
# NEURAL NETWORK PROJECT I

# SVM for Classification of Spam Email Messages

SUBMITTED BY

NAME: LIN KANG

STUDENT NUMBER: A0191818W

EMAIL ADDRESS: e0344070@u.nus.edu

# 1. Requirement of SVM project

**Task 1:** Write a MATLAB (M-file) program to compute the discriminant function g (▪), if one exists, for the following SVMs, using the training set provided:

i. A hard-margin SVM with the linear kernel

$$\mathrm{K}(x_1, x_2) = {x_1}^T x_2 \tag{1}$$

ii. A hard-margin SVM with a polynomial kernel

$$\mathrm{K}(x_1, x_2) = ({x_1}^T x_2 + 1)^p \tag{2}$$

where the values of p are listed in Table 1.

iii. A soft-margin SVM with a polynomial kernel as given in Equation (2) above, and with the values for p and C as listed in Table 1.

Note that a MATLAB function quadprog (available in the Optimization Toolbox) can be used to solve constraint optimization problems.

**Task 2:** Write a MATLAB (M-file) program to implement the SVMs with the discriminant functions obtained in Task 1. Apply these SVMs to classify the given training set and test set and report the classification results in Table 1 by filling the entries indicated by "?". Discuss the results and their implications, including issues related to the admissibility of the kernels and the existence of optimal hyperplanes for the three types of SVMs listed in Task 1 above.

**Task 3:** Design a SVM of your own. This SVM can be one of the three types specified in Task 1 above (i.e., hard-margin with linear kernel, hard-margin with polynomial kernel, and soft-margin with polynomial kernel), or one with your own choice of kernel. Using the given training set, compute the discriminant function g(▪) of the SVM. Implement the resulting SVM in a MATLAB M-file program. This program will be used to classify the evaluation set as part of the assessment.

# 2. Results of SVM classification

| **Type of SVM** | **Training accuracy** | | | | **Test accuracy** | | | |
|---|---|---|---|---|---|---|---|---|
| **Hard margin with Linear kernel** | 0.5885 | | | | 0.5592 | | | |
| **Hard margin with polynomial kernel** | P = 2 | P = 3 | P = 4 | P = 5 | P = 2 | P = 3 | P = 4 | P = 5 |
| | 0.5485 | 0.6255 | 0.8150 | 0.7845 | 0.5443 | 0.6016 | 0.8366 | 0.8008 |
| **Soft margin with polynomial kernel** | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 |
| **P = 1** | 0.9240 | 0.8440 | 0.8260 | 0.8115 | 0.9141 | 0.8405 | 0.8229 | 0.8040 |
| **P = 2** | 0.8925 | 0.8895 | 0.8865 | 0.8845 | 0.8900 | 0.8854 | 0.8783 | 0.8724 |
| **P = 3** | 0.8265 | 0.7610 | 0.7470 | 0.7220 | 0.8359 | 0.7741 | 0.7591 | 0.7350 |
| **P = 4** | 0.6190 | 0.5980 | 0.5980 | 0.5960 | 0.6484 | 0.6243 | 0.6230 | 0.6230 |
| **P = 5** | 0.5960 | 0.5960 | 0.5960 | 0.5960 | 0.6191 | 0.6204 | 0.6204 | 0.6198 |

Table 1 RESULTS OF SVM CLASSIFICATION

# 3. Discussion about the result on table 1

## 3.1 Hard margin with linear kernel

From the table 1, it shows that the accuracy of training set and test set is quite low which is less than 60%. This result is not enough to classify the spam email accurately. Since the discriminant function is trained by the training data, the accuracy of training set is slightly higher than the test set. In conclusion, the hard-margin SVM with the linear kernel is not suitable for classification of spam email.

What's more, when apply the Mercer condition, the result is false, which means that the Gram matrix is not positive semi-definite. (Mercer condition false)

## 3.2 Hard margin with polynomial kernel

For the SVM with hard margin with polynomial kernel, the accuracy is quite low when the order (parameter 'p') of polynomial kernel is low. And with the increasing of order, both the training accuracy and the test accuracy increase and meet the top value when the order of polynomial kernel equal to 4. After that, the accuracy decrease slightly to 78% and 80% when the order of polynomial kernel equal to 5.

