Machine Learning Assignment 5 Report

111062117, Hsiang-Sheng Huang

November 23, 2024

1. Why ReLU is Preferred Over Sigmoid in Convolutional Blocks? (1%)

ReLU is preferred over Sigmoid due to:

- **Avoiding Vanishing Gradients:** ReLU maintains non-zero gradients for positive inputs, enabling deeper networks to train efficiently.
- **Efficiency:** ReLU involves simple thresholding ($f(x) = \max(0, x)$), while Sigmoid requires exponentials, making it computationally slower.

2. CNN Architecture Design and Parameter Choices (2%)

Architecture Design

Two convolutional layers, two max-pooling layers, and two fully connected layers:

- **Conv1:** 3×3 filters, 16 channels, stride 1, padding 1.
- **MaxPool1:** 2×2 pooling, stride 2.
- **Conv2:** 3×3 filters, 32 channels, stride 1, padding 1.
- **MaxPool2:** 2×2 pooling, stride 2.
- **Dense1:** 64 neurons (ReLU).
- **Dense2:** 1 neuron (Sigmoid).

Parameter Choices

These choices balance computational cost, model complexity, and feature extraction.

- **Filter Size:** 3×3 provides a good balance of spatial resolution and efficiency.
- **Pooling Size:** 2×2 reduces dimensions while retaining important features, minimizing overfitting risk.
- **Stride:** Stride 1 in Conv layers retains spatial dimensions, while 2 in pooling layers reduces computation.

3. Comparison of Learnable Parameters Between CNN and Lab4 NN (2%)

CNN Model (Bias Ignored)

- **Conv1:** $3 \times 3 \times 1 \times 16 = 144$ weights.
- **Conv2:** $3 \times 3 \times 16 \times 32 = 4608$ weights.
- **Dense1:** $2048 \times 64 = 131072$ weights.
- **Dense2:** $64 \times 1 = 64$ weights.

Total Parameters: $135,888$

Lab4 NN Model (Bias Ignored)

- Layer 1: $28 \times 28 \times 128 = 100,352$ weights.
- Layer 2: $128 \times 128 = 16,384$ weights.
- Layer 3: $128 \times 64 = 8,192$ weights.
- Layers 4-7: $64 \times 64 = 4,096 \times 4 = 16,384$ weights.
- Layers 8-11: $64 \times 32 = 2,048 \times 4 = 8,192$ weights.
- Layers 12-15: $32 \times 8 = 256 \times 4 = 1,024$ weights.
- Layer 16: $8 \times 4 = 32$ weights.

Total Parameters: $150,560$

Comparison

- **CNN Model:** $135,888$ parameters.
- **Lab4 NN Model:** $150,560$ parameters.

The CNN model uses fewer parameters due to shared weights in convolutional layers, making it more efficient for image data.