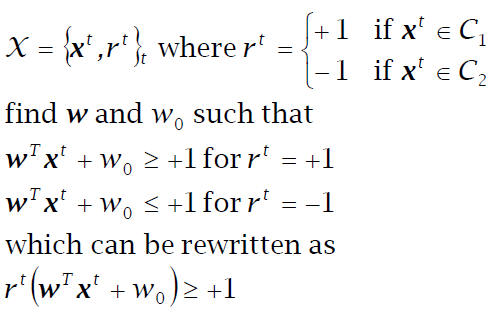
**Report for Homework 6**

**1.Introduction**

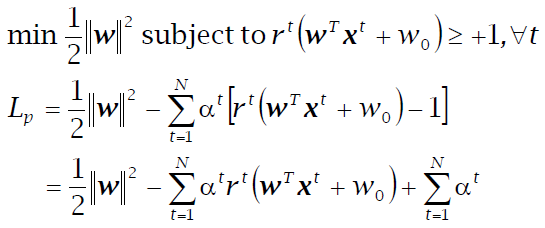
## In this homework, we deal with the problems about linear discrimination using [the toolbox LibSVM.](http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html) First, we study the LibSVM tutorial and use the functions embedded to implement linear discrimination algorithm on the optdigits dataset. Then we use the linear function as kernel function based on the tutorial and recognize the digits in this dataset. Also we use different kernel functions such as RB and Sigmoid to recognize the digits in the dataset. Moreover, we use grid search procedure to select out the best parameters for the kernel functions. And we obtain the confusion matrix to express the accuracy of the classifiers based on different function for the validation data. Also, we do dimensional reduction on the dataset and compare the accuracy and number of supporting vectors for different dimensions such as 64,32,16,8 and 4. Finally, we draw out the digits on the screen using the classifier based on RB function.

**2.Methods**

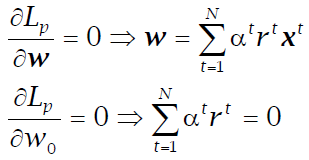
For Problem 2, we study the tutorial for LibSVM and work on the linear SVM to implement support vectors machine. In order to get the optimal separating hyperplane , the parameters in linear discrimination should satisfy the requirements below.



Therefore, in order to obtain the maximum margin to the discrimination boundary ,this linear discrimination model converts to a convex optimization problem to find a unique solution for the model. Then we focus on the Lagrangian dual problem of the original problem shown below.

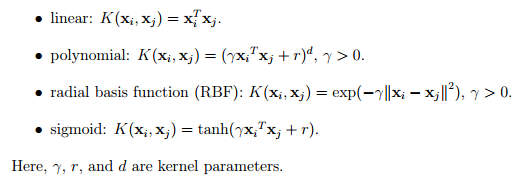


Using the KTT conditions of the convex optimization problem, we can get the parameters **W** and Lagrangian multipliers **α** . Most **α** are 0 and only a small number have **α** > 0 , they are the support vectors.



And in order to find out the relatively best parameters for the model ,we use cross validation for the training data to select out the parameters which bring the highest accuracy for the validation data. Also we take the grid search procedure to scan for the parameters. Each parameter has their step size in the search loop in order to find out the range in which the best parameters stand. Also because of long computation time, we can adjust the step size to save some time while doing simulation. Also, confusion matrix is introduced to express the accuracy for the classifier.

For Problem 4, there are other kernel functions in LibSVM, in this homework, I take RB and Sigmoid function. The functions in LibSVM are listed below.



For RB function, there are two parameters **c** and **γ** . So in grid search procedure, we use two 'for' loop to search out a pair of **c** and **γ** to achieve the highest accuracy. Because we have a large search scale and step size, thus in order to get even better accuracy , we also use a smaller search range to find out the accuracy around this previous pair of **c** and **γ** .

And for Sigmoid function, there are three parameters **c** , **γ** and **r**. So there are three 'for' loop in our code. And it may take a lot of time. We plot out the accuracy and supporting vectors based on different **c** , **γ** and **r** around the previous best ones from grid search procedure. And we plot out the accuracy and number of supporting vectors based on different kernel functions and parameters.

