IE 534, homework 2

Shuhui Guo

The objective of this homework is training a convolution neural network to classify handwritten numbers in the MNIST dataset. The dataset is divided into training set and testing set. The convolution neural network is built on the training set and the performance is evaluated on the testing set. The input image is $X \in \mathbb{R}^{d\times d}$, where $d = 28$. The output is a vector of probabilities for the potential images of size $K = 10$. The convolution network will have a single hidden layer with multiple channels of size $C = 5$, thus there is an array of C filters in the convolution layer. Let each filter have the size $k_y \times k_x = 3 \times 3$, the filters are given by a variable $K \in \mathbb{R}^{k_y\times k_x\times C}$. And the feature maps for the hidden layer are represented by $H \in \mathbb{R}^{(d-k_y+1)\times(d-k_x+1)\times C}$.

The architecture of the single layer convolution network with multiple channels is:

$$\begin{aligned} Z_{:,:,p} &= X_{:,:} * K_{:,:,p} \\ H_{:,:,p} &= \sigma(Z_{:,:,p}) \\ U_k &= W_{k,:,:,:} \cdot H + b_k \\ f(x;\theta) &= F_{softmax}(U) \end{aligned}$$

where $W_{k,:,:,:} \cdot H = \sum_{i,j,m} W_{k,i,j,p} H_{i,j,p}$, and ReLU unit $z^+ = max(0, z)$ is chosen to be the nonlinearity $\sigma(z)$. The parameters of this architecture are $W \in \mathbb{R}^{K\times(d-k_y+1)\times(d-k_x+1)\times C}$, $b \in \mathbb{R}^K$, $K \in \mathbb{R}^{k_y\times k_x\times C}$. These parameters are collected in $\theta = \{K, W, b\}$.

To efficiently minimize the cross-entropy error, stochastic gradient descent algorithm is used. Since running an epoch is time-consuming, the number of epochs is set to be 3 in this homework. The training accuracy of each epoch is 0.93123, 0.96493, 0.97175, respectively. And the accuracy on the testing set is 0.962.