

# Exercises

Paul Scemama

September 2022

## 1 Chapter 1

### 1.1 Error minimization

*Note:* although the exercise asks for a an arbitrary  $w_i$ , I will do it for the entire  $\mathbf{w}$ , 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  $\mathbf{w}$ .  $\frac{\partial E}{\partial \mathbf{w}}$  is the derivative of a function with respect to a vector, and thus the output will be a vector.

$$\frac{\partial E}{\partial \mathbf{w}} = \frac{\partial E}{\partial y(x_n, \mathbf{w})} \cdot \frac{\partial y(x_n, \mathbf{w})}{\partial \mathbf{w}} \quad (\text{Chain rule})$$

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

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

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

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

So,

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

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

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

So each row can be written as

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

where  $i$  indicates the row.

Set derivative equal to 0 and rearrange

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

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

$$\sum_{j=0}^M A_{ij} w_j = T_i \quad (3)$$

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

## 1.2 Error minimization with regularization

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

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

$$\frac{\partial \tilde{E}}{\partial w_i} = \frac{\partial \tilde{E}}{\partial y} \cdot \frac{\partial y}{\partial w_i} + \frac{\lambda}{2} \frac{\partial \mathbf{w}^T \mathbf{w}}{\partial w_i} \quad (\text{Chain rule})$$

Let's split the derivative up in parts,

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

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

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

And so together,

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

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

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

Distributing  $x_n^i$  and rearranging,

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

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.

### 1.3 Sum & product rules of probability

Boxes:

$r$  (red)  $\rightarrow$  3 apples, 4 os, 3 limes  
 $b$  (blue)  $\rightarrow$  1 apple, 1 o, 0 limes  
 $g$  (green)  $\rightarrow$  3 apples, 3 os, 4 limes

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.

#### 1.3.1 What is the probability of selecting an apple?

By the *sum rule*,

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

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

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

**1.3.2** Given you select an orange, what is the probability that it came from the green box?

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

By the *sum rule*,

$$p(fruit = o) = \sum_{box \in Boxes} p(o | box) = p(o, \textcolor{red}{r}) + p(o, \textcolor{blue}{b}) + p(o, \textcolor{green}{g}) \quad (6)$$

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

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

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

## 1.4