EE5907R
Pattern Recognition CA2

Name: Shao Guoliang
Matric NO.: A0152113Y

1. PCA

PCA uses the variance of each orthogonal basis to do the dimension reduction.

The algorithm is as follows:

First, we do the data pre-processing by centralize the training data and testing data, means we should subtract the mean of 784 dimension from the original training data and testing data;

Second, we calculate the co-variance matrix S of training data and find the eigenvector and eigenvalue of S by "svd" method. In the matlab code, we use

$$[eigvector, eigvalue] = svd(S, 'econ');$$

Third, we sort the eigenvector from the largest corresponding eigenvalue to the lowest;

Fourth, we choose the first 40,80 and 200 eigenvector to form a transformation matrix, and process the dimension reduction by multiply the original data set with the transformation matrix;

Finally, we use the nearest neighbor method to do the classification.

(1) PCA visualization

Fig.1.1 and Fig.1.2 show the projection of 2-D and 3-D dimension reduction results respectively. Fig.1.3 and Fig.1.4 indicate several orthogonal basis extracted from PCA analysis.
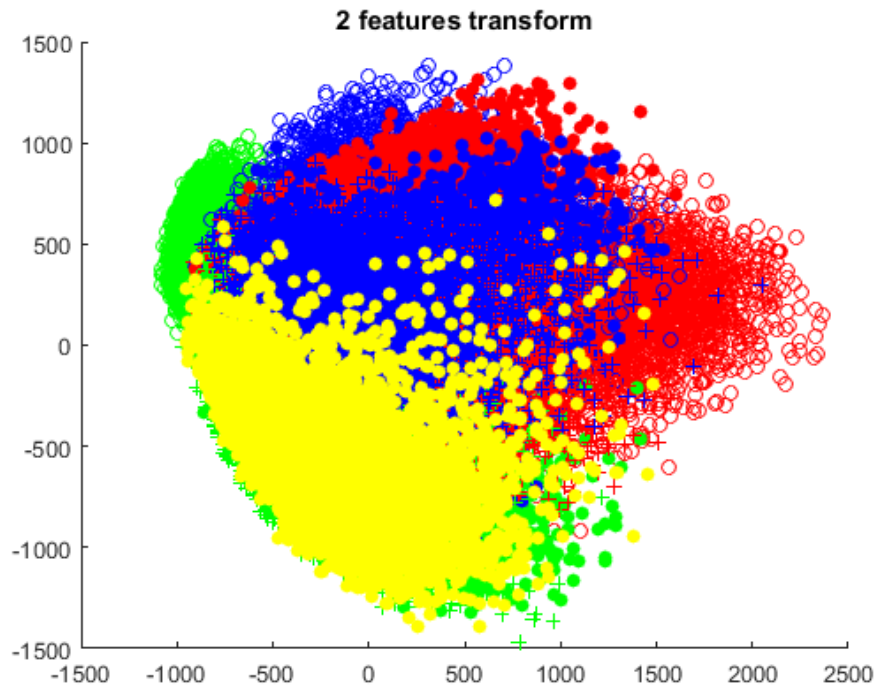

