SOLUTION MANUAL FOR PATTERN RECOGNITION AND MACHINE LEARNING

EDITED BY

ZHENGQI GAO

Information Science and Technology School
Fudan University

NOV.2017

0.1 Introduction

Problem 1.1 Solution

We let the derivative of *error function* E with respect to vector $\mathbf{w}$ equals to $\mathbf{0}$, (i.e. $\frac{\partial E}{\partial \mathbf{w}} = 0$), and this will be the solution of $\mathbf{w} = \{w_i\}$ which minimizes *error function* E. To solve this problem, we will calculate the derivative of E with respect to every w_i , and let them equal to 0 instead. Based on (1.1) and (1.2) we can obtain :

=>

$$\frac{\partial E}{\partial w_i} = \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\} x_n^i = 0$$

=>

$$\sum_{n=1}^{N} y(x_n, \mathbf{w}) x_n^i = \sum_{n=1}^{N} x_n^i t_n$$

=>

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

=>

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

=>

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

If we denote $A_{ij} = \sum_{n=1}^{N} x_n^{i+j}$ and $T_i = \sum_{n=1}^{N} {x_n}^i t_n$, the equation above can be written exactly as (1.222), Therefore the problem is solved.

Problem 1.2 Solution

This problem is similar to Prob.1.1, and the only difference is the last term on the right side of (1.4), the penalty term. So we will do the same thing as in Prob.1.1 :

=>

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

=>

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

=>

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

where

$$\delta_{ji} \begin{cases} 0 & \text{j} \neq \text{i} \\ 1 & \text{j} = \text{i} \end{cases}$$

Problem 1.3 Solution

This problem can be solved by *Bayes' theorem*. The probability of selecting an apple $P(a)$:

$$P(a) = P(a|r)P(r) + P(a|b)P(b) + P(a|g)P(g) = \frac{3}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 = 0.34$$

Based on *Bayes' theorem*, the probability of an selected orange coming from the green box $P(g|o)$:

$$P(g|o) = \frac{P(o|g)P(g)}{P(o)}$$

We calculate the probability of selecting an orange $P(o)$ first :

$$P(o) = P(o|r)P(r) + P(o|b)P(b) + P(o|g)P(g) = \frac{4}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 = 0.36$$

Therefore we can get :

$$P(g|o) = \frac{P(o|g)P(g)}{P(o)} = \frac{\frac{3}{10} \times 0.6}{0.36} = 0.5$$

Problem 1.4 Solution

This problem needs knowledge about *calculus*, especially about *Chain rule*. We calculate the derivative of $P_y(y)$ with respect to y, according to (1.27) :

$$\frac{dp_y(y)}{dy} = \frac{d(p_x(g(y))|g^{'}(y)|)}{dy} = \frac{dp_x(g(y))}{dy}|g^{'}(y)| + p_x(g(y))\frac{d|g^{'}(y)|}{dy} \qquad (*)$$

The first term in the above equation can be further simplified:

$$\frac{dp_x(g(y))}{dy}|g^{'}(y)| = \frac{dp_x(g(y))}{dg(y)}\frac{dg(y)}{dy}|g^{'}(y)| \qquad (**)$$

If $\hat{x}$ is the maximum of density over x, we can obtain :

$$\frac{dp_x(x)}{dx}\Big|_{\hat{x}} = 0$$

Therefore, when $y = \hat{y}, s.t. \hat{x} = g(\hat{y})$, the first term on the right side of $(**)$ will be 0, leading the first term in $(*)$ equals to 0, however because of the existence of the second term in $(*)$, the derivative may not equal to 0. But when linear transformation is applied, the second term in $(*)$ will vanish, (e.g. $x = ay + b$). A simple example can be shown by :

$$p_x(x) = 2x, \quad x \in [0,1] \quad => \quad \hat{x} = 1$$

And given that:

$$x = sin(y)$$

Therefore, $p_y(y) = 2\sin(y)|\cos(y)|$, $y \in [0, \frac{\pi}{2}]$, which can be simplified :

$$p_y(y) = \sin(2y), \quad y \in [0, \frac{\pi}{2}] \quad => \quad \hat{y} = \frac{\pi}{4}$$

