Homework 4

B99202064 Kuan Hou Chan

1 Ans: (B)

Since deterministic noise is the error between f and hypothesis, then if f is fixed and we choice a hypothesis $H^{'}$ that order is lower than H, , assume that f is 5^{th} order target and H is 4th order and $H^{'}$is 3rd order, then we use a 3rd order hypothesis to fit the 5th order target, the deterministic noise will increase.
Hence in general, deterministic noise will increase.

2 Ans: C

$H(10,0,3) = \sum_{q=0}^{10} w_q L_q = w_0 L_0 + w_1 L_1 + w_2 L_2$
$H(10,0,4) = \sum_{q=0}^{10} w_q L_q = w_0 L_0 + w_1 L_1 + w_2 L_2 + w_3 L_3$
$H_2 = w_0 L_0 + w_1 L_1 + w_2 L_2 + w_3 L_3$
Hence, $H(10,0,3) \cap H(10,0,4) = H_2$

3 Ans: A,D

$$E_{aug}(w) = E_{in}(w) + \frac{\lambda}{N} w^T w$$

Using GD update rules, $w(t+1) = w(t) - \eta \nabla E_{aug}(w(t))$
$\nabla E_{aug} = \nabla E_{in} + \frac{\lambda}{N} \nabla (w^T w) = \nabla E_{in} + \frac{2\lambda}{N}$
$\therefore w(t+1) = w(t) - \eta(\nabla E_{in} + \frac{2\lambda}{N}) = (1 - \frac{2\lambda\eta}{N}) w(t) - \eta \nabla E_{in}(w(t))$

4 Ans: B,C

1. If $||w_{reg}|| \geq ||w_{lin}||$ means that $||w_{lin}|| \leq \sqrt{C}$ since $||w_{reg}|| \leq \sqrt{C}$, it is contradiction because $||w_{lin}||$ will have a constraint condition that is unreasonable for regularization since we want to regularize the $||w_{reg}||$. Besides that if $||w_{reg}|| \leq \sqrt{C}$ then $||w_{lin}|| = ||w_{reg}||$ and $\lambda = 0$ because $||w_{lin}|| \in H(C)$.
Hence (B) is true and (A) and (E) is false.
2. As λ increase, then C will decrease implies that $||w_{reg}||$ will decrease since $||w_{reg}|| \leq \sqrt{C}$. Then $||w_{reg}||$ is non-increasing function of λ .
Hence (C) is true and (D) is false.

5 Ans: (C) $\sqrt{9+4\sqrt{6}}$

$h_0(x) = b_0$
Case 1:$(-1,0),(1,0)$, leave $(\rho,1)$
$h_0(x) = 0 = b_0 \Rightarrow y = 0$, distance $= 1$
Case 2:$(-1,0),(\rho,1)$, leave $(1,0)$
$h_0(x) = \frac{1}{2} = b_0 \Rightarrow y = \frac{1}{2}$, distance $= 1\frac{}{2Case3:}$(1,0),$(\rho,1)$, leave $(-1,0)$
$h_0(x) = \frac{1}{2} = b_0 \Rightarrow y = \frac{1}{2}$, distance $= 1\frac{}{2}$h$_1(x) = a_1x + b_1 = y$
$E_{h_0} = \frac{1}{3}(1 + \frac{1}{4} + \frac{1}{4})$
Case 1: $(-1,0),(1,0)$, leave $(\rho,1)$
$a_1 = 0, b_1 = 0 \Rightarrow y = 0$, distance $= 1$
Case 2:$(-1,0),(\rho,1)$, leave $(1,0)$
$a_1 = \frac{1}{\rho+1}, b_1 = a_1 = \frac{1}{\rho+1} \Rightarrow \frac{1}{\rho+1}x - y + \frac{1}{\rho+1} = 0$
$distance^2 = (\frac{1}{\rho+1} * 1 + \frac{1}{\rho+1})^2 = \frac{4}{(\frac{1}{\rho+1})^2}$
Case 3:$(1,0),(\rho,1)$, leave $(-1,0)$
$a_1 = \frac{1}{\rho-1}, b_1 = a_1 = -1 * \frac{1}{\rho-1} \Rightarrow \frac{1}{\rho-1}x - y - \frac{1}{\rho-1} = 0$
$distance^2 = (|\frac{-1}{\rho-1} * 1 - \frac{-1}{\rho-1}|)^2 = \frac{4}{(\frac{1}{\rho-1})^2}$
$E_{h_1} = \frac{1}{3}(1 + \frac{4}{(\frac{1}{\rho+1})^2} + \frac{4}{(\frac{1}{\rho-1})^2})$
Let $E_{h_0} = E_{h_1}$ we can obtain $\rho^4 - 18\rho^2 = 15$ and because $\rho > 0$ and solve the equation obtain $\rho = \sqrt{9+4\sqrt{6}}$

