Loss function analysis

Pan Wen

October 2023

1 Introduction

In this task, we will analyze the impact of four different loss functions on the model

```
nn.L1Loss
nn.CrossEntropyLoss
FocalLoss(gamma = 2)
FocalLoss(gamma = 0.5)
```

With hyperparameters

```
# random seed
SEED = 1
NUM_CLASS = 10
# Training
BATCH_SIZE = 128
NUM_EPOCHS = 30
EVAL_INTERVAL=1
SAVE_DIR = './log'
# Optimizer
LEARNING_RATE = 1e-1
MOMENTUM = 0.9
STEP=5
GAMMA=0.5
```

2 L1 loss

2.1 definition

The L1 loss, also known as the mean absolute error (MAE), is a type of loss function used in regression tasks. It measures the average absolute difference between the predicted and actual values. Mathematically, it is defined as:

$$loss(x, y) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|$$

Using L1 loss for classification would not be appropriate because it does not reflect the nature of the problem. L1 loss measures the absolute differences

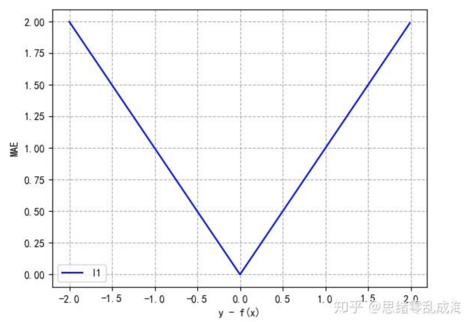

Figure 1: L1loss

between predicted and true values, which doesn't make sense when dealing with discrete class labels.

2.2 test result

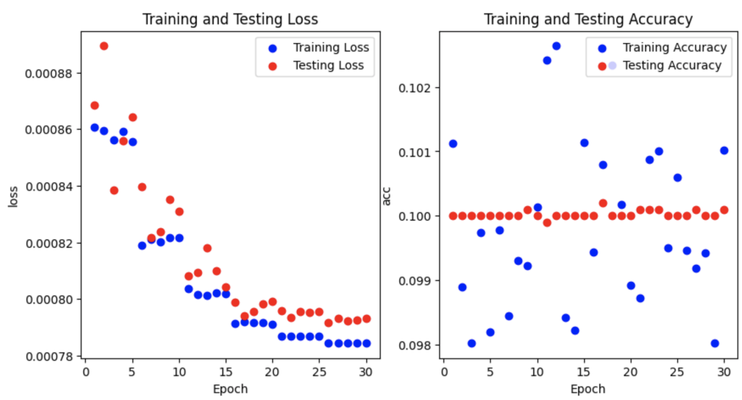

Figure 2: loss and accuracy for L1 loss

figure 2 shows that although the loss continues to decrease as training progresses, the accuracy has not increased and remains around 0.1, proving that L1 loss is not suitable for classification problems

3 Cross Entropy loss

3.1 definition

Cross-entropy loss, often referred to simply as "cross-entropy," is a loss function used in machine learning and particularly in classification tasks. It measures the performance of a classification model whose output is a probability value between 0 and 1.

The cross-entropy loss between predicted and true probability distributions y and $\hat{y}$,respectively, is defined as:

$$H(y, \hat{y}) = -\sum_i y_i log(\hat{y}_i)$$

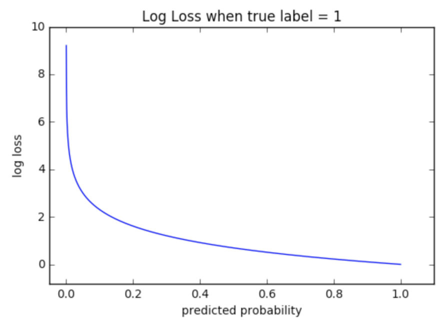

Figure 3: Cross Entropy loss

3.2 advantage and shortage

there are many advantage when using cross entropy loss

(1)Cross-entropy loss penalizes the model more when it makes confident incorrect predictions.

