Exercises

Paul Scemama

September 2022

1 Chapter 1

{**Exercise: 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,

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

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

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.

{**Exercise: 1.2**}

Note: this exercise is an extension of **1.1**, however I will do this for an arbitrary w_i.

$$\widetilde{E} = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \boldsymbol{w}) - t_n\}^2 + \frac{\lambda}{2}||\boldsymbol{w}||^2 \tag{4}$$

$$\frac{\partial \widetilde{E}}{\partial w_i} = \frac{\partial \widetilde{E}}{\partial y} \cdot \frac{\partial y}{\partial w_i} + \frac{\lambda}{2}\frac{\partial \boldsymbol{w}^T \boldsymbol{w}}{\partial w_i} \tag{Chain rule}$$

Let's split the derivative up in parts,

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

$$\frac{\partial y}{\partial w_i} = x_n^i$$

$$\frac{\lambda}{2}\frac{\partial \boldsymbol{w}^T \boldsymbol{w}}{\partial w_i} = \lambda w_i$$

And so together,

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

Substituting in the definition of $y(x_n, \boldsymbol{w}$ and setting to equal zero,

$$\frac{\partial \widetilde{E}}{\partial w_i} = \sum_{n=1}^{N} \{ \sum_{j=0}^{M} w_j x_n^j - t_n \} x_n^i + \lambda w_i = 0$$

Distributing x_n^i and rearranging,

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

So,

$$\sum_{j=0}^{M} A_{ij} w_j + \lambda w_i = T_i$$

where $A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j}$; $T_i = \sum_{n=1}^{N} (x_n)^i t_n$, as desired.

{**Exercise: 1.3**}

Boxes:

$$\begin{aligned} r \text{ (red)} &\rightarrow 3 \text{ apples, } 4 \text{ os, } 3 \text{ limes} \\ b \text{ (blue)} &\rightarrow 1 \text{ apple, } 1 \text{ o, } 0 \text{ limes} \\ g \text{ (green)} &\rightarrow 3 \text{ apples, } 3 \text{ os, } 4 \text{ limes} \end{aligned}$$

Box chosen at random with probabilities,

$$\begin{aligned} p(r) &= 0.2 \\ p(b) &= 0.2 \\ p(g) &= 0.6 \end{aligned}$$

and then a fruit is chosen at random.

1.3a. What is the probability of selecting an apple?

By the *sum rule*,

$$p(fruit = a) = \sum_{box \in Boxes} p(a, box) = p(a, r) + p(a, b) + p(a, g) \tag{5}$$

By the *product rule* on each term in (5),

$$\begin{aligned} p(a) &= p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g) \\ &= \frac{3}{10} \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{1}{5} + \frac{3}{10} \cdot \frac{3}{5} \\ &= \frac{17}{50} \end{aligned}$$

1.3b. Given the fruit you select is an o, what is the probability that it came from the green box?

$$\begin{aligned} p(box = g | fruit = o) &= \frac{p(o|g)p(g)}{p(fruit = o)} \\ &= \frac{3/10 \cdot 6/10}{p(fruit = o)} \end{aligned}$$

By the *sum rule*,

$$p(fruit = o) = \sum_{box \in Boxes} p(o|box) = p(o, r) + p(o, b) + p(o, g) \tag{6}$$

By the *product rule* on eoch term in (6),

$$\begin{aligned} p(o) &= p(o|r)p(r) + p(o|b)p(b) + p(o|g)p(g) \\ &= \frac{4}{10} \cdot \frac{2}{10} + \frac{1}{2} \cdot \frac{2}{10} + \frac{3}{10} \cdot \frac{6}{10} \\ &= \frac{9}{25} \end{aligned}$$

So $p(box = g | fruit = o) = \frac{18/100}{36/100} = \frac{1}{2}$.

Exercise: 1.4