Statistics and Data Analysis in MATLAB

Lecture 4: Model fitting

Kendrick Kay
Washington University in St. Louis
February 28, 2014
$$\text{squared error} = \sum_{i=1}^{n} (d_i - m_i)^2$$

**High squared error**

- **Data**
- **Model**
- **Residuals**

$$y = \frac{1}{4}x - \frac{1}{4}$$

**Low squared error**

- **Data**
- **Model**
- **Residuals**

$$y = \frac{1}{2}x + 0$$
Geometric interpretation of linear regression

$$y = Xw + n$$

$$\begin{bmatrix} y \\ x_1 \\ x_2 \\ n \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 1 & 0 \\ 2 & 1 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + \begin{bmatrix} -2.5 \\ 2.5 \\ 1 \end{bmatrix}$$

Data (y)
Residuals (n)
Optimal model fit ($w_1x_1 + w_2x_2$)

Kendrick Kay, Washington Univ. in St. Louis
Ordinary least-squares (OLS) solution

regressors $\cdot$ residuals $= 0$
regressors $\cdot$ (data $-$ model fit) $= 0$

$$X^T (y - X\hat{w}_{OLS}) = 0$$

$$X^T y - X^T X\hat{w}_{OLS} = 0$$

$$\hat{w}_{OLS} = (X^T X)^{-1} X^T y$$
Fitting nonlinear models using local, iterative optimization

\[ y = ax^n \]
Linear models do not have local minima

Nonlinear models may have local minima

Local minimum = Global minimum

Local minimum

Global minimum

Kendrick Kay, Washington Univ. in St. Louis
Maximum likelihood implies squared error

\[
\text{likelihood}(d \mid m) = \prod_{i=1}^{n} p(d_i) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(d_i - m_i)^2}{2\sigma^2}}
\]

arg \max_m \left( \text{likelihood}(d \mid m) \right)

= arg \max_m \left( \text{log-likelihood}(d \mid m) \right)

= arg \min_m \left( \text{negative-log-likelihood}(d \mid m) \right)

= arg \min_m \left( -\log \left( \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(d_i - m_i)^2}{2\sigma^2}} \right) \right)

= arg \min_m \left( -\log \left( \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} + \frac{(d_i - m_i)^2}{2\sigma^2} \right) \right)

= arg \min_m \left( \sum_{i=1}^{n} \frac{(d_i - m_i)^2}{2\sigma^2} \right)

= arg \min_m \left( \sum_{i=1}^{n} (d_i - m_i)^2 \right)

= arg \min_m \left( \text{squared error} \right)
Squared error vs. absolute error

- Squared error: \[ \sum_{i=1}^{n} (d_i - m_i)^2 \]
- Absolute error: \[ \sum_{i=1}^{n} |d_i - m_i| \]