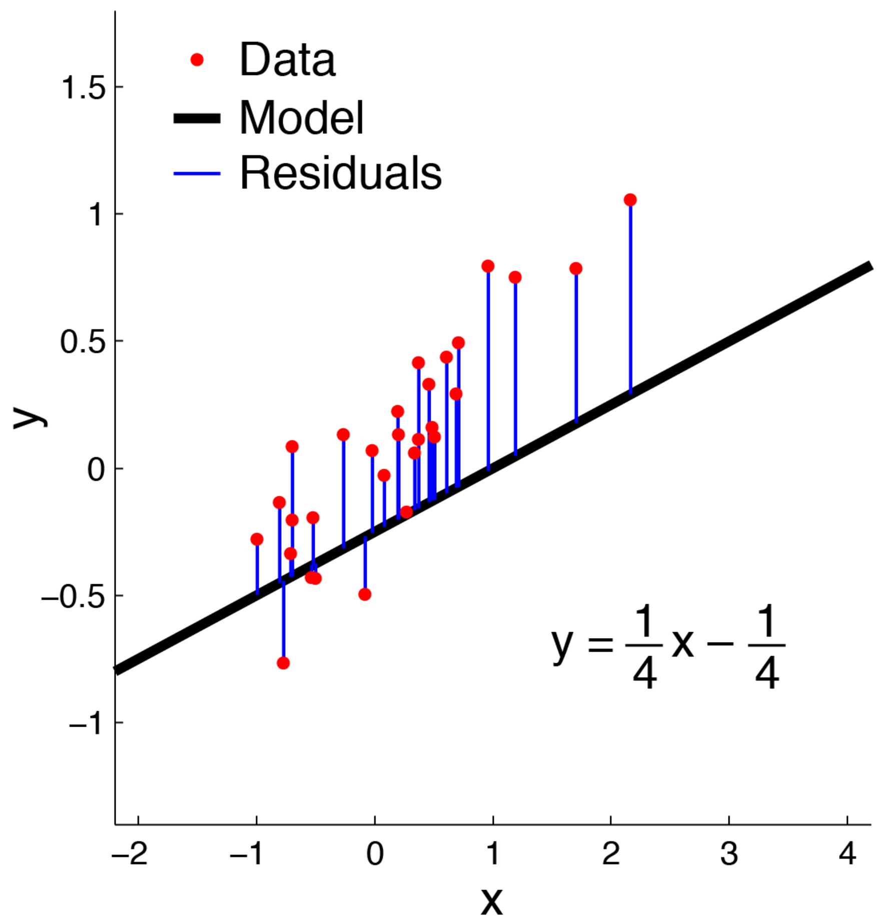
# Statistics and Data Analysis in MATLAB