(2)It can effectively handle multi-class classification problems.

(3)Using the sigmoid activation function with mean squared error (MSE) loss for classification tasks can lead to the vanishing gradient problem, especially when the model makes highly confident, but incorrect, predictions. Cross-entropy loss doesn't suffer from this issue to the same extent.

(4)The gradients of cross-entropy loss with respect to the model parameters are well-defined and lead to efficient optimization. This means that the loss function is easy to use with various optimization algorithms.

But Cross Entropy Loss also has drawbacks. When the distribution of categories is imbalanced, Cross Entropy Loss may lean towards the majority of categories, leading to a decrease in the performance of the model on a few categories.

3.3 test result

figure 4, it can be seen that the loss of the training and testing sets is constantly decreasing, and the accuracy of the training and testing sets is constantly increasing. This indicates that the model updates parameters during training to improve the accuracy of prediction

Figure 4: Loss and accuracy for Cross Entropy Loss

4 Focal loss

4.1 definition

Focal Loss is a modified loss function designed to address class imbalance in binary classification problems. The focal loss is particularly useful in tasks where the majority of the samples belong to one class, leading to a class imbalance problem.

γ is the focusing parameter that allows the loss to down-weight easy examples and up-weight hard examples. It modulates the effect of the cross-entropy term. Generally speaking, a smaller γ value makes the model more sensitive to easily classified samples, while a larger γ value makes the model more focused on difficult to classify samples. The focal loss is defined as:

$$FL(y, \hat{y}) = -\alpha(1-\hat{y})^{\gamma} log(\hat{y})$$

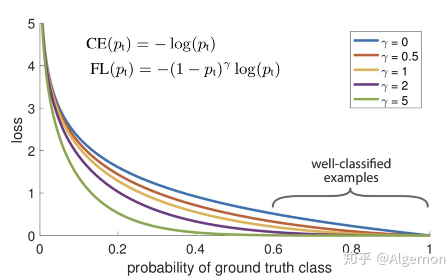

Figure 5: Focal loss

The key idea behind focal loss is to down-weight the contribution of well-classified examples (those with high predicted probabilities) during training, focusing more on the harder examples. This helps to address the problem of class imbalance.

4.2 advantage and shortage

there are many advantage when using focal loss

(1)Focal Loss reduces the emphasis on easily classified samples, making the model more focused on difficult to classify samples.

(2)Focal Loss was originally proposed for tasks such as dense object detection, where there are far more easily found background samples than target samples, making Focal Loss particularly effective.

there are also many disadvantage of focal loss

(1)Focal Loss introduces two hyper parameters γ and α that need to be adjusted and optimized during the training process

(2)The effectiveness of Focal Loss depends on the characteristics of the dataset and the nature of the classification problem, and may not always be superior to Cross Entropy Loss.

4.3 test result when $\gamma = 2$

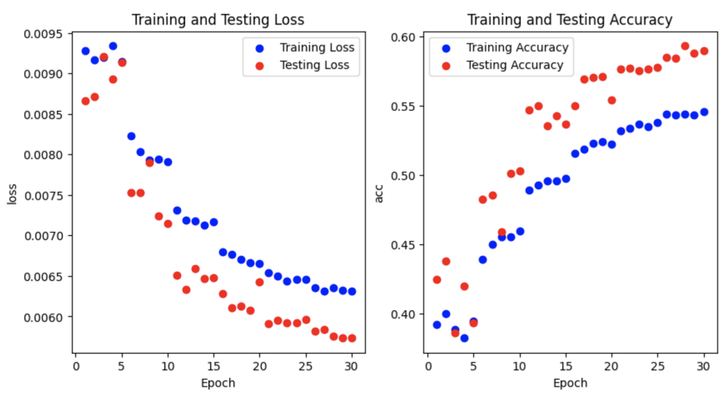

Figure 6: Focal loss $\gamma = 2$

The figure 6 shows the loss and accuracy of each epoch when γ is equal to 2. It can be seen that as the training progresses, the loss of the training and testing sets continues to decrease, while the accuracy continues to increase

