

Solutions Manual to Pattern Recognition and Machine Learning

Hiromichi Inawashiro

June 7, 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 to 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 to 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,

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

The right hand side can be written as

$$E((f(x))^2 - 2f(x)Ef(x) + (Ef(x))^2) = E(f(x))^2 - (Ef(x))^2. \quad (1.19)$$

Therefore,

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

1.6

By the definition,

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

The right hand side can be written as

$$Exy - E(xEy) - E(yEx) + E(ExEy) = Exy - ExEy. \quad (1.22)$$

The right hand side can be written as

$$\int xyp(x, y)dxdy - \int xp(x)dx \int yp(y)dy. \quad (1.23)$$

If x and y are independent, by the definition,

$$f(x, y) = f(x)f(y). \quad (1.24)$$

Then,

$$\int xyp(x, y)dxdy = \int p(x)dx \int p(y)dy. \quad (1.25)$$

Therefore,

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

1.7

Let

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx. \quad (1.27)$$

Then

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x^2 + y^2)\right) dx dy. \quad (1.28)$$

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

$$\int_0^{\infty} \int_0^{2\pi} \exp\left(-\frac{1}{2\sigma^2}r^2\right) \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} dr d\theta = 2\pi \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}r^2\right) r dr. \quad (1.29)$$

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

$$2\pi\sigma^2 \int_0^{\infty} \exp\left(-\frac{1}{2}s^2\right) s ds = 2\pi\sigma^2 \left[-\exp\left(-\frac{1}{2}s^2\right)\right]_0^{\infty}. \quad (1.30)$$

Therefore,

$$I = (2\pi\sigma^2)^{\frac{1}{2}}. \quad (1.31)$$

By the definition,

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

Then

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = (2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx. \quad (1.33)$$

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

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}t^2\right) dt = (2\pi\sigma^2)^{-\frac{1}{2}} I. \quad (1.34)$$

Therefore,

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1. \quad (1.35)$$

1.8

If x is under the Gaussian distribution with mean μ and variance σ^2 , then

$$\mathbb{E}x = \int_{-\infty}^{\infty} x \mathcal{N}(x|\mu, \sigma^2) dx. \quad (1.36)$$

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

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx. \quad (1.37)$$

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

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} (y + \mu) \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy. \quad (1.38)$$

Since

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} y \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy = 0, \quad (1.39)$$

and

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \mu \exp\left(-\frac{1}{2\sigma^2}y^2\right) dy = \mu \int_{-\infty}^{\infty} \mathcal{N}(y|\mu, \sigma^2) dy, \quad (1.40)$$

we have

$$\mathbb{E}x = \mu. \quad (1.41)$$

By the definition,

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1 \quad (1.42)$$

can be written as

$$(2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = 1. \quad (1.43)$$

Differentiating both sides with respect to σ^2 gives

$$\begin{aligned} & (2\pi)^{-\frac{1}{2}} \left(-\frac{1}{2}\right) (\sigma^2)^{-\frac{3}{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ & + (2\pi\sigma^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{2} (\sigma^2)^{-2} (x-\mu)^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = 0. \end{aligned} \quad (1.44)$$

The left hand side can be written as

$$\begin{aligned} -\frac{1}{2}(\sigma^2)^{-1} \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx + \frac{1}{2}(\sigma^2)^{-2} \int_{-\infty}^{\infty} (x - \mu)^2 \mathcal{N}(x|\mu, \sigma^2) dx \\ = -\frac{1}{2}(\sigma^2)^{-1} + \frac{1}{2}(\sigma^2)^{-2} \text{var} x. \end{aligned} \quad (1.45)$$

Therefore,

$$\text{var} x = \sigma^2. \quad (1.46)$$

1.9

By the definition,

$$\mathcal{N}(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (1.47)$$

Setting its derivative to zero gives

$$0 = (2\pi\sigma^2)^{-\frac{1}{2}} \left(-\frac{1}{\sigma^2}(x - \mu)\right) \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right). \quad (1.48)$$

Therefore, the mode is given by μ .