However, it is quite obvious :

$$\hat{x} \neq \sin(\hat{y})$$

Problem 1.5 Solution

This problem takes advantage of the property of expectation:

$$\begin{aligned} var[f] &= \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] \\ &= \mathbb{E}[f(x)^2 - 2f(x)\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2] \\ &= \mathbb{E}[f(x)^2] - 2\mathbb{E}[f(x)]^2 + \mathbb{E}[f(x)]^2 \\ => var[f] &= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 \end{aligned}$$

Problem 1.6 Solution

Based on (1.41), we only need to prove when x and y is independent, $\mathbb{E}_{x,y}[xy] = \mathbb{E}[x]\mathbb{E}[y]$. Because x and y is independent, we have :

$$p(x,y) = p_x(x)\,p_y(y)$$

Therefore:

$$\begin{aligned} \int\int xyp(x,y)\,dx\,dy &= \int\int xyp_x(x)p_y(y)\,dx\,dy \\ &= (\int xp_x(x)\,dx)(\int yp_y(y)\,dy) \\ => \mathbb{E}_{x,y}[xy] &= \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

Problem 1.7 Solution

This problem should take advantage of *Integration by substitution*.

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} exp(-\frac{1}{2\sigma^2}x^2 - \frac{1}{2\sigma^2}y^2)\,dx\,dy \\ &= \int_0^{2\pi}\int_0^{+\infty} exp(-\frac{1}{2\sigma^2}r^2)r\,dr\,d\theta \end{aligned}$$

Here we utilize :

$$x = r\cos\theta, \quad y = r\sin\theta$$

Based on the fact :

$$\int_0^{+\infty} exp(-\frac{1}{2\sigma^2})r\,dr = -\sigma^2 exp(-\frac{r^2}{2\sigma^2})|_0^{+\infty} = -\sigma^2(0-(-1)) = \sigma^2$$

Therefore, I can be solved :

$$I^2 = \int_0^{2\pi} \sigma^2\,d\theta = 2\pi\sigma^2, \quad => I = \sqrt{2\pi}\sigma$$

And next,we will show that Gaussian distribution $\mathcal{N}(x|\mu,\sigma^2)$ is normalized, (i.e. $\int_{-\infty}^{+\infty}\mathcal{N}(x|\mu,\sigma^2)\,dx = 1$) :

$$\begin{aligned}
\int_{-\infty}^{+\infty}\mathcal{N}(x|\mu,\sigma^2)\,dx &= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}\,dx \\
&= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}y^2\}\,dy \quad (y = x-\mu) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} exp\{-\frac{1}{2\sigma^2}y^2\}\,dy \\
&= 1
\end{aligned}$$

Problem 1.8 Solution

The first question will need the result of Prob.1.7 :

$$\begin{aligned}
\int_{-\infty}^{+\infty}\mathcal{N}(x|\mu,\sigma^2)x\,dx &= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}x\,dx \\
&= \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}y^2\}(y+\mu)\,dy \quad (y = x-\mu) \\
&= \mu\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}y^2\}\,dy + \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}y^2\}y\,dy \\
&= \mu + 0 = \mu
\end{aligned}$$

The second problem has already be given hint in the description. Given that :

$$\frac{d(fg)}{dx} = f\frac{dg}{dx} + g\frac{df}{dx}$$

We differentiate both side of (1.127) with respect to σ^2, we will obtain :

$$\int_{-\infty}^{+\infty}(-\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4})\mathcal{N}(x|\mu,\sigma^2)\,dx = 0$$

Provided the fact that $\sigma \neq 0$, we can get:

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

So the equation above has actually proven (1.51), according to the definition:

$$var[x] = \int_{-\infty}^{+\infty} (x-\mathbb{E}[x])^2 \mathcal{N}(x|\mu,\sigma^2)\,dx$$

