SENG 474
Yidan Long
V00898887
July 16. 2020

Assignment 2

I. Logistic Regression

1.1 Data Explanation

Before any implementations, I first adjusted the training set size by extracting classes of 5 and 7 (for 12,000 training examples in total) from the original Fashion-MNIST data.

Parameters	Values
Regularization parameter c	10e-5 to 10e5 (10 values in total)
Maximum iterations	10,000
Number of Training examples	12,000
Number of Test set examples	2,000

Table 1: Chosen parameters for logistic regression

The line plot of changes of training and test errors with varied regularization parameter C is shown in section 1.2 (figure 1), and the graph takes the parameters exposed in Table 1 above.

The reason that I chose these C values is that they clearly illustrate the regime of underfitting and overfitting. The underfitting happens when C is ranged from 10e-5 to 10e-4, since the training set is larger than test error at that range in a decreasing trend, and overfitting happens when C is ranged from 10e-4 to 10e2 where the test error increases while the training error still decreases.

1.2 Graph and Analysis

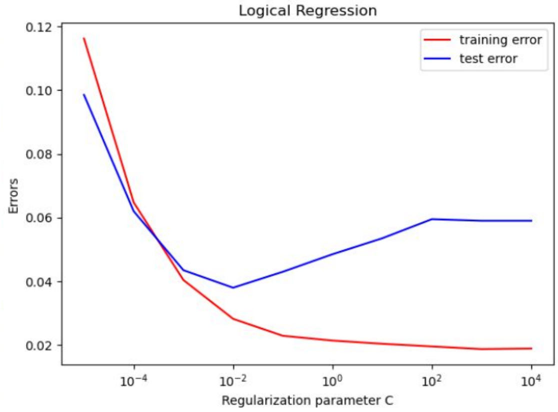

Figure 1: Changes of training and test error as varying the regularization parameter C

According to the line graph in Figure 1, the logistic regression model achieves a performance of 40% when regularization parameter C values 10e-2. While the training error and test error both decreases rapidly at first and transformed into a flat decrement as C value surges, the training error reaches the lowest point when the regularization parameter C is 10e-2. After C reaches 10e-2, the test error starts to gradually bounce up, and the error increases from about 0.04 to 0.06. When C ranges from 10e2 to 10e3, with a very little decrement on the line, the error drops from about 0.06 to 0.05.

It is interesting to conclude that before C values 10e-4, the training error is always slightly larger than the test error when underfitting happens. But after the intersection, the overfitting happens, the test error became always larger than the training error.

II. Linear kernel SVM

2.1 Data Explanation

I used the same adjusted training set in the previous logistic regression part (Section 1) which contains 12,000 training examples in total.

Parameters	Values
Regularization parameter c	10e-6 to 10e5 (11 values in total)
Maximum iterations	4,000
Number of Training examples	12,000
Number of Test set examples	2,000

Table 2: Chosen parameters for linear kernel SVM

The line plot of changes of training and test errors with varied regularization parameter C is shown below (figure 2), unlike the values that I used for logistic regression, I used different values of C in linear kernel SVM part (11 values of regularization parameter C in total), as shown in Table 2 above.

2.2 Graph and Analysis

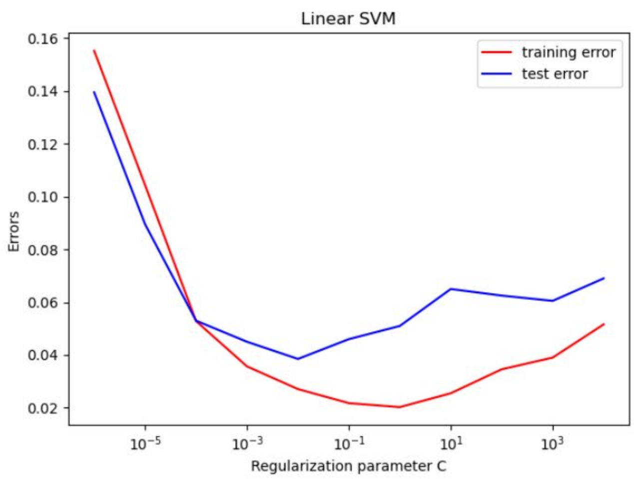

Figure 2: Changes of training and test error as varying the regularization parameter C

The line graph Figure 2 illustrates that the linear SVM model achieves a performance of 39% when C is 10e-2. While the training error first dips suddenly from C values 10e-6 to 10e0 and error drops from about 0.16 to 0.055, the trend transformed to a slow decrement, after training error reaches the lowest point, about 0.02 when C is 10e0, it starts to increase and error bounced up from about 0.02 to 0.05.