## Lecture 4:
## Model fitting

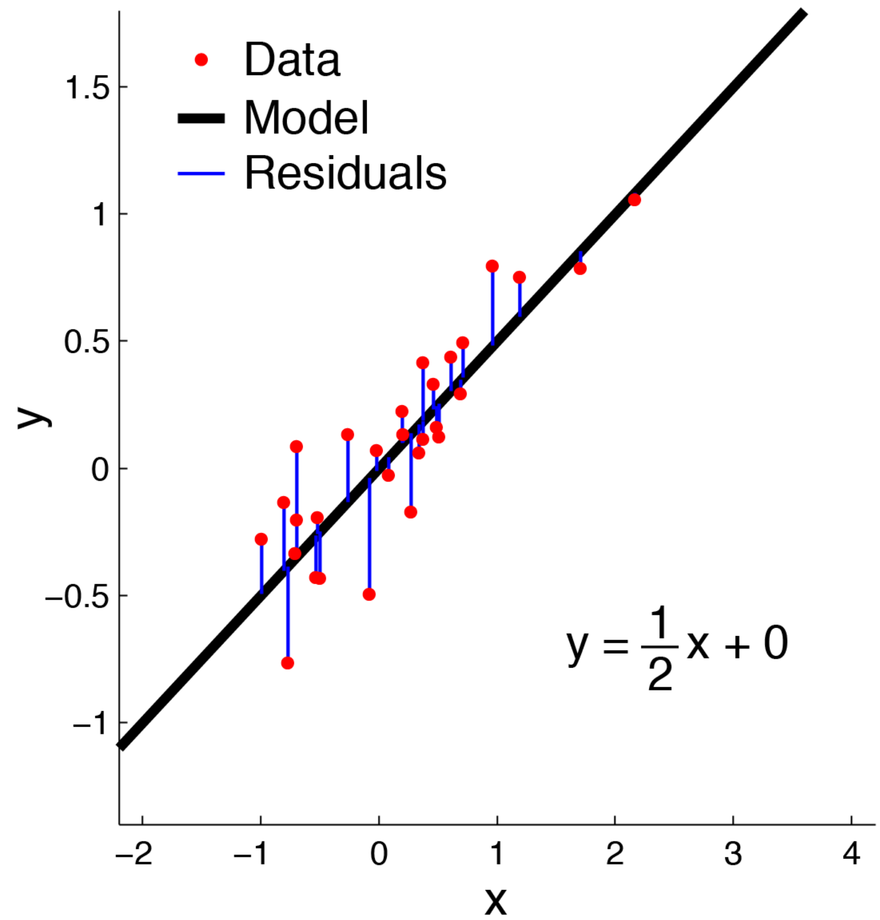
Kendrick Kay
Washington University in St. Louis
February 28, 2014