4.4 test result when $\gamma = 0.5$

The figure 6 shows the loss and accuracy of each epoch when γ is equal to 0.05. It can be seen that as the training progresses, the loss of the training and testing sets continues to decrease, while the accuracy continues to increase, but the accuracy will be little higher than $\gamma = 2$

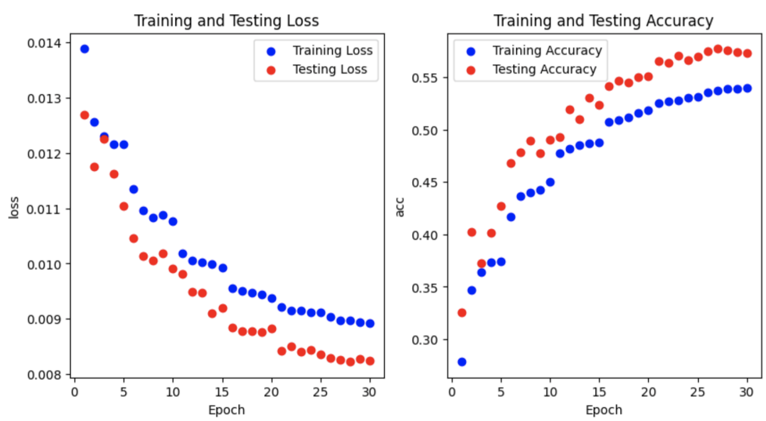

Figure 7: Focal loss $\gamma = 0.5$

5 summary

5.1 the last accuracy of model for different function

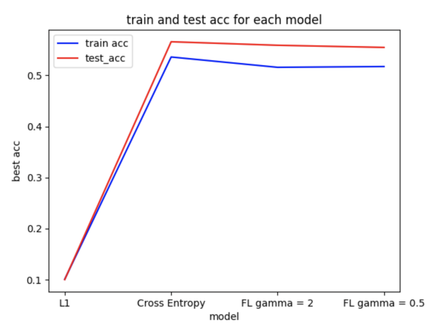

Figure 8: the last accuracy of model for different function

From the eighth figure, we can see the accuracy of the test and training sets for each type of loss function. From the figure, it can be seen that L1loss is the least suitable for image classification, and its accuracy is lower than other loss functions. The effects of CE loss and Focal loss are almost identical, and the difference in accuracy may be due to different training sets, Cross Entropy loss might be the best loss function for picture classification

5.2 the changing accuracy of each loss function

From the graph, it can be seen that the accuracy of L1loss is much lower than the other three loss functions, while Focal loss is lower at gamma=2 than Focal

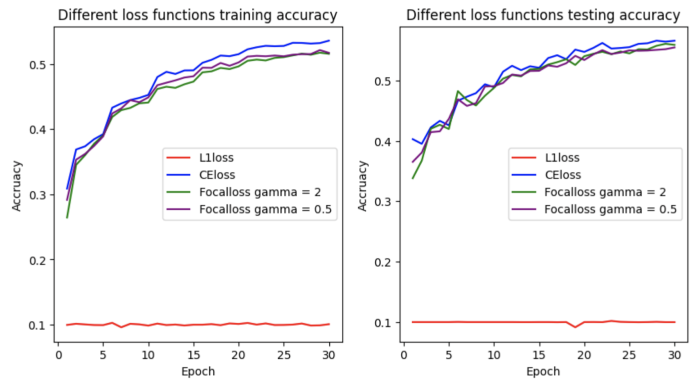

Figure 9: the changing accuracy of each loss function

loss at gamma=0.5 and Cross Entropy Loss. The accuracy of Focal loss is very close to that of Cross Entropy Loss at gamma=0.5

