Cross-validation. The practice is partitioning a sample of data into subsamples such that analysis is initially performed on a single subsample, while further subsamples are retained "blind" in order for subsequent use in confirming and validating the initial analysis.

Tied in with the idea of production error.

Loss function $L(y, \hat{y})$

E.g. $(y - \hat{y})^2$

Production error, $\Delta$

$E \left( y - \hat{y} \right)^2$

in classification problem

Prob$\left\{ \hat{y} \neq y \right\}$

$E \left( L(y, \hat{y}) \right)$ error rate
e.g. multiple regression

Data \( y_i, X_i \) 

Future observation \( y_0, X_0 \) 

\[ y_0 = X_0^T \beta + \epsilon_0 \]

Aggregate prediction error \( \Delta \)

\[ \Delta = \sum_i E(y_i - \hat{y}_i)^2 \]

\[ \hat{y}_0 = X_0^T \hat{\beta} \]

\[ y_0 - \hat{y}_0 = X_0^T (\beta - \hat{\beta}) + \epsilon_0 \]

\[ E(y_0 - \hat{y}_0)^2 = X_0^T (X_0 X_0)^{-1} X_0 0^2 + \epsilon_0^2 \]

\( \epsilon_0 \) is over \( X_1, \ldots, X_n \)

\[ \epsilon_0 = \frac{1}{n} \sum_i E(X_i^T X_i)^{-1} X_i X_i^T \sigma^2 + \epsilon_0^2 \]

\[ = \left( \frac{n}{n+1} \right) \sigma^2 \]

Estimate by \( \left( \frac{n}{n+1} \right) \sigma^2 \)
Estimate, deleting cases in turn

\[ \hat{\Delta}_{cv} = \frac{1}{m} \sum_{i=1}^{m} (y_{i} - \hat{x}_{i}^T \hat{\beta})^2 \]

\[ \hat{\beta} - \beta_{-i} = (X^T X)^{-1} \hat{x}_{i} (y_{i} - \hat{x}_{i}^T \hat{\beta}) / (1 - H_{ii}) \]

\[ H_{ii} : \text{leverage} \quad H = X(X^T X)^{-1} X^T \]

\[ \hat{\Delta}_{cv} = \frac{1}{m} \sum_{i=1}^{m} (y_{i} - \hat{x}_{i}^T \hat{\beta})^2 / (1 - H_{ii})^2 \]

Here, \( m(\hat{x}_{i}) = \hat{x}_{i}^T \hat{\beta} \)

In case of \( EY = m(\hat{x}, \lambda) \) and

\[ \gamma = \hat{m}(\hat{x}, \lambda) = H(\lambda) y \]

have

\[ \hat{\Delta}_{cv} = \frac{1}{m} \sum_{i=1}^{m} (y_{i} - \hat{m}(\hat{x}_{i}, \lambda))^2 / (1 - H_{ii}(\lambda))^2 \]

Can be used to choose \( \lambda \).
(a) Simulated data with underlying model m(x); (b) LOESS fits with $\alpha = \frac{1}{3}, \frac{2}{3}$

(a) LOESS fits based on $\alpha \in \left\{ \frac{1}{3}, \frac{2}{3} \right\}$ (dashed) and on cross-validated $\hat{\alpha}_{CV}(5) = 0.15$ (solid); (b) $\hat{\Delta}_{CV}(\alpha)$ for $0.11, \ldots, 0.4$
Cross-validation in S plus:

\[
\begin{align*}
\text{predict} & \quad \text{tree} (\text{design}) \\
\text{bcv} & \quad \text{MASS} \\
\text{ucv} & \quad \text{MASS} \\
\text{gcv, glm} & \quad \text{boot} \\
\text{gmlm, cv} & \quad \text{class} \\
\text{funcfit} & \quad \text{tree} \\
\text{xpred, rpart} & \quad \text{rpart} \\
\text{evalval} & \quad \text{suppclass} \\
\text{cv, tree} & \quad \text{tree} \\
\end{align*}
\]

\[\text{mgcv} ?\]