**ECE593 Introduction to Machine Learning Homework 6**

**Linear discrimination**

Bo Lin

Introduction

SVMs (Support Vector Machines) are a useful technique for data classification. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one “target value” (i.e. the class labels) and several “attributes” (i.e. the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

Given a training set of instance-label pairs  whereand , the support vector machines require the solution of the following optimization problem:



Here training vectors  are mapped into a higher dimensional space by the function .SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space.  is the penalty parameter of the error term.

 is the kernel function. Among the four basic kernel functions, this homework chose linear and radial basis function:



1. Study LibSVM

The LibSVM is an integrated software for multi-class support vector classification. After compiling the “make.m” function, it will create “svmtrain” and “svmpredict” for training and testing the SVM model.

The inputs of “svmtrain” function are an m by n matrix of m traning instances with n features, i.e. optdigits data points and an m by 1 vector of training labels. In addition, a string of training option is required to clarify the training kernel type and parameters. The “svmtrain” function returns a model which can be used for future prediction. It is a structure and is organized as [Parameters, nr\_class, totalSV, rho, Label, ProbA, ProbB, nSV, sv\_coef, SVs]

The inputs of “svmpredict” function are testing optdigits with the same format as inputs of “svmtrain”. In addition, the model that returned by “svmtrain” function should also be input to the “svmpredict” function. The “svmpredict” function has three outputs. The first one is a vector of predicted labels. The second output, accuracy, is a vector including accuracy, mean squared error and squared correlation coefficient. The third is a matrix containing decision values or probability estimates.

1. Using SVM to recognize optdigits

# Feature Scaling

Scaling before applying SVM is very important. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. This is the exact situation with “optdigits”. Another advantage is to avoid numerical difficulties during the calculation. The linearly scaling applied in this homework has two choices, when “scale = 0”, each attribute is scaled to the range [0, 1], when “scale = 1”, each attribute is scaled to the range [-1, 1].

The formula applied in this homework is:



# Cross-Validation

In machine learning, it is common strategy that separate the data set into training set and testing set. The prediction accuracy obtained from the testing set more precisely reflects the performance on classifying an independent dataset. This homework uses cross-validation as an improved version of above method. In v-fold cross-validation, the whole training data is first randomly divided into v subsets of equal (or similar) size. Sequentially one subset is tested using the classifier trained on the remaining v-1 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. For the homework, the number of folds is chosen as 5, and the final accuracy is the average accuracy testing among all five folds.

# Grid-Search

For RBF kernel, there are two parameters C and . It is not known beforehand which C and  are the best for this data set. A “grid-search” is applied after cross-validation. First a coarse grid search is applied. C and  are designed equal to  where for C, k = -5,-3,…,15 and for ,k=-15,-13,…,3. After identifying a better region on the grid, a finer grid search on that region will be conducted using a smaller interval of k=0.25.

# Dimensionality Reduction

Due to the fact that the simulation takes a lot of time, and also due to the simplicity of principle component analysis (PCA) method. This homework choose PCA as the method for dimensionality reduction. The function is enclosed in “pca1.m”. The dimension of the data, originally 64 dimension, is reduced to [4 8 16 32].

3. Simulation Result

# Compare Linear SVM with RBF Kernel SVM



Figure 1

From Fig.1 we can see that over all, the RBF SVM kernel always has a better validation accuracy than linear SVM kernel. This is probably due to the fact that using linearly separable hyperplane for high dimension data like opdigits is not as suitable as curve classifiers such as RBF SVM.

# Compare Parameters of Kernel Function with Accuracy



Figure 2



Figure 3

Fig. 2 and Fig. 3 are demonstration of the above grid search method. These grid search are simulated using RBF kernel function with 16 dimension data. The color bar shows the validation accuracy. From the loose grid search, the best accuracy area is chosen for fine grid search. And finally by comparing the accuracy, the best pair of C and is chosen.

# Compare Dimensionality with Accuracy

Also from Fig.1, we can see that as dimension increases from 4, the validation accuracy increases. This is true for both linear SVM and RBF SVM. As dimension reaches above 16, the accuracy seems wouldn’t improve anymore. This shows that there are some data points that are probably mislabeled at the first place.

# Compare # of Support Vectors with Parameters



Figure 4

From Fig. 4, we can see the number of support vectors change with the change in parameters of the kernel function. This is also a demonstration performed in 16 dimension RBF kernel function. But the pattern of Fig. 4 is similar with the inverse of Fig.2. As accuracy increases, the number of support vectors decreases.

# 5. Compare # of Support Vectors with Dimension



Figure 5

Fig. 5 compares the different number of support vectors in different dimension between the two different kernel functions. We can see that in dimension 8, both kernel functions has the smallest number of support vectors. As the dimension increases above 8, the number of support vectors for both kernel functions increases. Compare with Fig. 1, we can see that as number of support vector increases, the validation accuracy also increases. This shows that in proper dimension (4 D is not discussed here since 4 D will omit a lot of details of the data), the number of support vector increases to find a more accurate hyperplane.

4. Conclusion

This homework presented two type of support vector machine classification using LibSVM toolbox. While using SVM, both cross-classification and grid-search are implemented to acquire more accurate parameters for the kernel function. By comparing the validation accuracy and number of support vectors between two kernel functions and among 5 different dimensions, we can see that for higher dimension, the validation accuracy is higher, and the number of support vector is larger.