Cross-Validation
WHAT'S WRONG WITH THIS?
Too variable!
A Note About Variability

\(X_1, \ldots, X_n\) and

\[E[X] = \mu\]

\[V[X] = \sigma^2\]

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

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

\[V[\bar{X}] = \frac{\sigma^2}{n}\]
$K$-Fold Cross-Validation

Fold 1

Fold 2

Fold 3

$\hat{f}(\text{Val})$

Fold 1

$\hat{f}(\text{Val})$

Fold 2

$\hat{f}(\text{Val})$

Fold 3

Model fit to "Estimation" data

Test

Train

Validation
Cross-Validation Metrics

\[
\text{METRIC}_{-\text{CV}_k} = \frac{1}{k} \sum_{i=1}^{k} \text{METRIC}_i
\]

\[
\text{SE}[\text{METRIC}_{-\text{CV}_k}] = \sqrt{\frac{1}{k-1} \sum_{i=1}^{k} (\text{METRIC}_i - \text{METRIC}_{-\text{CV}_k})^2}
\]
ONE-SE Rule

"CV-Error"

Most Flexible - k - Least Flexible

At Neighbors
Which $K$?

Popular \{
\begin{align*}
  k &= 5 \\  k &= 10 \\
\end{align*}
\}

\begin{align*}
  k &= "n" \\
  \text{"leave-one-out"}
\end{align*}

Less computing