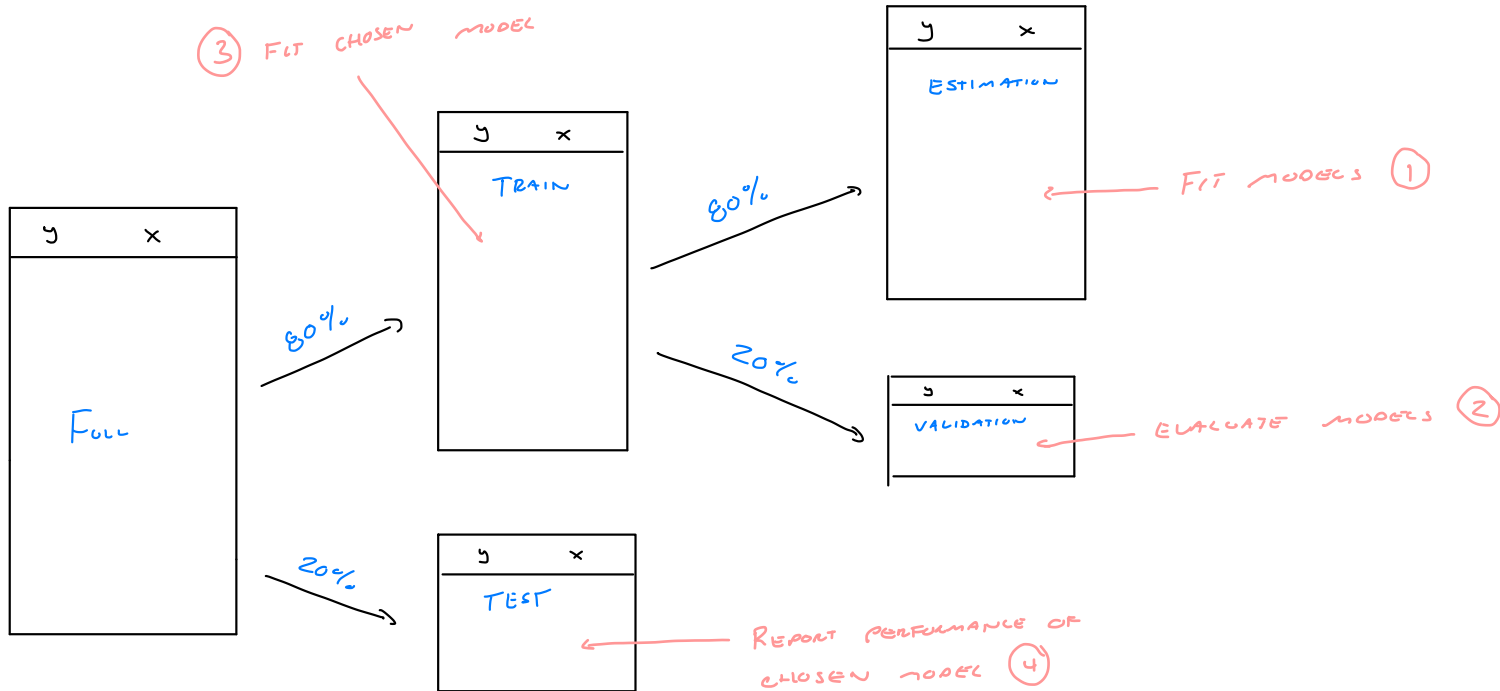


CROSS-VALIDATION



WHAT'S WRONG WITH THIS?
Too variable!

