Assignment 5

AIDAN KISER
4/16/24 '2'

Question 1: Neural Networks

a) Show that the derivative of:

$$P(C_k | \phi) = y_k(\phi) = \frac{\exp(a_k)}{\Sigma_j \exp(a_j)}$$

is:

$$\frac{\partial y_k}{\partial a_j} = y_k(I_{kj} - y_j)$$

- I_{kj} is 1 when $k=j$ and 0 otherwise.

We must calculate the derivative in both cases:

- When $k=j$:

$$\frac{\partial y_k}{\partial a_k} = \frac{\partial}{\partial a_k}\left(\frac{\exp(a_k)}{\Sigma_j \exp(a_j)}\right)$$

Using quotient rule:

$$\frac{\exp(a_k)\Sigma_j \exp(a_j) - \exp(a_k)\exp(a_k)}{(\Sigma_j \exp(a_j))^2}$$

$$= \frac{\exp(a_k)}{\Sigma_j \exp(a_j)} - \left(\frac{\exp(a_k)}{\Sigma_j \exp(a_j)}\right)^2 = y_k - y_k^2 = y_k$$

Which simplifies to $y_k(1 - y_k)$ because $y_k = \frac{\exp(a_k)}{\Sigma_j \exp(a_j)}$

- When $k \neq j$:

Since a_k does not depend on a_j when $k \neq j$, the derivative is:

$$-\left(\frac{\exp(a_k)\exp(a_j)}{(\Sigma_j \exp(a_j))^2}\right) = -y_k\left(\frac{\exp(a_j)}{\Sigma_j \exp(a_j)}\right) = -y_k y_j$$

• Combining the two results using I_{kj}:

$$\boxed{\frac{\partial y_k}{\partial a_j} = y_k (I_{kj} - y_j)}$$

b) Given the following negative log-likelihood function defined as the loss function:

$$E(w_1, \dots, w_K) = -\ln P(T \mid w_1, \dots, w_K) = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk}$$

Show that the derivative of E is:

$$\nabla_{w_j} E(w_1, \dots, w_K) = -\sum_{n=1}^{N} (y_{nj} - t_{nj}) \phi_n$$

• Calculate the derivative of the -log-likelihood with w_j.

$$\frac{\partial E}{\partial w_j} = -\sum_{n=1}^{N} \sum_{k=1}^{K} \frac{\partial}{\partial w_j} (t_{nk} \ln y_{nk})$$

$$= -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(\frac{1}{y_{nk}}\right)\left(\frac{\partial y_{nk}}{\partial w_j}\right)$$

• Find $\frac{\partial y_{nk}}{\partial w_j}$ using softmax derivative:

$$\frac{\partial y_{nk}}{\partial a_j} = y_{nk} (I_{kj} - y_{nj})$$

chain rule:

$$\frac{\partial y_{nk}}{\partial w_j} = \left(\frac{\partial y_{nk}}{\partial a_j}\right)\left(\frac{\partial a_j}{\partial w_j}\right)$$

• Since $a_j = w_j^T \phi_n$, its derivative is just ϕ_n, therefore:

$$\frac{\partial y_{nk}}{\partial w_j} = y_{nk} (I_{kj} - y_{nj}) \phi_n$$

• Plug into derivative of E:

$$\frac{\partial E}{\partial w_j} = -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \frac{1}{y_{nk}} y_{nk} (I_{kj} - y_{nj}) \phi_n$$

Since I_{kj} is 1 when $k = j$ and 0 otherwise, and t_{nk} is 0 for all k except the true class, this reduces to:

$$\boxed{= -\sum_{n=1}^{N} (t_{nj} - y_{nj})\,\phi_n}$$