$$\text{squared error} = \sum_{i=1}^{n} \left(d_i - m_i\right)^2$$

## High squared error




## Low squared error

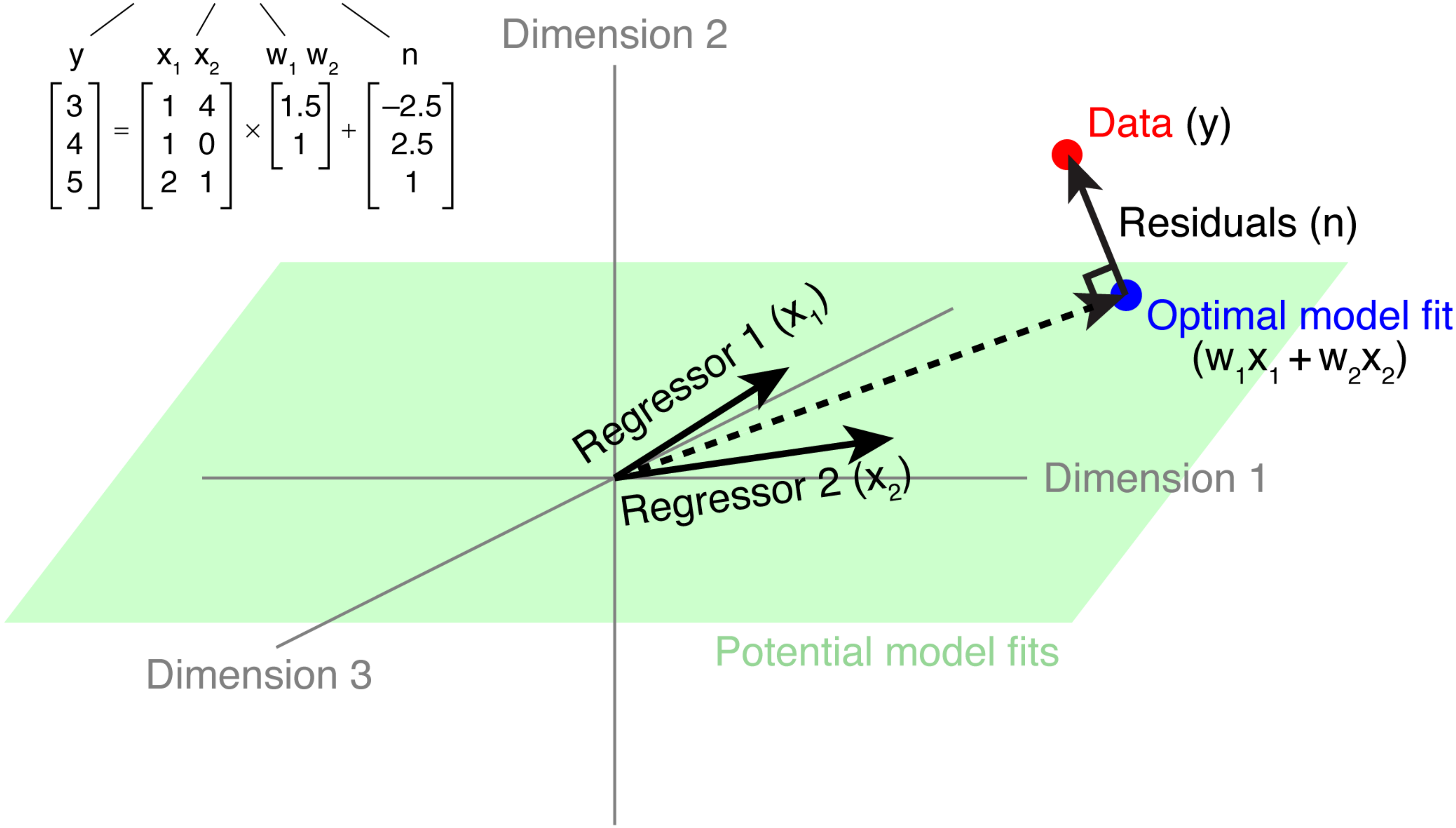



*Kendrick Kay, Washington Univ. in St. Louis*

# Geometric interpretation of linear regression

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{n}$$

$$\underset{y}{\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}} = \underset{x_1 \; x_2}{\begin{bmatrix} 1 & 4 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}} \times \underset{w_1 \; w_2}{\begin{bmatrix} 1.5 \\ 1 \end{bmatrix}} + \underset{n}{\begin{bmatrix} -2.5 \\ 2.5 \\ 1 \end{bmatrix}}$$




