Lec 37

Recap:

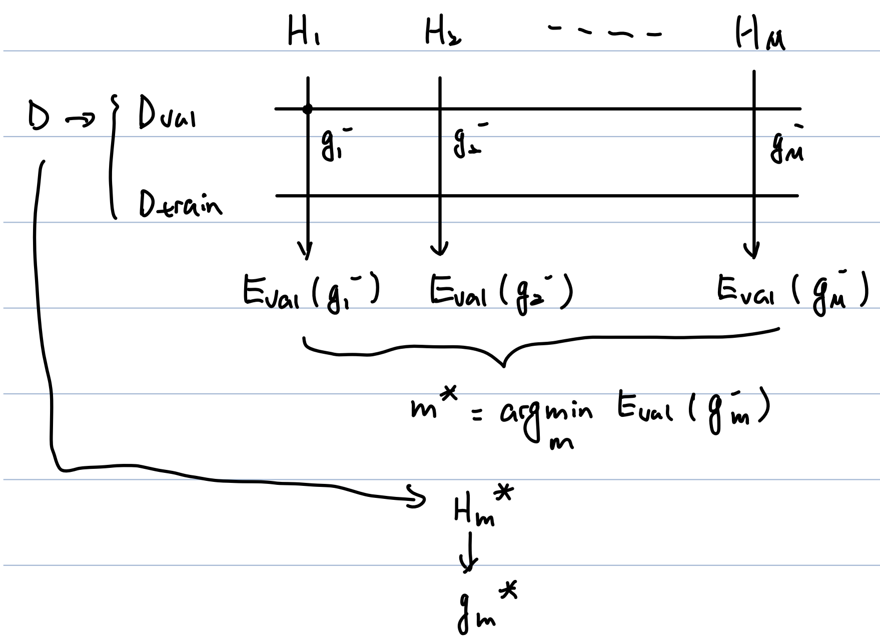
↳ Validation:

$|D_{train}| = N - k$ $\qquad$ $D_{train} \longrightarrow \boxed{A} \longrightarrow g^-$

$|D_{val}| = k$ $\qquad$ $E_{val}(g^-) = \frac{1}{k} \sum_{(x, y_n) \in D_{val}} e(g^-(\underline{x}), y_n)$

$$\underbrace{E_{val}(g^-) \approx}_{\text{large } k} \underbrace{E_{out}(g^-) \approx}_{\text{small } k} E_{out}(g)$$

↳ Model selection:



↳ Hope:

$$E_{val}(g_{m^*}^-) \approx E_{out}(g_m^-) \approx E_{out}(g_{m^*})$$

? ↓ want to prove. $\qquad$ small k

Previously: $E_{val}(g_m^-) \approx E_{out}(g_m^-)$, $\forall m$

⇐ ① unbiased

② consistent

③ Hoeff dim

$$E_{out}(g_m^-) \le E_{val}(g_m^-) + \sqrt{\frac{1}{2k}\log\frac{2}{\delta}} \quad \text{(linear binary)}$$

problem: does not work for m^*.

We have chosen m^* by D_{val}.

↳ Recall: Use D to choose best $g \in H$

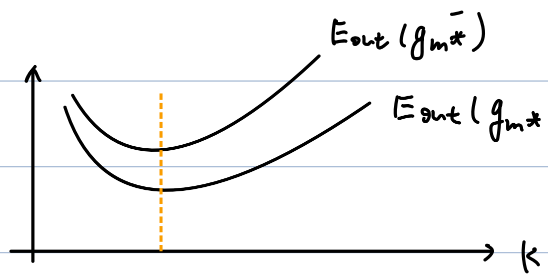
Let $G \triangleq \{g_1^-, g_2^-, \cdots g_M^-\}$

Given without using D_{val}, use D_{val} to choose $\bar{g}_m^* \in G$

Hoeffding + union bound ⇒ with prob $1-\delta$

$$\Rightarrow E_{out}(g_{m^*}^-) \le E_{val}(g_{m^*}^-) + \sqrt{\frac{1}{2k}\log\frac{2M}{\delta}}$$

$$\Pr\{|E_{val}(g_{m^*}^-) - E_{out}(g_{m^*}^-)|\} > \varepsilon$$
$$\le 2Me^{-2k\varepsilon^2} = \delta$$



