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

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

So

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

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

Hence

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

$\square$

Problem 2

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

$$\begin{aligned}
\frac{\partial^2 F}{\partial A^2} &= \frac{\partial}{\partial A}\left(\frac{1}{N}\sum_{n=1}^{N}-y_n p_n^T z_n\right) &(5)\\
&= \frac{1}{N}\sum_{n=1}^{N}-y_n\left(\frac{-y_n\exp\left(\xi_n\right)\left(1+\exp\left(\xi_n\right)\right)z_n-y_n\left(\exp\left(\xi_n\right)\right)^2 z_n}{\left(1+\exp\left(\xi_n\right)\right)^2}\right)z_n &(6)\\
&= \frac{1}{N}\sum_{n=1}^{N}-y_n\left(\frac{-y_n\exp\left(\xi_n\right)z_n}{\left(1+\exp\left(\xi_n\right)\right)^2}\right)z_n &(7)\\
&= \frac{1}{N}\sum_{n=1}^{N}\left(y_n\right)^2\left(\frac{\exp\left(\xi_n\right)}{1+\exp\left(\xi_n\right)}\left(1-\frac{\exp\left(\xi_n\right)}{1+\exp\left(\xi_n\right)}\right)\right)z_n^2 &(8)\\
&= \frac{1}{N}\sum_{n=1}^{N}z_n^2 p_n\left(1-p_n\right) &(9)
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial^2 F}{\partial A\partial B} &= \frac{\partial}{\partial A}\left(\frac{1}{N}\sum_{n=1}^{N}-y_n p_n^T\right)=\frac{1}{N}\sum_{n=1}^{N}-y_n\left(\frac{-y_n\exp\left(\xi_n\right)z_n}{\left(1+\exp\left(\xi_n\right)\right)^2}\right) &(10)\\
&= \frac{1}{N}\sum_{n=1}^{N}z_n p_n\left(1-p_n\right) &(11)\\
\frac{\partial^2 F}{\partial B\partial A} &= \frac{\partial}{\partial B}\left(\frac{1}{N}\sum_{n=1}^{N}-y_n p_n^T z_n\right)=\frac{1}{N}\sum_{n=1}^{N}-y_n\left(\frac{-y_n\exp\left(\xi_n\right)}{\left(1+\exp\left(\xi_n\right)\right)^2}\right)z_n &(12)\\
&= \frac{1}{N}\sum_{n=1}^{N}z_n p_n\left(1-p_n\right) &(13)\\
\frac{\partial^2 F}{\partial B^2} &= \frac{\partial}{\partial B}\left(\frac{1}{N}\sum_{n=1}^{N}-y_n p_n^T\right)=\frac{1}{N}\sum_{n=1}^{N}-y_n\left(\frac{-y_n\exp\left(\xi_n\right)}{\left(1+\exp\left(\xi_n\right)\right)^2}\right) &(14)\\
&= \frac{1}{N}\sum_{n=1}^{N}p_n\left(1-p_n\right) &(15)
\end{aligned}$$

Hence, we have

$$H\left(F\right)=\frac{1}{N}\sum_{n=1}^{N}\left[\begin{array}{cc} z_n^2 p_n\left(1-p_n\right) & z_n p_n\left(1-p_n\right)\\ z_n p_n\left(1-p_n\right) & p_n\left(1-p_n\right)\end{array}\right] \tag{16}$$

$\square$

Problem 3

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

$$\lim_{\gamma\to\infty} \exp\left(-\gamma \left\|\mathbf{x}_n - \mathbf{x}_m\right\|^2\right) = 0 \tag{17}$$

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

And $\boldsymbol{\beta}$ is

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

$\square$

Problem 4

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

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

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

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

So P_2 is equivalent to

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

with no constraints.

