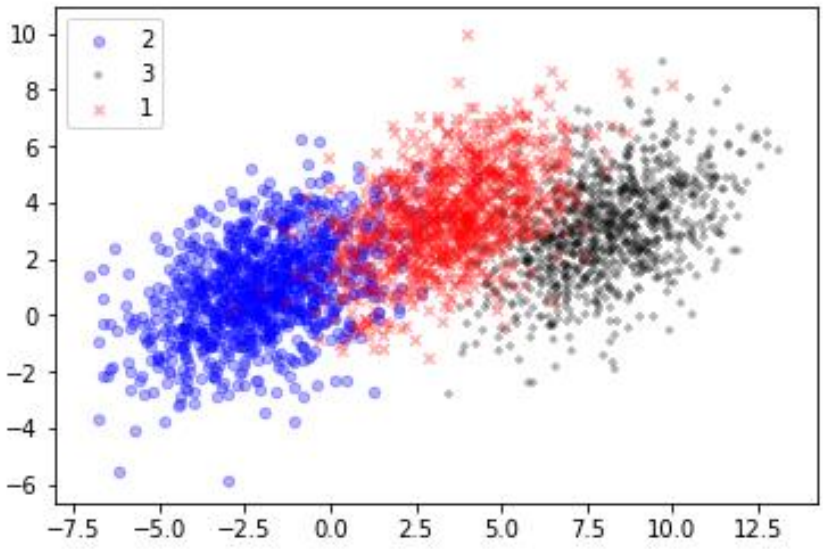
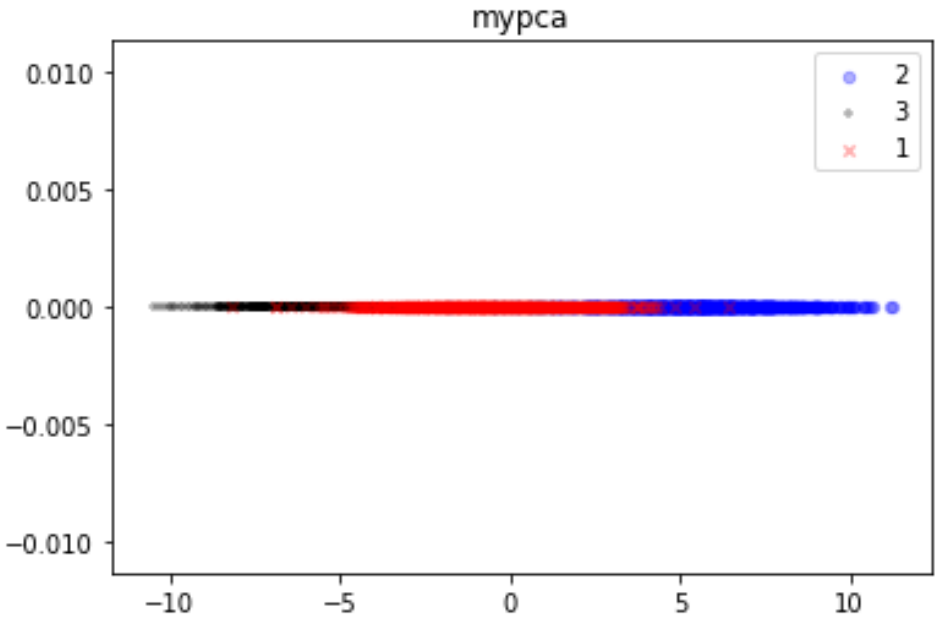
Part II

(a) Warming up. Plot the first 1000 samples of each category.



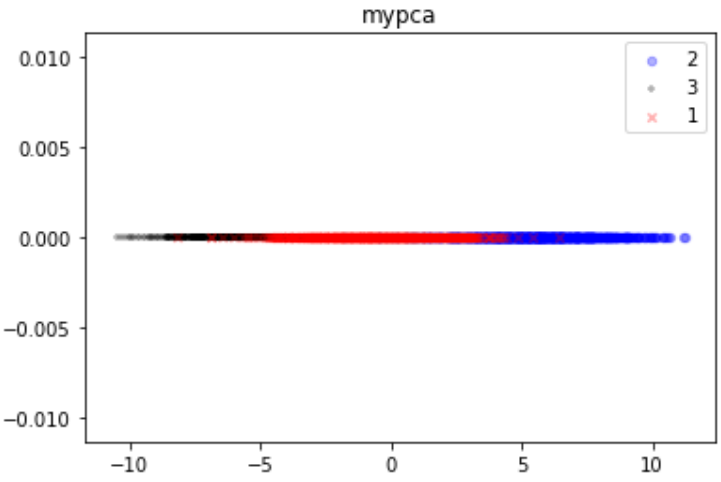
(b) Assume that the first 1000 samples of each category are training samples. We first perform dimension reduction to the training samples (i.e. from two dimensions to one dimension) with PCA method. Please plot the projected points of the training samples along the first PCA axis.



(c) Assume that the rest of the samples in each category are target samples requesting for classification. Please use PCA method and the nearest-neighbor classifier to classify these samples, and then compute the misclassification rate of each category.

