Lec 36 Validation (ch 4.3)

1. Intro:

↳ For supervised learning:

- $\mathcal{D} = \{(x_1, y_1) \cdots (x_N, y_N)\}$

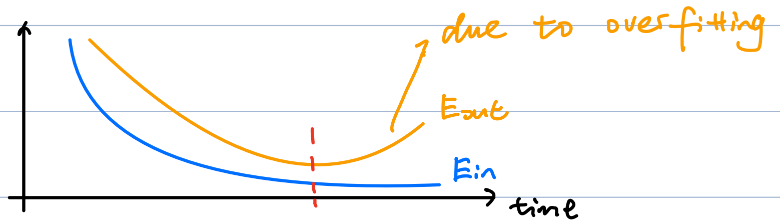
$$\mathcal{D} \longrightarrow \boxed{\text{Learning Algo}} \longrightarrow g$$

- Training error:

$$E_{in}(g) = \frac{1}{N}\sum_{n=1}^{N} e(g(x_n), y_n)$$

- Test error:

$$E_{out}(g) = \mathbb{E}_x[e(g(x), y)].$$

- 

stop here is good. But how to know this pt?

↳ Want to estimate $E_{out}(g)$ using training data.

<u>Problem:</u> $E_{in}(h)$ is a good estimation for $E_{out}(h)$ (Hoeffding)

when h is given <u>independent</u> of $\mathcal{D}$

But not necessarily good for final hypo. g: VC dim

bias-var trade off

<u>Idea:</u> use validation dataset

$$\underset{N \text{ samples}}{\mathcal{D}} \longrightarrow \begin{cases} \mathcal{D}_{train} : & N-K \text{ samples} \\ \mathcal{D}_{valid} : & K \text{ samples} \end{cases}$$

$$\mathcal{D}_{train} \longrightarrow \boxed{\text{Learing Algo}} \longrightarrow \boxed{g^-}$$

2. Validation

↳ $\underset{N \text{ samples}}{\mathcal{D}} \longrightarrow \begin{cases} \mathcal{D}_{train} : & N-K \text{ samples} \\ \mathcal{D}_{valid} : & K \text{ samples} \end{cases}$

$$\mathcal{D}_{train} \longrightarrow \boxed{\begin{array}{c}\text{Learing} \\ \text{Algo}\end{array}} \longrightarrow \boxed{g^-}$$

↳ $E_{val}(g^-) = \frac{1}{K} \sum\limits_{(x_n, y_n) \in \mathcal{D}} e(g^-(x_n), y_n$

We want : $E_{val}(g^-) \approx E_{out}(g^-) \approx E_{out}(g)$

↳ Prf $E_{val}(g^-) \approx E_{out}(g^-)$:

- Properties :

① $\mathbb{E}_{\mathcal{D}val}[E_{val}(g^-)] = E_{out}(g^-)$

⇒ E_{val} is an unbiased estimate for $E_{out}(g^-)$

② $Var[E_{val}(g^-)] = \frac{1}{K}(\sigma^2)$, where

$\sigma^2 = var[e(g^-(x), y)]$

As $K \to \infty$, $Var \to 0$

∴ $E_{val}(g^-)$ is a consistent estimate for $E_{out}(g^-)$

③ With prob. $1-\delta$,

$E_{out}(g^-) \leq E_{val}(g^-) + \sqrt{\frac{1}{2K} \log \frac{2}{\delta}}$

// $O(\frac{1}{\sqrt{K}})$ for binary classification

- Fact :

Relation b/t $E_{val}(g^-)$ and $E_{out}(g^-)$ is nearly identical to

〃 〃 $E_{in}(h)$ and $E_{out}(h)$

- Prf:

Denote $\mathcal{D}_{val} = \{(x_1, y_1) \cdots (x_K, y_K)\}$

$\because$ Property ①:

$$\mathbb{E}_{\mathcal{D}val}[E_{val}(g^-)] = \mathbb{E}_{val}[\underbrace{\frac{1}{K}\sum_{n=1}^{K} e(g^-(x_n), y_n)}_{E_{val}(g^-)}]$$

$$= \frac{1}{K}\sum_{n=1}^{K} \mathbb{E}_{x_n}\underbrace{[e(g^-(x_n), y_n)]}_{\text{def. of } E_{out}(g^-)}$$

$$= E_{out}(g^-)$$

$\because$ Property ②:

$$Var[E_{val}(g^-)] = Var[\frac{1}{K}\sum_{n=1}^{K} e(g^-(x_n), y_n)]$$

$$= \frac{1}{K^2}\sum_{n=1}^{K}\sigma^2$$

$$= \frac{1}{K}\sigma^2$$

$\because$ Property ③:

g^- is independent of $\mathcal{D}_{val}$

$\Rightarrow$ Hoeffding bound is valid with K samples

↳ How to select K?

$E_{val}(g^-) \approx E_{out}(g^-)$ for large K

$E_{out}(g^+) \approx E_{out}(g)$ for small K

In practise $K \approx \frac{N}{5}$ is reasonable $\Rightarrow$ 20%

3. <u>Model selection by Validation:</u>

↳ <u>Problem:</u> Given a dataset $\mathcal{D}$,

select the best model from $\mathcal{H}_1 \cdots \mathcal{H}_M$

e.g. $\mathcal{H}_i$: the class of ith-order polynomial

<u>step 1:</u>

train each model $\mathcal{H}_m$ on the training set $\mathcal{D}_{train}$ to output the final hypo. g_m^-

<u>step 2:</u>

use D_{val} to compute $E_{val}(g_m^-)$ for $1 \le m \le M$

<u>step 3:</u>

select the best model $\mathcal{H}_{m^*}$

$$m^* = \underset{1 \le m \le M}{\arg\min} \; E_{val}(g^-_m)$$

(can do better if have time)

<u>step 4:</u>

use the complete dataset to train $\mathcal{H}_{m^*}$ and output the final hypo. g_{m^*}.