

Exercise 2.1 c

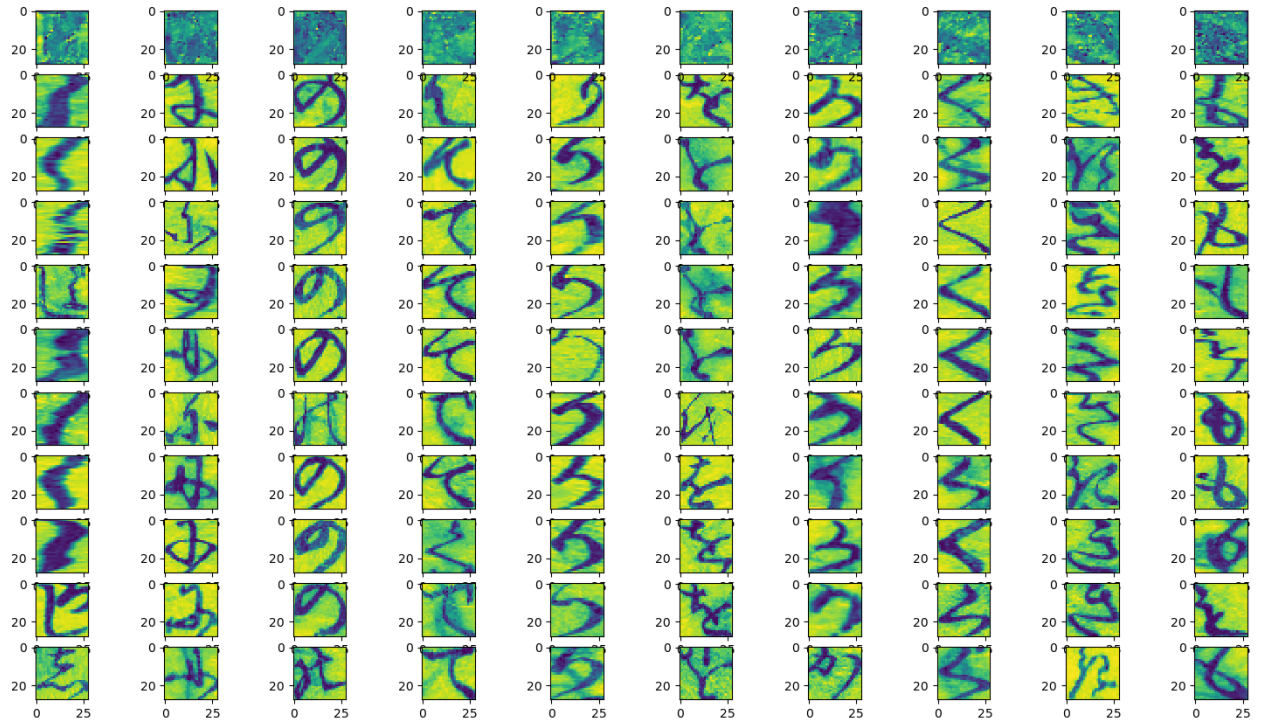


Figure 1: Visualization of the weights vectors and example data images. The darker parts in the visualization of the weights vectors look similar to the real example images.

Exercise 2.2

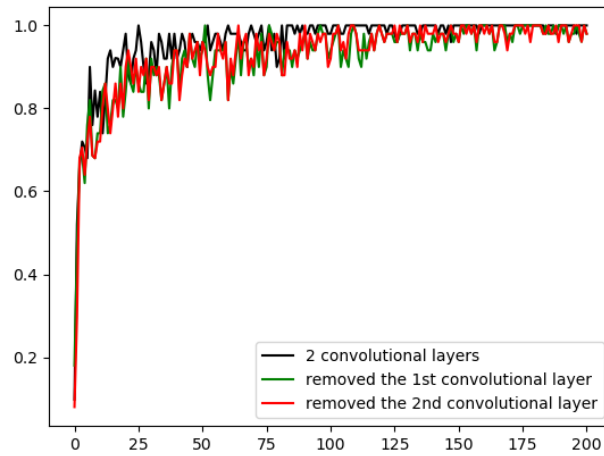


Figure 2: (c). Visualization of training accuracies while removing convolutional layers

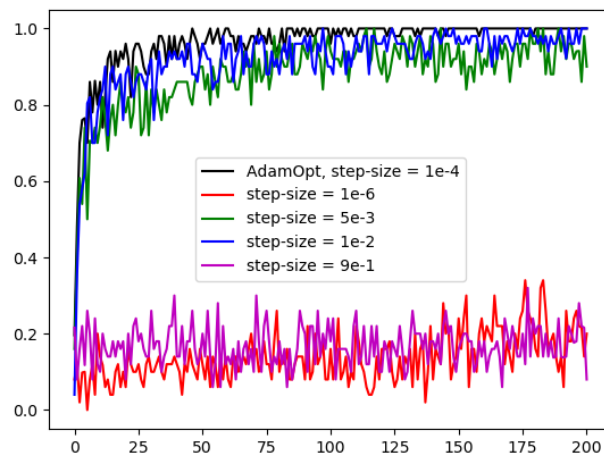


Figure 3: (d). Visualization of training accuracies with changing the step size. Too small step size will make the gradient descent progress so slowly that within the iterations it cannot reach or come out of a local minimum; Too large step size will make the gradient descent travel between two sides of a (local) minimum and never reach it.

