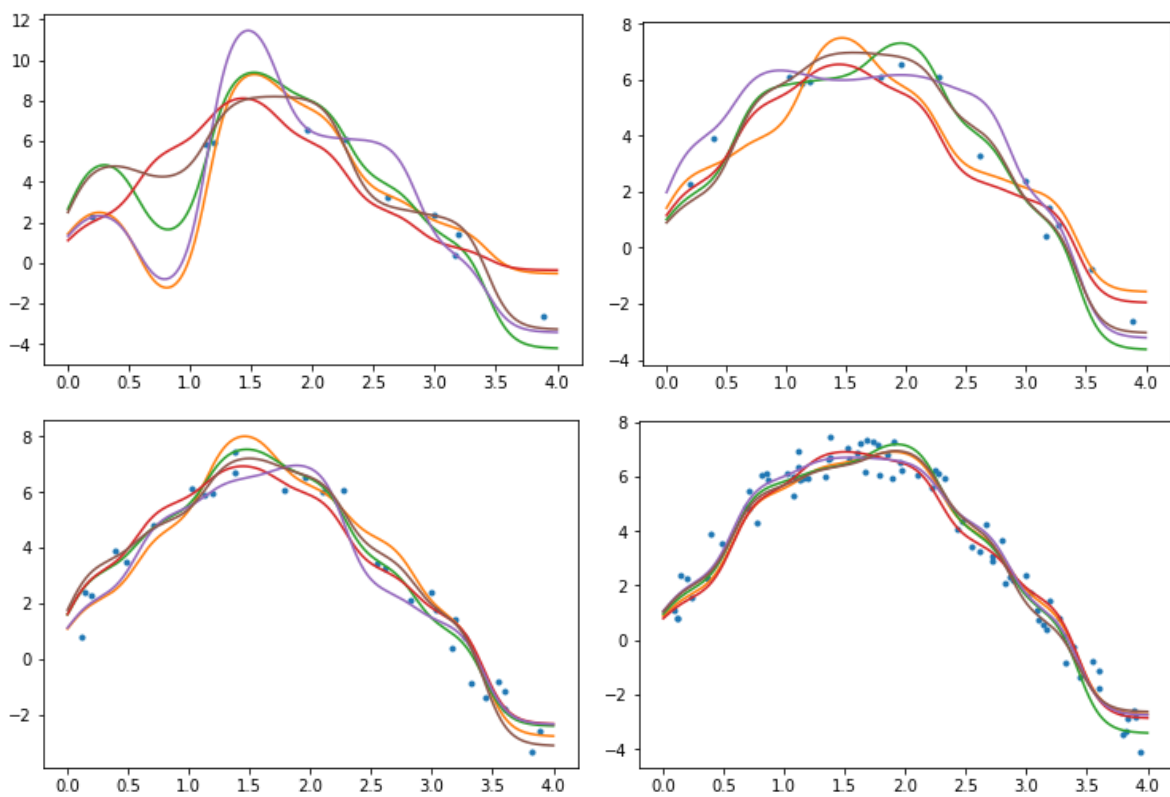
SN 和 mN 分別如圖所示，左半部分為 SN，右半部分為 mN。

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

根據 N 的不同($N=10,15,30,80$)所得到的五條曲線圖分別如下頁所示：

圖片依序分別是，左上 $N=10$ ，右上 $N=15$ ，左下 $N=30$ ，右下 $N=80$ 。

N 代表的是資料筆數的多寡，隨著資料筆數越多，圖形越趨於平緩，五條曲線的差異也降得越低（五條曲線變得比較密合），像一開始只有十筆資料時（左上圖），隨機抽出的五個 Weight 差異非常大，所以五條線條非常散亂，但是當資料筆數逐漸增加到 80 筆的時候（右下圖），可以看到 Weight 之間差異比較不那麼大，整體曲線起伏也比較平緩。

