The dilemma about K

The following chain of reasoning:

$$E_{\text{out}}(g) \approx E_{\text{out}}(g^-) \approx E_{\text{val}}(g^-)$$

(small K) (large K)

highlights the dilemma in selecting K:

Can we have K both small and large? ☺

Leave one out

$N - 1$ points for training, and 1 point for validation!

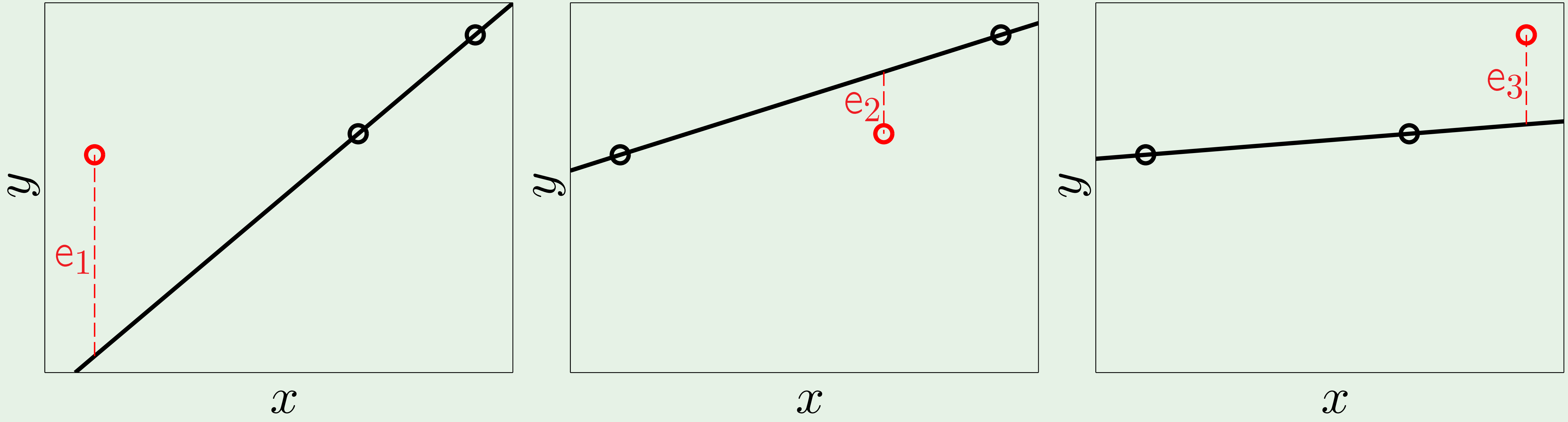
$$\mathcal{D}_n = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{n-1}, y_{n-1}), \cancel{(\mathbf{x}_n, y_n)}, (\mathbf{x}_{n+1}, y_{n+1}), \ldots, (\mathbf{x}_N, y_N)$$

Final hypothesis learned from $\mathcal{D}_n$ is g_n^-

$$\mathbf{e}_n = E_{\text{val}}(g_n^-) = \mathbf{e}\left(g_n^-(\mathbf{x}_n), y_n\right)$$

cross validation error: $E_{\text{cv}} = \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \mathbf{e}_n$