Also we do dimensional reduction on the dataset. In this homework, we use PCA to reduce the dimension to 32,16,8 and 4. And we use linear ,RB and sigmoid function for the data in these dimensions to get different classifiers, thus we can compare the accuracy and number of supporting vectors for different dimensions.

For Problem 5, we just reduce the dimension of the dataset to 2, and by using the classifier based on RB function which we have trained, the predicted digits for the image can be obtained. So we can plot out the real digits and predicted digits for the 2D data .

**3. Results**

For Problem 2, we study the tutorial for support vectors machine and choose the linear function. We use grid search procedure and cross validation to find out the best parameters. And we study the accuracy based on different parameter **c** as shown in Fig.1. And we find the best **c** is around 0.008. The number of SVs is shown in Fig.2.

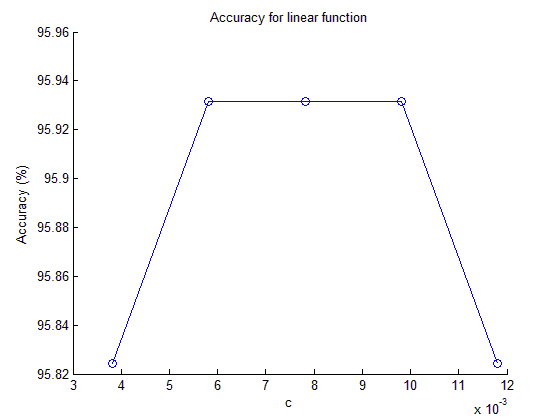


Fig.1 Accuracy for different parameters on Linear function

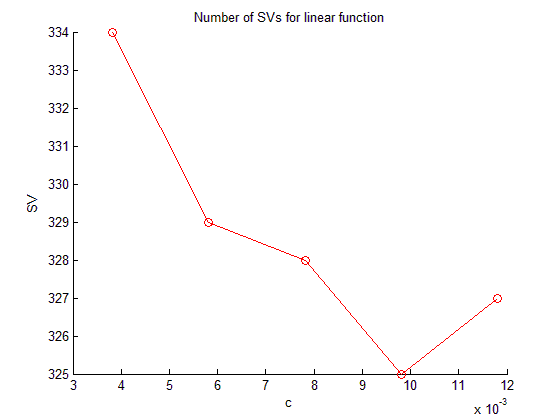


Fig.2 Number of SVs for different parameters on Linear function

Based on the Fig.2 above , we can see that when **c** is around 0.008, the accuracy is lowest, but the number of SVs is around 328 which is not the lowest number. After dimensional reduction, we can draw out the relationship between accuracy, number of SVs and dimension shown in Fig.3 and Fig.4. We can find that when the dimension is 32, the highest accuracy can be achieved ,about 96%. However, at the dimension of 16, we can find the smallest number of supporting vectors, around 260.