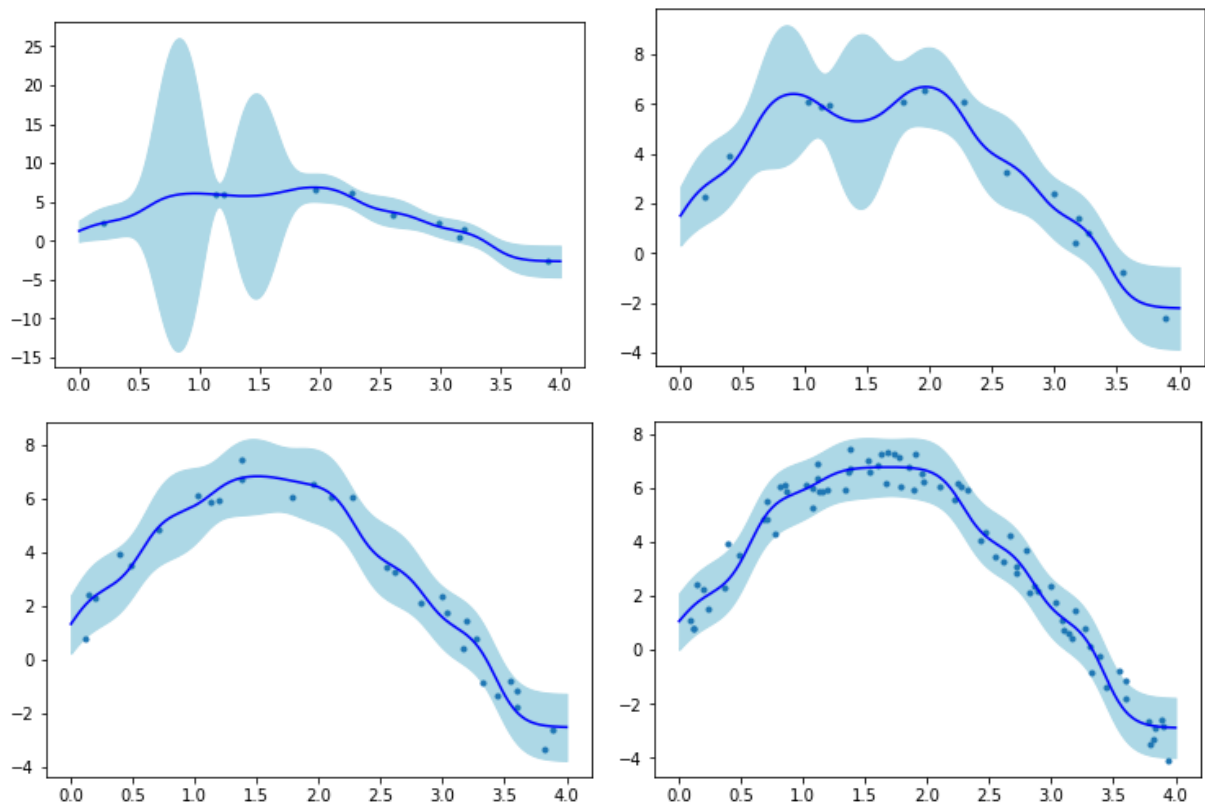


3. Similar to Fig. 3.8, please plot the predictive distribution of target value t and show the mean curve and the region of variance with one standard deviation on either side of the mean curve.

根據 N 的不同($N=10,15,30,80$)，所得到的五條曲線由 mean 加減一個標準差所涵蓋的區域為淺藍色的部分，而中間藍色線段為這五條曲線曲平均：

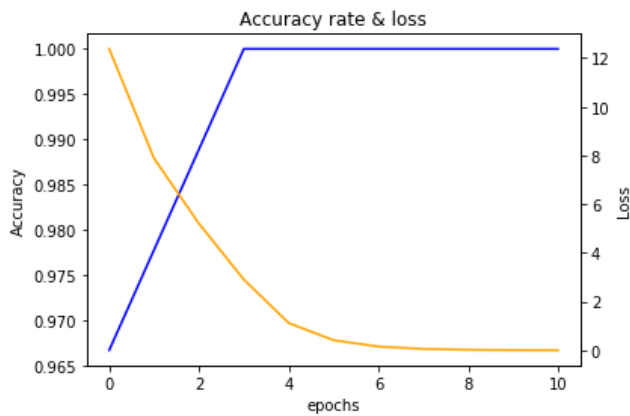
圖片依序分別是，左上 $N=10$ ，右上 $N=15$ ，左下 $N=30$ ，右下 $N=80$ 。

