Programming 2

Ruiwan Xu

1. Perceptron algorithm

Figure 1 and Figure 2 shows the training error rate and testing error rate of the original data, data with l1 norm and data with l2 norm, where the black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm. From the figures we can observe that data with l1 norm works best, because the training and testing error rate is relatively low.

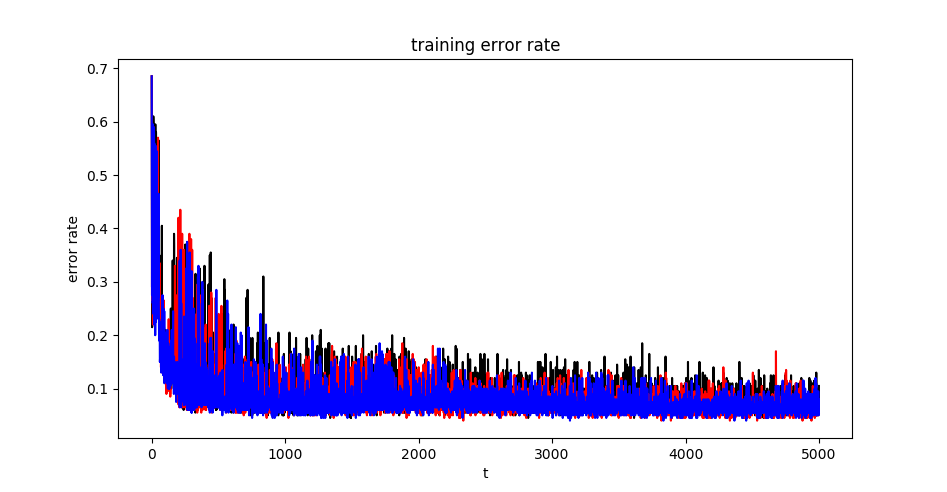


Figure 1: Training error rate of Perceptron Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.

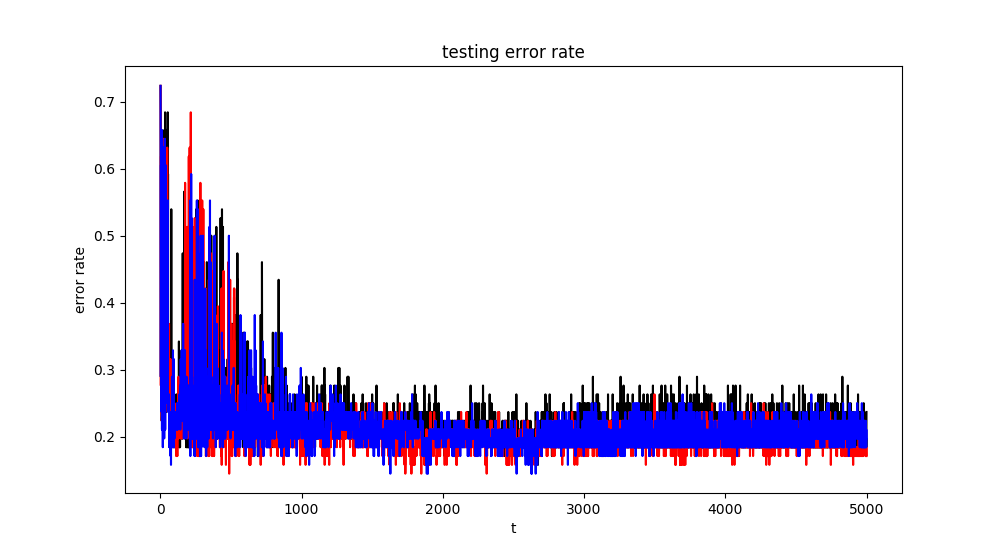


Figure 2: Testing error rate of Perceptron Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.

Figure 3 and Figure 4 shows the training and testing error rate of Averaged Perceptron Algorithm where the black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.

It is obviously that that the Averaged Perceptron Algorithm works better than the Perceptron Algorithm. The error rate is smoother with less vibrations due to the average of the historical weights.