Simultaneously, the test error also decreases in a fast way at the beginning when C ranged from 10e-6 to 10e-4, and error drops from about 0.14 to 0.05, then it slows down the decrement from about 0.05 to 0.04 at the range c values from 10e-4 to 10e-2. Starting from 10e-2, the test error transformed into an increasing trend, the error changes from about 0.04 to 0.07. However, when C ranged from 10e1 to 10e3, the test error slightly decreases from 0.065 to 0.060.

It is interesting to conclude that before C values 10e-4, the training error is always slightly larger than the test error. But after the intersection happens when C is 10e-4, the test error became comparably larger than training error, which is similar to the trend in the logistic regression graph in section 1.2 (Figure 1).

III. K-folds

3.1 Logistic Regression Using k-folds:

3.1.1 Data Explanation

I used 6 as my k value to select the optimal value of the logistic regression regularization parameter c, and the line graph of changes of training and test error is shown below in Figure 3.

Parameters	Values
K value	6
Regularization parameter c	10e-5 to 10e5
Number of training set	2,000
Optimal value of c	10e-2

Table 3: Chosen parameters for logistic regression via k-fold cross-validation

3.1.2 Graph and Analysis

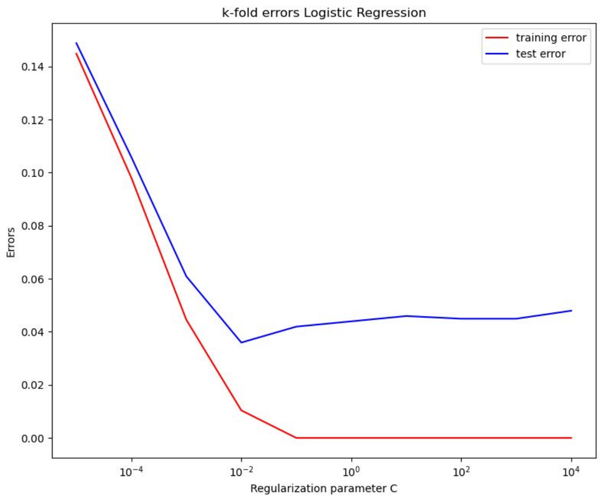

Figure 3: Training and test error changes as varying the regularization parameter C in logistic regression

The line graph gives information about changes in training and test errors using k-fold cross-validation on logistic regression. The test error reaches the lowest point when C is 10e-2 which is about 40%. The test error is always larger than the training error as varying regularization parameter C, as C increases, the interval also increases.

The training error drops fast as C value increases to 10e-2. The test error shares the same trend when C ranged from 10e-5 to 10e-2. Starts from C values 10e-2, both the training error and test error perform a flat trend. However, training error increases from about 0.04 to 0.045, and test error decreases from about 0.02 to 0.015.

The optimal regularized logistic regression classifier is regularization parameter C values 10e-2 since the test error reaches the lowest point (performance of 40%).

3.1.3 Comparison Analysis

By comparing the two training and test error line graph in section 1.2(figure1) and in section 3.1.2(figure 3). I conclude that after using k-fold validation, the best performance is about 40%. However, the difference is that intersection of training and test error did not happen in logistic regression using k-fold validation. Also, the test error in the regular logistic regression model bounced up a bit from the lowest point of 0.04 to 0.06 while test error using k-fold cross validation remains about 40%.

3.2 Linear SVM Using k-folds:

I still used 6 as my k value to select the optimal value of the linear SVM regularization parameter, and the Linear SVM line graph is shown below in Figure 4.

