
Machine Learning

Answer Sheet for Homework 6

Da-Min HUANG

R04942045

Graduate Institute of Communication Engineering, National Taiwan University

December 23, 2015

Problem 1

With the definition of z_n , rewrite the equation

$$\min_{A,B} F(A, B) = \frac{1}{N} \sum_{n=1}^N \ln(1 + \exp(-y_n(Az_n + B))) \quad (1)$$

So

$$\frac{\partial F}{\partial A} = \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{\exp(-y_n(Az_n + B))}{1 + \exp(-y_n(Az_n + B))} \right)^T z_n = \frac{1}{N} \sum_{n=1}^N -y_n p_n^T z_n \quad (2)$$

$$\frac{\partial F}{\partial B} = \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{\exp(-y_n(Az_n + B))}{1 + \exp(-y_n(Az_n + B))} \right)^T = \frac{1}{N} \sum_{n=1}^N -y_n p_n^T \quad (3)$$

Hence

$$\nabla F(A, B) = \frac{1}{N} \sum_{n=1}^N [-y_n z_n p_n, -y_n p_n]^T \quad (4)$$

□

Problem 2

Use the result of Problem 1 and define $\exp(-y_n(Az_n + B)) = \exp(\xi_n)$, we have

$$\frac{\partial^2 F}{\partial A^2} = \frac{\partial}{\partial A} \left(\frac{1}{N} \sum_{n=1}^N -y_n p_n^T z_n \right) \quad (5)$$

$$= \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{-y_n \exp(\xi_n) (1 + \exp(\xi_n)) z_n - y_n (\exp(\xi_n))^2 z_n}{(1 + \exp(\xi_n))^2} \right) z_n \quad (6)$$

$$= \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{-y_n \exp(\xi_n) z_n}{(1 + \exp(\xi_n))^2} \right) z_n \quad (7)$$

$$= \frac{1}{N} \sum_{n=1}^N (y_n)^2 \left(\frac{\exp(\xi_n)}{1 + \exp(\xi_n)} \left(1 - \frac{\exp(\xi_n)}{1 + \exp(\xi_n)} \right) \right) z_n^2 \quad (8)$$

$$= \frac{1}{N} \sum_{n=1}^N z_n^2 p_n (1 - p_n) \quad (9)$$

where $y_n^2 = 1$ since $y_n \in \{-1, +1\}$.

The other term is

$$\frac{\partial^2 F}{\partial A \partial B} = \frac{\partial}{\partial A} \left(\frac{1}{N} \sum_{n=1}^N -y_n p_n^T \right) = \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{-y_n \exp(\xi_n) z_n}{(1 + \exp(\xi_n))^2} \right) \quad (10)$$

$$= \frac{1}{N} \sum_{n=1}^N z_n p_n (1 - p_n) \quad (11)$$

$$\frac{\partial^2 F}{\partial B \partial A} = \frac{\partial}{\partial B} \left(\frac{1}{N} \sum_{n=1}^N -y_n p_n^T z_n \right) = \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{-y_n \exp(\xi_n)}{(1 + \exp(\xi_n))^2} \right) z_n \quad (12)$$

$$= \frac{1}{N} \sum_{n=1}^N z_n p_n (1 - p_n) \quad (13)$$

$$\frac{\partial^2 F}{\partial B^2} = \frac{\partial}{\partial B} \left(\frac{1}{N} \sum_{n=1}^N -y_n p_n^T \right) = \frac{1}{N} \sum_{n=1}^N -y_n \left(\frac{-y_n \exp(\xi_n)}{(1 + \exp(\xi_n))^2} \right) \quad (14)$$

$$= \frac{1}{N} \sum_{n=1}^N p_n (1 - p_n) \quad (15)$$

Hence, we have

$$H(F) = \frac{1}{N} \sum_{n=1}^N \begin{bmatrix} z_n^2 p_n (1 - p_n) & z_n p_n (1 - p_n) \\ z_n p_n (1 - p_n) & p_n (1 - p_n) \end{bmatrix} \quad (16)$$

□

Problem 3

As $\gamma \rightarrow \infty$, we have

$$\lim_{\gamma \rightarrow \infty} \exp(-\gamma \|\mathbf{x}_n - \mathbf{x}_m\|^2) = 0 \quad (17)$$

So K should be a zero matrix with size $N \times N$, which is $\mathbf{0}_{N \times N}$.

And β is

$$\beta = (\lambda I + K)^{-1} \mathbf{y} = \lambda^{-1} \mathbf{y} \quad (18)$$

□

Problem 4

If $|y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| > \epsilon$, then

$$\begin{cases} |y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| - \epsilon = \xi_n^\wedge \text{ and } \xi_n^\vee = 0, & \text{if } y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b > 0 \\ |y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| - \epsilon = \xi_n^\vee \text{ and } \xi_n^\wedge = 0, & \text{if } y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b < 0 \end{cases} \quad (19)$$

and if $|y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| \leq \epsilon$, then $\xi_n^\wedge = 0$ and $\xi_n^\vee = 0$. Hence, we have

$$(\xi_n^\wedge)^2 + (\xi_n^\vee)^2 = (\max(0, |y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| - \epsilon))^2 \quad (20)$$

So P_2 is equivalent to

$$\min_{b, \mathbf{w}} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N (\max(0, |y_n - \mathbf{w}^T \phi(\mathbf{x}_n) - b| - \epsilon))^2 \right) \quad (21)$$

with no constraints.

□

Problem 5

□

Problem 6

□

Problem 7

☐

Problem 8

☐

Problem 9

☐

Problem 10

☐

Problem 11

☐

Problem 12

☐

Problem 13

☐

Problem 14

☐

Problem 15



Problem 16



Problem 17



Problem 18



Problem 19



Problem 20



Reference

- [1] Lecture Notes by Hsuan-Tien LIN, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.