Where $\mathbb{E}[x] = \mu$ has already been proved. Therefore :

$$var[x] = \sigma^2$$

Finally,

$$\mathbb{E}[x^2] = var[x] + \mathbb{E}[x]^2 = \sigma^2 + \mu^2$$

Problem 1.9 Solution

Here we only focus on (1.52), because (1.52) is the general form of (1.42). Based on the definition : The maximum of distribution is known as its mode and (1.52), we can obtain :

$$\begin{aligned}
\frac{\partial \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \mathbf{x}} &= -\frac{1}{2}[\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1})^T](\mathbf{x}-\boldsymbol{\mu})\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \\
&= -\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})
\end{aligned}$$

Where we take advantage of :

$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}+\mathbf{A}^T)\mathbf{x} \quad \text{and} \quad (\boldsymbol{\Sigma}^{-1})^T = \boldsymbol{\Sigma}^{-1}$$

Therefore,

$$\text{only when} \quad \mathbf{x} = \boldsymbol{\mu}, \; \frac{\partial \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})}{\partial \mathbf{x}} = 0$$

Note: You may also need to calculate *Hessian Matrix* to prove that it is maximum. However, here we find that the first derivative only has one zero point. Based on the description in the problem, this point should be maximum point.

Problem 1.10 Solution

We will solve this problem based on the definition of *expectation*, *variation* and *independence*.

$$\begin{aligned}
\mathbb{E}[x+z] &= \int\int (x+z)p(x,z)\,dx\,dz \\
&= \int\int (x+z)p(x)p(z)\,dx\,dz \\
&= \int\int xp(x)p(z)\,dx\,dz \;+\; \int\int zp(x)p(z)\,dx\,dz \\
&= \int(\int p(z)\,dz)xp(x)\,dx \;+\; \int(\int p(x)\,dx)zp(z)\,dz \\
&= \int xp(x)\,dx \;+\; \int zp(z)\,dz \\
&= \mathbb{E}[x] \;+\; \mathbb{E}[z]
\end{aligned}$$

$$\begin{aligned}
var[x+z] &= \int\int (x+z-\mathbb{E}[x+z])^2 p(x,z)\,dx\,dz \\
&= \int\int \{(x+z)^2 - 2(x+z)\mathbb{E}[x+z]) + \mathbb{E}^2[x+z]\}\,p(x,z)\,dx\,dz \\
&= \int\int (x+z)^2\,p(x,z)\,dx\,dz - 2\mathbb{E}[x+z]\int\int (x+z)p(x,z)\,dx\,dz + \mathbb{E}^2[x+z] \\
&= \int\int (x+z)^2\,p(x,z)\,dx\,dz - \mathbb{E}^2[x+z] \\
&= \int\int (x^2+2xz+z^2)p(x)p(z)\,dx\,dz - \mathbb{E}^2[x+z] \\
&= \int(\int p(z)dz)x^2\,p(x)dx + \int\int 2xz\,p(x)p(z)\,dx\,dz + \int(\int p(x)dx)z^2\,p(z)dz - \mathbb{E}^2[x+z] \\
&= \mathbb{E}[x^2]+\mathbb{E}[z^2]-\mathbb{E}^2[x+z]+\int\int 2xz\,p(x)p(z)\,dx\,dz \\
&= \mathbb{E}[x^2]+\mathbb{E}[z^2]-(\mathbb{E}[x]+\mathbb{E}[z])^2+\int\int 2xz\,p(x)p(z)\,dx\,dz \\
&= \mathbb{E}[x^2]-\mathbb{E}^2[x]+\mathbb{E}[z^2]-\mathbb{E}^2[z]-2\mathbb{E}[x]\mathbb{E}[z]+2\int\int xz\,p(x)p(z)\,dx\,dz \\
&= var[x]+var[z]-2\mathbb{E}[x]\mathbb{E}[z]+2(\int xp(x)\,dx)(\int zp(z)\,dz) \\
&= var[x]+var[z]
\end{aligned}$$

Problem 1.11 Solution

Based on prior knowledge that μ_{ML} and σ^2_{ML} will decouple. We will first calculate μ_{ML} :

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

We let :

$$\frac{d(ln\,p(\mathbf{x}|\,\mu,\sigma^2))}{d\mu} = 0$$

Therefore :