$\square$

Problem 5

The first term is of course

$$\frac{\partial}{\partial\beta_m}\left(\frac{1}{2}\sum_{m=1}^{N}\sum_{n=1}^{N}\beta_n\beta_m K(\mathbf{x}_n, \mathbf{x}_m)\right) = \sum_{n=1}^{N}\beta_n K(\mathbf{x}_n, \mathbf{x}_m) \tag{22}$$

With $\mathbf{w}_*$, rewrite the result of Problem 4,

$$\text{something} = C\sum_{n=1}^{N}\left(\max\left(0, \left|y_n - \sum_{m=1}^{N}\beta_m K(\mathbf{x}_n, \mathbf{x}_m) - b\right| - \epsilon\right)\right)^2 \tag{23}$$

$$= C\sum_{n=1}^{N}\left(\max\left(0, |y_n - s_n| - \epsilon\right)\right)^2 \tag{24}$$

Consider the following cases.

1. $|y_n - s_n| \geq \epsilon$.

   Then we have

$$\frac{\partial}{\partial \beta_m}\left(\text{something}\right) = \frac{\partial}{\partial \beta_m}\left(C\left(|y_n - s_n| - \epsilon\right)^2\right) \tag{25}$$

$$= \left(2C\left(|y_n - s_n| - \epsilon\right)\right)\frac{\partial}{\partial \beta_m}|y_n - s_n| \tag{26}$$

$$= -2C\left(|y_n - s_n| - \epsilon\right)\operatorname{sign}\left(y_n - s_n\right)\frac{\partial s_n}{\partial \beta_m} \tag{27}$$

$$= -2C\left(|y_n - s_n| - \epsilon\right)\operatorname{sign}\left(y_n - s_n\right)K\left(\mathbf{x}_n, \mathbf{x}_m\right) \tag{28}$$

2. $|y_n - s_n| < \epsilon$.

   Then we have

$$\frac{\partial}{\partial \beta_m}\left(\text{something}\right) = 0 \tag{29}$$

So we have

$$\frac{\partial F\left(b, \boldsymbol{\beta}\right)}{\partial \beta_m} = \sum_{n=1}^{N}\beta_n K\left(\mathbf{x}_n, \mathbf{x}_m\right) - 2C\sum_{n=1}^{N}[\![|y_n - s_n| \geq \epsilon]\!]\left(|y_n - s_n| - \epsilon\right)\operatorname{sign}\left(y_n - s_n\right)K\left(\mathbf{x}_n, \mathbf{x}_m\right) \tag{30}$$

□

Problem 6

First, we have

$$e_t = \frac{1}{M}\sum_{m=1}^{M}\left(g_t\left(\tilde{\mathbf{x}}_m\right)\right)^2 - 2g_t\left(\tilde{\mathbf{x}}_m\right)\tilde{y}_m + \left(\tilde{y}_m\right)^2 \tag{31}$$

And e_0 is

$$e_0 = \frac{1}{M}\sum_{m=1}^{M}\left(0\right)^2 - 2\cdot 0\cdot \tilde{y}_m + \left(\tilde{y}_m\right)^2 = \frac{1}{M}\sum_{m=1}^{M}\left(\tilde{y}_m\right)^2 \tag{32}$$

where we have used that $g_0\left(\mathbf{x}\right) = 0$, $\forall \mathbf{x}$.
So e_t can be rewritten as

$$e_t = e_0 + \frac{1}{M}\sum_{m=1}^{M}\left(g_t\left(\tilde{\mathbf{x}}_m\right)\right)^2 - 2g_t\left(\tilde{\mathbf{x}}_m\right)\tilde{y}_m = e_0 + s_t - \frac{2}{M}\sum_{m=1}^{M}g_t\left(\tilde{\mathbf{x}}_m\right)\tilde{y}_m \tag{33}$$

Hence,

$$\sum_{m=1}^{M} g_t\left(\tilde{\mathbf{x}}_m\right) \tilde{y}_m = \frac{M}{2}\left(e_0 + s_t - e_t\right) \tag{34}$$

$\square$

Problem 7

$\square$

Problem 8

$\square$

Problem 9

$\square$

Problem 10

$\square$

Problem 11

$\square$

Problem 12

$\square$

Problem 13

$\square$

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.