Howework 5

Lingyu Zhou

2024.4.10

2.

1.

$$
\begin{aligned}
\bar{y}(x) &= \mathbb{E}[y] \\
&= \mathbb{E}[f(x) + \epsilon] \\
&= \mathbb{E}[f(x)] + \mathbb{E}[\epsilon] \\
&= f(x) \text{ , since } f \text{ is deterministic}
\end{aligned}
$$

2.

It measures the expected error of the expected classifier and the actual value given data x. It should be 0 when $\bar{h}_k(x)$ is the ideal expected classifier for data point x.

3.

$$
\begin{aligned}
\text{EPE}_k(x) &= \mathbb{E}_{D,(x,y)}[(y - h_k(x))^2] \\
&= \mathbb{E}_{D,(x,y)}[(y - \bar{y}(x) + \bar{y}(x) - h_k(x))^2] \\
&= \mathbb{E}_{D,(x,y)}[(y - \bar{y}(x))^2 + 2(y - \bar{y}(x))(\bar{y}(x) - h_k(x)) + (\bar{y}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_D[\mathbb{E}_{(x,y)}[2(y - \bar{y}(x))(\bar{y}(x) - h_k(x))]] + \mathbb{E}_{D,(x,y)}(\bar{y}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_D[2(\bar{y}(x) - h_k(x))\mathbb{E}_{(x,y)}[(y - \bar{y}(x))]] + \mathbb{E}_{D,(x,y)}(\bar{y}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_D[2(\bar{y}(x) - h_k(x))(\mathbb{E}_{(x,y)}[y] - \bar{y}(x)) + \mathbb{E}_{D,(x,y)}(\bar{y}(x) - h_k(x))^2] \\
&:= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_D[2(\bar{y}(x) - h_k(x))(\mathbb{E}[y|x] - \mathbb{E}[y|x]) + \mathbb{E}_{D,(x,y)}(\bar{y}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{D,(x,y)}[(\bar{y}(x) - \bar{h}(x) + \bar{h}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{D,(x,y)}[(\bar{y}(x) - \bar{h}(x))^2 + 2(\bar{y}(x) - \bar{h}(x))(\bar{h}(x) - h_k(x)) + (\bar{h}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{(x,y)}[(\bar{y}(x) - \bar{h}(x))^2] + \mathbb{E}_{D,(x,y)}[2(\bar{y}(x) - \bar{h}(x))(\bar{h}(x) - h_k(x)) + \mathbb{E}_{D,(x,y)}[(\bar{h}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{(x,y)}[(\bar{y}(x) - \bar{h}(x))^2] + \mathbb{E}_{(x,y}[\mathbb{E}_D[2(\bar{y}(x) - \bar{h}(x))(\bar{h}(x) - h_k(x))]] + \mathbb{E}_{D,(x,y)}[(\bar{h}(x) - h_k(x))^2] \\
&:= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{(x,y)}[(\bar{y}(x) - \bar{h}(x))^2] + \mathbb{E}_{(x,y}[\mathbb{E}_D[2(\bar{y}(x) - \bar{h}(x))(\bar{h}(x) - \bar{h}(x))]] + \mathbb{E}_{D,(x,y)}[(\bar{h}(x) - h_k(x))^2] \\
&= \mathbb{E}_{(x,y)}[(y - \bar{y}(x))^2] + \mathbb{E}_{(x,y)}[(\bar{y}(x) - \bar{h}(x))^2] + \mathbb{E}_{D,(x,y)}[(\bar{h}(x) - h_k(x))^2] \\
&= \underbrace{\mathbb{E}_{D,(x,y)}[(y(x) - \bar{y}(x))^2]}_{\text{noise}} + \underbrace{\mathbb{E}_{D,(x,y)}[(\bar{y}(x) - \bar{h}(x))^2]}_{\text{bias}} + \underbrace{\mathbb{E}_{D,(x,y)}[(\bar{h}(x) - h_k(x))^2]}_{\text{variance}}
\end{aligned}
$$

4.