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

And because:

$$\frac{d(ln\, p(\mathbf{x}|\mu,\sigma^2))}{d\sigma^2} = \frac{1}{2\sigma^4}(\sum_{n=1}^{N}(x_n-\mu)^2 - N\sigma^2)$$

We let :

$$\frac{d(ln\, p(\mathbf{x}|\mu,\sigma^2))}{d\sigma^2} = 0$$

Therefore :

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

Problem 1.12 Solution

It is quite straightforward for $\mathbb{E}[\mu_{ML}]$, with the prior knowledge that x_n is i.i.d. and it also obeys Gaussian distribution $\mathcal{N}(\mu,\sigma^2)$.

$$\mathbb{E}[\mu_{ML}] = \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N} x_n] = \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N} x_n] = \mathbb{E}[x_n] = \mu$$

For $\mathbb{E}[\sigma_{ML}^2]$, we need to take advantage of (1.56) and what has been given in the problem :

$$\begin{aligned}
\mathbb{E}[\sigma_{ML}^2] &= \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2] \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}(x_n-\mu_{ML})^2] \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}(x_n^2-2x_n\mu_{ML}+\mu_{ML}^2)] \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}x_n^2] - \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}2x_n\mu_{ML}] + \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}\mu_{ML}^2] \\
&= \mu^2+\sigma^2 - \frac{2}{N}\mathbb{E}[\sum_{n=1}^{N}x_n(\frac{1}{N}\sum_{n=1}^{N}x_n)] + \mathbb{E}[\mu_{ML}^2] \\
&= \mu^2+\sigma^2 - \frac{2}{N^2}\mathbb{E}[\sum_{n=1}^{N}x_n(\sum_{n=1}^{N}x_n)] + \mathbb{E}[(\frac{1}{N}\sum_{n=1}^{N}x_n)^2] \\
&= \mu^2+\sigma^2 - \frac{2}{N^2}\mathbb{E}[(\sum_{n=1}^{N}x_n)^2] + \frac{1}{N^2}\mathbb{E}[(\sum_{n=1}^{N}x_n)^2] \\
&= \mu^2+\sigma^2 - \frac{1}{N^2}\mathbb{E}[(\sum_{n=1}^{N}x_n)^2] \\
&= \mu^2+\sigma^2 - \frac{1}{N^2}[N(N\mu^2+\sigma^2)]
\end{aligned}$$

Therefore we have:

$$\mathbb{E}[\sigma_{ML}^2] = (\frac{N-1}{N})\sigma^2$$

Problem 1.13 Solution

This problem can be solved in the same method used in Prob.1.12 :

$$\begin{aligned}
\mathbb{E}[\sigma_{ML}^2] &= \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2] \quad \text{(Because here we use } \mu \text{ to replace } \mu_{ML}\text{)} \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}(x_n-\mu)^2] \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}(x_n^2-2x_n\mu+\mu^2)] \\
&= \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}x_n^2] - \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}2x_n\mu] + \frac{1}{N}\mathbb{E}[\sum_{n=1}^{N}\mu^2] \\
&= \mu^2+\sigma^2-\frac{2\mu}{N}\mathbb{E}[\sum_{n=1}^{N}x_n]+\mu^2 \\
&= \mu^2+\sigma^2-2\mu^2+\mu^2 \\
&= \sigma^2
\end{aligned}$$

Note: The biggest difference between Prob.1.12 and Prob.1.13 is that the mean of Gaussian Distribution is known previously (in Prob.1.13) or not (in Prob.1.12). In other words, the difference can be shown by the following equations:

$$\mathbb{E}[\mu^2] = \mu^2 \quad (\mu \text{ is determined, i.e. its } \textit{expectation} \text{ is itself, also true for } \mu^2)$$

$$\mathbb{E}[\mu_{ML}^2] = \mathbb{E}[(\frac{1}{N}\sum_{n=1}^{N}x_n)^2] = \frac{1}{N^2}\mathbb{E}[(\sum_{n=1}^{N}x_n)^2] = \frac{1}{N^2}N(N\mu^2+\sigma^2) = \mu^2+\frac{\sigma^2}{N}$$