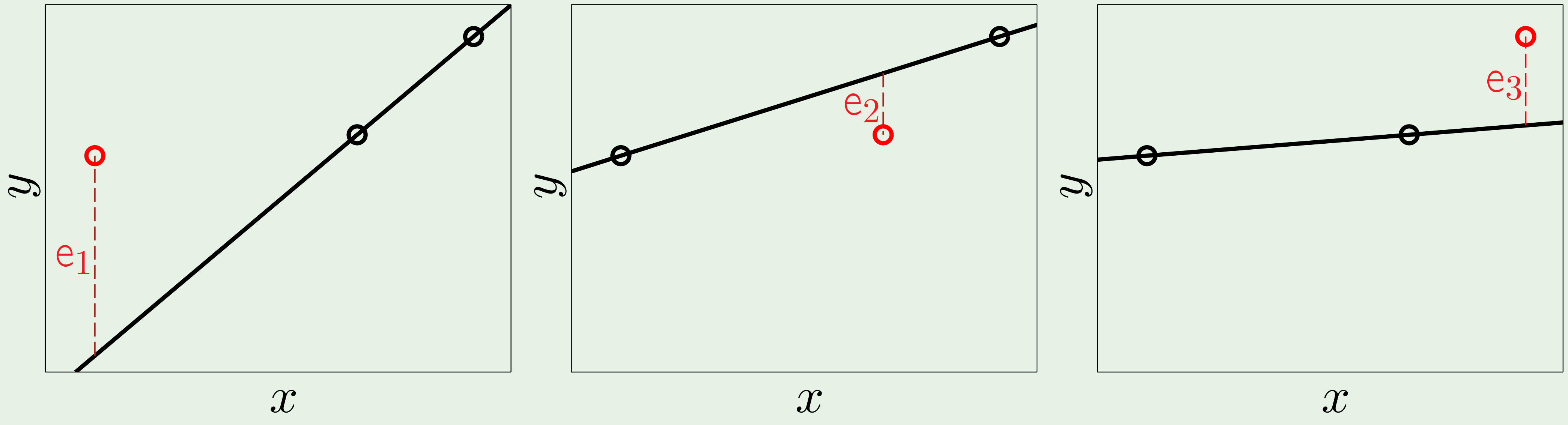
Illustration of cross validation



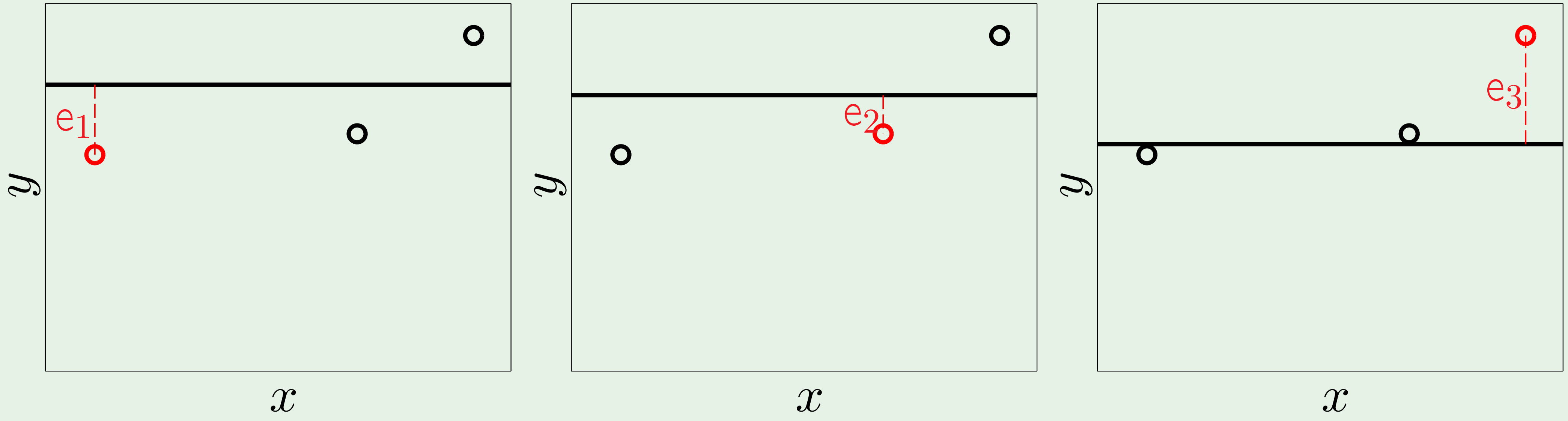
$$E_{\rm cv} = \frac{1}{3}\left(e_1 + e_2 + e_3 \right)$$

