Validation versus regularization

In one form or another, $E_{\text{out}}(h) = E_{\text{in}}(h) + \text{overfit penalty}$

Regularization:

$$E_{\text{out}}(h) = E_{\text{in}}(h) + \underbrace{\text{overfit penalty}}_{\text{regularization estimates this quantity}}$$

Validation:

$$\underbrace{E_{\text{out}}(h)}_{\text{validation estimates this quantity}} = E_{\text{in}}(h) + \text{overfit penalty}$$

Analyzing the estimate

On out-of-sample point $(\mathbf{x}, y)$, the error is $\mathbf{e}(h(\mathbf{x}), y)$

Squared error: $\left(h(\mathbf{x}) - y\right)^2$

Binary error: $[\![h(\mathbf{x}) \neq y]\!]$

$$\mathbb{E}\left[\mathbf{e}(h(\mathbf{x}), y)\right] = E_{\text{out}}(h)$$

$$\text{var}\left[\mathbf{e}(h(\mathbf{x}), y)\right] = \sigma^2$$

From a point to a set

On a validation set $(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_K, y_K)$, the error is $E_{\text{val}}(h) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{e}(h(\mathbf{x}_k), y_k)$

$$\mathbb{E}\left[E_{\text{val}}(h)\right] = \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[\mathrm{e}(h(\mathbf{x}_k), y_k)\right] = E_{\text{out}}(h)$$

$$\text{var}\left[E_{\text{val}}(h)\right] = \frac{1}{K^2}\sum_{k=1}^{K}\text{var}\left[\mathrm{e}(h(\mathbf{x}_k), y_k)\right] = \frac{\sigma^2}{K}$$

$$E_{\text{val}}(h) = E_{\text{out}}(h) \pm O\left(\frac{1}{\sqrt{K}}\right)$$

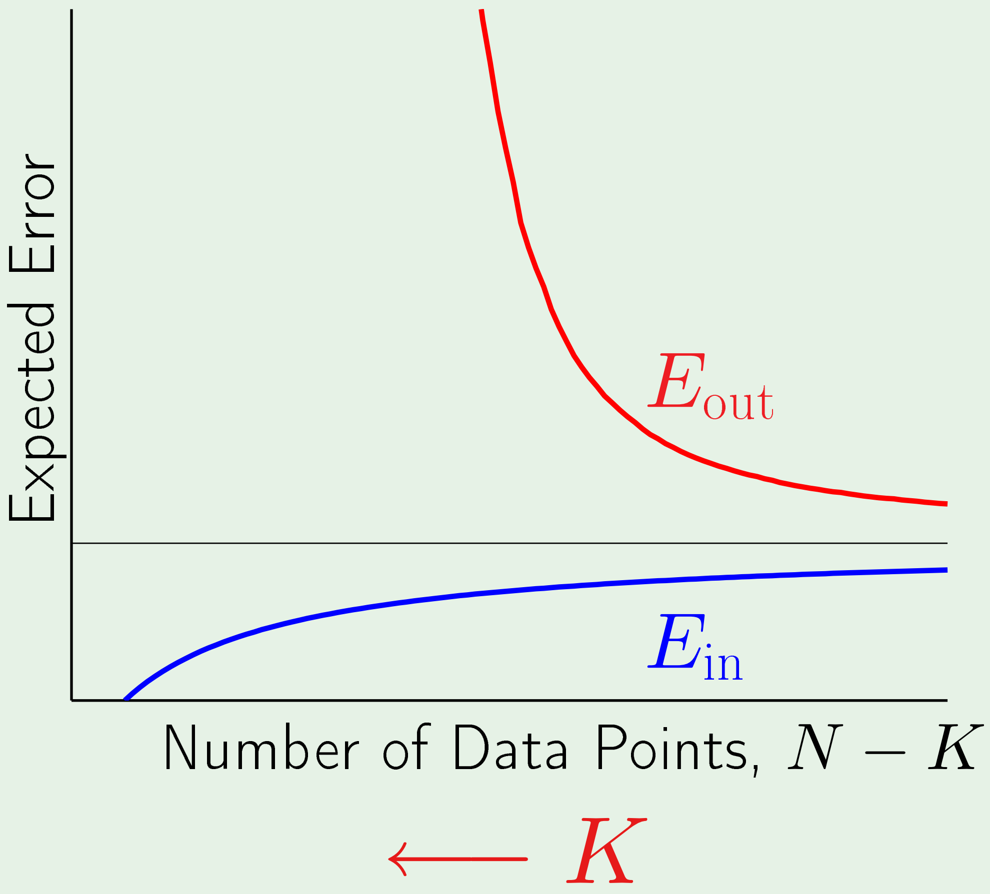
K is taken out of N

Given the data set $\mathcal{D} = (\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_N, y_N)$

$\underbrace{K \text{ points}}_{\mathcal{D}_{\text{val}}} \rightarrow$ validation $\qquad \underbrace{N - K \text{ points}}_{\mathcal{D}_{\text{train}}} \rightarrow$ training

$O\left(\frac{1}{\sqrt{K}}\right)$: Small $K \implies$ bad estimate

Large $K \implies$?