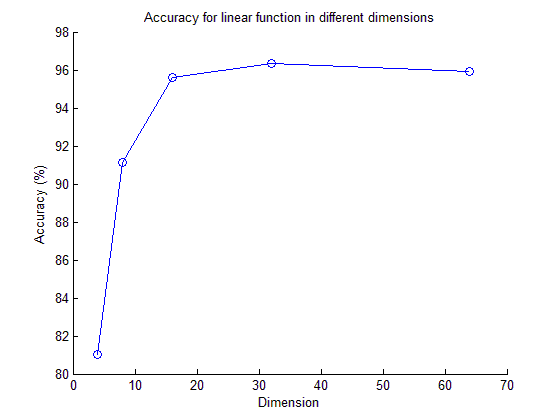


Fig.3 Accuracy for different dimensions on Linear function

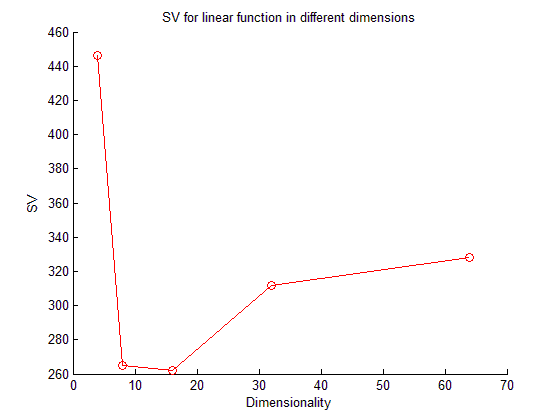


Fig.4 Number of SVs for different dimensions on Linear function

For Problem 3, we use RB function to train the model based on the best parameters **c** and **γ**. Then we use a smaller step size to plot out the relationship between accuracy, SVs and different parameters . Fig. 5 and 6 show the accuracy for different parameters. We find that **c** has little impact on the accuracy with the best **γ .**

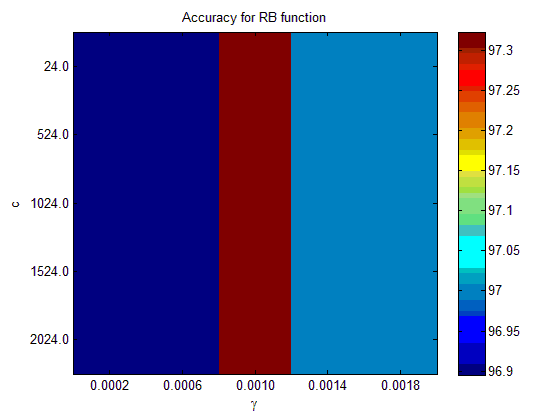


Fig.5 Accuracy for different parameters using imagesc

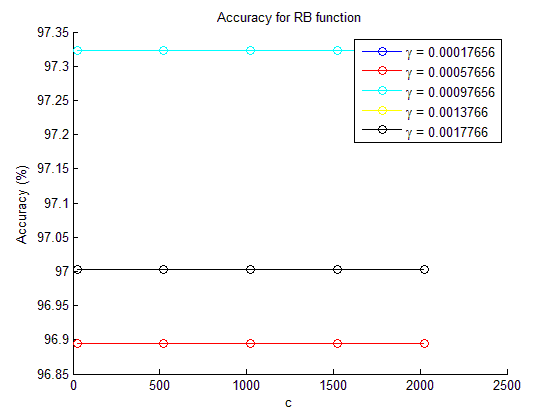


Fig.6 Accuracy for different parameters

Fig.7 and 8 shows the number of SVs based on different parameters for RB function. We find that **c** also has little influence on the number of SVs. And the number of SVs is very sensitive to **γ** . The smallest **γ** can get the smallest number of supporting vectors.

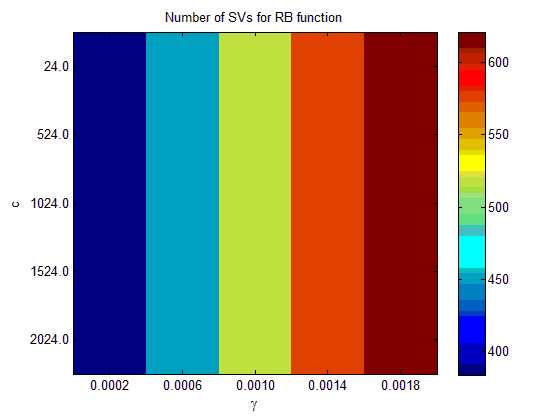


Fig.7 Number of SVs for different parameters on RB function using imagesc

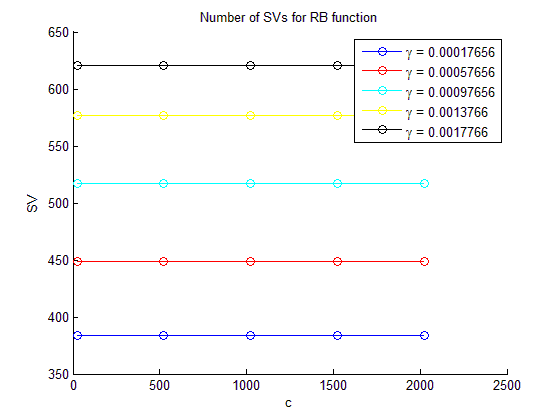


Fig.8 Number of SVs for different parameters on RB function

Also we use PCA to do dimensional reduction for 32,16,8 and 4. We compare the accuracy and number of SVs for different dimensions. Fig.9 and 10 shows the accuracy and number of SVs for different dimensions. We can see that when the dimension is 64, the highest accuracy can be achieved, about 97%. And for dimension 16 and 32 ,we can get almost the same accuracy as that for dimension 64. And we can see that when the dimension is 8, we get the smallest number of supporting vectors, about 310.

