DASC 521 Homework 06: One-Versus-All Support Vector Classification

Ender Erkaya

December 2021

1 Introduction

In this homework, we apply SVM Kernel classification for multiclass data using binary classification algorithm demonstrated in lab08. To use binary classification for multiclass problem, two approaches are present: "One-vs-Other", "One-vs-All" One-vs-Other approach aims to develop classifiers for any two chosen class among K class. The number of classifiers is therefore: $comb(K, 2) = \frac{K(K-1)}{2}$. One-vs-All approach applied in this homework aims to develop binary classifiers for each class c versus its complement, other classes $\neq c$. To use the binary SVM Kernel algorithm developed in lab08, we recreated y^c_{train} for each class c by denoting $y_{train}(y_{train} == c) = 1$ as "1" and $y_{train}(y_{train} \neq c) = -1$ as "-1" otherwise. At the end, K different classifiers are developed for each class and we pick the maximum one for each K score function.

2 SVM Kernel Classifier

We use dual convex optimization setting to solve SVM Kernel Classifier using Gaussian Kernel as a similarity measure.

Dual Problem in Matrix-Vector Form

$$
\begin{aligned}
\text{minimize} \quad & -\mathbf{1}^\top \boldsymbol{\alpha} + \frac{1}{2}\boldsymbol{\alpha}^\top((yy^\top) \odot \mathbf{K})\boldsymbol{\alpha} \\
\text{with respect to} \quad & \boldsymbol{\alpha} \in R^N \\
\text{subject to} \quad & \boldsymbol{y}^\top \boldsymbol{\alpha} = 0 \\
& \mathbf{0} \leq \boldsymbol{\alpha} \leq C\mathbf{1} \\
\text{where} \quad & C \in R_+
\end{aligned} \tag{1}
$$

3 Confusion Matrix for Training Data

The algorithm is trained by solving quadratic programming using cvx solver. The confusion matrices obtained for test and training data are given below:

y_train y_predicted	1	2	3	4	5
0	207	1	0	9	0
1	2	199	1	1	0
2	0	1	204	6	0
3	0	1	4	185	1
4	0	0	0	0	178

Figure 1: Confusion Matrix for Training Data

4 Confusion Matrix for Test Data

y_test y_predicted	1	2	3	4	5
1	641	23	3	137	9
2	43	714	27	40	4
3	4	39	666	90	10
4	100	32	69	541	16
5	12	2	6	15	757

Figure 2: Confusion Matrix for Test Data

5 Accuracy vs Regularization Parameter C

SVM algorithm is applied choosing different regularization parameter C with Gaussian kernel width parameter $s = 10$. Accuracy is calculated as:

$$Accuracy = \frac{trace(confusion\ matrix)}{N}$$

The obtained plot is given below in figure 3:

According to plot, for better generalization, ie better test performance, the regularization parameter $C \approx 10$. It means regularization parameter has optimum. Further, forcing regularization reduces the accuracy on test data, without changing the training data accuracy. Hence, for better generalization regularization parameter should be adjusted properly, by checking accuracy on test data.

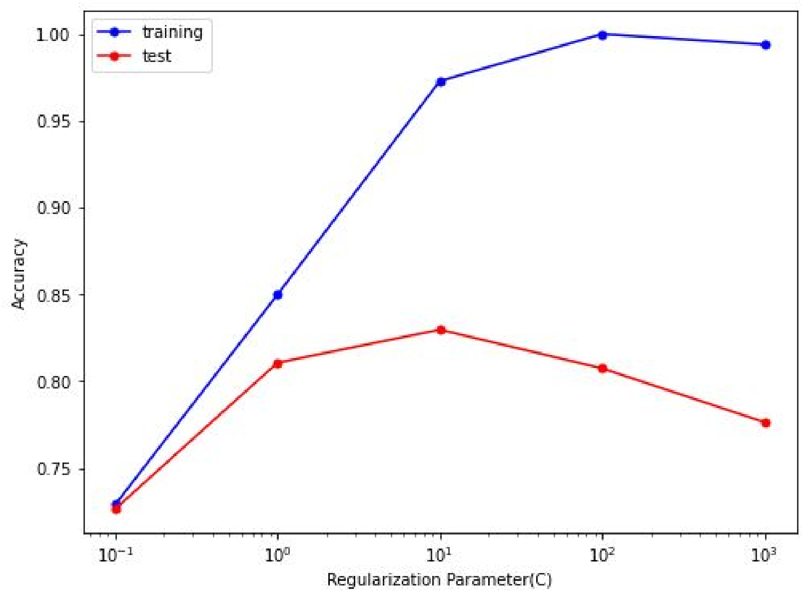

Figure 3: Accuracy vs Regularization Parameter(C)