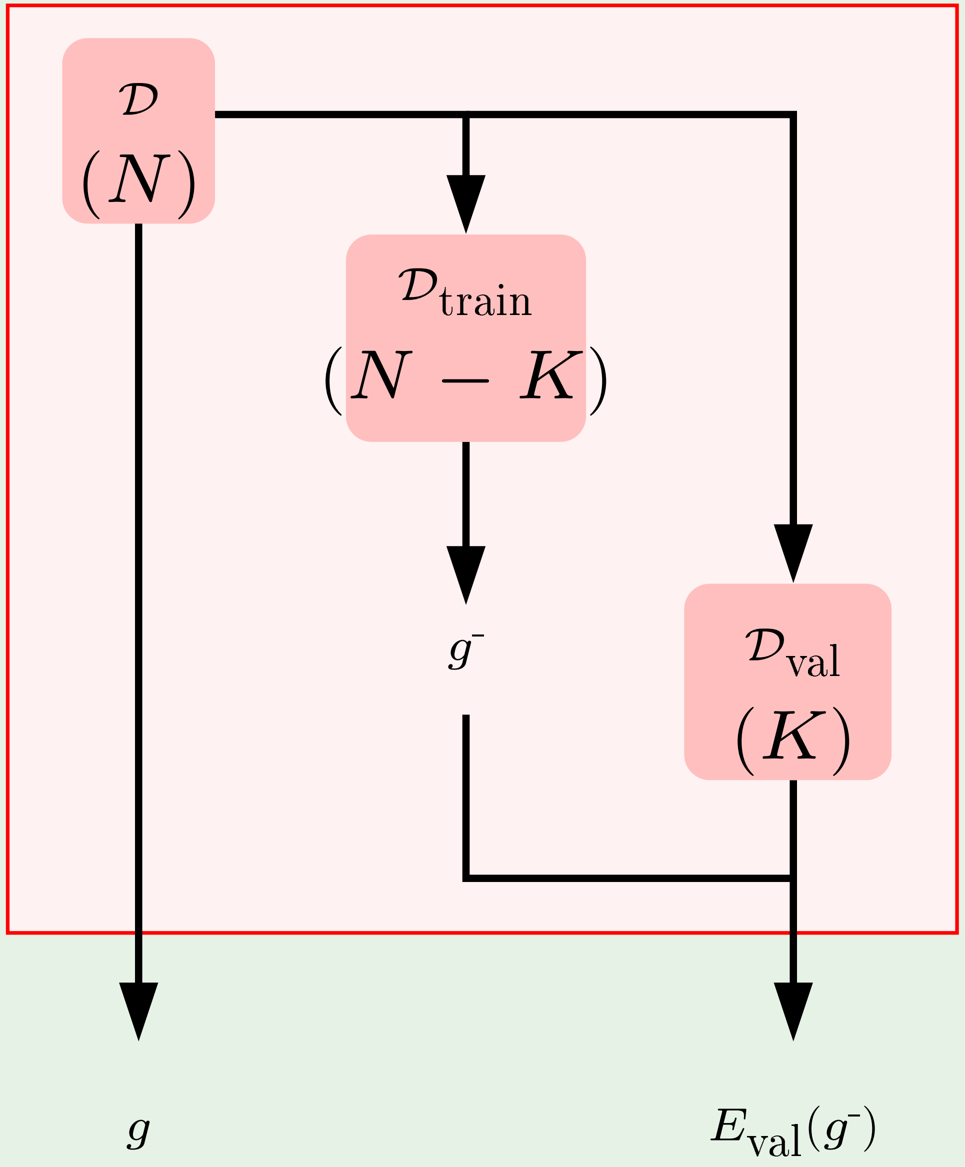


K is put back into N

$$\mathcal{D} \longrightarrow \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$$
$$\downarrow \qquad \downarrow \qquad \downarrow$$
$$N \qquad N-K \qquad K$$

$$\mathcal{D} \Longrightarrow g \qquad \mathcal{D}_{\text{train}} \Longrightarrow g^-$$

$E_{\text{val}} = E_{\text{val}}(g^-)$ Large $K \Longrightarrow$ bad estimate!



