

We suggested to split the data into a training and a test set. This comes with some drawbacks:

- we may not have enough data;
- we are really focusing on a few data points as a test set

	training	test
step/fold	A	B

25% D

1)

2)

3)

4)

$$err_1 = \frac{1}{|B|} \sum_{i \in B} (y_i - \hat{f}^{A,B,C}(x_i))^2$$

$$err_2 = \frac{1}{|C|} \sum_{i \in C} (y_i - \hat{f}^{A,B,D}(x_i))^2$$

$$err_3 = \frac{1}{|B|} \sum_{i \in B} (y_i - \hat{f}^{A,C,D}(x_i))^2$$

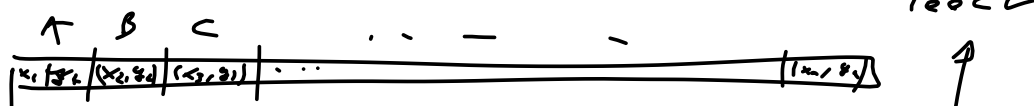
$$err_4 = \frac{1}{|A|} \sum_{i \in A} (y_i - \hat{f}^{B,C,D}(x_i))^2$$

$$err_{cv} = \frac{1}{4} \sum_{j=1}^4 err_j$$

randomly

CROSS-VALIDATION

- we split the data in K folds, of approximately same size
- K is arbitrary, the most used values are $K=5, K=10, K=n$



particular case one observation per fold; deterministic

We want to find the best model complexity by cross-validation

tuning

	$p=1$	$p=2$	...	$p=20$
fold 1	$err_1^{p=1}$	$err_1^{p=2}$	...	$err_1^{p=20}$
fold 2	$err_2^{p=1}$	$err_2^{p=2}$	...	$err_2^{p=20}$
...	...	...	...	...
fold K	$err_K^{p=1}$	$err_K^{p=2}$	...	$err_K^{p=20}$
	$err_{cr}^{p=1}$	$err_{cv}^{p=2}$	$\bigcirc \dots$ $p=f$	$err_{cv}^{p=20}$

$err_1^p = \frac{1}{|E|} \sum_{i \in E} (y_i - \hat{f}_{p,r}^{(k-1)})^2$

$\leftarrow p^*$ the p corresponding to the smallest error