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

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

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 :

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

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 root. 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:

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

We let :

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

Therefore :

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

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

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

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:

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

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

Problem 1.14 Solution

This problem is quite similar to the fact that *any function $f(x)$ can be written into the sum of an odd function and an even function*. If we let:

$$w_{ij}^S = \frac{w_{ij}+w_{ji}}{2} \quad \text{and} \quad w_{ij}^A = \frac{w_{ij}-w_{ji}}{2}$$

It is obvious that they satisfy the constraints described in the problem, which are :

$$w_{ij} = w_{ij}^S + w_{ij}^A\,, \quad w_{ij}^S = w_{ji}^S\,, \quad w_{ij}^A = -w_{ji}^A$$

To prove (1.132), we only need to simplify it :

$$\begin{aligned}\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}x_i x_j &= \sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij}^S + w_{ij}^A)x_i x_j \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^S x_i x_j + \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^A x_i x_j\end{aligned}$$

Therefore, we only need to prove that the second term equals to 0, and here we use a simple trick: we will prove twice of the second term equals to 0 instead.

$$\begin{aligned}2\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^A x_i x_j &= \sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij}^A + w_{ij}^A)x_i x_j \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D}(w_{ij}^A - w_{ji}^A)x_i x_j \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^A x_i x_j - \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ji}^A x_i x_j \\ &= \sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}^A x_i x_j - \sum_{j=1}^{D}\sum_{i=1}^{D} w_{ji}^A x_j x_i \\ &= 0\end{aligned}$$

Therefore, we choose the coefficient matrix to be symmetric as described in the problem. Considering about the symmetry, we can see that if and only if for $i = 1,2,...,D$ and $i \le j$, w_{ij} is given, the whole matrix will be determined. Hence, the number of independent parameters are given by :

$$D + D - 1 + ... + 1 = \frac{D(D+1)}{2}$$

Note: You can view this intuitively by considering if the upper triangular part of a symmetric matrix is given, the whole matrix will be determined.

Problem 1.15 Solution

This problem is a more general form of Prob.1.14, so the method can also be used here: we will find a way to use $w_{i_1 i_2 ... i_M}$ to represent $\widetilde{w}_{i_1 i_2 ... i_M}$.

We begin by introducing a mapping function:

$$F(x_{i1}x_{i2}...x_{iM}) = x_{j1}x_{j2}...,x_{jM}$$

$$s.t. \bigcup_{k=1}^{M} x_{ik} = \bigcup_{k=1}^{M} x_{jk}, \quad \text{and} \quad x_{j1} \ge x_{j2} \ge x_{j3} ... \ge x_{jM}$$

It is complexed to write F in mathematical form. Actually this function does a simple work: it rearranges the element in a decreasing order based on its subindex. Several examples are given below, when $D = 5$, $M = 4$:

$$F(x_5x_2x_3x_2) = x_5x_3x_2x_2$$

$$F(x_1x_3x_3x_2) = x_3x_3x_2x_1$$

$$F(x_1x_4x_2x_3) = x_4x_3x_2x_1$$

$$F(x_1x_1x_5x_2) = x_5x_2x_1x_1$$

After introducing F, the solution will be very simple, based on the fact that F will not change the value of the term, but only rearrange it.

$$\sum_{i_1=1}^{D}\sum_{i_2=1}^{D}\cdots\sum_{i_M=1}^{D} w_{i_1i_2\ldots i_M}x_{i1}x_{i2}\ldots x_{iM} = \sum_{j_1=1}^{D}\sum_{j_2=1}^{j_1}\cdots\sum_{j_M=1}^{j_{M-1}} \widetilde{w}_{j_1j_2\ldots j_M}x_{j1}x_{j2}\ldots x_{jM}$$

where $$\widetilde{w}_{j_1j_2\ldots j_M} = \sum_{w\in\Omega} w$$

$$\Omega = \{w_{i_1i_2\ldots i_M} \,\big|\, F(x_{i1}x_{i2}\ldots x_{iM}) = x_{j1}x_{j2}\ldots x_{jM},\ \forall x_{i1}x_{i2}\ldots x_{iM}\,\}$$

By far, we have already proven (1.134). *Mathematical induction* will be used to prove (1.135) and we will begin by proving $D = 1$, i.e. $n(1,M) = n(1,M-1)$. When $D = 1$, (1.134) will degenerate into $\widetilde{w}x_1^M$, i.e., it only has one term, whose coefficient is govern by $\widetilde{w}$ regardless the value of M.