*Kendrick Kay, Washington Univ. in St. Louis*

# Ordinary least-squares (OLS) solution

$$\text{regressors} \cdot \text{residuals} = 0$$

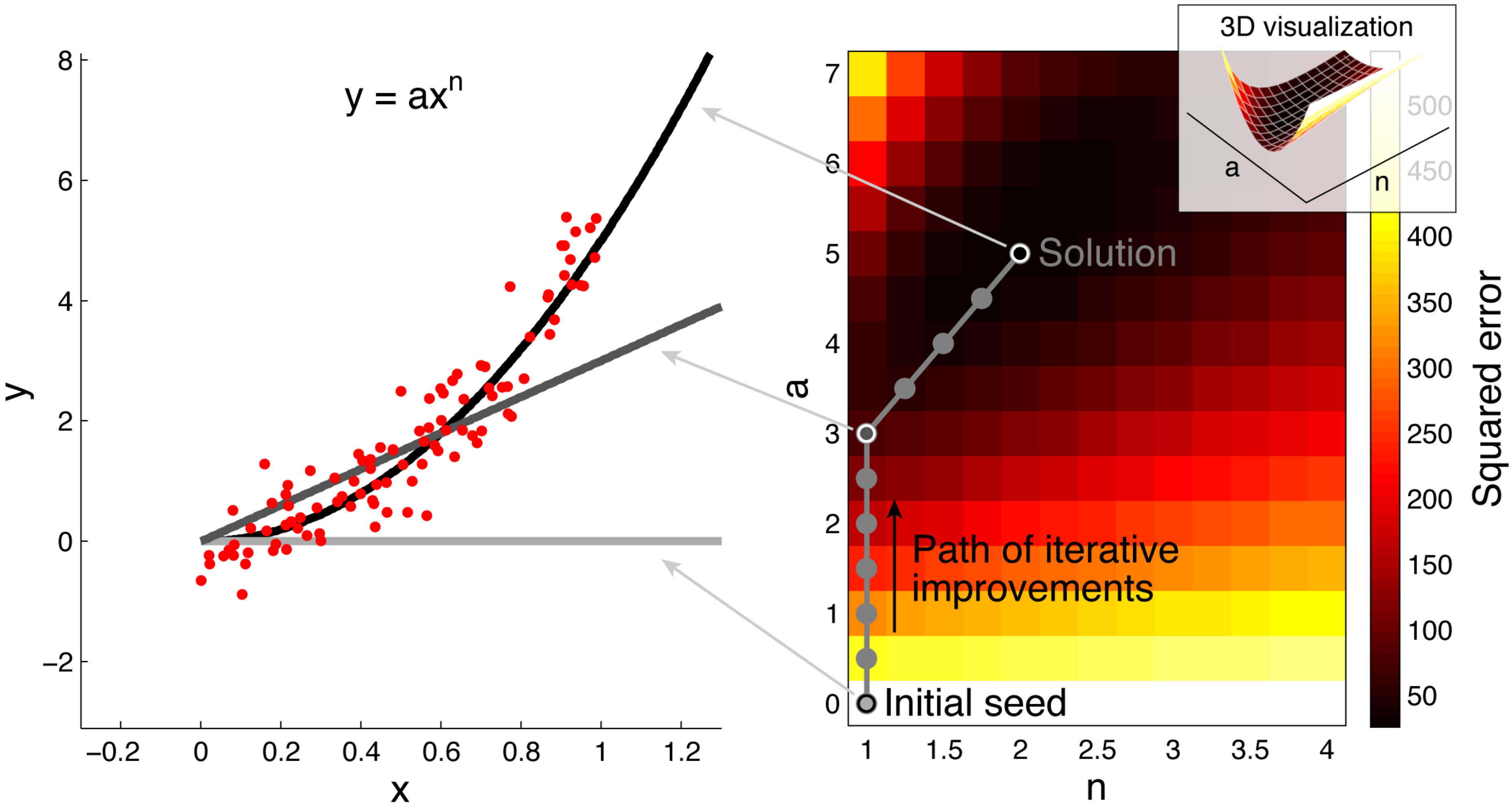
$$\text{regressors} \cdot (\text{data} - \text{model fit}) = 0$$

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

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

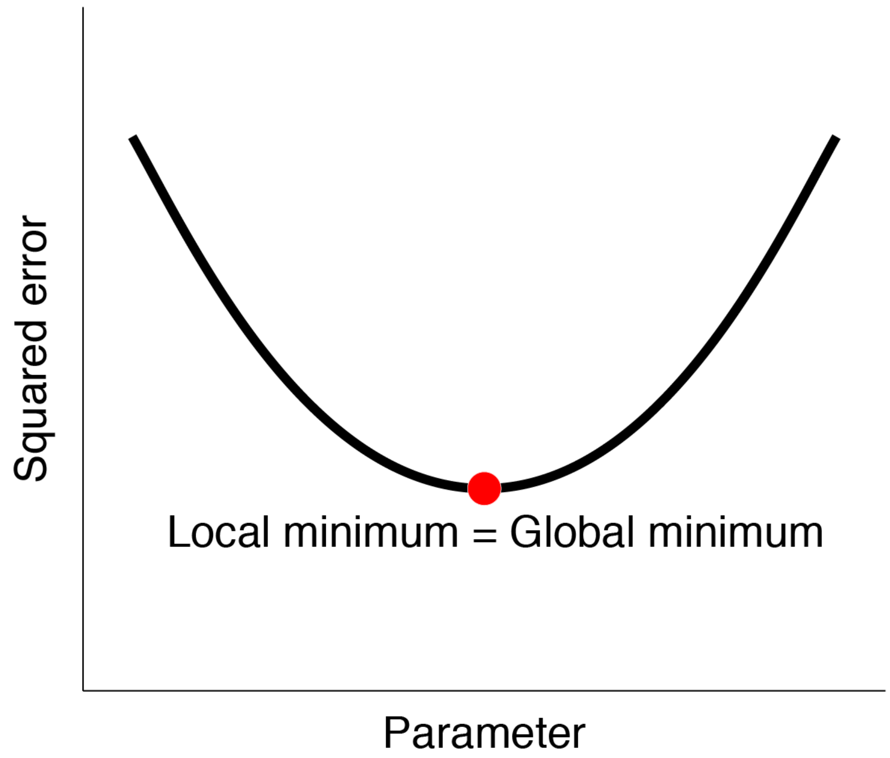
$$\hat{\mathbf{w}}^{\mathrm{OLS}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

