Machine Learning

Answer Sheet for Homework 4

Da-Min HUANG

R04942045

Graduate Institute of Communication Engineering, National Taiwan University

November 7, 2015

Problem 1

Deterministic error is the difference between best $h^* \in H$ and f. If $H' \subset H$, then the complexity of H' is lower than H in general. Hence, in general, the deterministic error increases.

$\square$

Problem 2

1.

$$
\begin{aligned}
H(10,0,3) \cap H(10,0,4) &= \left\{\sum_{i=0}^{2} w_q L_q(x)\right\} \cap \left\{\sum_{i=0}^{3} w_q L_q(x)\right\} && (1)\\
&= \left\{\sum_{i=0}^{2} w_q L_q(x)\right\} = H_2 && (2)
\end{aligned}
$$

2.

$$
\begin{aligned}
H(10,0,3) \cup H(10,1,4) &= \left\{\sum_{i=0}^{2} w_q L_q(x)\right\} \cup \left\{\sum_{i=0}^{3} w_q L_q(x) + \sum_{i=4}^{10} L_q(x)\right\} && (3)\\
&= \left\{\sum_{i=0}^{3} w_q L_q(x) + \sum_{i=4}^{10} L_q(x)\right\} && (4)
\end{aligned}
$$

3.

$$H(10,1,3) \cap H(10,1,4) = \left\{ \sum_{i=0}^{2} w_q L_q(x) + \sum_{i=3}^{10} L_q(x) \right\} \cap \left\{ \sum_{i=0}^{3} w_q L_q(x) + \sum_{i=4}^{10} L_q(x) \right\} \tag{5}$$

$$= \left\{ \sum_{i=0}^{2} w_q L_q(x) + \sum_{i=4}^{10} L_q(x) \right\} \tag{6}$$

4.

$$H(10,0,3) \cup H(10,0,4) = \left\{ \sum_{i=0}^{2} w_q L_q(x) \right\} \cup \left\{ \sum_{i=0}^{3} w_q L_q(x) \right\} \tag{7}$$

$$= \left\{ \sum_{i=0}^{3} w_q L_q(x) \right\} = H_3 \tag{8}$$

$\square$

Problem 3

We have

$$\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \eta \nabla E_{\text{aug}}(\mathbf{w}_t) \tag{9}$$

where $\nabla E_{\text{aug}}(\mathbf{w}_t)$ is

$$\nabla E_{\text{aug}}(\mathbf{w}_t) = \frac{\partial}{\partial \mathbf{w}_t^T} \left(E_{\text{in}}(\mathbf{w}_t) + \frac{\lambda}{N} \mathbf{w}_t^T \mathbf{w}_t \right) = \frac{\partial E_{\text{in}}(\mathbf{w}_t)}{\partial \mathbf{w}_t^T} + \frac{2\lambda}{N} \mathbf{w}_t \tag{10}$$

Hence, we have

$$\mathbf{w}_{t+1} \leftarrow \left(1 - \frac{2\eta\lambda}{N}\right) \mathbf{w}_t - \eta \nabla E_{\text{in}}(\mathbf{w}_t) \tag{11}$$

$\square$

Problem 4

Since $\mathbf{w}_{\text{lin}}$ is the optimal solution for the plain-vanilla linear regression, we have

$$E_{\text{in}}(\mathbf{w}_{\text{lin}}) \leq E_{\text{in}}(\mathbf{w}_{\text{reg}}(\lambda)) \tag{12}$$

Also, $\mathbf{w}_{\text{reg}}(\lambda)$ is the optimal solution for $E_{\text{aug}}(\mathbf{w})$, we have

$$E_{\text{aug}}(\mathbf{w}_{\text{reg}}(\lambda)) \leq E_{\text{aug}}(\mathbf{w}_{\text{lin}}) \tag{13}$$

So, we have

$$E_{\text{in}}\left(\mathbf{w}_{\text{reg}}(\lambda)\right) + \frac{\lambda}{N}\mathbf{w}_{\text{reg}}^{T}(\lambda)\,\mathbf{w}_{\text{reg}}(\lambda) \le E_{\text{in}}\left(\mathbf{w}_{\text{lin}}\right) + \frac{\lambda}{N}\mathbf{w}_{\text{lin}}^{T}\mathbf{w}_{\text{lin}} \tag{14}$$

$$0 \le E_{\text{in}}\left(\mathbf{w}_{\text{reg}}(\lambda)\right) - E_{\text{in}}\left(\mathbf{w}_{\text{lin}}\right) \le \frac{\lambda}{N}\left(\left\|\mathbf{w}_{\text{lin}}\right\|^2 - \left\|\mathbf{w}_{\text{reg}}(\lambda)\right\|^2\right),\ \forall\lambda \tag{15}$$

Hence, we have $\|\mathbf{w}_{\text{lin}}\| \ge \|\mathbf{w}_{\text{reg}}(\lambda)\|$ if $\lambda > 0$.
Since this inequality holds for all λ and $\|\mathbf{w}_{\text{lin}}\|$ is not a function of λ. We know that $\|\mathbf{w}_{\text{reg}}(\lambda)\|$ is a non-increasing function of λ for $\lambda \ge 0$.

□

Problem 5

For constant model with three points $A(-1,0)$, $B(\rho,1)$ and $C(1,0)$.

$$\frac{1}{3}\left(\underbrace{\left(0-\frac{1}{2}\right)^2}_{\text{leave } A} + \underbrace{(1-0)^2}_{\text{leave } B} + \underbrace{\left(0-\frac{1}{2}\right)^2}_{\text{leave } C}\right) = \frac{1}{2} \tag{16}$$

For linear model. Leave A, we get line $y = \frac{1}{\rho-1}(x-1)$; leave B, we get line $y = 0$; leave C, we get line $y = \frac{1}{\rho+1}(x+1)$. So the error is

$$\frac{1}{3}\left(\left(0-\left(\frac{-2}{\rho-1}\right)\right)^2 + (1-0)^2 + \left(0-\frac{2}{\rho+1}\right)^2\right) \tag{17}$$

Then we have

$$\frac{1}{3}\left(\frac{4}{\rho^2-2\rho+1} + 1 + \frac{4}{\rho^2+2\rho+1}\right) = \frac{1}{2} \Rightarrow \rho = \pm\sqrt{9+4\sqrt{6}} \tag{18}$$

Since $\rho > 0$, we have $\rho = \sqrt{9+4\sqrt{6}}$.

□

Problem 6

□

Problem 7

□

Problem 8

□

Problem 9

□

Problem 10

□

Problem 11

□

Problem 12

□

Problem 13

□

Problem 14

□

Problem 15

$\square$

Problem 16

$\square$

Problem 17

$\square$

Problem 18

$\square$

Problem 19

$\square$

Problem 20

$\square$

Reference

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