Homework 2 Question 1

Yanlin Li, Student #: 1003770305

February 18, 2021

(a)

There are 7840 parameters.

(b)

Let:

$$y_{nk} = \frac{exp(w_k^T x_n)}{\sum_{j=0}^{9} exp(w_j^T x_n)}$$

The likelihood function is:

$$p(T|X, w) = \prod_{n=1}^{N} [\prod_{k=0}^{9} y_{nk}^{t_{nk}}]$$

The cross-entropy loss is:

$$E(w) = -ln(p(T|X, w)) = -\sum_{n=1}^{N} [\sum_{k=0}^{9} t_{nk} \ln y_{nk}]$$

Take the gradient, we have:

$$\nabla_{w_k} E(w) = \frac{dE(w)}{dw_k}$$

When $k = j$:

$$\frac{d(\ln y_{nk})}{dw_j} = x_n(1 - \frac{exp(w_j^T x_n)}{\sum_{i=0}^{9} exp(w_i^T x_n)})$$

When $k \neq j$:

$$\frac{d(\ln y_{nk})}{dw_j} = -x_n \frac{exp(w_j^T x_n)}{\sum_{i=0}^{9} exp(w_i^T x_n)}$$

Plug in the above values and the value of y_{nk}:

$$\nabla_{w_k} E(w) = \sum_{n=1}^{N} (y_{nk} - t_{nk}) x_n$$

(c)

I used 10000 samples, and the final results are:
1. Training accuracy is: 0.8679
2. Test accuracy is: 0.834

(d)

Here is the figure:

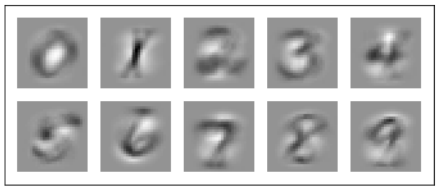

Figure 1: weights

The codes begin on the next page.