COSE474 Project #1: MLP Implementation

2021320322 윤민서

In this project, it is important to implement the answer of short quiz and backpropagation in the lecture as a code. An important part of the forward pass is numpy's "keepdims" argument, which is responsible for maintaining the dimension of the multi-dimensional array in which the operations are carried out during max or sum operations. If the "keepdims" argument is not used, the N-1-dimensional array is returned when an operation is performed on the N-dimensional array, which also affects the later calculation results, resulting in an error. An important part of the backward pass is to align the dimension of the array to be calculated. We can use numpy's method called "reshape" to align dimensions.

After implementing the forward pass and the backward pass, a function of tuning the hyperparameter was implemented. The grid search method was used to find an appropriate combination, considering using values higher than a default value for the parameter of the train function. This is because the default value was 0.000005 for reg, 0.001 for learning rate, and 100 for epochs, which were considered lower than the appropriate hyperparameter value. However, if the regularization is too high, errors such as divided by zero and overflow occur, so you should be more careful about finding appropriate values.

As a result of using a lot of time to tune the hyperparameter, it was possible to obtain an accuracy that greatly exceeded the target test set accuracy of 36% (Test accuracy: 51.3%). The best combination of hyperparameters obtained through the validation set is hidden size: 200, learning rate: 0.001, reg: 0.1, epochs: 2000. Unlike the initial problem that the default value of the learning rate was low, the result that the default value of 0.001 was the most appropriate value was obtained. In addition, when using the same hyperparameter, if the epoch increases, the accuracy of the validation set increases.

If it was an environment where more time could be invested in learning, I think a higher accuracy could be obtained. However, due to the nature of the project with a deadline, I think it is also a task to find a good hyperparameter as possible within a given time.