A NOTE ABOUT VARIABILITY

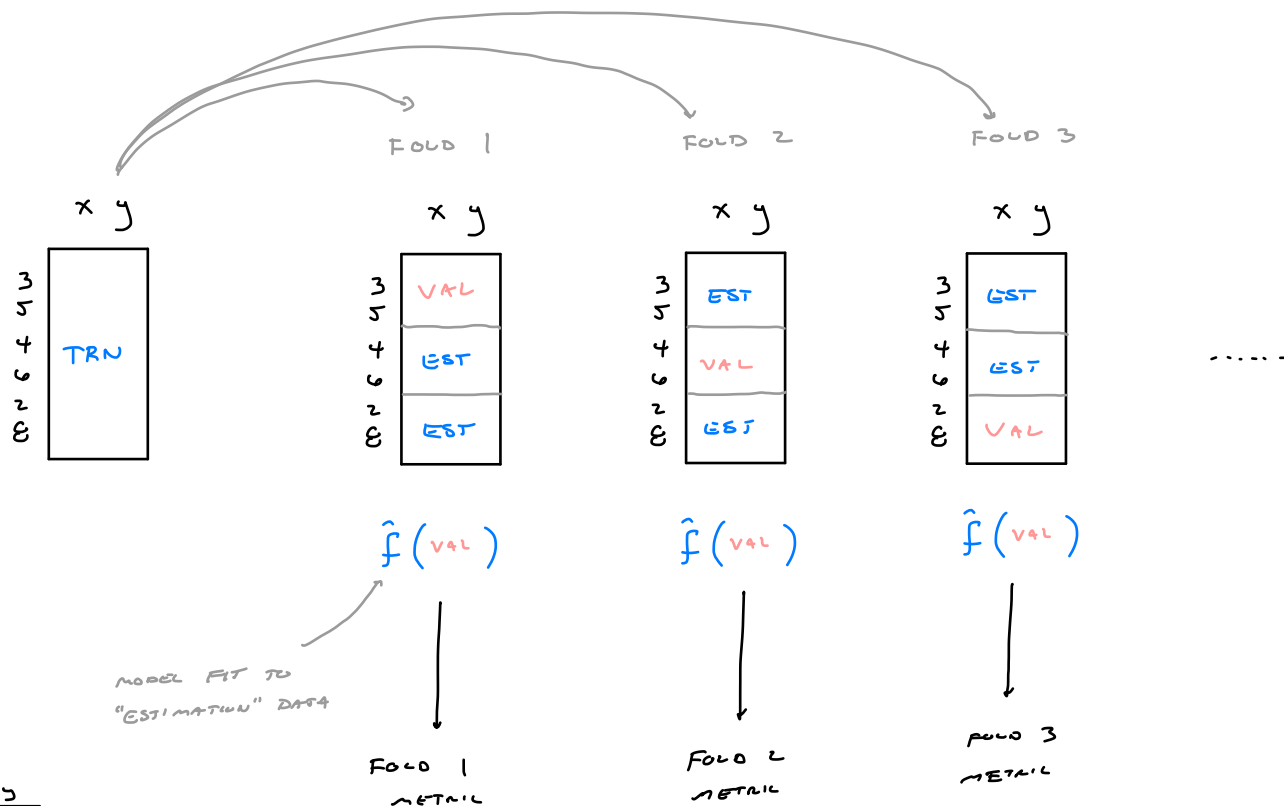
$$X_1, \dots, X_n \quad \text{i.i.d.} \quad \begin{aligned} E[X] &= \mu \\ V[X] &= \sigma^2 \end{aligned}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E[\bar{X}] = \mu$$

$$V[\bar{X}] = \sigma^2 / n$$

K-FOLD CROSS-VALIDATION



1
7

x	y

TST

CROSS-VALIDATION METRICS

RMSE
MAE
ACC
BTL



METRIC - CV_K

=

$\frac{1}{K}$



K FOLDS

$\sum_{i=1}^K$

METRIC_i



VALUE OF METRIC FOR FOLD i

SE[METRIC - CV_K]

$\frac{1}{K-1}$

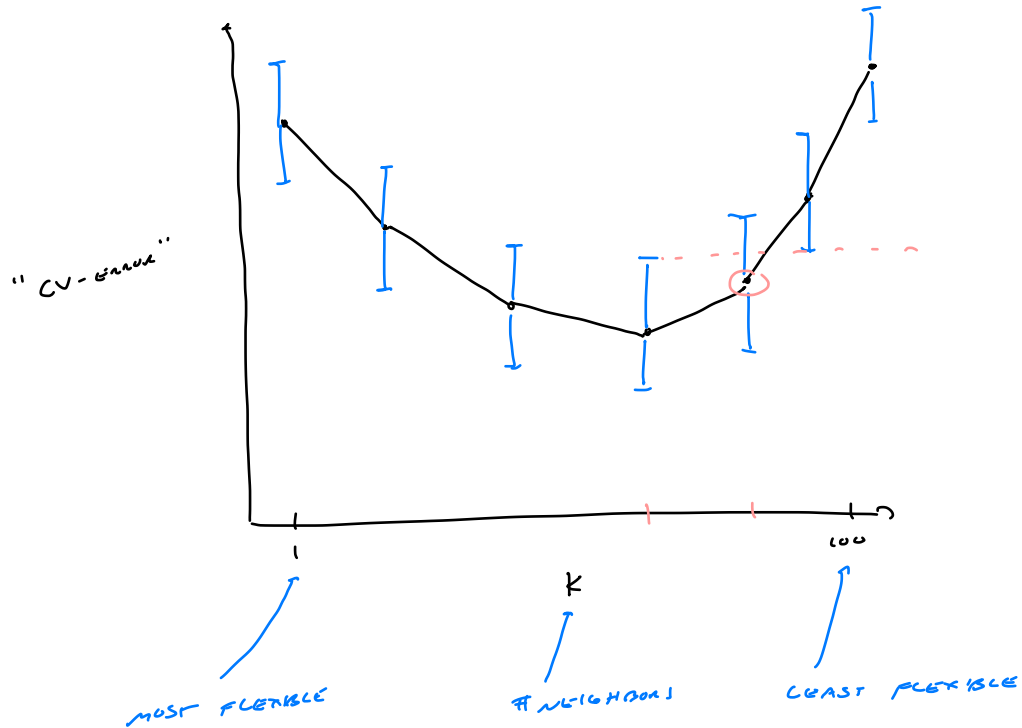
$\sum_{i=1}^K$

(METRIC_i - METRIC - CV_K)²



SAMPLE SD OF FOLD METRICS

ONE-SE RULE



WHICH K ?

Popular $\left\{ \begin{array}{l} K = 5 \\ K = 10 \end{array} \right.$ \leftarrow LESS COMPUTING

$K = "n"$ \leftarrow "LEAVE-ONE-OUT"