Solutions Manual to
Pattern Recognition and Machine Learning

Hiromichi Inawashiro

June 8, 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 to 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 to zero gives

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

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,

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

The right hand side can be written as

$$\mathrm{E}xy - \mathrm{E}\left(x\mathrm{E}y\right) - \mathrm{E}\left(y\mathrm{E}x\right) + \mathrm{E}\left(\mathrm{E}x\mathrm{E}y\right) = \mathrm{E}xy - \mathrm{E}x\mathrm{E}y. \tag{1.22}$$

The right hand side can be written as

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

If x and y are independent, by the definition,

$$f(x,y) = f(x)f(y). \tag{1.24}$$

Then,

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

Therefore,

$$\mathrm{cov}(x,y) = 0. \tag{1.26}$$

1.7

Let

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

Then

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

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

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

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

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

Therefore,

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

By the definition,

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

Then

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

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

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

Therefore,

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

1.8

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

$$\mathrm{E}x = \int_{-\infty}^{\infty} x\mathcal{N}\left(x|\mu, \sigma^2\right) dx. \tag{1.36}$$

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

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

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

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

Since

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

and

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

we have

$$\mathrm{E}x = \mu. \tag{1.41}$$

By the definition,

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

can be written as

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

Differentiating both sides with respect to σ^2 gives

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

The left hand side can be written as

$$-\frac{1}{2}\left(\sigma^2\right)^{-1}\int_{-\infty}^{\infty}\mathcal{N}\left(x|\mu,\sigma^2\right)dx+\frac{1}{2}\left(\sigma^2\right)^{-2}\int_{-\infty}^{\infty}(x-\mu)^2\mathcal{N}\left(x|\mu,\sigma^2\right)dx$$
$$=-\frac{1}{2}\left(\sigma^2\right)^{-1}+\frac{1}{2}\left(\sigma^2\right)^{-2}\mathrm{var}x. \tag{1.45}$$

Therefore,

$$\mathrm{var}x=\sigma^2. \tag{1.46}$$

1.9

By the definition,

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

Setting its derivative to zero gives

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

Therefore, the mode is given by μ.

Similarly, by the definition,

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

Setting its derivative to zero gives

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

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

1.10

By the definition,

$$\mathrm{E}(x+y)=\int\int(x+y)p(x,y)dxdy. \tag{1.51}$$

The right hand side can be written as

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

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

$$\mathrm{E}x+\mathrm{E}y. \tag{1.53}$$

Therefore,

$$\mathrm{E}(x+y)=\mathrm{E}x+\mathrm{E}y. \tag{1.54}$$

Similarly, by the definition,

$$\mathrm{var}(x+y)=\mathrm{E}\left(x+y-\mathrm{E}(x+y)\right)^2 \tag{1.55}$$

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

$$\begin{aligned}\mathrm{E}\left(x-\mathrm{E}x\right)^2+2\mathrm{E}\left(\left(x-\mathrm{E}x\right)\left(y-\mathrm{E}y\right)\right)+\mathrm{E}\left(y-\mathrm{E}y\right)^2\\ =\mathrm{var}x+2\mathrm{cov}(x,y)+\mathrm{var}y.\end{aligned} \tag{1.56}$$

If x and y are independent, then

$$\mathrm{cov}(x,y)=0, \tag{1.57}$$

by 1.6. Therefore,

$$\mathrm{var}(x+y)=\mathrm{var}x+\mathrm{var}y. \tag{1.58}$$

1.11

To maximise

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

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

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

Therefore,

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

1.12

If x_m and x_n are independent, then

$$\mathrm{E}x_m x_n = \mathrm{E}x_m \mathrm{E}x_n. \tag{1.62}$$

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,

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

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,

$$\mathrm{E}x_m x_n = \mu^2 + \delta_{mn}\sigma^2. \tag{1.64}$$

Here, since

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

we have

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

Therefore,

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

Similarly, since

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

we have

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

The right hand side can be writen as

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

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

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

while the second and third terms can be writen as

$$-2\mathrm{E}\mu_{\mathrm{ML}}^2 + \mathrm{E}\mu_{\mathrm{ML}}^2 = -\mathrm{E}\mu_{\mathrm{ML}}^2. \tag{1.72}$$

Here,

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

The right hand side can be written as

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

Therefore,

$$\mathrm{E}\mu_{\mathrm{ML}}^2 = \mu^2 + \frac{1}{N}\sigma^2. \tag{1.75}$$

Thus,

$$\mathrm{E}\sigma_{\mathrm{ML}}^2 = \frac{N-1}{N}\sigma^2. \tag{1.76}$$

1.13

It is clear that

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

The right hand side can be writen as

$$\frac{1}{N}\sum_{n=1}^{N}\mathrm{E}\left(x_n^2-2\mu x_n+\mu^2\right) = \frac{1}{N}\sum_{n=1}^{N}\mathrm{E}x_n^2 - \frac{2\mu}{N}\sum_{n=1}^{N}\mathrm{E}x_n + \mu^2. \tag{1.78}$$

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

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

while the second term can be writen as

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

Therefore,

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

1.14

Let

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

Then

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

Here,

$$\sum_{i=1}^{D}\sum_{j=1}^{D}w_{ij}^{\mathrm{A}}x_ix_j=\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij}-w_{ji})x_ix_j. \tag{1.84}$$

The right hand side can be written as

$$\frac{1}{2}\left(\sum_{i=1}^{D}\sum_{j=1}^{D}w_{ij}x_ix_j-\sum_{i=1}^{D}\sum_{j=1}^{D}w_{ji}x_ix_j\right)=0. \tag{1.85}$$

Therefore,

$$\sum_{i=1}^{D}\sum_{j=1}^{D}w_{ij}^{\mathrm{A}}x_ix_j=0. \tag{1.86}$$

Additionally,

$$\sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij} x_i x_j = \sum_{i=1}^{D} \sum_{j=1}^{D} \left(w_{ij}^{\mathrm{S}} + w_{ij}^{\mathrm{A}} \right) x_i x_j. \tag{1.87}$$

The right hand side can be written as

$$\sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij}^{\mathrm{S}} x_i x_j + \sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij}^{\mathrm{A}} x_i x_j = \sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij}^{\mathrm{S}} x_i x_j, \tag{1.88}$$

where the result above is used. Therefore,

$$\sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij} x_i x_j = \sum_{i=1}^{D} \sum_{j=1}^{D} w_{ij}^{\mathrm{S}} x_i x_j. \tag{1.89}$$

Finally, since the matrix w_{ij}^{S} is $D \times D$ symmetric matrix, its number of independent parameters is $\frac{D(D+1)}{2}$.