從這四張圖片可以很清楚看到剛剛 Q1-2 裡所寫的現象，隨著資料筆數增加，Weight 變得比較平穩以後圖形裡面淺藍色區塊也減少。可以看到其實資料都會落於淺藍色區段之內。



2 Logistic Regression (60%)

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



左圖中橘色部分為 Loss，而藍色部分為 Accuracy。可以看到隨 Loss 下降，Accuracy 則往上竄升，直到 Accuracy 達到 1，而 Loss 趨近於零。

2. Show the classification result of test data.

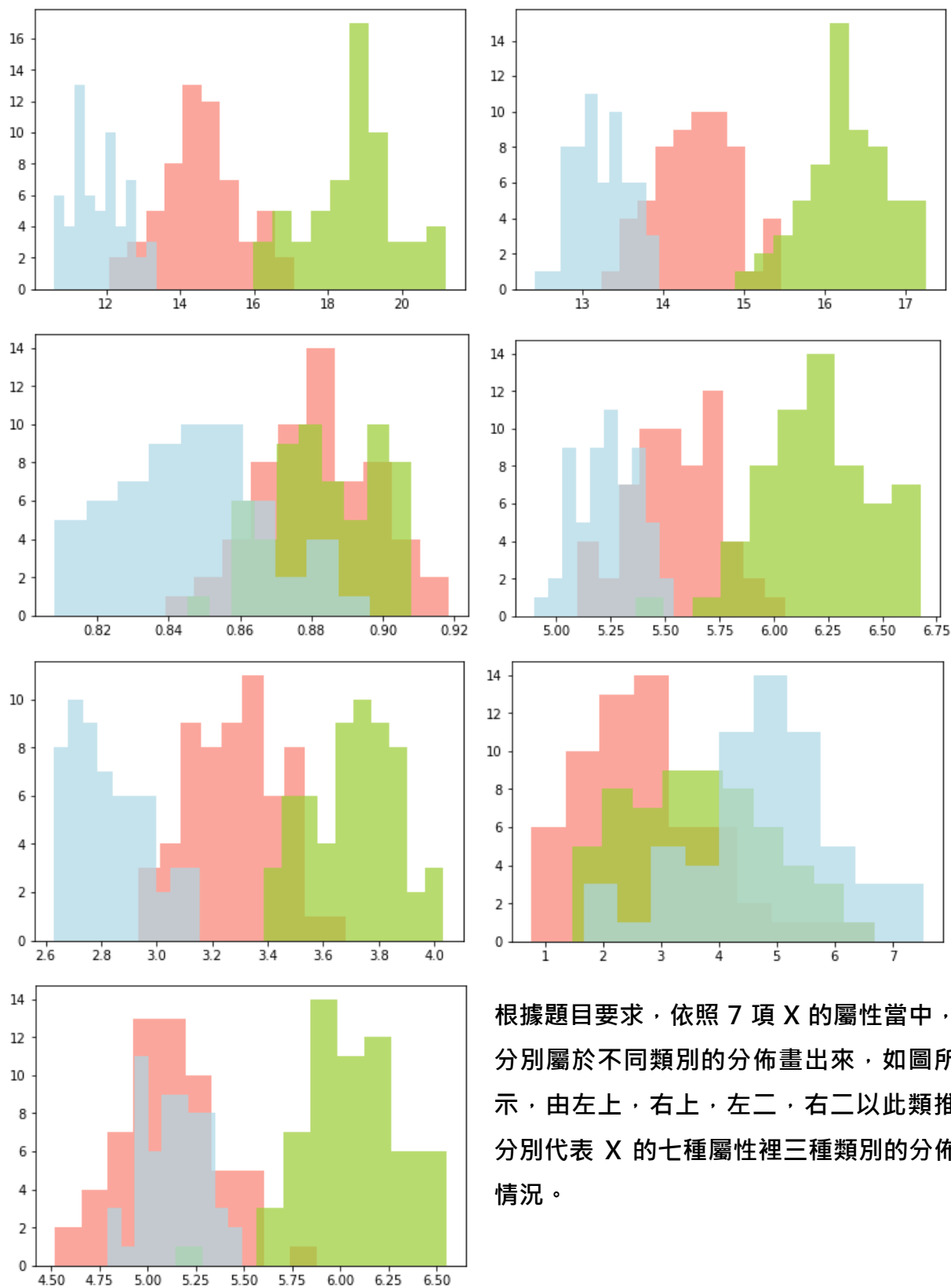
predict - NumPy array

	0	1	2
0	0	0	1
1	1	0	0
2	0	0	1
3	0	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	0	0	1
8	0	0	1
9	0	0	1
10	0	1	0
11	0	1	0
12	1	0	0
13	0	1	0
14	1	0	0
15	1	0	0
16	0	1	0
17	0	1	0
18	0	1	0
19	1	0	0
20	1	0	0
21	0	0	1
22	1	0	0
23	1	0	0
24	1	0	0
25	1	0	0
26	1	0	0
27	1	0	0
28	1	0	0
29	1	0	0

```
In [9]: print(predict)
[[0. 0. 1.]
 [1. 0. 0.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 0. 1.]
 [0. 1. 0.]
 [0. 1. 0.]
 [1. 0. 0.]