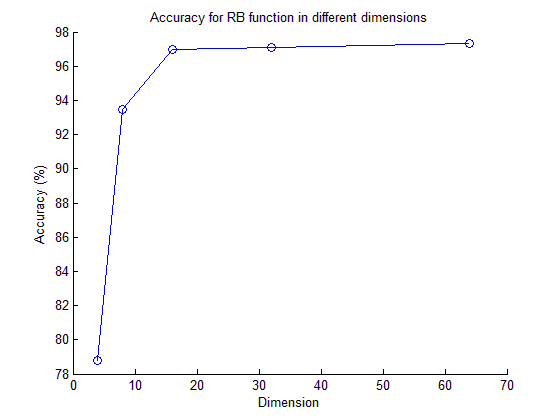


Fig.9 Accuracy for different dimensions on RB function

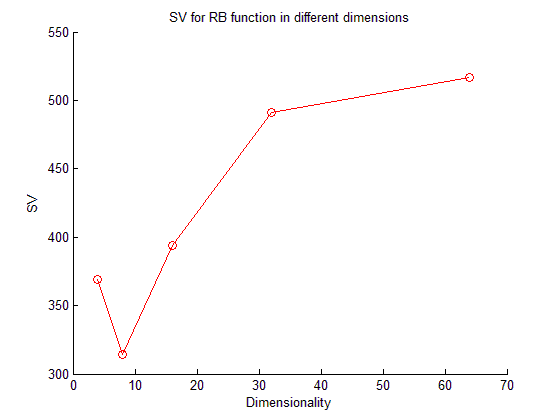


Fig.10 Number of SVs for different dimensions on RB function

For sigmoid function, there are three parameters, we find that **r** has little impact in the accuracy and number of SVs in our range of simulation. We plot out the accuracy and number of SVs based on different **c** , **γ** and **r**  shown in Figure 11,12,13 and 14. We find that for the two **r** , we get the same graph for **c** and **γ** . Also we can see that when **γ = 2.05e-5 ,** we can get the highest accuracy about 96.2%. However , for number of SVs , when **γ = 5e-5**, the lowest number of SVs can be achieved.

