Project 1

20233508 DoHyoung Lee

1. Verifying the Model Selection

This model adopts 30-sized minibatch with 100 epochs and learning rate is 0.00015. There are two suggestions to improve the performance of model and to prevent overfitting: one is early-stopping and the other is changing learning rate. Early-stopping halts training model before the gap of train loss and validation loss increases. The model is set to stop when the training accuracy is above 95% and the validation accuracy differs from the training accuracy by 3 percentage points or less. Changing learning rate term decreases the learning rate then the model stops stepping around the global minima and fall into the fit. The learning rate in this model decays 0.9 times when the acquired validation accuracy remains unchanged (exactly, the gap is under 0.1 percent points). As a result, this model halts when the training accuracy was 95.278% and the validation accuracy is 92.727%, and this validation accuracy verifies this model is suitable for our image classification task.

A. Benefit of Using Convolution Layers and Max Pooling Layers

Convolution layers is efficient to process unstructured data, especially image processing and classification tasks. In unstructured data, it is hard to extract features since we do not know where the specific patterns can be features are will appear in the image. In addition, not all patterns are same by distortion or individual differences even though they represent the same features. Convolution layers can detect patterns by filters which name is kernel and extract features from images. For example, in this model, there are 32 output channels in the first convolution layer. This means this layer scans the image for the presence of 32 different patterns and their locations.

The number of channels are also important since when we use too many channels it will cause overfitting similar with the case when we use too many features to classify data. I used 32 channels as an input of fully connected layer (i.e. 8 times compared with the skeleton code), the training accuracy was 100%, however, the validation accuracy is underneath 92%.

Pooling layers down samples each channel. Max pooling layers returns maximum value among the stride grid. It reduces complexity of computation.

B. Benefit of Using the Hierarchical Structure of Multiple Layers.

Multiple layer perceptron (MLP) can supply nonlinear decision boundary. The boundary of abnormality is represented complicatedly on the feature space and MLP can operate 'nonlinear regression' to minimize the loss with boundary of abnormality.

2. Important Hyperparameters of My Simulation

In my simulation, the first important thing is learning rate. It starts with 0.00015, and it is enough value to prevent the model fall into the local minima. After the model finds global minima fit (I assumed that 10 epochs are sufficient), learning rate decays when the validation accuracy changes less then 0.1 percent points. The model which does not use learning rate decaying part shows 93.1 percent, higher than that of our model, but the loss is 0.0496 while that of our model is 0.0293. The smaller loss shows that our model can operate image classification with less variance, and this is why I selected our model as the best model.

The second component is epoch. I set the model halts when the training accuracy is over 95% (hence, prevent underfitting) and the gap of training accuracy and validation accuracy is under 3 percent points (hence, prevent overfitting). If I exclude this early stopping term, the train accuracy converges to almost 100% while the validation accuracy reaches peak, 94 percent, and fall down. Loss also increases after the epochs our best model halts. These two terms imply overfitting is occurred.

3. Learning Curves of My Best Model

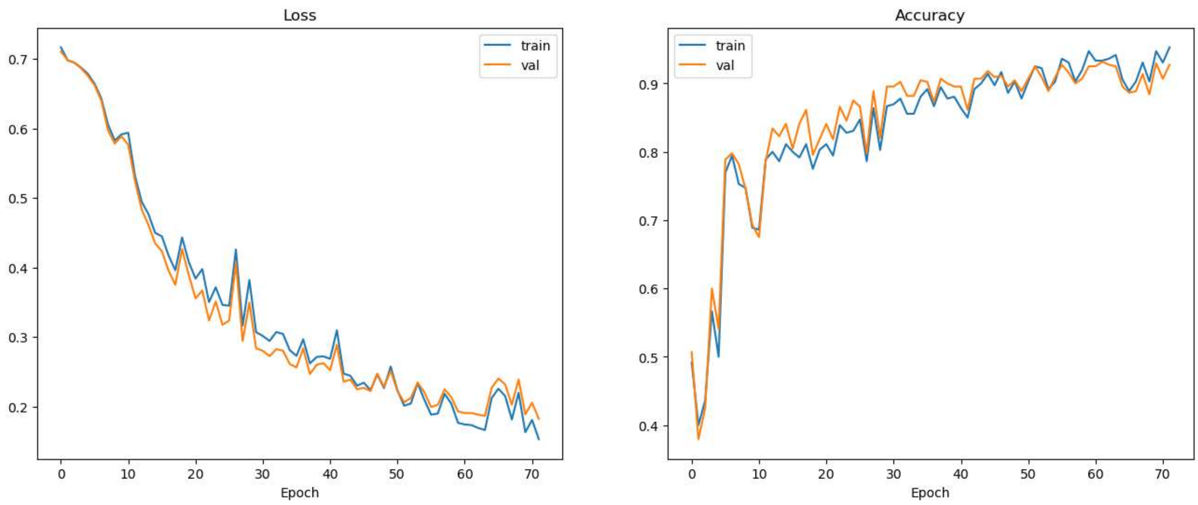

Figure 1 Loss Curve (left) and Accuracy Curve (right) of Train and Validation Data

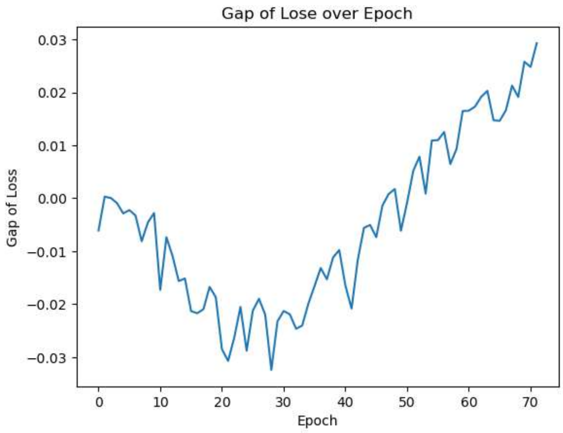

Figure 2 The Gap of Loss (validation – train) over Epoch

A. Is It Overfitted?

Figure 1 is learning curve of loss and accuracy. To determine whether the model is overfitted or not, I plotted the gap of loss (figure 2). This value decreases to almost 30 epoch and increases. Although the criteria of early stopping are about the accuracy, and we know loss and accuracy has low correlation, the gap of loss was lower than that of other models. Compared with other models, this model shows significantly low loss while conserve fairly high accuracy (over 90 percent), and this means this model does not be overfitted.

B. How to Prevent Overfitting?

There are several methods to prevent overfitting. One is early stopping, which is used in our model. As training data training continues, the model tends to fit only the training data. Early stopping prevents overfitting by make model halts before the model starts to fit only the training data.

Second, regularization can also prevent overfitting. When the model is overfitted, that means the decision boundary uses too much terms to avoid making a single exception for all training data. Regularization can reduce terms then allows 'broader' decision boundary. This boundary may decrease training accuracy, but this can be undone in a similar fashion to reverting to a previous epoch.

4. Why the Model Could Not Achieve the Perfect Performance?

Training data is similar with previous years' questions. Nobody can know what the questions will be this year. This is why the model cannot achieve the perfect performance. The test (or validation) data should not be used to training, so we cannot expect the model make right decision to never seen, test, data always.