

Solutions Manual to Pattern Recognition and Machine Learning

Hiromichi Inawashiro

June 4, 2024

1 Introduction

1.1

To minimise

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N (y(x_n, \mathbf{w}) - t_n)^2, \quad (1.1)$$

setting its derivative as zero gives

$$\mathbf{0} = \sum_{n=1}^N \frac{\partial y(x_n, \mathbf{w})}{\partial \mathbf{w}} (y(x_n, \mathbf{w}) - t_n). \quad (1.2)$$

Substituting

$$y(x_n, \mathbf{w}) = \sum_{j=0}^M w_j x_n^j \quad (1.3)$$

gives

$$0 = \sum_{n=1}^N x_n^i \left(\sum_{j=0}^M w_j x_n^j - t_n \right). \quad (1.4)$$

Therefore,

$$\sum_{j=0}^M A_{ij} w_j = T_i \quad (1.5)$$

where

$$\begin{aligned} A_{ij} &= \sum_{n=1}^N x_n^{i+j}, \\ T_i &= \sum_{n=1}^N x_n^i t_n. \end{aligned} \quad (1.6)$$

1.2

To minimise

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N (y(x_n, \mathbf{w}) - t_n)^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2, \quad (1.7)$$

setting its derivative as zero gives

$$\mathbf{0} = \sum_{n=1}^N \frac{\partial y(x_n, \mathbf{w})}{\partial \mathbf{w}} (y(x_n, \mathbf{w}) - t_n) + \lambda \mathbf{w}. \quad (1.8)$$

Substituting

$$y(x_n, \mathbf{w}) = \sum_{j=0}^M w_j x_n^j \quad (1.9)$$

gives

$$0 = \sum_{n=1}^N x_n^i \left(\sum_{j=0}^M w_j x_n^j - t_n \right) + \lambda w_i. \quad (1.10)$$

Therefore,

$$\sum_{j=0}^M \tilde{A}_{ij} w_j = T_i \quad (1.11)$$

where

$$\begin{aligned} \tilde{A}_{ij} &= \sum_{n=1}^N x_n^{i+j} + \lambda \delta_{ij}, \\ T_i &= \sum_{n=1}^N x_n^i t_n. \end{aligned} \quad (1.12)$$

1.3

Let a , o and l be the events where an apple, orange and lime are selected respectively. The probability that an apple is selected is given by

$$p(a) = p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g). \quad (1.13)$$

Substituting $p(a|r) = \frac{3}{10}$, $p(r) = \frac{1}{5}$, $p(a|g) = \frac{1}{2}$, $p(r) = \frac{1}{5}$, $p(a|g) = \frac{3}{10}$ and $p(g) = \frac{3}{5}$ gives

$$p(a) = \frac{17}{50}. \quad (1.14)$$

If an orange is selected, the probability that it came from the green box is given by

$$p(g|o) = \frac{p(g, o)}{p(o)}. \quad (1.15)$$

Here,

$$\begin{aligned} p(g, o) &= p(o|g)p(g), \\ p(o) &= p(o|r)p(r) + p(o|b)p(b) + p(o|g)p(g). \end{aligned} \quad (1.16)$$

Substituting $p(o|r) = \frac{2}{5}$, $p(r) = \frac{1}{5}$, $p(o|b) = \frac{1}{2}$, $p(b) = \frac{1}{5}$, $p(o|g) = \frac{3}{10}$ and $p(g) = \frac{3}{5}$ gives $p(g, o) = \frac{9}{50}$ and $p(o) = \frac{9}{25}$. Therefore,

$$p(g|o) = \frac{1}{2}. \quad (1.17)$$

1.5

By the definition,

$$\text{var } f(x) = \text{E} (f(x) - \text{E} f(x))^2. \quad (1.18)$$

The right hand side can be written as

$$\text{E} ((f(x))^2 - 2f(x)\text{E} f(x) + (\text{E} f(x))^2) = \text{E} (f(x))^2 - (\text{E} f(x))^2. \quad (1.19)$$

Therefore,

$$\text{var } f(x) = \text{E} (f(x))^2 - (\text{E} f(x))^2. \quad (1.20)$$

1.6

By the definition,

$$\text{cov}(x, y) = \text{E} xy - \text{E} x \text{E} y. \quad (1.21)$$

Here,

$$\text{E} xy = \int xyf(x, y) dx dy \quad (1.22)$$

If x and y are independent, by the definition

$$f(x, y) = f(x)f(y), \quad (1.23)$$

then

$$\int xyf(x, y) dx dy = \int f(x) dx \int f(y) dy. \quad (1.24)$$

Therefore,

$$\text{E} xy = \text{E} x \text{E} y, \quad (1.25)$$

and thus

$$\text{cov}(x, y) = 0. \quad (1.26)$$