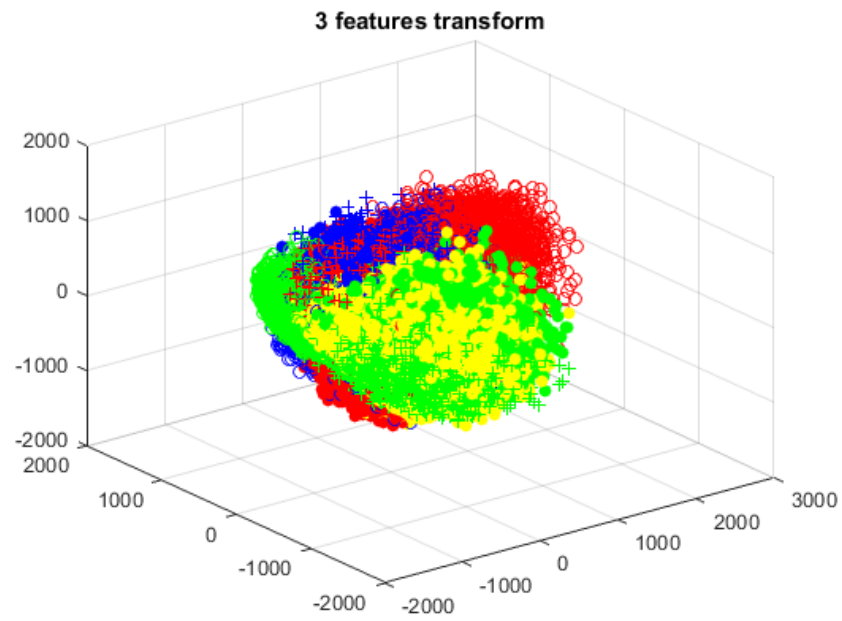
Fig.1.1 training data projection to 2-dimension
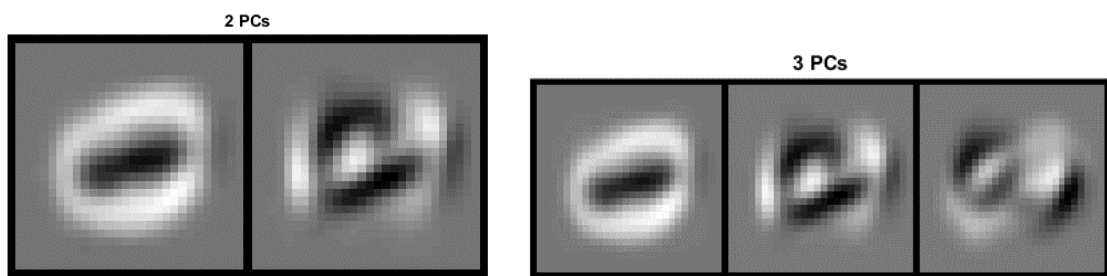
Fig.1.2 training data projection to 3-dimension

Fig.1.3 First 2 PCs and 3 PCs visualization

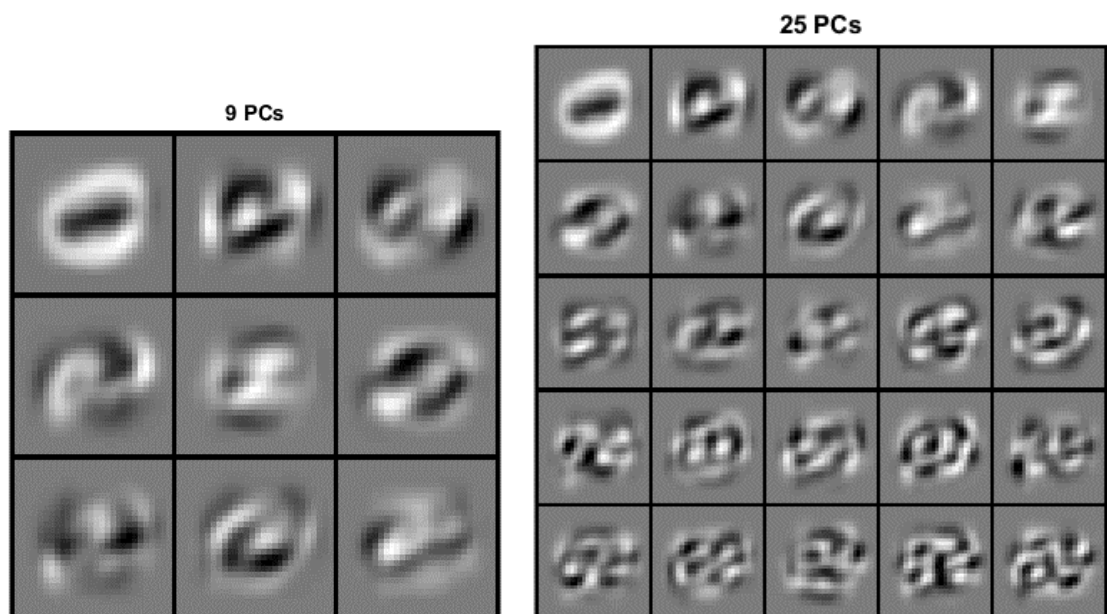
Fig.1.3 First 9 PCs and 25 PCs visualization

(2) PCA classification accuracy

NO. PCs	Classification error
40	0.0273
80	0.027
200	0.0311

We can conclude from the error results that the higher energy conservation does not guarantee a higher performance of testing error.