 [0. 1. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [0. 1. 0.]
 [0. 1. 0.]
 [0. 1. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [0. 0. 1.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]
 [1. 0. 0.]]
```

左圖中為 test 資料放下去以後依照前意題中以訓練好的 $\text{Weight}_{\text{new}}$ 所估計出來的分類情形。

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

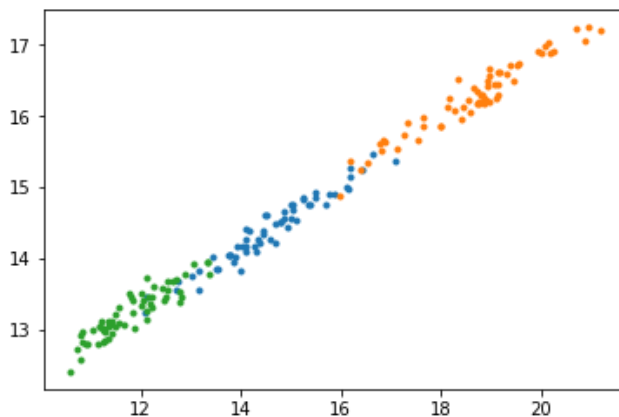


根據題目要求，依照 7 項 X 的屬性當中，分別屬於不同類別的分佈畫出來，如圖所示，由左上，右上，左二，右二以此類推，分別代表 X 的七種屬性裡三種類別的分佈情況。

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

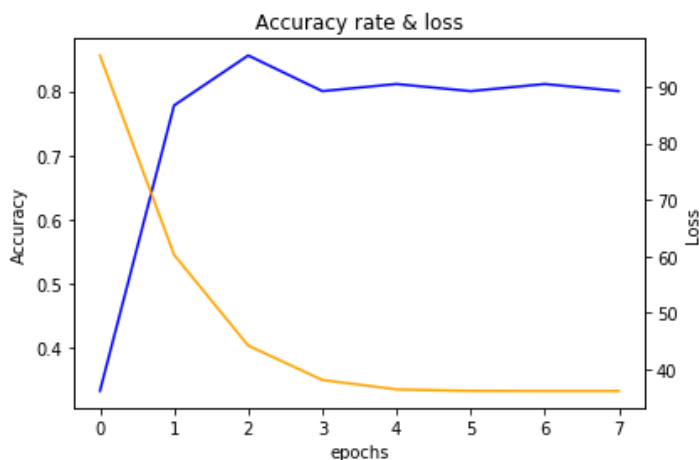
原因是因為本題的 Curve 是一個 Convex 圖形，因此可以透過這支程式當中的 training 過程逐步找到 global minimum。

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



根據題目要求繪製出最有貢獻度的 X 屬性，能夠讓資料分得最開的屬性。依照第三小題當中的類別長條圖分佈中，可以看到屬性 1 和 2 是可以讓資料分佈得最分散的屬性，也就是說能夠讓長條圖們重疊較少的。因此依照屬性 1 和 2 所畫出的類別分佈圖如左圖所示。

