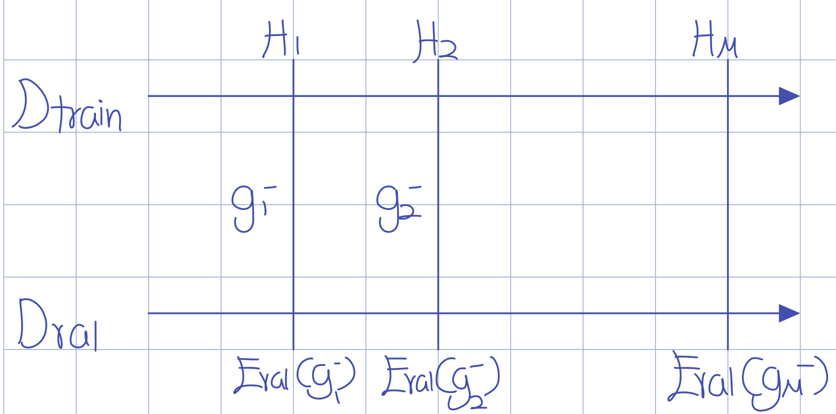
Model Selection

Given dataset $\mathcal{D}$

M candidate hypothesis classes $\mathcal{H}_1, \dots, \mathcal{H}_M$

Select best hypothesis class for $\mathcal{D}$

$$\mathcal{D} \to \begin{cases} \mathcal{D}_{train} \text{ (N−K Samples)} \\ \mathcal{D}_{val} \text{ (K Samples)} \end{cases}$$



Pick hypothesis class with smallest $E_{val}(g^-)$

Selected hypothesis class

$$m^* = \underset{m \in \{1, \dots, M\}}{\operatorname{argmin}} E_{val}(g_m^-)$$

Hypothesis Class $\mathcal{H}_{m^*}$

$\mathcal{D}$ (N Points) $\longrightarrow$ Learning Algorithm $\mathcal{H}_{m^*}$ $\longrightarrow$ $g_{m^*}(\;)$

Final Output Hypothesis

Estimate $E_{out}(g_{m^*})$

Completed $E_{val}(g^-_{m_1}), \dots, E_{val}(g^-_{M})$

Selected smallest $E_{val}(g^-_m)$

Generate g_{m^*}

$$E_{out}(g_{m^*}) \approx E_{out}(g^-_{m^*}) \approx \underset{\text{Computed}}{E_{val}(g^-_{m^*})}$$

Under what condition is

$$E_{out}(g^-_{m^*}) \approx E_{val}(g^-_{m^*})$$

Last lecture $M=1$

$$E_{out}(g^-_1) \le E_{val}(g^-_1) + \sqrt{\frac{1}{2K}\log\frac{2}{\delta}}$$

Same argument shows

$$E_{out}(g^-_2) \le E_{val}(g^-_2) + \sqrt{\frac{1}{2K}\log\frac{2}{\delta}}$$

$$\vdots$$

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

$$m^* = \underset{m\in\{1,\dots,M\}}{\operatorname{argmin}} E_{val}(g^-_m)$$

$$\cancel{E_{out}(g^-_{m^*}) \le E_{val}(g^-_{m^*}) + \sqrt{\frac{1}{2K}\log\frac{2}{\delta}}}$$

g_{m^*} is not independent D_{val}

Since g_{m^*} is not fixed hypothesis for D_{val}, but depends o

$\{E_{val}(g_m^-)\}_{m=1,\dots,M}$

Apply union bound argument

$$\Pr(|E_{val}(g_{m^*}^-) - E_{out}(g_{m^*}^-)| \geq \varepsilon) \leq \Pr\left(\bigcup_{m=1}^{M} |E_{val}(g_{m^*}^-) - E_{out}(g_{m^*}^-)| \geq \varepsilon\right)$$

$$\leq 2Me^{-2K\varepsilon^2}$$

$$E_{out}(g_m^*) \leq E_{val}(g_m^*) + \sqrt{\frac{1}{2K}\log\frac{2M}{\delta}}$$

Key Point, Penalty term only depend on the number of candidate hypothesis, not on VC dimension

Alternative to model selection.

$\boxed{K = N/5}$

$$H^* = H_1 \cup \dots \cup H_M$$

Select

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

$$E_{out}(\hat{g}) \leq E_{in}(\hat{g}) + O\left(\sqrt{\frac{\log d_{VC}(H^*)}{2N}}\right)$$

Very Large

Cross Validation (Don't Split)

Leave-one-out CV

$D = \{ x_1, \ldots, x_N \}$

Auxilary datasets

$D^1_{train} = \{ x_2, \ldots, x_N \}$

$D^1_{val} = \{ x_1 \}$

$D^1_{train} \xrightarrow[LA]{} g_1^- \xrightarrow[D^1_{val}]{} E_{val}(g_1^-)$

$D^2_{train} = \{ x_1, x_3, \ldots, x_N \}$

$D^2_{val} = \{ x_2 \}$

$D^2_{train} \xrightarrow[LA]{} g_2^- \xrightarrow[D^2_{val}]{} E_{val}(g_2^-)$

$$E^{CV} = \frac{1}{N}\left(E^1_{val}(g_1^-) + \ldots + E^N_{val}(g_n^-) \right)$$

Cross validation error

$E^{CV} \approx E_{out}(g)$

$E_D [E^{CV}] = E_{out}(g)$ 80-20 Rule