(3) Determine the value of d

In order to preserve 95% energy of the training data, we should compare the sum of first A dimension eigenvalue with the total sum of eigenvalue. After calculating it, we have A = 154 so that the selected PCs preserve the 95% of the original information energy.

The error of 154 PCs is 0.0306.

Usually we will choose the following criterion:

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > Threshold(0.9 \ or \ 0.95)$$

Other criterion may come from cross-validation training error rate that gives an classification accuracy over 90%.

2. LDA

Compared with PCA, LDA analysis calculate co-variance matrix within class as well as between classes, and take the derivation of Fisher Criterion to form a eigenvalue decomposition problem.

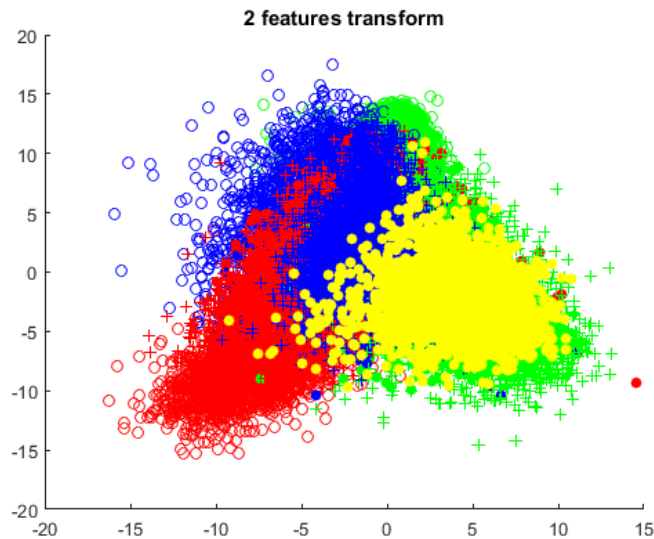Fig.2.1 and Fig.2.2 show the projection with dimensionality of 2 and 3 respectively.

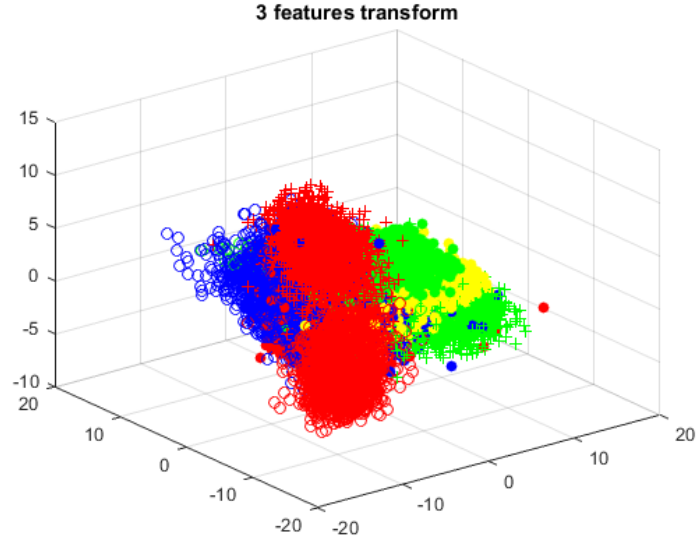Fig.2.1 2-dimension projection of LDA

Fig.2.2 3-dimension projection of LDA

The training error is indicated in the following chart:

Dimensionality	Classification error
2	0.5259
3	0.3336
9	0.1050

In the LDA analysis, the co-variance matrix between classes is the sum of the outer product of two vectors from mean of corresponding label data set $(\mu_i - \mu)$ and $(\mu_j - \mu)$, so the rank of S_b is at most $C - 1$, where C is the number of total labels. In this assignment, $C = 10$, so the maximum of dimensionality decomposition of $S_w^{-1}S_b$ is 9.

3. SVM

In this assignment, we use the libsvm toolkit to do the classification. First we do the dimension reduction by using PCA, and use 40PCs, 80PCs, and 200PCs reduction as the input of SVM. In addition, we do the SVM using linear kernel and radial base kernel respectively.