Similarly, by the definition,

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (1.49)$$

Setting its derivative to zero gives

$$\mathbf{0} = -(2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} (\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1})^\top) (\mathbf{x} - \boldsymbol{\mu}) \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (1.50)$$

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

1.10

By the definition,

$$\mathbb{E}(x + y) = \int \int (x + y) p(x, y) dx dy. \quad (1.51)$$

The right hand side can be written as

$$\int x \left(\int p(x, y) dy \right) dx + \int y \left(\int p(x, y) dx \right) dy = \int xp(x) dx + \int yp(y) dy. \quad (1.52)$$

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

$$Ex + Ey. \quad (1.53)$$

Therefore,

$$E(x + y) = Ex + Ey. \quad (1.54)$$

Similarly, by the definition,

$$\text{var}(x + y) = E(x + y - E(x + y))^2 \quad (1.55)$$

By the result above and the definition, the right hand side can be written as

$$\begin{aligned} E(x - Ex)^2 + 2E((x - Ex)(y - Ey)) + E(y - Ey)^2 \\ = \text{var}x + 2\text{cov}(x, y) + \text{var}y. \end{aligned} \quad (1.56)$$

If x and y are independent, then

$$\text{cov}(x, y) = 0, \quad (1.57)$$

by 1.6. Therefore,

$$\text{var}(x + y) = \text{var}x + \text{var}y. \quad (1.58)$$

1.11

To maximise

$$\ln p(\mathbf{x}|\mu, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2, \quad (1.59)$$

setting the partial derivatives with respect to μ and σ^2 to zero gives

$$\begin{aligned} 0 &= \frac{1}{\sigma^2} \sum_{n=1}^N (x_n - \mu), \\ 0 &= -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{n=1}^N (x_n - \mu)^2. \end{aligned} \quad (1.60)$$

Therefore,

$$\begin{aligned}\mu_{\text{ML}} &= \frac{1}{N} \sum_{n=1}^N x_n, \\ \sigma_{\text{ML}}^2 &= \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{\text{ML}})^2.\end{aligned}\tag{1.61}$$

1.12

If x_m and x_n are independent, then

$$\mathbb{E}x_mx_n = \mathbb{E}x_m\mathbb{E}x_n.\tag{1.62}$$

If they are samples from the Gaussian distribution with mean μ and variance σ^2 , the right hand side is given by μ^2 . On the other hand, by the definition,

$$\mathbb{E}x_n^2 = \text{var}x_n + (\mathbb{E}x_n)^2.\tag{1.63}$$

If x_n is a sample from the Gaussian distribution with mean μ and variance σ^2 , the right hand side is given by $\sigma^2 + \mu^2$. Therefore,

$$\mathbb{E}x_mx_n = \mu^2 + \delta_{mn}\sigma^2.\tag{1.64}$$

Here, since

$$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n,\tag{1.65}$$

we have

$$\mathbb{E}\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N \mathbb{E}x_n.\tag{1.66}$$

Therefore,

$$\mathbb{E}\mu_{\text{ML}} = \mu.\tag{1.67}$$

Similarly, since

$$\sigma_{\text{ML}}^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{\text{ML}})^2,\tag{1.68}$$

we have

$$\mathbb{E}\sigma_{\text{ML}}^2 = \frac{1}{N} \sum_{n=1}^N \mathbb{E}(x_n - \mu_{\text{ML}})^2.\tag{1.69}$$

The right hand side can be written as

$$\frac{1}{N} \sum_{n=1}^N \mathbb{E} (x_n^2 - 2\mu_{\text{ML}}x_n + \mu_{\text{ML}}^2) = \frac{1}{N} \sum_{n=1}^N \mathbb{E} x_n^2 - \frac{2}{N} \mathbb{E} \left(\mu_{\text{ML}} \left(\sum_{n=1}^N x_n \right) \right) + \mathbb{E} \mu_{\text{ML}}^2. \quad (1.70)$$