Model selection using CV

Linear:

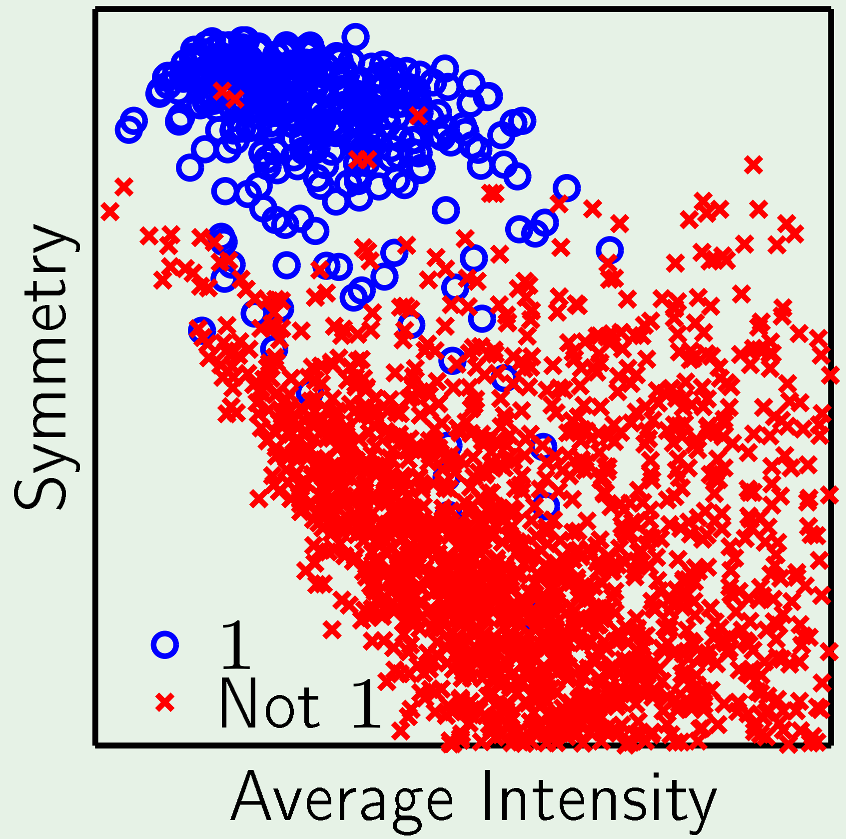


Constant:

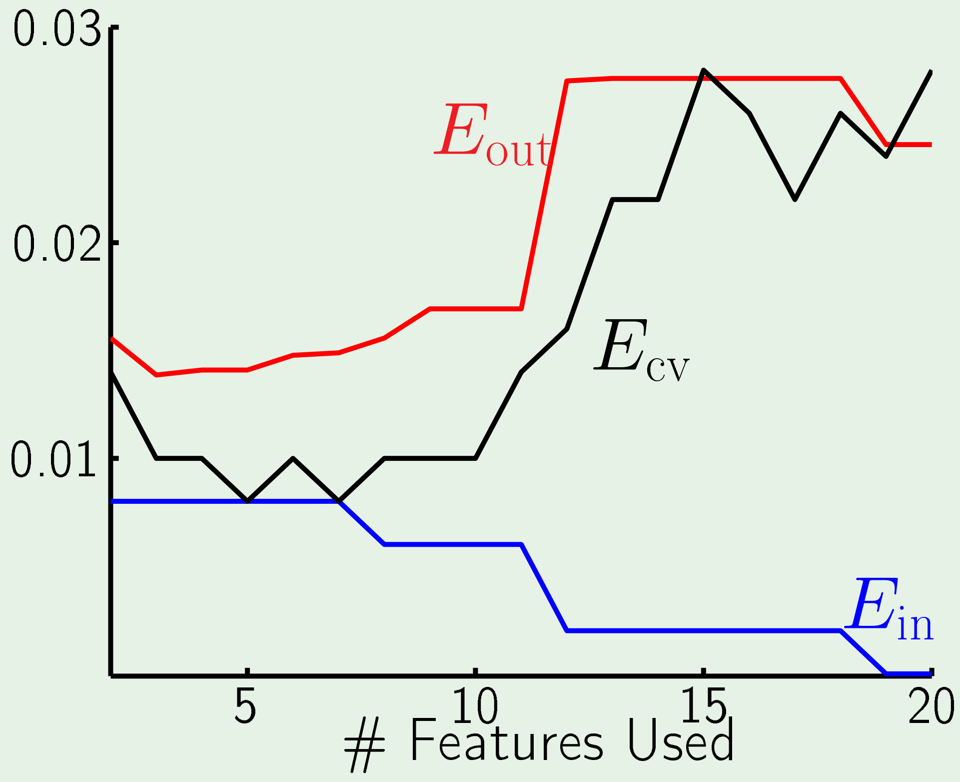


Cross validation in action

Digits classification task



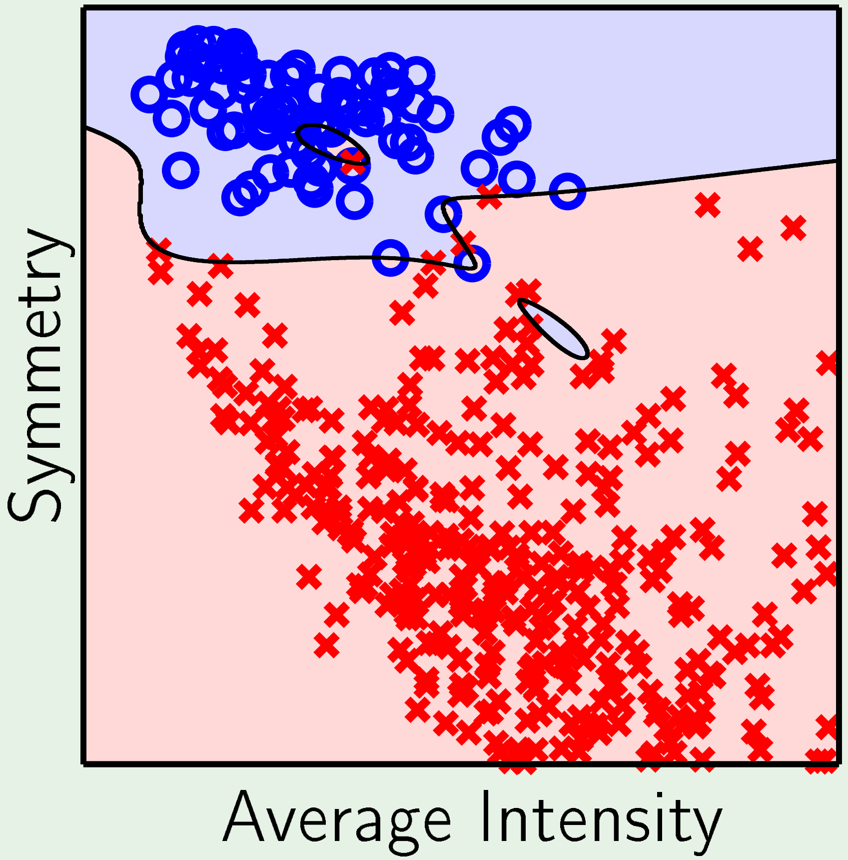
Different errors



$$(1, x_1, x_2) \rightarrow (1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, \ldots, x_1^5, x_1^4x_2, x_1^3x_2^2, x_1^2x_2^3, x_1x_2^4, x_2^5)$$

