HW to Chapter 10 “Normalization and Optimization Methods”

Non-programming Assignment

1. What is normalization and why is it needed?

Normalization is the process of scaling input data to ensure that all features have similar ranges. This is important because large differences in scale between features can lead to inefficient learning. For example, a feature with much larger values could dominate the gradients, leading to imbalanced updates in gradient-based optimization. Normalization helps avoid this by scaling data so that the gradient descent works more efficiently across all features.

1. What are vanishing and exploding gradients?

Vanishing gradients occur when the gradients become extremely small as they propagate back through the layers during backpropagation, particularly with deep networks. This leads to very slow learning or no learning at all in the earlier layers.

Exploding gradients happen when the gradients grow exponentially during backpropagation, which can destabilize the learning process and cause the weights to grow too large, resulting in model divergence.

1. What Adam algorithm and why is it needed?

Adam (Adaptive Moment Estimation) is an optimization algorithm designed to adjust the learning rate for each parameter individually. It uses running averages of both the gradients (first moment) and the squared gradients (second moment) to control the learning rate dynamically. Adam helps achieve faster convergence in deep neural networks compared to traditional methods like standard gradient descent.

1. How to choose hyperparameters?

Choosing hyperparameters such as the learning rate, batch size, and regularization strength is typically done through experimentation. Techniques like grid search, random search, or more advanced methods like Bayesian optimization can be used. A common approach is to start with default values (e.g., learning rate = 0.001 for Adam) and fine-tune based on the performance of the model on a validation set.