The result shows that for hard margin with polynomial kernel the suitable choose of order could improve the appearance of SVM, Too high order could also affect the result.

When apply the Mercer condition, the judgement given by MATLAB shows that when p equal to 2, 3, 4, the kernel fulfill the Mercer condition and when p equal to 5, it could not fulfill the Mercer condition.

| **Mercer condition** | **P = 2** | **P = 3** | **P = 4** | **P = 5** |
|---|---|---|---|---|
| **Ture/False** | Ture | Ture | Ture | False |

Table 2 Mercer condition judgement of polynomial kernel

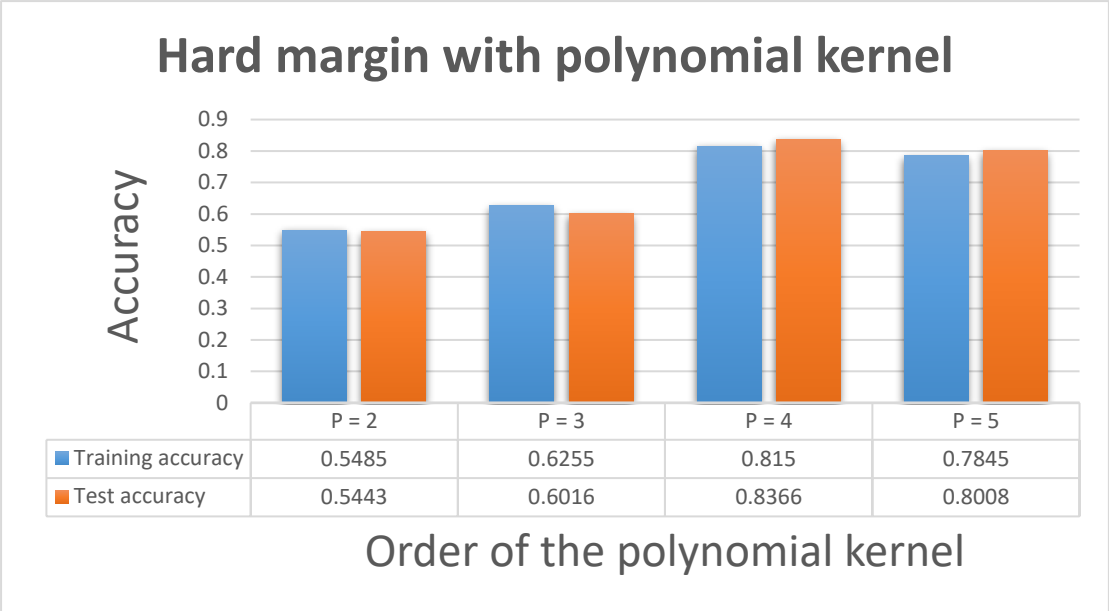


Figure 1 Hard margin with polynomial kernel

### 3.3 Soft margin with polynomial kernel

The figure 2 and 3 below show the result of SVM with soft margin with polynomial kernel. It's obvious that the parameters including the order of polynomial kernel and the margin parameters C affects the accuracy of training set and test set. Also the result of both these set are similar to each other.

When focusing on the parameter p, which represents the order of polynomial kernel, with the increasing of p, the accuracy of both training set and test set decrease slightly. It shows that lower order polynomial kernel with soft margin has higher accuracy compared to the high order polynomial kernel with soft margin. On the other hand, the parameter C, which represents the soft margin size, also affects the accuracy. When the size of soft margin increase (parameter C increase), the accuracy decreases which means that suitable small soft margin could help to improve the appearance of SVM.

Also when apply the Mercer condition. The result is the same as SVM2, since they use the same kernel.

| **Mercer condition** | **P = 1** | **P = 2** | **P = 3** | **P = 4** | **P = 5** |
|---|---|---|---|---|---|
| **Ture/False** | Ture | Ture | Ture | Ture | False |