3.2.1 Data Explanation

Parameters	Values
K value	6
Regularization parameter c	10e-5 to 10e5
Number of training set	2,000
Optimal value of c	10e-2

Table 4: Chosen parameters for linear kernel SVM via k-fold cross-validation

3.2.2 Graph and Analysis

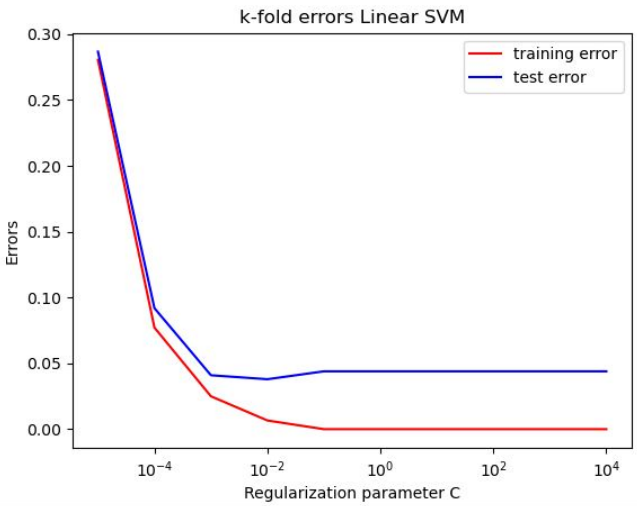

Figure 4: Training and test error changes as varying the regularization parameter C in linear kernel SVM

Both the training error and test error plummet rapidly from 0.30 to 0.08 when C increases from 10e-5 to 10e-4, then the test error drops to 0.04 while training error drops to 0.03 when C is 10e-3. After two trends of rapid decrement, the test error drops to 0.04 when C is 10e-2, which is the best performance in linear kernel SVM. Then the test error rises a bit to 0.048 and remains flat.

The optimal regularized logistic regression classifier is regularization parameter C values 10e-2 since the test error dives to the lowest point (performance of 40%).

3.2.3 Comparison Analysis

After comparing the two training and test error line graph in section 2.2(figure 2) and in section 3.2.2(figure4). I conclude that after using k-fold validation, the overall trend is similar, both abruptly declines at first. The training error and test error did not intersect through the whole trend when using k-fold cross-validation. However, in the regular mode of linear SVM, the errors intersect before the test error reaches the lowest point. The difference between the two trends using different approaches is that both the training error and test error gradually surged when not using k-fold cross-validation.

IV. K-fold Gaussian Kernel SVM

4.1 Data Explanation

Parameters	Values
K value	5
Regularization parameter c	10e-5 to 10e3
Scale parameter gamma	10e-5 to 10e3
Number of training set	2,000

Table 5: Chosen parameters for gaussian kernel SVM via k-fold cross-validation

4.2 Graph and Analysis

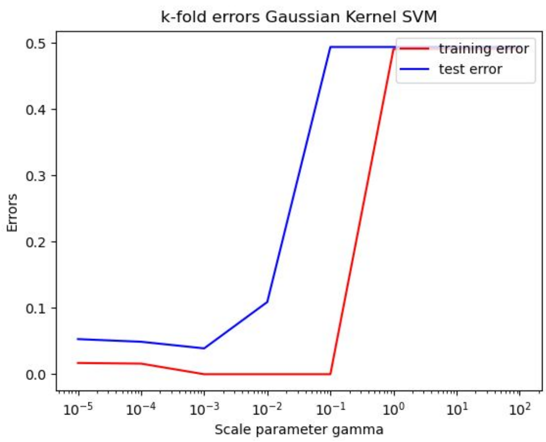

Figure 5: Changes of training and test error as varying the regularization parameter C and scale parameter gamma

As shown in the Figure 5, it is clear to see that the test error surges from about 0.39 to 0.5. When gamma value from 10e-3 to 10e-2, the increasing trend in not too steady. However, when gamma increases to 10e-1, there is a significant rise happens. The optimal gamma value is 10e-3 where the test error is the lowest, about 0.39, and the best regularization parameter C value is 10.

4.3 Comparison Analysis

Compare to the test error of linear SVM with regularization parameter tuned via cross-validation on part 3, as shown in Table 6 below, the errors are very close.

SVM model	Best performance
Linear SVM	40% when C values 10e-3
Gaussian SVM	39% when gamma values 10e-1

Table 6: Comparison of test error using k-fold cross-validation on either linear SVM varying regularization parameter C, and Gaussian SVM varying regularization parameter C and scale parameter gamma