Weicong Feng

Professor Chia-Ling Tsai

CSC74020 Machine Learning

11/12/2021

Assignment3

Report

# Solution

## Logistic Regression with Regularization

Logistic regression is one member of linear family and commonly used in multiply classification and possibility proximation. Usually, it can’t be shown:

where

Regularization is main approach to limit the model become too complicated to generalize. I use weight decay in this assignment. The error measure with weight decay is shown in next session.

## Error Measure

Without Regularization, the error measure should be:

and the gradient is:

With weight decay, the error measure should be:

Where or

## Feature Reduction

I use PCA to reduce feature dimensions:

A picture containing text, clock, watch

Description automatically generated

A picture containing text, clock

Description automatically generated

# 5-fold Cross Validation

## Experiment

### Preprocess:

After loading data, the PCA, polynomial transformation and standard scaling are applied on the data before feeding to the model. Here, I set the n\_component of PCA as 0.99, and the degree of polynomial transformation is 10.

### Folding:

After preprocessing and before training the model, 5-fold is applied on the training data set. The training data set is split to 5 subsets. One of them is used as validation data set in each time, others are combined to new training data set.

### Train:

Because we need to observe the model performance in various weight decay parameter of lambda, I create a list or array of lambda. Then train the model with different lambda value in a loop, the return of loss in and loss out are stored in two lists, l\_in and l\_out.

I use two formular versions to train, one being same as our textbook, another being common online. The loss function and gradient function on textbook are shown above, the online version is:

Text

Description automatically generated

A picture containing diagram

Description automatically generated

### Visualize：

After getting the mean of the loss in and loss out in 5-folds, we can obtain the average of loss in and loss out in various lambda. However, only the last value of loss is used to plot the chart as the loss values during the training are meaningless. In the charts, x-axis is lambda value, y-axis is the loss value. The loss in is plotted as blue color while the loss out is plotted as red.

## Result

Chart, line chart

Description automatically generated

Online formular

Chart, line chart

Description automatically generated

Online formular

Chart

Description automatically generated

Textbook formular

Chart, line chart

Description automatically generated

Textbook formular

## Discussion

At first, I think the two different formular should be equivalent, only the form is different. However, the experiment shows different behaviors.

With uniform regularizer, both formulars work similarly, the charts show almost two upward sloping parallel lines. Only a discernible drop appears at the beginning with tiny value of lambda. This is because the model is not complicated enough, the regularization can’t improve the model performance. Instead, as the lambda rises, the model pays less attention to the original loss value and lead to both losses goes up.

With low-order regularizer, two formular versions work variously. The textbook formular version works similar to the uniform regularizer. However, the online formular version shows obvious performance improvement at tiny value of lambda. I believe this is because low-order regularizer is more effective in suppressing high-order terms, but the model is not too complicated. So, the weight decay turns to harm the model performance when the lambda grows further.

In addition, that the gap between loss in and loss out is close enough is another reason why the regularization doesn’t show distinct improvement in loss out.