DD2424 Deep Learning in Data Science

Yuchen Gao

March 28, 2022

Introduction

This is the report for assignment 1 bonus exercise, the improvement for implementation of basic feed forward network and backward propagation. The comparison between the application of sigmoid and softmax activation function.

Improving model

Improvements implementation

1. The flip trick to turn the pictures upside down every 2 epochs to avoid over-fitting
2. The learning rate decay, set a decay parameter to reduce the learning rate every 10 epochs so that the optimization at the beginning is faster, and when it comes to converge, it can be slow down.
3. Loading all of the batches to train, this increase the size of training batch, but it will cost the compute efficiency.

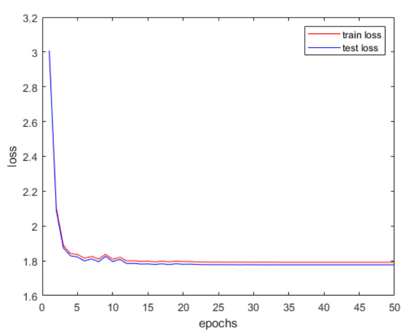

Figure 1: Loss function for performance improved version setting.

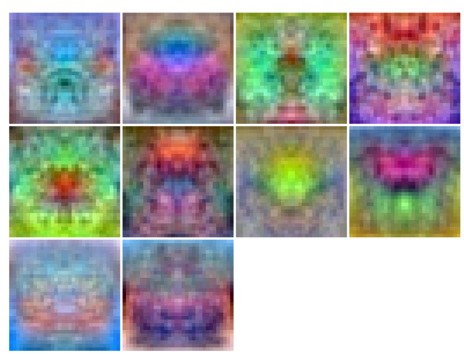

Figure 2: Template for performance improved version setting.

During the model improvement, I found that the most efficient way to increase accuracy is to randomly shuffle data at the beginning of each epoch.

Accuracy matrix

Best accuracy	0.45

Table 1: The best accuracy performance

Binary cross-entropy

The derivative:

$$\begin{aligned} \frac{\partial l_{cross}}{\partial p_i} = \frac{-y_i}{p_i} + \frac{1-y_i}{1-p_i} = \frac{p_i - y_i}{p_i(1-p_i)} \\ \frac{\partial p_i}{\partial s_i} = p_i * (1-p_i) \\ \frac{\partial l_{cross}}{\partial s_i} = \frac{\partial l_{cross}}{\partial p_i} * \frac{\partial p_i}{\partial s_i} \\ = p_i - y_i \end{aligned} \tag{1}$$

From the deduction above, we can conclude that the gradient computation result is identical to the categorical cross-entropy, so there is no need to re-write the gradient descent function.

The graphs

Cross-entropy loss

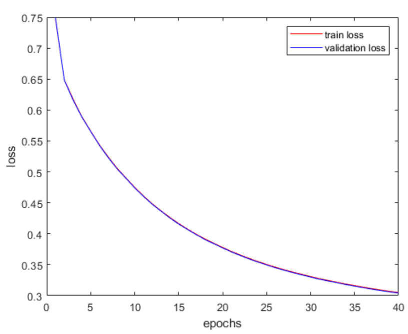

Figure 3: Loss function for binary cross-entropy.

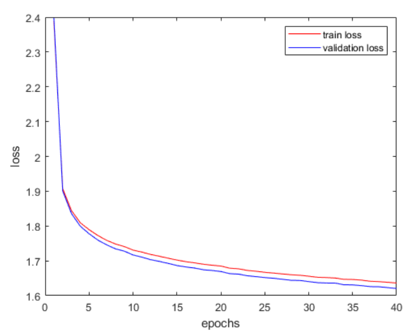

Figure 4: Loss function for categorical cross-entropy.

Comments

1. From the loss graph above, we can deduct the binary cross-entropy has less likely to come across over-fitting problem
2. The categorical loss function converges much faster than the binary loss
3. The binary cross-entropy has the ultimate accuracy around 0.4528 on the test set, the the accuracy of categorical cross-entropy is 0.4587(with random shuffle at the beginning of each epoch). This means the performance is generally the same when we are dealing with this specific task, but the binary approach can be used when the classes are not strictly opposed from each other, especially when they have something in common and the instances may even belong to number of categories.

Histogram compare

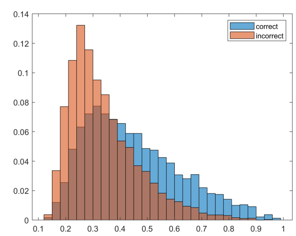

Figure 5: Histogram for categorical loss.

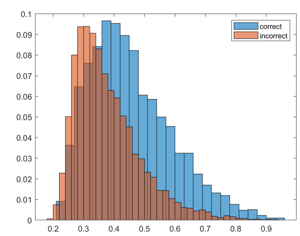

Figure 6: Histogram for binary loss.

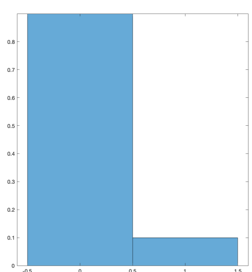

Figure 7: Histogram for the ground truth label matrix.

The ground truth label matrix shows the distribution of the appearance of 0 and 1, the more Figure 6 is similar to the Figure 7, the better performance is. As the binary make judgements about each element in the matrix, the determination procedure is executed one after another, so the distribution is closer to 0, as each part contributes a little to the decision instead of decide alone.