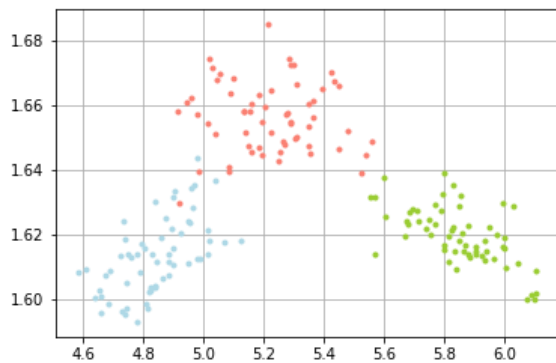
6. Use the variables you choose in (5) and redo (1) and (2).



根據第三小題和第五小題的圖形，推斷出屬性 1 和屬性 2 是有貢獻度的分類依據。因此憑藉這兩個屬性來對資料做分類，可以看到在七個 epoch 內依然達到收斂。雖然 Loss 收斂了，但是 Accuracy 卻沒有一開始那麼好了。從上一小題的圖片當中可以看到，分類之間還是有些微重疊，因此 Accuracy 較第一小題下降一些也屬於正常。

一小題下降一些也屬於正常。

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.



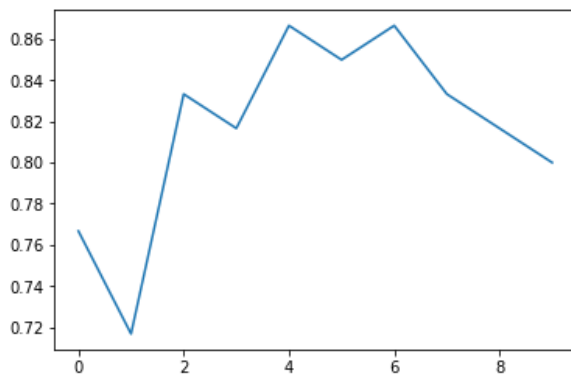
左圖為本題所要求之圖片，依照

Fisher's linear discriminant

所繪製出來的結果。

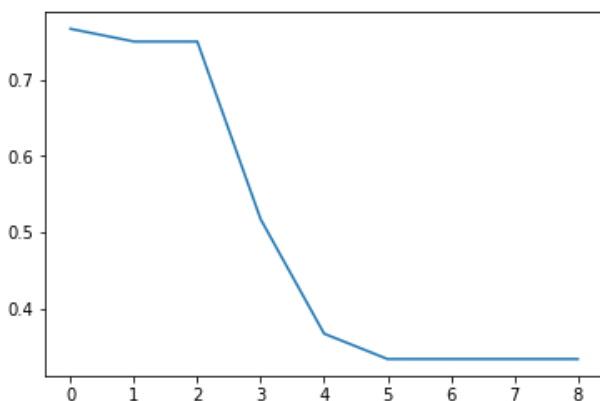
3 Nonparametric Methods (Bonus Question 30%)

1. K -Nearest-Neighbor Classifier



依照 KNN，讓 K 由 1-10 做分類所產生的 Accuracy，如左圖所示。

2. Alternative solution by fixing the distance and determining the K from training data



此圖是依照上一題的 KNN 進行修改，改成不是取前 K 個與自己相近的鄰居做投票決定分類，而是以和自身距離在 V ($V=2-10$) 之內的點來投票，而其結果的 Accuracy 如左圖所示。可以看到在第五個 epoch 之後 Accuracy 降得極低，原因是因為，在計算的時候，原始得 KNN 會取前 K 名與自己較近的

的鄰居做投票，但是在改版以後取與自己的距離小於 V 的都可以投票。觀察過每筆資料間的距離，最大就是 7.多，因此在 $V=5-10$ 的時候，幾乎所有樣本都可以投票，而每個分類出現的次數都是 50，因而導致最後幾次分類器都只會全部都猜同一個類別的狀況(如果使用 argmax 就會全部都猜 1)而導致正確率大約持平 0.333333。