2. <u>Alternative to Model Selection:</u>

↳ Merge all hypo. into one big H

↳ $H = H_1 \cup H_2 \cup \cdots \cup H_m$

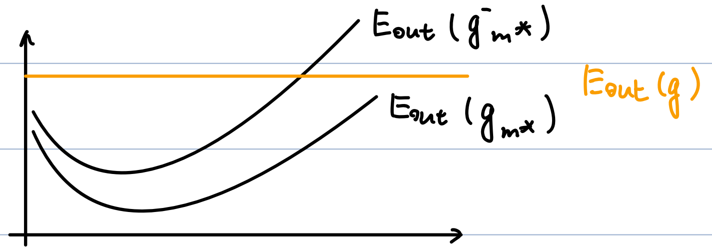
$$D \to \boxed{A} \to g$$

$$g = \underset{h \in H}{\arg\min}\ E_{in}(h)$$

↳ with prob. $1-\delta$,

$$E_{out}(g) \le E_{in}(g) + \theta\left(\sqrt{\frac{d_{vc}(H)\log(\frac{N}{\delta})}{N}}\right)$$

↳ $d_{vc}(H)$ usually is very large



Cross Validation

↳ Avoid the dilemma in selecting K.

1. Leave-one-out (LOO) cross validation:

↳ Portion D in training set of size N-1, and validation of size 1

↳ N ways to do that:

For $n = 1, 2, \cdots N$,

- Let training dataset be:

$$D_n = \{(x_1, y_1)(x_2, y_2), \cdots \cancel{(x_n, y_n)}, \cdots (x_N, y_N)\}$$

- And validation dataset:

$$D_n^{val} \{(x_n, y_n)\}.$$

validation error

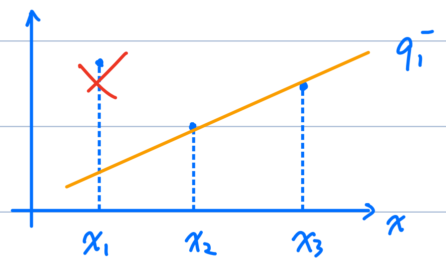
$$D_n \longrightarrow \boxed{A} \xrightarrow{g_n^-} \boxed{D_n^{val}} \longrightarrow e_n = e(g_n^-(x_n), y_n)$$

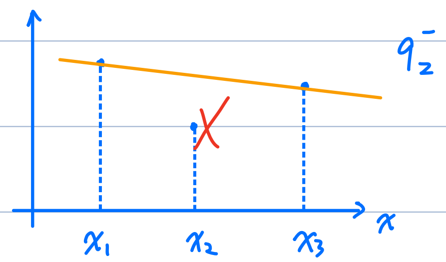
- After repeating for $n = 1, 2, \cdots N$,

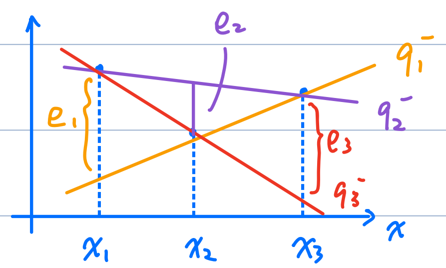
  the cross-validation error:

$$E_{cv} = \frac{1}{N} \sum_{n=1}^{N} e_n$$

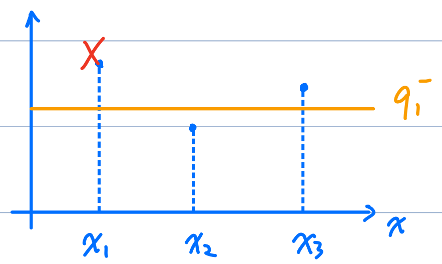
e.g. $d=1$, $N=3$ (Linear Regression)

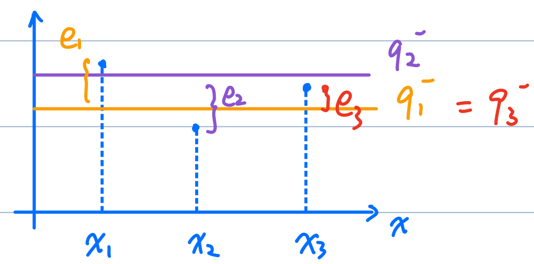






e.g. Constant function, $d=2$. $N=3$.





This one is prefered since less error.