EE655000 Machine Learning HW1

107060011 電機資訊學士班 涂皓鈞

Part 2. Computer assignment

3. The following shows the result of testing set with MAP classifier:

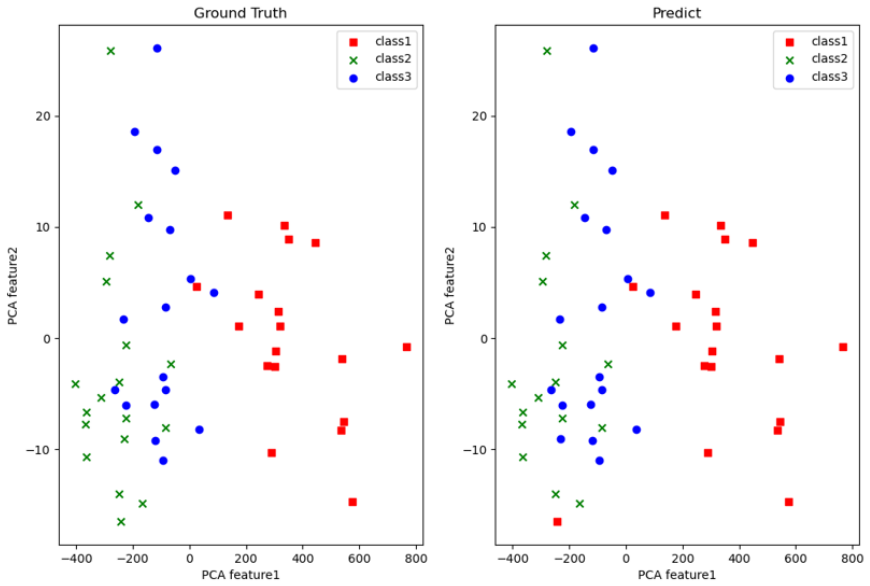

Fig. 1: 2D plots of the data distribution in testing set

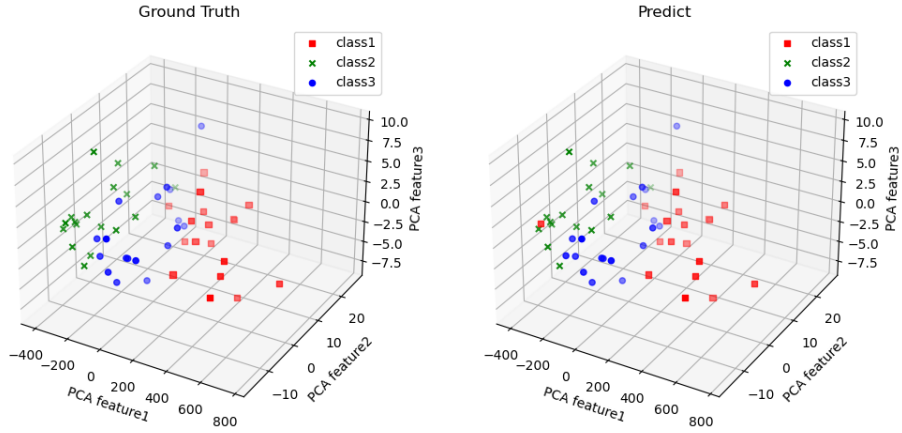

Fig. 2: 3D plots of the data distribution in testing set

```
In [1]: runfile('C:/Users/User/Desktop/ml/HW1/HW1.py', wdir='C:/Users/User/Desktop/ml/HW1')
Loading data...
Number of samples: 178
Split dataset.
Save train/test dataset.
Start MAP implementation.
------------------------------
Accuracy=0.9629629629629629
```

Fig. 3: Accuracy in testing set

4. The following table shows the accuracy among different combinations of label numbers in testing set with MAP classifier.

	Class 1	Class 2	Class 3	Accuracy
Combination 1	18	18	18	96.3%
Combination 2	18	20	16	96.3%
Combination 3	18	22	14	98.1%
Combination 4	18	24	12	98.1%
Combination 5	18	26	10	96.3%
Combination 6	18	28	8	96.3%
Combination 7	18	30	6	96.3%
Combination 8	18	32	4	94.4%

Table 1: Accuracy in testing set with different combinations of class numbers

Since total number of labels in this dataset is (class1, class2, class3) = (59, 71, 48), the setting of label numbers follow # of class 2 $\geq$ class 1 $\geq$ class 3, and the quantity of testing set is fixed to 54. From the result, **there is a small difference between different combinations of label numbers (i.e. different prior distribution)**. I think there are two main reasons. First, **13 features of the dataset may strongly follow Gaussian distribution**, inducing high accuracy even when the label number is small. Second, there are 13 features to predict the label. **It's likely that 13 features are sufficient to determine the correct label even in a small quantity of labels**. Next, the following shows the difference between # of features and the accuracy.

# of features	Accuracy
13	96.3%
12	96.3%
11	94.4%
10	94.4%
9	92.6%
8	92.6%
7	92.6%
6	88.9%
5	83.3%
4	81.5%
3	77.7%

2	75.9%
1	66.7%

Table 2: Accuracy in testing set with different # of features

From Table 2, we can observe that there is an accuracy drop when # of features is decreased. When we only use one feature as reference, the accuracy is 66.7%, **suggesting that the feature fit Gaussian distribution to some extent.** Hence, there is a stable accuracy change during the decreased # of features.

In conclusion, I think the prior distribution may take great effect on the prediction of the classifier. In real cases, the collected data may not always follow the specific distribution, so the effect of prior distribution will emerge. Hence, MAP classifier may achieve higher accuracy than MLE classifier since it takes prior distribution into account.