Therefore, we have proven when $D = 1$, $n(D,M) = 1$. Suppose (1.135) holds for D, let's prove it will also hold for $D+1$, and then (1.135) will be proved based on *Mathematical induction*.

Let's begin based on (1.134):

$$\sum_{i_1=1}^{D+1}\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} \widetilde{w}_{i_1i_2\ldots i_M}x_{i1}x_{i2}\ldots x_{iM} \tag{$*$}$$

We divide $(*)$ into two parts based on the first summation: the first part is made up of $i_i = 1,2,\ldots,D$ and the second part $i_1 = D+1$. After division, the first part corresponds to $n(D,M)$, and the second part corresponds to $n(D+1,M-1)$. Therefore we obtain:

$$n(D+1,M) = n(D,M) + n(D+1,M-1) \tag{$**$}$$

And given the fact that (1.135) holds for D :

$$n(D,M) = \sum_{i=1}^{D} n(i,M-1)$$

Therefore,we substitute it into $(**)$

$$n(D+1,M)=\sum_{i=1}^{D}n(i,M-1)+n(D+1,M-1)=\sum_{i=1}^{D+1}n(i,M-1)$$

We will prove (1.136) in a different but simple way. We rewrite (1.136) in *Permutation and Combination* view:

$$\sum_{i=1}^{D}C_{i+M-2}^{M-1}=C_{D+M-1}^{M}$$

Firstly, We expand the summation.

$$C_{M-1}^{M-1}+C_{M}^{M-1}+...C_{D+M-2}^{M-1}=C_{D+M-1}^{M}$$

Secondly, we rewrite the first term on the left side to C_M^M, because $C_{M-1}^{M-1}=C_M^M=1$. In other words, we only need to prove:

$$C_{M}^{M}+C_{M}^{M-1}+...C_{D+M-2}^{M-1}=C_{D+M-1}^{M}$$

Thirdly, we take advantage of the property : $C_N^r=C_{N-1}^r+C_{N-1}^{r-1}$. So we can recursively combine the first term and the second term on the left side, and it will ultimately equal to the right side.

(1.137) gives the mathematical form of $n(D,M)$, and we need all the conclusions above to prove it. When $M=1$,it is obvious $n(D,1)=D$, because in this case (1.134) will only have D terms if we expand it. When $M=2$, it degenerates to Prob.1.14, so $n(D,2)=\frac{D(D+1)}{2}$ is also obvious. Suppose (1.137) holds for $M-1$, let's prove it will also hold for M.

$$\begin{aligned}
n(D,M) &= \sum_{i=1}^{D}n(i,M-1) \quad \text{(based on (1.135))}\\
&= \sum_{i=1}^{D}C_{i+M-2}^{M-1} \quad \text{(based on (1.137) holds for } M-1 \text{)}\\
&= C_{M-1}^{M-1}+C_{M}^{M-1}+C_{M+1}^{M-1}...+C_{D+M-2}^{M-1}\\
&= (C_{M}^{M}+C_{M}^{M-1})+C_{M+1}^{M-1}...+C_{D+M-2}^{M-1}\\
&= (C_{M+1}^{M}+C_{M+1}^{M-1})...+C_{D+M-2}^{M-1}\\
&= C_{M+2}^{M}...+C_{D+M-2}^{M-1}\\
&\quad ...\\
&= C_{D+M-1}^{M}
\end{aligned}$$

By far, all have been proven.