1) Linear kernel results(Accuracy):

NO. PCs	Penalty C							
	C=10		C=1		C=1e-1		C=1e-2	
	Train	Test	Train	Test	Train	Test	Train	Test
40	93.1417%	93.2200%	93.0800%	93.2500%	92.6300%	93.2200%	91.2233%	91.6300%
80	94.61%	94.25%	94.3817%	94.2400%	93.6617%	93.7700%	91.9233%	92.3100%
200	95.8350%	94.3900%	95.1867%	94.7400%	94.0867%	94.0500%	92.1833%	92.5000%

We can conclude from the classification results that

(1) A higher penalty value gives a higher classification accuracy. But actually a much higher value of penalty will also cause the reduction of classification accuracy. We can observe that, for 40PCs and 200 PCs, the highest accuracy achieved when C=1.

(2) Fix the penalty value, we can see that more principle components can give a higher classification accuracy.

2) Radial base kernel results (Accuracy):

| NO. PCs	Penalty C							
	C=10		C=1		C=1e-1		C=1e-2	
	Train	Test	Train	Test	Train	Test	Train	Test
40	96.4883%	96.34%	93.6433%	94.11%	90.9317%	91.54%	84.0467%	84.85%
80	95.4833%	95.29%	93.26%	93.76%	90.1683%	90.73%	75.5733%	76.5%
200	94.4433%	94.4%	92.33%	92.6%	87.8683%	88.9%	45.7283%	45.75%

We can conclude from the classification results that

(1) Contrast to the linear kernel, a higher penalty value will give a higher classification accuracy using the radial base kernel. This may result from the high sensitivity of penalty value using radial base kernel results. Besides, the difference between different penalty value is very big.

(2) Contrast to the linear kernel, the lower PCs gives a better classification performance when fix the penalty value. So the highest accuracy occurs at 40PCs trained with penalty value of 10.

4. CNN

In this assignment, we use the 'matconvnet-1 (1).0-beta24' as the training platform to proceed the convolutional neural network. We change the architecture of the number of nodes in convolution layer, number of convolution layer, number of nodes in the fully connected layer and type of fully connected layer. The training results are shown in Fig. 4.1.
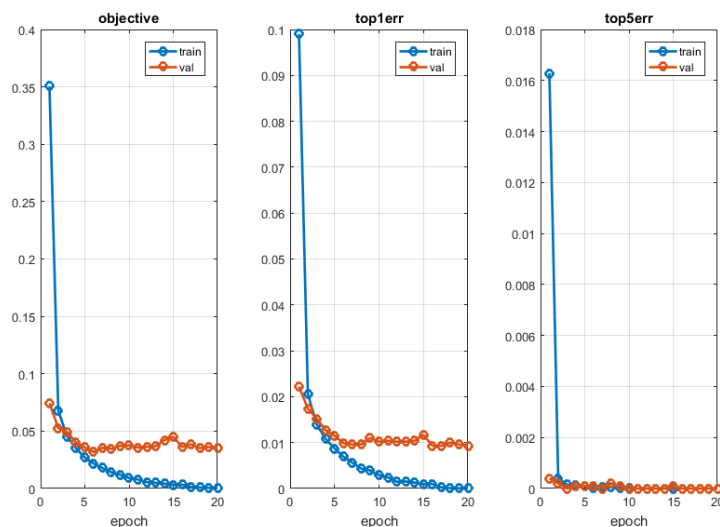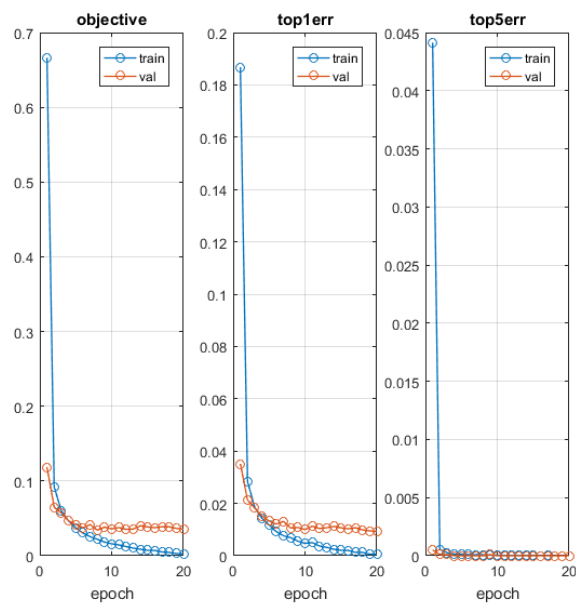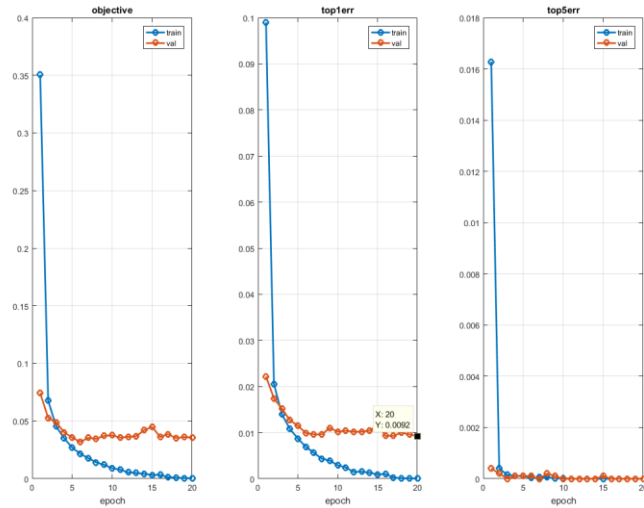