$$
\begin{aligned}
\mathbb{E}_{D,(x,y)}[(\bar{y}(x) - y(x))^2] &= \mathbb{E}_{D,(x,y)}\left[(\mathbb{E}[f(x) + \epsilon] - f(x) - \epsilon)^2\right] \\
&= \mathbb{E}_{D,(x,y)}\left[(f(x) + 0 - f(x) - \epsilon)^2\right] \\
&= \mathbb{E}_{D,(x,y)}\left[\epsilon^2\right] \\
&= VAR(\epsilon) - \mathbb{E}[\epsilon]^2 \\
&= \sigma^2
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{D,(x,y)}[(h(x)-\bar{h}(x))^2] &= \mathbb{E}\left[\left(\frac{1}{k}\sum_{l=1}^{k} f(x_l)+\epsilon_l-\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} f(x_l)+\epsilon_l\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{k}\sum_{l=1}^{k} f(x_l)+\epsilon_l-\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\left[f(x_l)+\epsilon_l\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{k}\sum_{l=1}^{k} f(x_l)+\epsilon_l-f(x_l)\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{k}\sum_{l=1}^{k}\epsilon_l\right)^2\right] \\
&= \frac{1}{k^2}\mathbb{E}\left[\left(\sum_{l=1}^{k}\epsilon_l\right)^2\right] \\
&= \frac{1}{k^2}\mathbb{E}\left[\left(\left(\sum_{l=1}^{k}\epsilon_l^2\right)+\epsilon_1\epsilon_2+\cdots+\epsilon_{k-1}\epsilon_k\right)^2\right] \\
&= \frac{1}{k^2}\left[\left(\mathbb{E}\left(\sum_{l=1}^{k}\epsilon_l^2\right)+\mathbb{E}[\epsilon_1\epsilon_2+\cdots+\epsilon_{k-1}\epsilon_k]\right)^2\right] \\
&= \frac{1}{k^2}\left[\left(\mathbb{E}\left(\sum_{l=1}^{k}\epsilon_l^2\right)+\mathbb{E}[\epsilon_1]\mathbb{E}[\epsilon_2]+\cdots+\mathbb{E}[\epsilon_{k-1}]\mathbb{E}[\epsilon_k]\right)^2\right], \epsilon \text{ independent with each other} \\
&= \frac{1}{k^2}\left[\sum_{l=1}^{k}\mathbb{E}[\epsilon_l^2]\right] \\
&= \frac{1}{k^2}\left[\sum_{l=1}^{k} VAR(\epsilon)-\mathbb{E}[\epsilon]^2\right] \\
&= \frac{1}{k^2}\left[k\sigma^2\right] \\
&= \frac{\sigma^2}{k}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{D,(x,y)}[(\bar{h}(x)-\bar{y}(x))^2] &= \mathbb{E}_{D,(x,y)}\left[\left(\frac{1}{k}\sum_{i=1}^{k} f(x_l)-\mathbb{E}[f(x)+\epsilon]\right)^2\right] \\
&= \mathbb{E}_{D,(x,y)}\left[\left(\frac{1}{k}\sum_{i=1}^{k} f(x_l)-\mathbb{E}[f(x)]+\mathbb{E}[\epsilon]\right)^2\right] \\
&= \mathbb{E}_{D,(x,y)}\left[\left(\frac{1}{k}\sum_{i=1}^{k} f(x_l)-f(x)\right)^2\right] \\
&= \left(\frac{1}{k}\sum_{i=1}^{k} f(x_l)-f(x)\right)^2, \text{ since } f(\cdot) \text{ is deterministic}
\end{aligned}
$$

Therefore:

$$
\text{EPE}_k(x) = \sigma^2+\frac{\sigma^2}{k}+\left(\frac{1}{k}\sum_{i=1}^{k} f(x_l)-f(x)\right)^2 = \sigma^2+\frac{\sigma^2}{k}+\left(\frac{f(x_1)+\cdots+f(x_k)}{k}-f(x)\right)^2
$$

3.

	Overfitting	Underfitting
increase the regularization	yes	no
decrease the regularization	no	yes
use less features	yes	no
use more features	no	yes
use a more complex model	no	yes
use a less complex model	yes	no

4.

In this question for simplicity, write 2-norm "$||\cdot||_2$" as "$||\cdot||$", $\vec{w}$ as w, $\vec{x}$ as x, keep $\bar{w}$

1.

Since $||a-b|| < ||a|| + ||b||$ and $||a-b|| \geq 0$, $||a-b||^2 < ||a||^2 + 2||a|| \cdot ||b|| + ||b||^2$.
Therefore$||w(D) - \bar{w}||^2 < ||w(d)||^2 + 2||w(D)|| \cdot ||\bar{w}|| + ||\bar{w}||^2$.
Given $||w(d)|| \leq B$, it follows $||w(D)||^2 \leq B^2$; slso since $\bar{w} = \mathbb{E}(w(D))$ and $\bar{w} \leq \max_D(w(D)) \leq B$, combined we have $||w(D) - \bar{w}||^2 \leq B^2 + 2B \cdot B + B^2 = 4B^2$

2.

$$
\begin{aligned}
h_D(x) - \bar{h}(x) &= w(D)^\top x - \mathbb{E}_D[h_D(x)] \\
&= w(D)^\top x - \mathbb{E}_D[w(D)^\top x] \\
&= w(D)^\top x - \mathbb{E}_D[w(D)]^\top x, \text{ since } x \text{ not depend on } D \\
&= w(D)^\top x - \bar{w}(D)^\top x \\
&= (w(D) - \bar{w}(D))^\top x
\end{aligned}
$$

Taking square and expectation on both sides gives:

$$
\begin{aligned}
\mathbb{E}\left[\left(h_D(x) - \bar{h}(x)\right)\right]^2 &= \mathbb{E}\left[\left((w(D) - \bar{w}(D)^\top x)\right)^2\right] \\
&\leq \mathbb{E}\left[(w(D) - \bar{w}(D))^\top (w(D) - \bar{w}(D))\, x^\top x\right], \text{ given} \\
&= \mathbb{E}\left[||w(D) - \bar{w}(D)||^2 \cdot ||x||^2\right] \\
&= \mathbb{E}\left[||w(D) - \bar{w}(D)||^2\right], \text{ given} \\
&\leq 4B^2, \text{by } \mathbf{1.}
\end{aligned}
$$