1   Question 1: conv2d

Done as instructed; tests passed.

2   Question 2: Pararellelization

Done as instructed and tested for correctness. Marginal improvement.

3   Question 3: Tweaking hyper-parameters

I first of all changed the batch size from 512 to 128 and then went with the following architecture:

- Conv2D layer: Filter size: 6×6, filter depth: 8, strides: $2, 2$, output: $12 \times 12 \times 8$

- LeakyReLU

- Conv2D layer: Filter size: 5×5, filter depth: 16, strides: $1, 1$, output: $8 \times 8 \times 16$

- LeakyReLU

- Conv2D layer: Filter size: 4×4, filter depth: 32, strides: $1, 1$, output: $5 \times 5 \times 32$

- LeakyReLU

- Conv2D layer: Filter size: 3×3, filter depth: 64, strides: $2, 2$, output: $2 \times 2 \times 64$

- LeakyReLU

- Conv2D layer: Filter size: 2×2, filter depth: 128, strides: $1, 1$, output: $1 \times 1 \times 128$

- LeakyReLU

- Flatten to fully connected linear layer: Size: 128

- LeakyReLU

- Fully connected linear layer: Size: 10

- SoftMax

Training over the same 50 epochs and overfitting a bit, a test set MSE of 0.053001 and a test set precision of 0.985377 were achieved with early stopping after 42 epochs. This is quite a bit better than the untuned test results after 50 epochs (MSE:0.107567 precision:0.966900).