Table 3 Mercer condition judgement of polynomial kernel

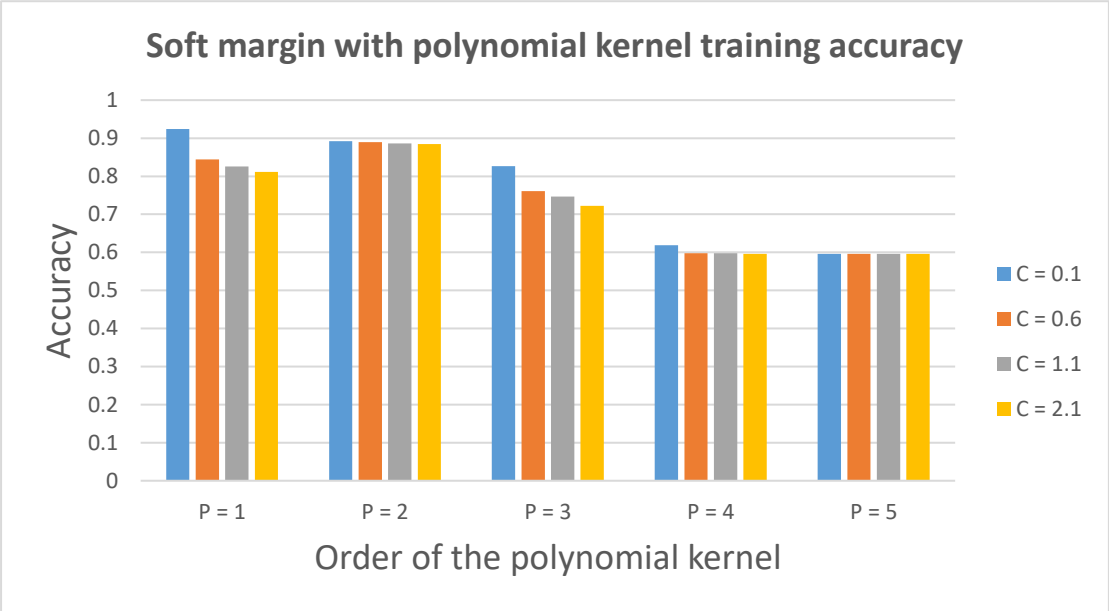


Figure 2 Soft margin with polynomial kernel training accuracy

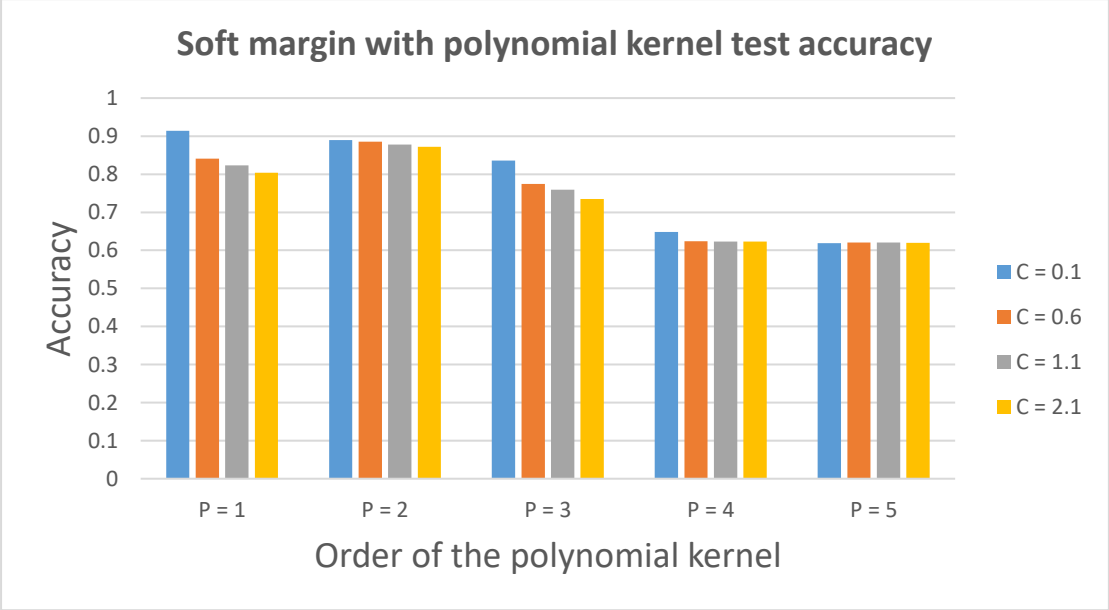


Figure 3 Soft margin with polynomial kernel test accuracy

# 4. Soft margin with gaussian kernel

## 4.1 Soft margin with gaussian kernel

In order to get the best SVM for spam email classification, there are two options. One is to use these three SVM and to improve the parameters to get the best SVM. The other option is to find another kind of kernel.