5.3 the changing precision of each loss function

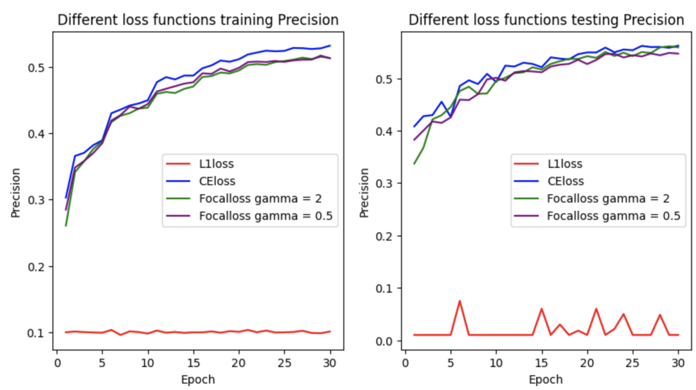

Figure 10: the changing precision of each loss function

Precision, is a performance metric used to evaluate the accuracy of a classification model. It gives the ratio of true positive predictions to the total number of instances predicted as positive. The difference in Precision and Accuracy is not significant, and the trend is basically the same. L1loss is the lowest, while Focal loss is inferior to Focal loss at gamma=0.5 and Cross Entropy Loss at gamma=2.

5.4 the changing recall of each loss function

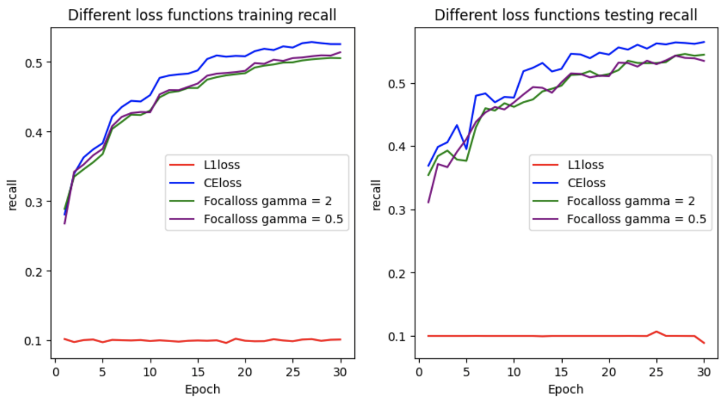

Figure 11: the changing recall of each loss function

Recall, is a performance metric used in the evaluation of classification models. It is particularly important in contexts where the consequences of false negatives are significant. From the graph, it can be seen that the highest recall rate is for crossEnterpyLoss, and the lowest is for L1loss Focal loss. The difference between gamma=2 and gamma=0.5 is not significant.

5.5 the changing f1 of each loss function

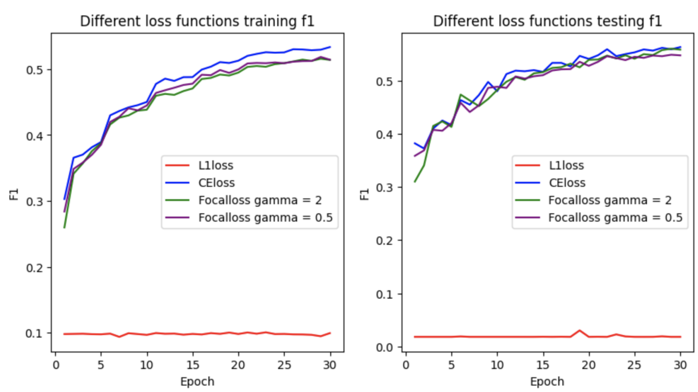

Figure 12: the changing f1 of each loss function

The F1 score is particularly useful for imbalanced datasets because it considers the performance of both positive and negative categories, and achieves a balance between positive and negative categories. The trend of f1 is basically the same as the trend of precision and accuracy