Using $\mathcal{D}_{\text{val}}$ more than once

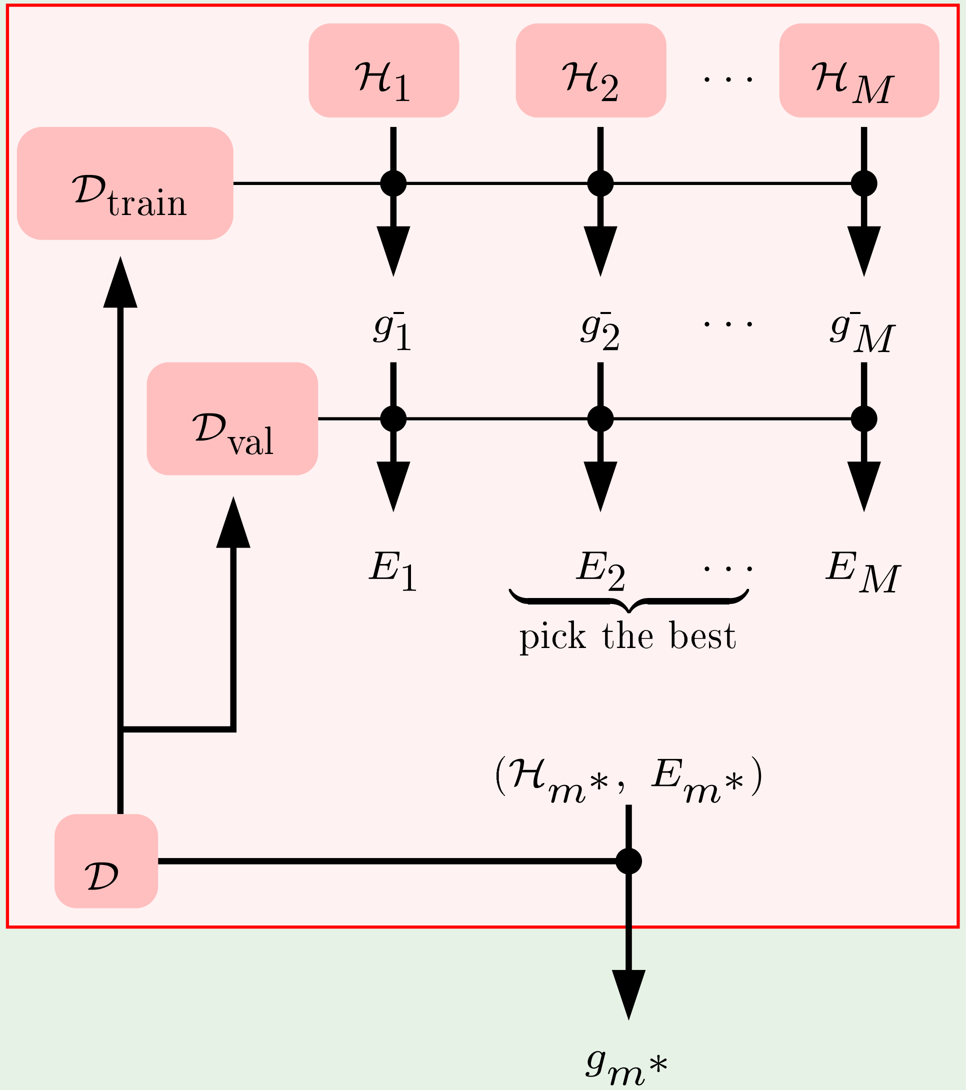
M models $\mathcal{H}_1, \ldots, \mathcal{H}_M$

Use $\mathcal{D}_{\text{train}}$ to learn g_m^- for each model

Evaluate g_m^- using $\mathcal{D}_{\text{val}}$:

$$E_m = E_{\text{val}}(g_m^-); \quad m = 1, \ldots, M$$

Pick model $m = m^*$ with smallest E_m

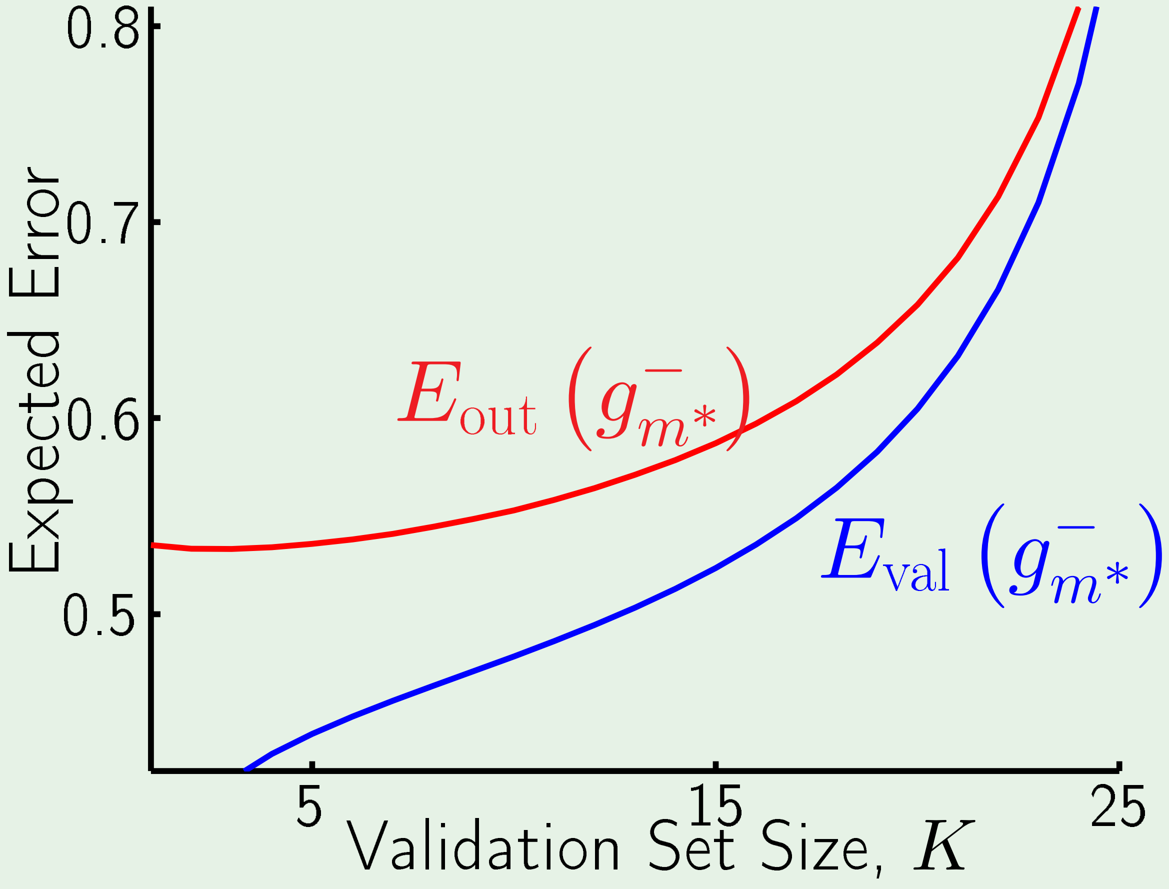


The bias

We selected the model $\mathcal{H}_{m^*}$ using $\mathcal{D}_{\mathrm{val}}$

$E_{\mathrm{val}}(g^-_{m^*})$ is a biased estimate of $E_{\mathrm{out}}(g^-_{m^*})$

Illustration: selecting between 2 models



How much bias

For $\color{red}{M}$ models: $\mathcal{H}_1, \dots, \mathcal{H}_M$

$\mathcal{D}_{\text{val}}$ is used for "training" on the **finalists model**:

$$\mathcal{H}_{\text{val}} = \{g_1^-, g_2^-, \dots, g_{\text{M}}^-\}$$

Back to Hoeffding and VC!

$$E_{\text{out}}(g_{m^*}^-) \le E_{\text{val}}(g_{m^*}^-) + O\left(\sqrt{\frac{\ln \color{red}{M}}{K}}\right)$$

regularization λ early-stopping T

Data contamination

Error estimates: $E_{\text{in}},\ E_{\text{test}},\ E_{\text{val}}$

Contamination: Optimistic (deceptive) bias in estimating E_{out}

Training set: totally contaminated

Validation set: slightly contaminated

Test set: totally 'clean'