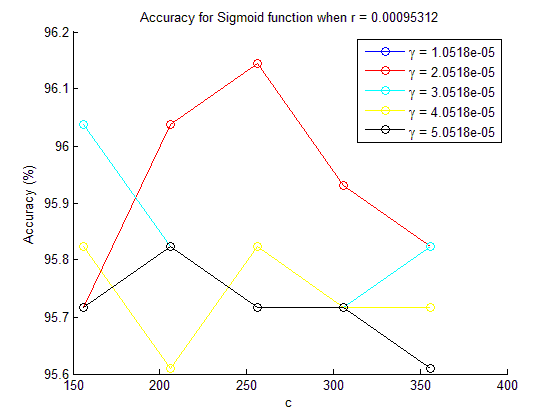


Fig.11 Accuracy for different parameters on sigmoid function

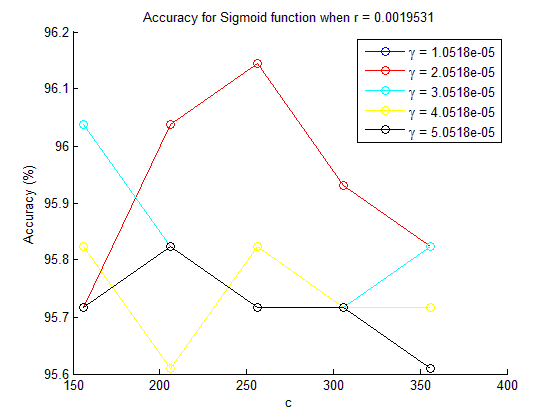


Fig.12 Accuracy for different parameters on sigmoid function

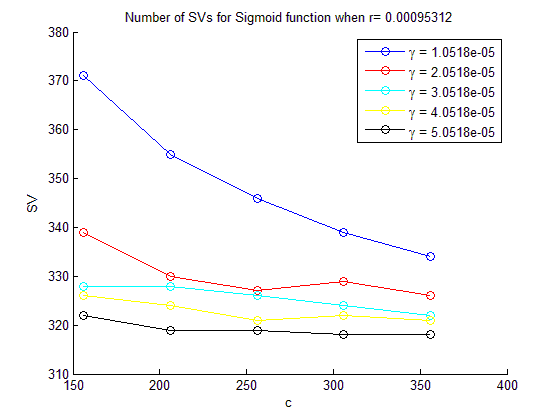


Fig.13 SVs for different parameters on sigmoid function

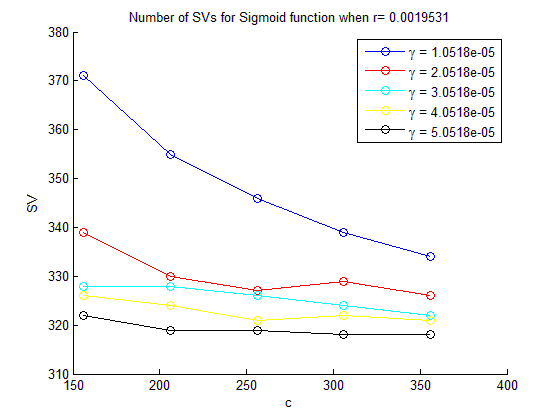


Fig.14 SVs for different parameters on sigmoid function

Also we do dimensional reduction based on PCA for sigmoid function. We find that for dimension 32 ,we can get the highest accuracy about 96%. And for dimension 16 , the smallest number of SVs ,about 260 can be achieved. And when the dimension is 64, the number of SVs is about 320.

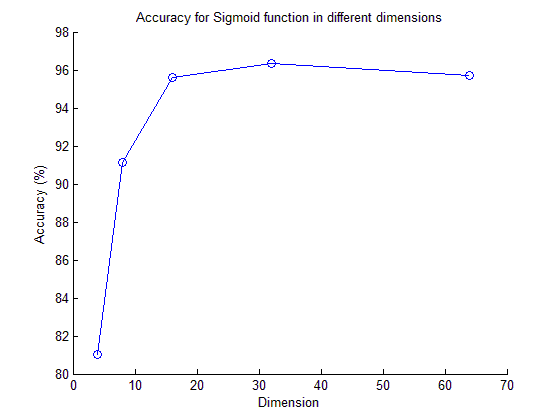


Fig.15 Accuracy for different dimensions on sigmoid function

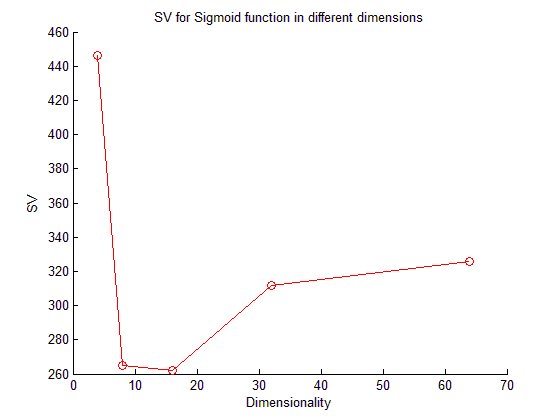


Fig.16 SVs for different dimensions on sigmoid function

For Problem 5, we draw out the predicted digits and real digits based on the validation data using the RB function. We find that only a few points are misclassified after comparing the real digits and predicted digits. The RB function can reach very high accuracy .