Exercise 2.3

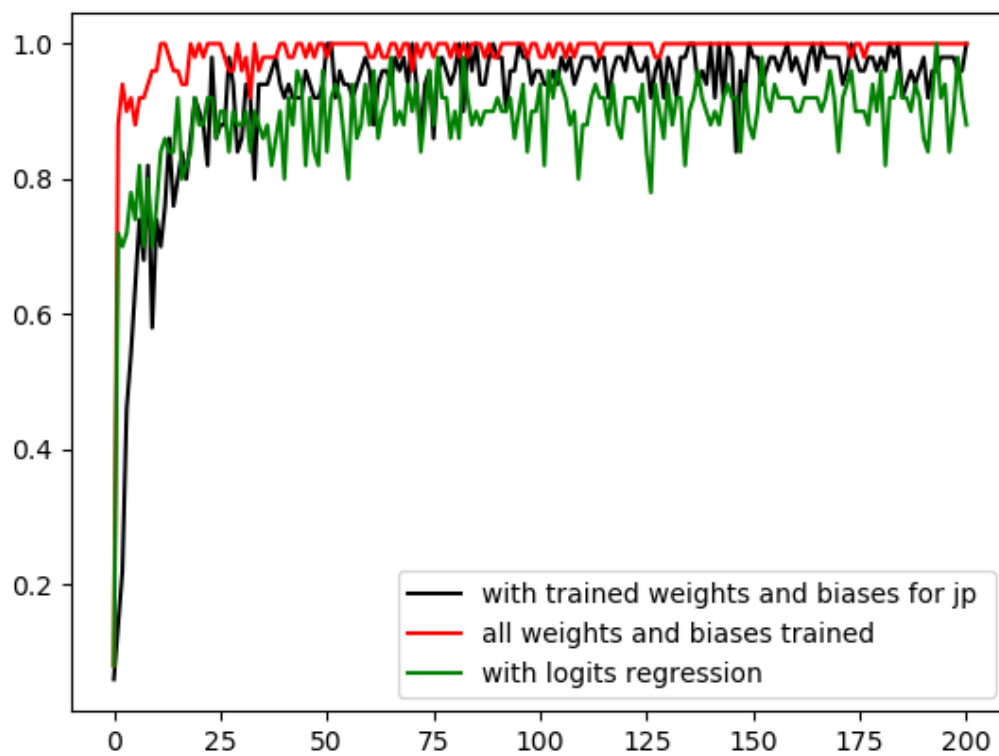


Figure 4: Visualization of training MNIST with different strategies.

Exercise 2.5

(a) $\mathbb{E}(\bar{x}) = \mathbb{E}\left(\frac{1}{L} \sum_{l=1}^L x_l\right) = \frac{1}{L} \sum_{l=1}^L \mathbb{E}(x_l) = \frac{1}{L} \sum_{l=1}^L \frac{\theta}{2} = \frac{\theta}{2}$

(b) With $\hat{x} := 2\bar{x}$ it is $\mathbb{E}(\hat{x}) = \mathbb{E}(2\bar{x}) = 2\mathbb{E}(\bar{x}) = \theta$.

(c) The standard error

$$\mathbb{E}((\hat{x} - \theta)^2) = \mathbb{E}((\hat{x} - \mathbb{E}(\hat{x}))^2) = \text{Var}(\hat{x}) = \text{Var}\left(\frac{2}{L} \sum_{l=1}^L x_l\right) = \frac{4}{L^2} \sum_{l=1}^L \text{Var}(x_l) = \frac{4}{L} \text{Var}(x_1)$$

clearly goes to zero for $L \rightarrow \infty$.

(d) Let $\tilde{x} := \hat{x} + \frac{1}{L}$. It is

$$\mathbb{E}(\tilde{x}) = \mathbb{E}\left(\hat{x} + \frac{1}{L}\right) = \theta + \frac{1}{L}$$

and

$$\mathbb{E}((\tilde{x} - \theta)^2) = \mathbb{E}\left(\left(\hat{x} + \frac{1}{L} - \theta\right)^2\right) = \mathbb{E}\left((\hat{x} - \theta)^2 + \frac{1}{L^2} + \frac{2}{L}(\hat{x} - \theta)\right) = \frac{4}{L} \text{Var}(x_1) + \frac{1}{L^2}$$

which goes to zero for $L \rightarrow \infty$.