

Machine Learning

Lab4 Report

- I. What are the key differences between sigmoid and softmax activation functions, and why did we choose them for binary and multi-class classification respectively?

Sigmoid 和 Softmax 都是常用的激活函數，但它們的主要區別在於輸出範圍和用途。

Sigmoid 函數的輸出範圍為 0 到 1，適合於單個輸出值的二分類問題，因為它可以將輸出解釋為「屬於某類別的概率」。而 Softmax 函數會將多個輸出值規範化，使它們的總和為 1，因此適合於多分類問題，可以將每個類別的輸出視為該類別的概率。在二分類問題中，模型的輸出是一個單一的概率值，因此使用 Sigmoid 函數來限制輸出範圍，使其可以解釋為目標類別的概率。而在多分類問題中，需要為每個類別提供一個概率，並且這些概率的總和必須為 1。Softmax 函數正好可以滿足這個要求，將多個輸出轉換成符合概率分佈的值，因此適合於多分類問題。

- II. Why does the loss oscillate during model training?

當學習率過大時，每次更新模型參數時，可能導致模型在接近最佳點時反復越過這個點，造成損失函數的震盪。或是使用小批量（mini-batch）進行梯度下降時，由於每個批次中的樣本並不完全相同，梯度計算會有所偏差，這也可能導致值的波動。

- III. How does changing the learning rate and batch size affect model training time?

學習率影響模型的參數更新速度。如果學習率過大，模型的每一步可能會跳過最優解的區域，導致模型不穩定；但如果學習率過小，模型更新慢，訓練時間會增加。

batch size 影響每次更新時使用的數據量。batch size 大能夠穩定地逼近目標方向，減少震盪，但計算需求較大，訓練時間長。

- IV. regression results

