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, $p_y(y) = 2sin(y)|cos(y)|$, $y \in [0, \frac{\pi}{2}]$, 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 is blablabla...