(a) 20-50-500-10 structure 0.0089
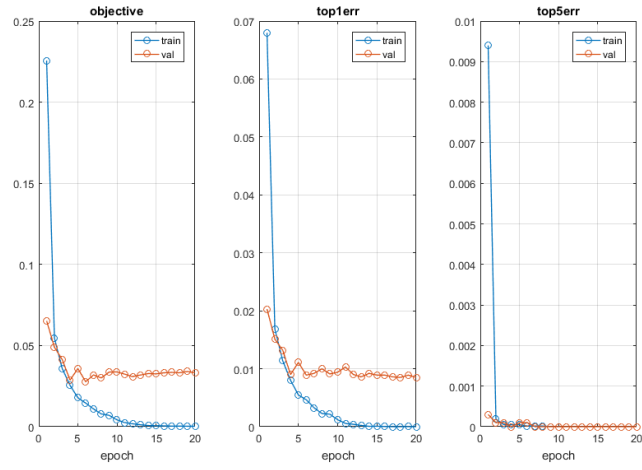
(b) 50-50-500-10 structure 0.0085
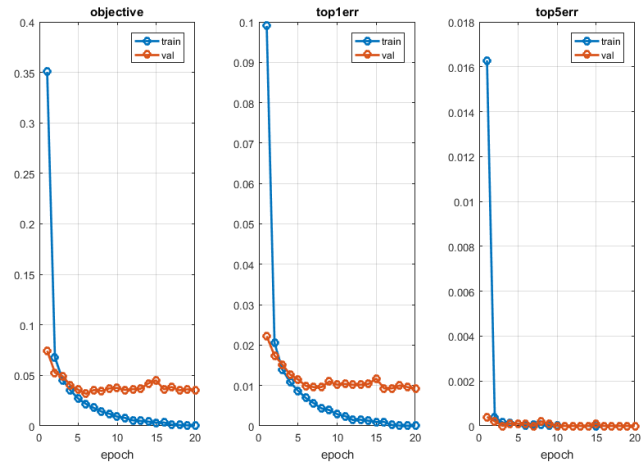
(c) 20-50-50-500-10 structure 0.0134

(d) 20-50-500-200-10 structure 0.093

(e) 20-50-100-10 structure 0.0092

Structure	Top1Error reported
20-50-500-10	0.0089
50-50-500-10	0.0085
20-50-50-500-10	0.0134
20-50-100-10	0.0092
20-50-500-200-10	0.0093
Change type	0.0072

Conclusion:

From the reported classification error of different neural network structures, we can draw a conclusion that:

1) Adding the number of nodes in the convolution layers can increase the accuracy. Adding nodes means increase the dimension of input information. (compared with 50-50-500-10)
2) Adding the number of layers will reduce the accuracy. (compared with 20-50-50-500-10)
3) Increase the number of nodes in fully connected layer will increase the accuracy. (compared with 20-50-100-10)
4) Increase the number of fully connected layer will decrease the error. (compared with 20-50-500-200-10)
5) Change the type of fully connected layer will increase the accuracy. (compared with 0.0072)