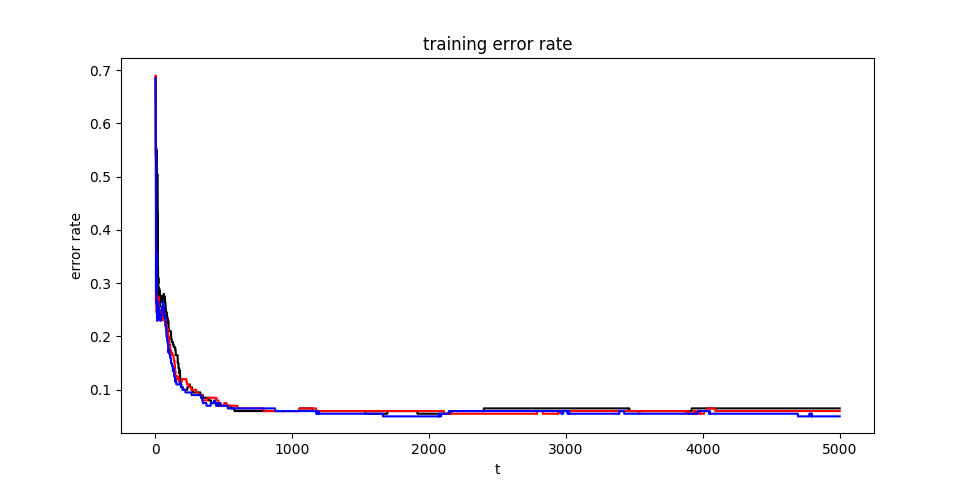


Figure 3: Training error rate of Averaged Perceptron Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.



Figure 4: Testing error rate of Averaged Perceptron Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.

1. Gradient Descent

Figure 5 and Figure 6 shows that training and testing error rate of gradient descent algorithm with respect to iterations where the black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm. The original data seems to work best. It converges more quickly than the other two and has the lowest training and testing error rate.

The gradient descent algorithm has an approximate training and testing error rate with perceptron algorithm. The gradient descent algorithm converges slower than perceptron due to the zigzag feature of the gradient descent which converges around 10000 iterations while perceptron converging around 5000 iterations.

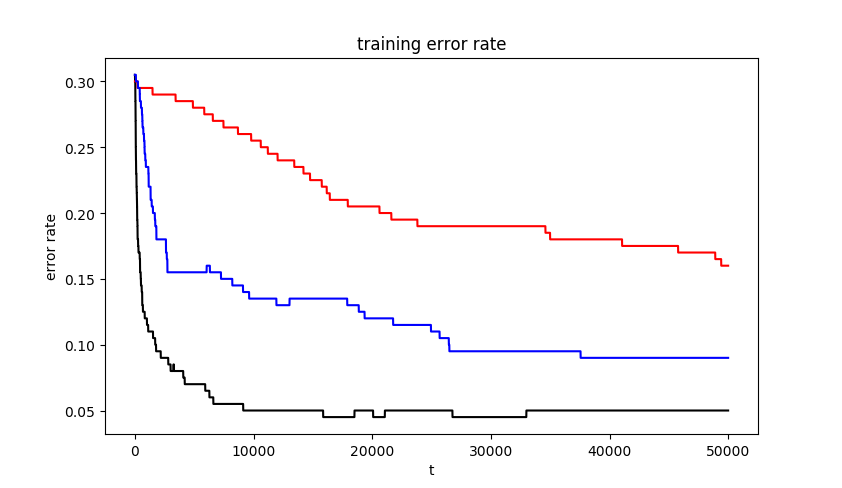


Figure 5: Training error rate of Gradient descent Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.



Figure 6: Testing error rate of Averaged Perceptron Algorithm: The black line denotes original data, the red line denotes data with l1 norm and the blue line denotes data with l2 norm.

