Solutions Manual to
Pattern Recognition and Machine Learning

Hiromichi Inawashiro

June 6, 2024

1 Introduction

1.1

To minimise

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

setting its derivative as zero gives

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

Substituting

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

gives

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

Therefore,

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{1.5}$$

where

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

1.2

To minimise

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

setting its derivative as zero gives

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

Substituting

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

gives

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

Therefore,

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

where

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

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

$$p(a) = p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g). \tag{1.13}$$

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

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

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

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

Here,

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

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,

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

1.5

By the definition,

$$\mathrm{var} f(x) = \mathrm{E}\left(f(x) - \mathrm{E}f(x)\right)^2. \tag{1.18}$$

The right hand side can be written as

$$\mathrm{E}\left((f(x))^2 - 2f(x)\mathrm{E}f(x) + (\mathrm{E}f(x))^2\right) = \mathrm{E}\left(f(x)\right)^2 - \left(\mathrm{E}f(x)\right)^2. \tag{1.19}$$

Therefore,

$$\mathrm{var} f(x) = \mathrm{E}\left(f(x)\right)^2 - \left(\mathrm{E}f(x)\right)^2. \tag{1.20}$$

1.6

By the definition,

$$\mathrm{cov}(x, y) = \mathrm{E}xy - \mathrm{E}x\mathrm{E}y, \tag{1.21}$$

where the right hand side can be written as

$$\int xyf(x, y)dxdy - \int xf(x)dx \int yf(y)dy. \tag{1.22}$$

If x and y are independent, by the definition

$$f(x, y) = f(x)f(y), \tag{1.23}$$

then

$$\int xyf(x, y)dxdy = \int f(x)dx \int f(y)dy. \tag{1.24}$$

Therefore,

$$\mathrm{cov}(x, y) = 0. \tag{1.25}$$

1.7

Let

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx. \tag{1.26}$$

Then

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x^2+y^2\right)\right) dxdy. \tag{1.27}$$

By the transformation from Cartesian coordinates (x, y) to polar coordinates (r, θ), the right hand side can be written as

$$\int_0^{\infty}\int_0^{2\pi} \exp\left(-\frac{1}{2\sigma^2}r^2\right) \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} drd\theta = 2\pi \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) rdr. \tag{1.28}$$

By the transformation $s = \frac{r}{\sigma}$, the right hand side can be written as

$$2\pi\sigma^2 \int_0^{\infty} \exp\left(-\frac{1}{2}s^2\right) sds = 2\pi\sigma^2 \left[-\exp\left(-\frac{1}{2}s^2\right)\right]_0^{\infty}. \tag{1.29}$$

Therefore,

$$I = \left(2\pi\sigma^2\right)^{\frac{1}{2}}. \tag{1.30}$$

By the definition,

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{1.31}$$

Then

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx. \tag{1.32}$$

By the transformation $t = x - \mu$, the right hand side can be written as

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}t^2\right) dt = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} I. \tag{1.33}$$

Therefore,

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = 1. \tag{1.34}$$

1.8

If x is under the Gaussian distribution, then

$$\mathrm{E}x = \int_{-\infty}^{\infty} x\mathcal{N}\left(x|\mu,\sigma^2\right) dx. \tag{1.35}$$

By the definition, the right hand side can be written as

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx. \tag{1.36}$$

By the transformation $y = x - \mu$, it can be written as

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} (y+\mu) \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy. \tag{1.37}$$

Since

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} y \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy = 0, \tag{1.38}$$

and

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \mu \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy = \mu \int_{-\infty}^{\infty} \mathcal{N}\left(y|\mu,\sigma^2\right) dy, \tag{1.39}$$

we have

$$\mathrm{E}x = \mu. \tag{1.40}$$

By the definition,

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = 1 \tag{1.41}$$

can be written as

$$\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = 1. \tag{1.42}$$

Differentiating both sides with respect to σ^2 gives

$$\begin{aligned} &(2\pi)^{-\frac{1}{2}} \left(-\frac{1}{2}\right) \left(\sigma^2\right)^{-\frac{3}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &+ \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{2}\left(\sigma^2\right)^{-2} (x-\mu)^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = 0. \end{aligned} \tag{1.43}$$

The left hand side can be written as

$$-\frac{1}{2}\left(\sigma^2\right)^{-1}\int_{-\infty}^{\infty}\mathcal{N}\left(x|\mu,\sigma^2\right)dx+\frac{1}{2}\left(\sigma^2\right)^{-2}\int_{-\infty}^{\infty}(x-\mu)^2\mathcal{N}\left(x|\mu,\sigma^2\right)dx$$
$$=-\frac{1}{2}\left(\sigma^2\right)^{-1}+\frac{1}{2}\left(\sigma^2\right)^{-2}\mathrm{var}x. \tag{1.44}$$

Therefore,

$$\mathrm{var}x=\sigma^2. \tag{1.45}$$

1.9

By the definition,

$$\mathcal{N}\left(x|\mu,\sigma^2\right)=\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{1.46}$$

Setting its derivative as zero gives

$$0=\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\left(-\frac{1}{\sigma^2}(x-\mu)\right)\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{1.47}$$

Therefore, the mode is given by μ.

Similarly, by the definition,

$$\mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=(2\pi)^{-\frac{D}{2}}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\intercal\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{1.48}$$

Setting its derivative as zero gives

$$\mathbf{0}=(2\pi)^{-\frac{D}{2}}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}\left(-\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\intercal\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{1.49}$$

Therefore, the mode is given by $\boldsymbol{\mu}$.