6 Ans: A,C

A:True. Since each week will have 2 predictions and along 5 week we can expand it like a tree that will have $2*2*2*2*2 = 32$ predictions.
B:False. Sender should target at least 32 people for make sure at least one person receive the correct prediction.
C:True. After the first game that will reduced the people from 32 to 16 and sender should target at least 16 people.
D:False. Sender should target at least 32 people for make sure at least one person receive the correct prediction.
E:False. A and C is true.

7 Ans: B

The cost: $32 + 16 + 8 + 4 + 2 + 1 = 63$
Earn $= 1000 - 63 * 10 = 370$

8 Ans: A

Since before i look at the data, i do some calculation and derive a credit approval function, then my hypothesis set will be 1 only

9 Ans: C

$N = 10000, \varepsilon = 0.01$
$P[E_{in}(h) - E_{out}(h) > \varepsilon] \leq 2e^{-2\varepsilon^2 N} = 0.271$

10 Ans: A,C

A: True. Since the bank applies it vague function $a(x)$ to some of customers, the data already choice by them self based on their own idea. Then the data is come from some knowledge. By sampling bias, the data is having some bias that when we training the data that will produce the same biased outcome. Hence if $a(x) = 1$ the situation should not happen. For improved the performance, and because $a(x)$ and $g(x)$ received the customers are iid and from the same distribution implies that the test set will have the same distribution of training set, and the bank is lucky enough so that Hoeffding

will work.
Since $g(x)$ is using $a(x)$ for error measure, and $g(x)$ is guarantee by Hoeffding, then we can use $g(x)$ to provide the Hoeffding bound and $a(x)$ to improved the performance, then the intersection of $a(x)$ and $g(x)$ will improve the overall performance.

11 Ans: D

$$min_w \frac{1}{N+K}[\sum_{n=1}^{N}(y_n - w^T x_n)^2 + \sum_{k=1}^{K}(\tilde{y}_k - w^T \tilde{x}_k)^2]$$

$$\frac{1}{N+K}[\sum_{n=1}^{N} 2(y_n - w^T x_n)(-x_n) + \sum_{k=1}^{K} 2(\tilde{y}_k - w^T \tilde{x}_k)(-\tilde{x}_k)] = 0$$

$$\sum_{n=1}^{N} 2(x_n y_n - w^T x_n x_n) = \sum_{k=1}^{K} 2(-\tilde{x}_k \tilde{y}_k + w^T \tilde{x}_k \tilde{x}_k)$$

$$(x^T y - w^T x^T x) = (-\tilde{x}^T \tilde{y} + w^T \tilde{x}^T \tilde{x})$$

$$w^T = (x^T x + \tilde{x}^T \tilde{x})^{-1}(x^T y + \tilde{x}^T \tilde{y})$$

12 Ans: B

$$w_{reg} = (X^T X + \lambda I)^{-1} X^T y$$

$$w_{reg} = w^T$$

$$(X^T X + \lambda I)^{-1} X^T y = (X^T X + \tilde{X}^T \tilde{X})^{-1}(X^T y + \tilde{X}^T \tilde{y})$$

$$(X^T X + \lambda I)^{-1} X^T y = (X^T X + \tilde{X}^T \tilde{X})^{-1} X^T y + (X^T X + \tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{y}$$

$$\tilde{y} = 0, \tilde{X} = \sqrt{\lambda} I$$

13 Bonus 21

From Q12, we know that for $w^T w, \tilde{x} = \sqrt{\lambda} I, \tilde{y} = 0$
then for $w^T \Gamma^T \Gamma w, w^T \tilde{x} = \sqrt{\lambda}\Gamma w \Rightarrow \tilde{x} = \sqrt{\lambda}\Gamma, \tilde{y} = 0$

14 Bonus 22

From Q12, we know that for $w^T w, \tilde{x} = \sqrt{\lambda} I, \tilde{y} = 0$
then for $||w - w_{hint}||^2, w^T \tilde{x} = \sqrt{\lambda}(w - w_{hint}) \Rightarrow \tilde{x} = \sqrt{\lambda}(1 - w^{-1} w_{hint}), \tilde{y} = 0$