First, I compare my PCA and sklearn's PCA, figure left the first 10 data. And they are absolutely the same. Figure right is the plot.

```
mypca [[ 0.80370801]
 [-1.22865829]
 [-0.61864528]
 [ 0.56751044]
 [ 0.27452158]
 [ 2.58828737]
 [-3.59424225]
 [ 1.65489044]
 [-0.59427054]
 [-0.61892269]]
skpca [[ 0.80370801]
 [-1.22865829]
 [-0.61864528]
 [ 0.56751044]
 [ 0.27452158]
 [ 2.58828737]
 [-3.59424225]
 [ 1.65489044]
 [-0.59427054]
 [-0.61892269]]
```



Then I compute the misclassification rate of each category. In this table, score is the number of right classifications for each category, and m_rate is the total misclassification rate.

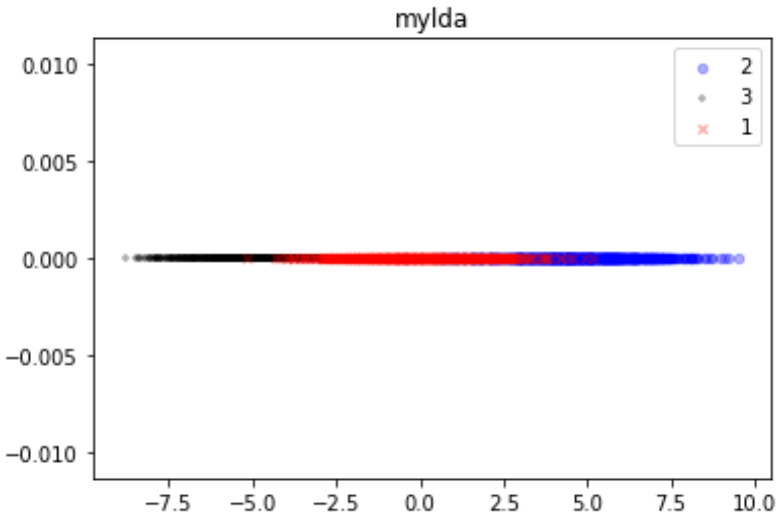
K	**1**	**3**	**5**	**7**	**9**	**11**	**13**	**15**
score1	**714**	**755**	**773**	**783**	**785**	**784**	**789**	**788**
score2	**907**	**908**	**909**	**905**	**917**	**914**	**903**	**910**
score3	**798**	**837**	**842**	**841**	**844**	**843**	**843**	**849**
m_rate	**0.193**	**0.166**	**0.158**	**0.157**	**0.151**	**0.153**	**0.155**	**0.151**

K	17	19	21	23	25	27	29	31
score1	803	814	813	815	813	802	789	790
score2	910	908	903	901	907	907	909	911
score3	852	844	856	852	854	855	859	857
m_rate	0.145	0.144	0.142	0.144	0.142	0.145	0.147	0.147

(d) Repeat (b) and (c) with FLD method

First, I compare my PCA and sklearn's PCA, figure left the first 10 data. Though they are not absolutely the same, the ratio of every data is the same. Figure right is the plot.

```
sklda [[ 1.29962975]
 [-0.44980795]
 [-0.4068433 ]
 [ 0.75624668]
 [-0.02416422]
 [ 1.90153379]
 [-0.70043047]
 [ 1.48362212]
 [-0.66780667]
 [-0.68864579]]
mylda [[ 2.04221262]
 [-0.70681936]
 [-0.63930556]
 [ 1.18835116]
 [-0.03797118]
 [ 2.98803279]
 [-1.10064266]
 [ 2.33133461]
 [-1.04937826]
 [-1.08212444]]
```



Then I compute the misclassification rate of each category. In this table, score is the number of right classifications for each category, and m_rate is the total misclassification rate. For this time, data of sklearn is not the same as mine, so I make table for both of them.

K	1	3	5	7	9	11	13	15
skscore1	790	840	845	857	861	875	878	875
skscore2	886	896	898	908	898	898	896	892
skscore3	899	922	922	926	924	927	928	922
sk_rate	0.142	0.114	0.112	0.103	0.105	0.100	0.099	0.104
myscore1	801	828	835	847	853	871	869	867
myscore2	880	897	892	901	901	897	897	897
myscore3	909	932	925	930	928	922	927	927
my_rate	0.137	0.114	0.116	0.107	0.106	0.103	0.102	0.103
K	17	19	21	23	25	27	29	31
skscore1	878	878	870	868	870	872	871	873
skscore2	896	899	900	902	902	900	897	894
skscore3	921	921	923	925	925	924	929	927
sk_rate	0.102	0.101	0.102	0.102	0.100	0.101	0.100	0.102
myscore1	868	865	866	858	862	861	863	866
myscore2	899	901	900	903	904	901	899	896
myscore3	930	929	933	932	933	933	934	930
my_rate	0.101	0.102	0.100	0.102	0.100	0.102	0.101	0.103

(e) Describe and interpret your findings by comparing the misclassification rates of (c) and (d).

First, the projection after dimension reduction was observed. The

classification effect of category 1 and category 3 of lda was slightly better than that of pca, with smaller overlapping regions and smaller inter-class spacing of category 1.

Then, misclassification rates were compared. In general, the classification error of lda was significantly smaller than that of pca. Lda reached the minimum value of 0.100 at k=25, and pca reached the minimum value of 0.142 at k=25. For each category of 1000 train data, the optimal value is obtained at the 25 nearest neighbor, which may have some regularity.

Lda can take advantage of the classification labels of the training data, so it can perform better than pca.