Exercises

Paul Scemama

September 2022

1 Chapter 1

1.1

Note: although the exercise asks for a an arbitrary w_i, I will do it for the entire $\boldsymbol{w}$, and then show that each row of the resultant solution vector corresponds to the necessary scalar equation mentioned in the book. The reason being is that I believe it is more instructive.

Get derivative of (1.2) *with respect to* $\boldsymbol{w}$. $\frac{\partial E}{\partial \boldsymbol{w}}$ is the derivative of a function with respect to a vector, and thus the output will be a vector.

$$\frac{\partial E}{\partial \boldsymbol{w}} = \frac{\partial E}{\partial y(x_n, \boldsymbol{w})} \cdot \frac{\partial y(x_n, \boldsymbol{w})}{\partial \boldsymbol{w}} \tag{Chain rule}$$

We will first compute $\frac{\partial E}{\partial y(x_n, \boldsymbol{w})}$.

$$\frac{\partial E}{\partial y(x_n, \boldsymbol{w})} = \sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\}$$

And $\frac{\partial y(x_n, \boldsymbol{w})}{\partial \boldsymbol{w}}$.

$$\frac{\partial y(x_n, \boldsymbol{w})}{\partial \boldsymbol{w}} = [\ 1\ x_n^1\ x_n^2\ x_n^3 ... x_n^M\ ]$$

So,

$$\frac{\partial E}{\partial \boldsymbol{w}} = \underbrace{\sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\}}_{scalar} \cdot \overbrace{\boldsymbol{x}_n}^{vector}$$

where $\boldsymbol{x}_n = [\ 1\ x_n^1\ x_n^2\ x_n^3 ... x_n^M\ ]$. Rewriting the above equation we get,

$$\begin{aligned}
\frac{\partial E}{\partial \boldsymbol{w}} &= \sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\} \cdot \boldsymbol{x}_n \\
&= \begin{bmatrix} \sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\} \cdot x_n^0 \\ ... \\ ... \\ ... \\ \sum_{n=1}^{N} \{y(x_n, \boldsymbol{w}) - t_n\} \cdot x_n^M \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} \{\sum_{j=0}^{M} w_j x_n^j - t_n\} \cdot x_n^0 \\ ... \\ ... \\ ... \\ \sum_{n=1}^{N} \{\sum_{j=0}^{M} w_j x_n^j - t_n\} \cdot x_n^M \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} \{\sum_{j=0}^{M} w_j x_n^j x_n^0 - t_n x_n^0\} \\ ... \\ ... \\ ... \\ \sum_{n=1}^{N} \{\sum_{j=0}^{M} w_j x_n^j x_n^M - t_n x_n^M\} \end{bmatrix}
\end{aligned}$$

So each row can be written as

$$\sum_{n=1}^{N}\{\sum_{j=0}^{M} w_j x_n^{(j+i)} - t_n x_n^i\}$$

where i indicates the row.

Set derivative equal to 0 *and rearrange*

$$\sum_{n=1}^{N}\sum_{j=0}^{M} w_j x_n^{(i+j)} = \sum_{n=1}^{N} t_n (x_n)^i \tag{1}$$

$$\sum_{j=0}^{M}\sum_{n=1}^{N} x_n^{(i+j)} w_j = \sum_{n=1}^{N} t_n (x_n)^i \tag{2}$$

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{3}$$

So to get each element of the vector that contains w_i's that minimize E, we solve (3) for $i = 0...M$. This equates to what Bishop wanted us to show.