1. Locally weighted Logistic Regression

The gradient is

The Hessian matrix is

, where

Figure 7, 8 and 9 shows the training error rate of locally weighted logistic regression of original data, data with l1 norm and data with l2 norm with respect to different. The black line denotes, the red line denotes, the blue line denotes, the cyan line denotes, the green line denotes, the magenta line denotes.

The training error rate is much lower than the perceptron, especially when, the training error is zero. When becomes larger, the training error will grow. Data with l2 norm converges at a fastest rate.

|  |  |
| --- | --- |
| Figure7: Training error rate of original data with respect to different tau's | Figure 8: Training error rate of data with l1 norm with respect to  different tau's |

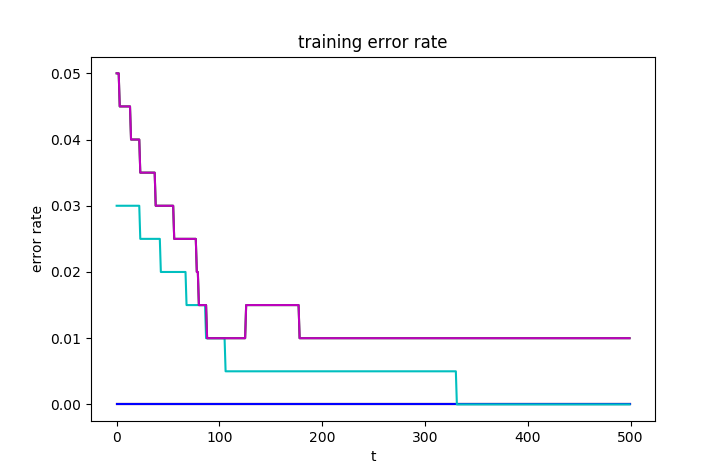


Figure 9: Training error rate of data with l2 norm with respect to different tau's

Figure 10, 11 and 12 shows the testing error rate of locally weighted logistic regression of original data, data with l1 norm and data with l2 norm with respect to different. The black line denotes, the red line denotes, the blue line denotes, the cyan line denotes, the green line denotes, the magenta line denotes.

When is relatively larger, the testing error rate is lower and larger converges faster which is opposite to the training error. It means that when is too small, it will lead to the over fitting of the training data.

The testing error converges around 0.1 of perceptron and around 0.2 of averaged perceptron. When using original data and data with l2 norm, the testing error of local weighted logistic regression is lower than perceptron. Also, data with l2 norm converges at the fastest rate. Due to the algorithm is time-consuming, only 500 iterations are taking into simulation.

|  |  |
| --- | --- |
| C:\Users\Administrator\Desktop\ori_test.png  Figure 10: Testing error rate of original data with  respect to different tau's | C:\Users\Administrator\Desktop\l1_test.png  Figure 11: Testing error rate of data with l1 norm with  respect to different tau's |

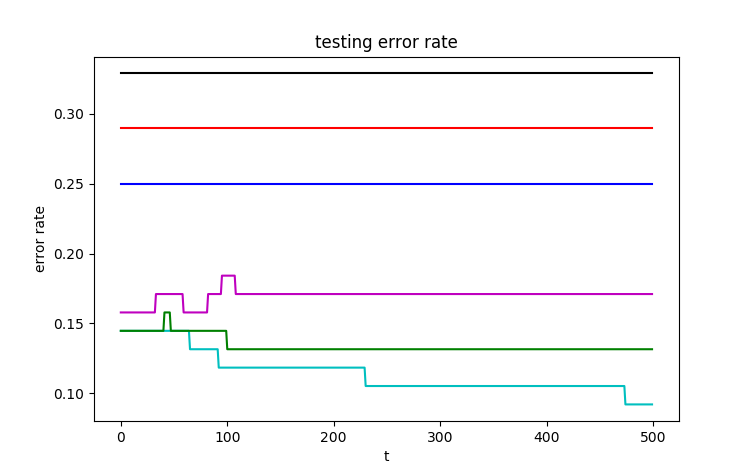


Figure 12: Testing error rate of data with l2 norm with respect to different tau's