With the help of internet, the gaussian kernel is recommend as the SVM kernel, the final result is shown below. It is obvious that the best result is that C=1 and $\sigma = 0.001$. The accuracy is around 90%.

| **Type of SVM** | **Training accuracy** | | | | **Test accuracy** | | | |
|---|---|---|---|---|---|---|---|---|
| **Soft margin with gaussian kernel** | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 |
| **σ = 5** | 0.8765 | 0.8785 | 0.8780 | 0.8810 | 0.8770 | 0.8828 | 0.8815 | 0.8854 |
| **σ = 1** | 0.8850 | 0.8755 | 0.8770 | 0.8775 | 0.8913 | 0.8750 | 0.8822 | 0.8802 |
| **σ = 0.5** | 0.8875 | 0.8770 | 0.8785 | 0.8780 | 0.8939 | 0.8789 | 0.8828 | 0.8867 |
| **σ = 0.1** | 0.8895 | 0.8855 | 0.8815 | 0.8810 | 0.8926 | 0.8887 | 0.8848 | 0.8809 |
| **σ = 0.05** | 0.8880 | 0.8880 | 0.8865 | 0.8860 | 0.8913 | 0.8926 | 0.8893 | 0.8919 |
| **σ = 0.01** | 0.8930 | 0.8955 | 0.8960 | 0.8945 | 0.8965 | 0.8997 | 0.8991 | 0.8978 |
| **σ = 0.005** | 0.8970 | 0.8985 | 0.8985 | 0.8950 | 0.8997 | 0.9010 | 0.9004 | 0.8984 |
| **σ = 0.001** | 0.8980 | 0.8980 | 0.8970 | 0.8970 | 0.9017 | 0.9017 | 0.9010 | 0.9004 |

| **Type of SVM** | Training accuracy | | | | Test accuracy | | | |
|---|---|---|---|---|---|---|---|---|
| **Soft margin with gaussian kernel** | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 | C = 0.1 | C = 0.6 | C = 1.1 | C = 2.1 |
| **σ = 0.0005** | 0.8320 | 0.8305 | 0.8600 | 0.8945 | 0.8424 | 0.8431 | 0.8730 | 0.8971 |
| **σ = 0.0001** | 0.8970 | 0.8965 | 0.8960 | 0.8965 | 0.9023 | 0.9023 | 0.9010 | 0.9010 |

Table 4 RESULTS OF SVM CLASSIFICATION

For Mercer condition, the Gram matrix is not positive semi-definite. (Mercer condition false)

## 4.2 Optimization of soft margin with polynomial kernel

The other way is to improve the parameter of SVM with soft margin with polynomial kernel. From the 3.3, it is obvious that with p=1 and little soft margin parameter C the SVM could reach higher accuracy.

From the table below, we could find that the optimal parameter for SVM with soft margin with polynomial kernel is p=1 and C=0.04. The accuracy could reach 92.7%. what's more, the kernel also fulfill the Mercer condition.

Finally, this project choose the SVM with soft margin with polynomial kernel and the parameter is p=1 and C=0.04. The file's name is task 3.

| **Type of SVM** | **Training accuracy** | **Test accuracy** |
|---|---|---|
| **Soft margin with polynomial kernel** | P=1 | |
| **C = 0.01** | 0.9205 | 0.9186 |
| **C = 0.02** | 0.9255 | 0.9206 |
| **C = 0.03** | 0.9255 | 0.9251 |
| **C = 0.04** | 0.9280 | 0.9271 |
| **C = 0.05** | 0.9280 | 0.9251 |
| **C = 0.06** | 0.9260 | 0.9219 |
| **C = 0.07** | 0.9260 | 0.9232 |
| **C = 0.08** | 0.9275 | 0.9232 |
| **C = 0.09** | 0.9265 | 0.9206 |
| **C = 0.10** | 0.9255 | 0.9193 |

Table 5 RESULTS OF SVM CLASSIFICATION