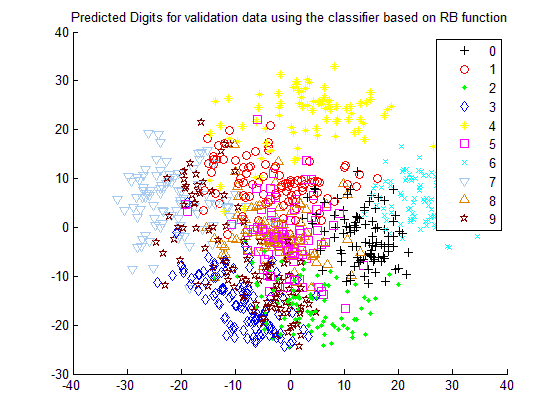


Fig.17 Predicted digits for validation data using RB function

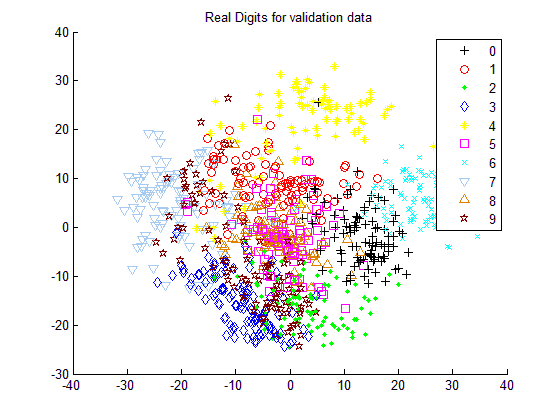


Fig.17 Real digits for validation data using RB function

**4. Discussion**

Based on the results in the section above, we can make a discussion on the classification accuracy and number of SVs based on different kernel functions. We find that by using kernel functions ,we may achieve higher accuracy than that using linear function, for example, in our simulation, the highest accuracy for linear function is about 95.9%, however , for RB function , the accuracy is about 97.3 % , and for sigmoid function, the accuracy is about 96.1% . This may be because the kernel function has more parameters, which account for more factors that may influence the accuracy. So the kernel function can better characterize the dataset ,thus leading to a better model for the classifier. We also study the parameters for the accuracy and number of SVs . For linear function, we find that the parameter **c** can significantly influence the accuracy and number of SVs, in our range of simulation , the accuracy is between 95.8% and 95.9% , and the number of SVs ranges from 325 to 334. Also for dimensional reduction, we find that for dimension 32, the highest accuracy can be achieved, about 96%. This may because when the dimension is 32, the dataset can be easily classified by the linear function based on the features of the new data. For RB function, we find that in our range of simulation, parameter **c** has little influence in the accuracy and number of SVs.And for the best **γ** we get from the grid search procedure, the accuracy can reach about 97.3% . But for the number of SVs, there are no dependency between SVs and accuracy. The number of SVs is very sensitive to the parameter **γ**, for different **γ** , it ranges from 380 to 620. Also we find the dimensional reduction can influence the accuracy, when the dimension is 64, the highest accuracy can be reached. Also when the dimension is 64, the corresponding number of SVs is the largest. For sigmoid function, we find that the parameter **r** has little influence in the accuracy and number of SVs. And we can see that the **γ** from the grid search procedure, is not the best any more. This is because the search range in the grid procedure may be too large. And the accuracy is very sensitive to the parameter **γ** , even a change of 1.0e-5 can get a different accuracy. For the number of SVs ,we also find the number is more sensitive to **γ** than **c .** For dimensional reduction **,** we can see that, for dimension 32, the higher accuracy can be achieved. However for 16 and 64,we can get the almost same accuracy. But for number of SVs, the results are very different. It ranges from 260 to 450 for different dimensions.

For Problem 5 , we compare the real digits and the predicted digits based on the classifier using RB function. We can see that the classifier can reach very high accuracy ,and only a few points for '0' and '9' are misclassified based on the RB function.

**5. Software listing and executable software**

This program uses MATLAB including 2 files, HW6.m and Sigmoid.m.

Please click and run HW6.m to get the graphs and results for all the problems. And for another kernel function (Sigmoid function), please click Sigmoid.m.