Homework Set 7

Problem 1 (Support vector machine)
The objective of this assignment is to calculate an SVM classifier. Consider the following training data:

$$\boldsymbol{x}_1 = \begin{bmatrix}-1\\1\end{bmatrix}\ y_1 = +1, \quad \boldsymbol{x}_3 = \begin{bmatrix}1\\1\end{bmatrix}\ y_3 = +1,$$
$$\boldsymbol{x}_2 = \begin{bmatrix}0\\1\end{bmatrix}\ y_2 = -1, \quad \boldsymbol{x}_4 = \begin{bmatrix}0\\0\end{bmatrix}\ y_4 = +1.$$

These data are not linearly separable. However, we can make them linearly separable by introducing the square of the first coordinate as an additional "feature". In other words, consider the feature map $\phi\colon \Re^2 \to \Re^3$ given by

$$\phi\left(\begin{bmatrix}\alpha\\\beta\end{bmatrix}\right) = \begin{bmatrix}\alpha\\\beta\\\alpha^2\end{bmatrix}$$

A kernel SVM will be used to perform the classification problem:

$$f(\varphi(\boldsymbol{x})) = sgn\left(\sum_{i=1}^{m} \lambda_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}_j) + b\right)$$

Answer the following questions:

a) Compute $\phi(\boldsymbol{x}_i)$ for $i = 1, 2, 3,$ and 4.

b) Compute the inner product $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\phi(\boldsymbol{x}_i))^{\mathrm{T}}\phi(\boldsymbol{x}_j)$ for all pairs (i, j).

c) Write down the Lagrangian function $L(\boldsymbol{w}, b, \lambda_i)$.

d) Find the values of λ_1, λ_2, λ_3, and λ_4 which maximize L subject to the constraint $\sum_{i=1}^{4} \lambda_i y_i = 0$. (Hint: use constraint to eliminate one of the λ_i, then maximize over the remaining three.)

e) Compute $\boldsymbol{w}$ and b.

f) Draw a picture showing the four training points together with the (curved) preimage of the optimal hyperplane in the original $[x\ , z]^{\mathrm{T}}$ space.

Problem 2 (Support vector machine)

MNIST database of handwritten digits

Recognizing hand written digits or characters is an interesting research topic. A dataset of hand written digits in the file "07HW2_digit.mat" can be downloaded from the course webpage. It is originally from the MNIST database (http://yann.lecun.com/exdb/mnist/). The original database consists of 60,000 training and 10,000 test images of different hand written digits (Fig. 1). To keep computation time more manageable, the numbers of the training and test samples in the dataset are reduced to 5,000 and 1,000, respectively, in the provided dataset. The images are 28×28 pixels in size and 256 in gray level.

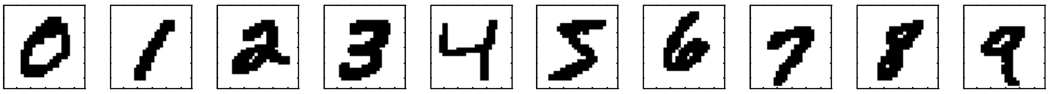

Figure 1, sample hand-written digit images

Load the file "07HW2_digit.mat" into MATLAB. You will see variables

1) `train0, train1, …, train9` as the training samples
2) `test0, test1, …, test9` as the test samples

The number in the name of each variable corresponds to the digit that the image belongs to. Each training variable contains 500 images of the same digit, and each test variable contains 100 images of the same digit. To display the 1st zero digit image, type the following commands in MATLAB:

```
>>colormap gray;
>>imagesc(~reshape(train0(1,:), 28, 28)');
```

LIBSVM

LIBSVM is a popular SVM tool. Download *LIBSVM* from http://www.csie.ntu.edu.tw/~cjlin/libsvm and unpack it to your working directory. Run "\matlab\make.m" to compile the source code into MEX (MATLAB executable) files. Now it is ready for use. *LIBSVM* develops SVM classifiers using a training sample variable and a label variable. For example, to train an SVM classifier using training variable `TRAINING` and the corresponding label variable `GROUP`, type in

```
>>model = svmtrain(GROUP, TRAINING)
```

The returned variable `SVMStruct` contains information of the trained classifier. Note that the training data in the variable `TRAINING` are usually normalized between 0 and 1 before it is fed into the classifier for training.

If soft-margin SVM classifiers with radial basis function (RBF) kernels are used, the performance of the classifiers depends on the choice of γ in the RBF kernel and the soft-margin penalty c. These two parameters needed to be assigned by the operator. Check the help of `svmtrain` for more details.

To calculate the classification accuracy using a test dataset `TEST` and the corresponding label variable `testGROUP`, type

```
>>[pGROUP Accuracy p_val] = svmpredict(testGROUP, TEST, model)
```

The returned variable `Accuracy` contains the accuracy of the classifier.

Perform the following tasks:

a) Develop an SVM classifier to differentiate the images of digit 0 to the images of digit 1. Use soft-margin SVM classifier and an RBF kernel. Try different combinations of $\gamma = [2^{-14}$, 2^{-12},…, $2^{-6}]$ and $c = [2^{-5}$, 2^{-3},…, $2^{3}]$, and use test data or cross-validation to find the best pair of γ and c.

b) Try linear and polynomial kernels with various kernel parameters. Compare and report the performance of the kernels on the test data.

c) Develop several SVM classifiers to recognize the images of all the digits. Note that SVM is a 2-class classifier. Implement multiple-class SVM techniques to fulfill the job. Do <u>NOT</u> use the built-in multiclass classification tools. Compare and report the performance of the kernels on the test data.