Introduction to Machine Learning

Lab 4: Neural Network

Student ID: 111060013   Name: 劉祐廷

1. **What are the key differences between sigmoid and softmax activation functions, and why did we choose them for binary and multi-class classification respectively?**
   The sigmoid function primarily aims to make the output approach 0 or 1 as much as possible, achieving a binary distribution effect. In contrast, the softmax function compresses an input vector into a probability distribution vector with a sum of 1. Therefore, the sigmoid function is more suitable for binary classification models, while the softmax function is better suited for multi-class classification models.

2. **Why does the loss oscillate during model training? (list at least 2 reasons)**
   When the learning rate is slightly too high, it can cause gradient descent to over-correct, leading to sudden increases in loss. With SGD, since each update of w is calculated using only a small portion of the data, the process is less stable, which can result in oscillations in the loss during training.

3. **How does changing the learning rate and batch size affect model training time?**
   When the learning rate is set higher, the training speed will be faster, but it's important to pair it with an appropriate decay rate to avoid issues where the learning rate becomes too large and causes the training to diverge. Additionally, a smaller batch size speeds up training; compared to larger batch sizes, a smaller batch size can achieve better results with fewer epochs.

4. **Regression results**