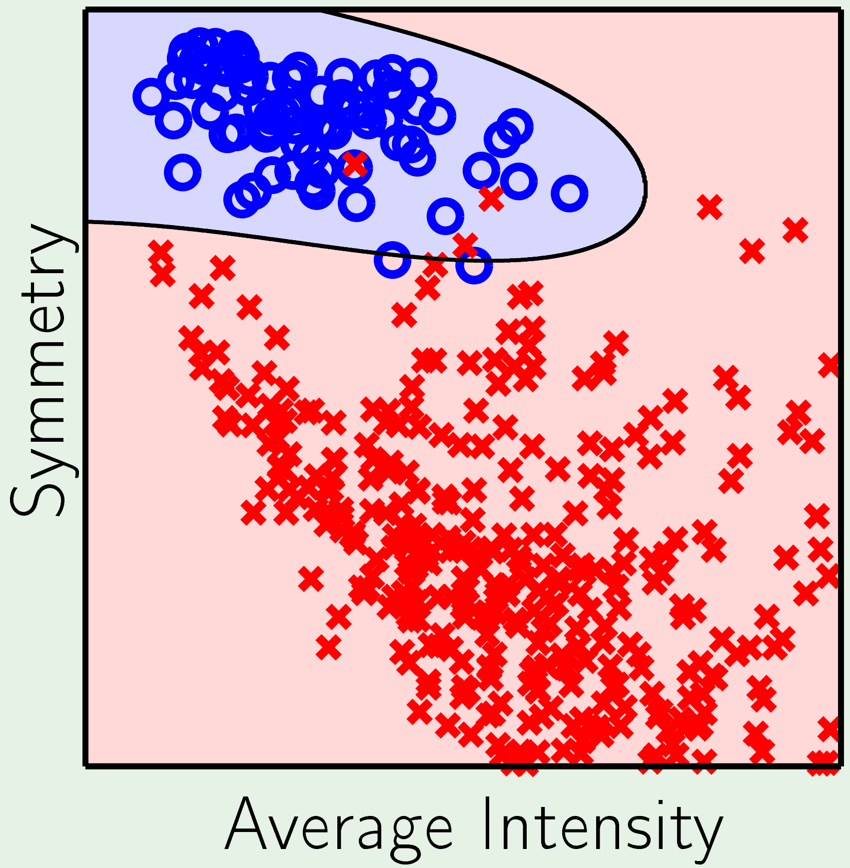
The result

without validation



$E_{\text{in}} = 0\%$ $E_{\text{out}} = 2.5\%$

with validation

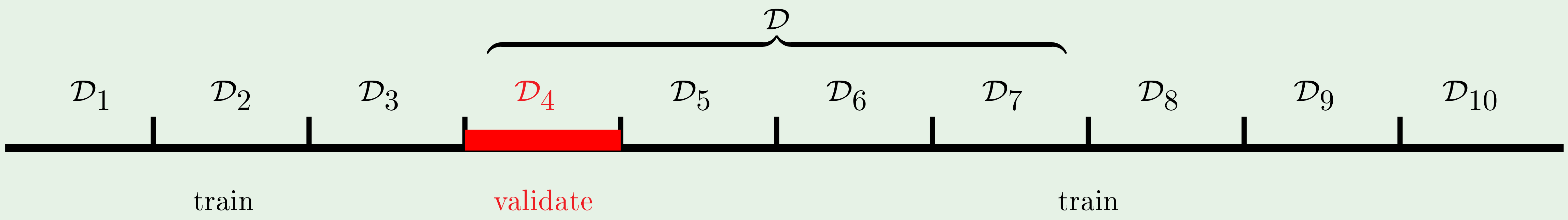


$E_{\text{in}} = 0.8\%$ $E_{\text{out}} = 1.5\%$

Leave more than one out

Leave one out: N training sessions on $N-1$ points each

More points for validation?



$\frac{N}{K}$ training sessions on $N-K$ points each

10-fold cross validation: $K = \frac{N}{10}$