Machine Learning Assignment 6 Report

111062117, Hsiang-Sheng Huang

December 19, 2024

Comparison Between Dense-only and RNN Model

Theoretically, RNNs should perform better on sine wave data due to their ability to handle sequential dependencies. However, in practice, the Dense-only model performed better in my implementation because it converged faster, while the RNN required careful tuning of the learning rate to address gradient-related issues, such as vanishing or exploding gradients.

Stacking Two Consecutive RNN Layers

If the first RNN layer is RNN(1, 16), its output shape will be (batch_size, timesteps, 16). To stack another RNN, the second layer should be RNN(16, units), where "units" is the number of hidden units. Stacking RNNs allows the model to learn more abstract temporal features but requires ensuring dimensional consistency between layers.

Effects of Larger Hidden Units in RNN Layer

Increasing hidden units improves the model's capacity to capture complex patterns but increases computational cost and can exacerbate gradient issues. Techniques like gradient clipping or adaptive learning rates are necessary to maintain stability.