The first term of the right hand side can be written as

$$\frac{1}{N} \sum_{n=1}^N (\mu^2 + \sigma^2) = \mu^2 + \sigma^2, \quad (1.71)$$

while the second and third terms can be written as

$$-2\mathbb{E} \mu_{\text{ML}}^2 + \mathbb{E} \mu_{\text{ML}}^2 = -\mathbb{E} \mu_{\text{ML}}^2. \quad (1.72)$$

Here,

$$\mathbb{E} \mu_{\text{ML}}^2 = \mathbb{E} \left(\frac{1}{N} \sum_{n=1}^N x_n \right)^2. \quad (1.73)$$

The right hand side can be written as

$$\frac{1}{N^2} \sum_{n=1}^N \mathbb{E} x_n^2 + \frac{2}{N^2} \sum_{1 \leq m < n \leq N} \mathbb{E} x_m x_n = \frac{1}{N} (\mu^2 + \sigma^2) + \frac{N-1}{N} \mu^2. \quad (1.74)$$

Therefore,

$$\mathbb{E} \mu_{\text{ML}}^2 = \mu^2 + \frac{1}{N} \sigma^2. \quad (1.75)$$

Thus,

$$\mathbb{E} \sigma_{\text{ML}}^2 = \frac{N-1}{N} \sigma^2. \quad (1.76)$$

1.13

It is clear that

$$\mathbb{E} \left(\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2 \right) = \frac{1}{N} \sum_{n=1}^N \mathbb{E} (x_n - \mu)^2. \quad (1.77)$$

The right hand side can be written as

$$\frac{1}{N} \sum_{n=1}^N \mathbb{E} (x_n^2 - 2\mu x_n + \mu^2) = \frac{1}{N} \sum_{n=1}^N \mathbb{E} x_n^2 - \frac{2\mu}{N} \sum_{n=1}^N \mathbb{E} x_n + \mu^2. \quad (1.78)$$

The first term of the right hand side can be written as

$$\frac{1}{N} \sum_{n=1}^N (\mu^2 + \sigma^2) = \mu^2 + \sigma^2, \quad (1.79)$$

while the second term can be written as

$$-\frac{2\mu}{N} \sum_{n=1}^N \mu = -2\mu^2. \quad (1.80)$$

Therefore,

$$\mathbb{E} \left(\frac{1}{N} \sum_{n=1}^N (x_n - \mu)^2 \right) = \sigma^2. \quad (1.81)$$

1.14

Let

$$\begin{aligned} w_{ij}^S &= \frac{1}{2}(w_{ij} + w_{ji}), \\ w_{ij}^A &= \frac{1}{2}(w_{ij} - w_{ji}). \end{aligned} \quad (1.82)$$

Then

$$\begin{aligned} w_{ij} &= w_{ij}^S + w_{ij}^A, \\ w_{ij}^S &= w_{ji}^S, \\ w_{ij}^A &= -w_{ji}^A. \end{aligned} \quad (1.83)$$

Therefore,

$$\sum_{i=1}^D \sum_{j=1}^D w_{ij} x_i x_j = \sum_{i=1}^D \sum_{j=1}^D (w_{ij}^S + w_{ij}^A) x_i x_j. \quad (1.84)$$

The right hand side can be written as

$$\sum_{i=1}^D \sum_{j=1}^D w_{ij}^S x_i x_j + \sum_{i=1}^D \sum_{j=1}^D w_{ij}^A x_i x_j = \sum_{i=1}^D \sum_{j=1}^D w_{ij}^S x_i x_j, \quad (1.85)$$

where

$$\sum_{i=1}^D \sum_{j=1}^D w_{ij}^A x_i x_j = 0, \quad (1.86)$$

which is clear from the definition, is used.

Finally, it is also clear from the definition that the number of independent parameters in the matrix w_{ij}^S is $\frac{D(D+1)}{2}$.