Rule of Thumb:

$$K = \frac{N}{5}$$

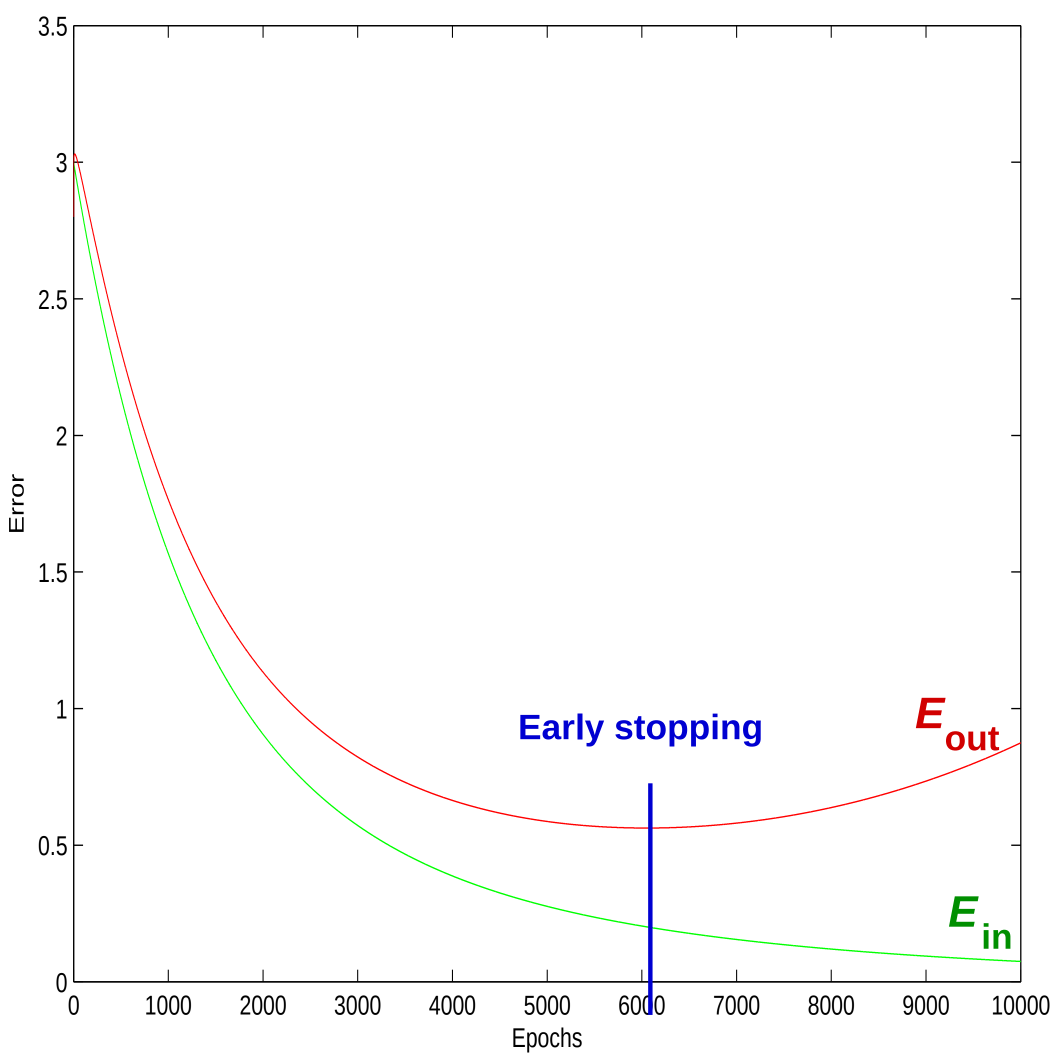
Why 'validation'

$\mathcal{D}_{\text{val}}$ is used to make learning choices

If an estimate of E_{out} affects learning:

the set is no longer a **test** set!

It becomes a validation set



What's the difference?

Test set is unbiased; validation set has optimistic bias

Two hypotheses h_1 and h_2 with $E_{\text{out}}(h_1) = E_{\text{out}}(h_2) = 0.5$

Error estimates $\mathbf{e}_1$ and $\mathbf{e}_2$ uniform on $[0, 1]$

Pick $h \in \{h_1, h_2\}$ with $\mathbf{e} = \min(\mathbf{e}_1, \mathbf{e}_2)$

$\mathbb{E}(\mathbf{e}) < 0.5$ optimistic bias