# Fitting nonlinear models using local, iterative optimization

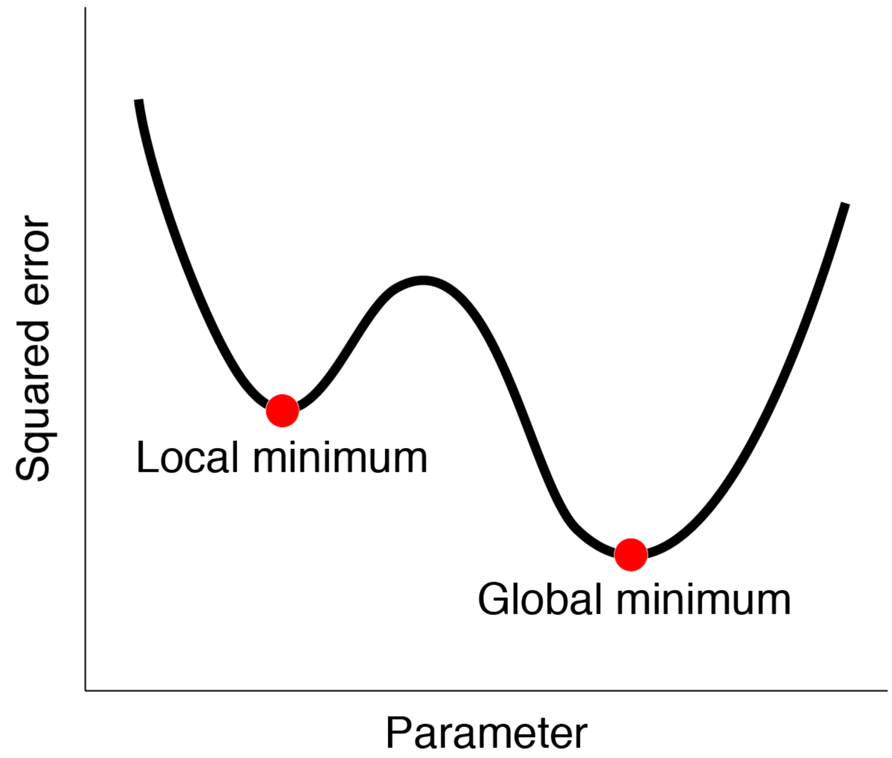



*Kendrick Kay, Washington Univ. in St. Louis*

## Linear models do not have local minima




## Nonlinear models may have local minima




*Kendrick Kay, Washington Univ. in St. Louis*

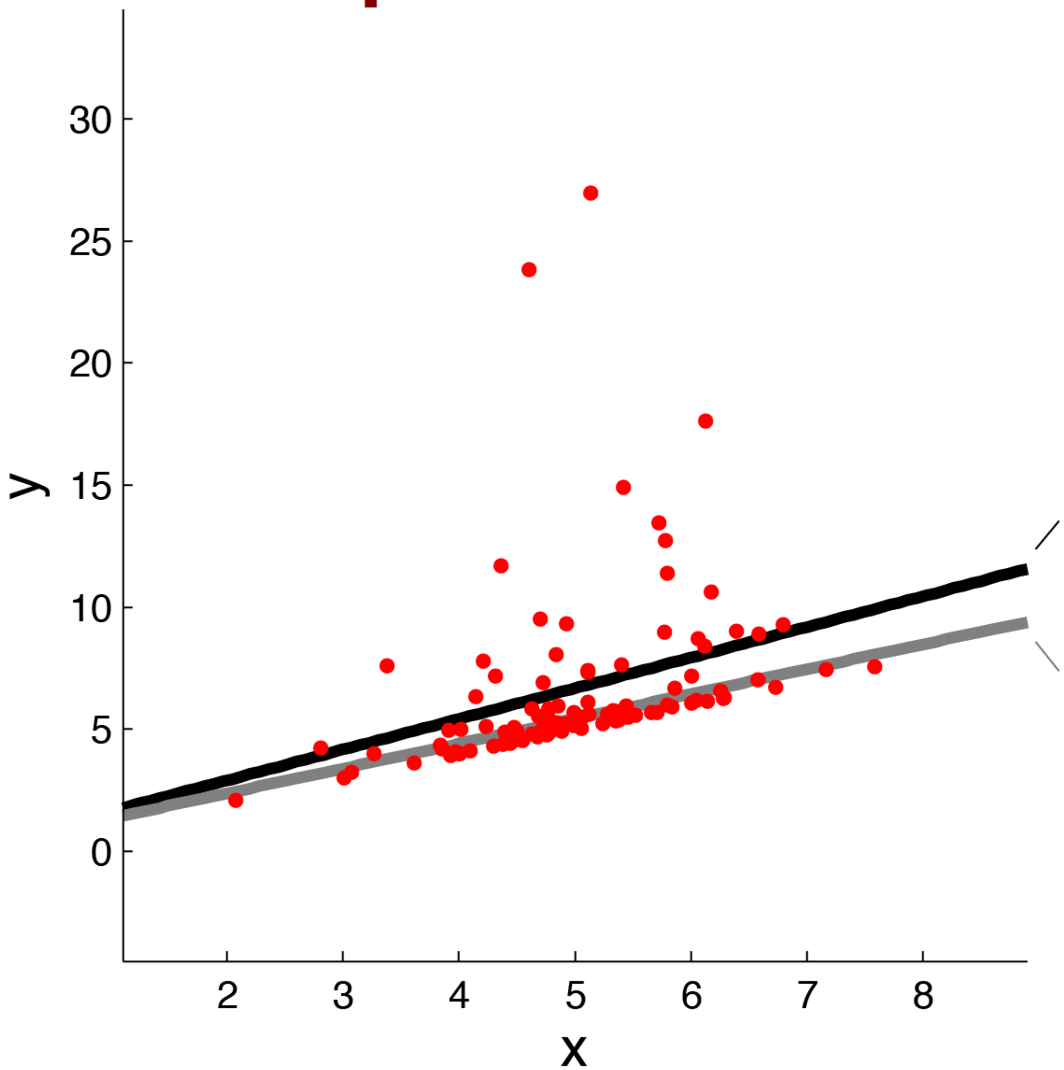
# Maximum likelihood implies squared error

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

$$\begin{aligned}
&\underset{m}{\arg\max}\left(\text{likelihood}(d \mid m)\right) \\
&= \underset{m}{\arg\max}\left(\text{log-likelihood}(d \mid m)\right) \\
&= \underset{m}{\arg\min}\left(\text{negative-log-likelihood}(d \mid m)\right) \\
&= \underset{m}{\arg\min}\left(-\log\left(\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(d_i - m_i)^2}{2\sigma^2}}\right)\right) \\
&= \underset{m}{\arg\min}\left(\sum_{i=1}^{n}\left(-\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \frac{(d_i - m_i)^2}{2\sigma^2}\right)\right) \\
&= \underset{m}{\arg\min}\left(\sum_{i=1}^{n} \frac{(d_i - m_i)^2}{2\sigma^2}\right) \\
&= \underset{m}{\arg\min}\left(\sum_{i=1}^{n} (d_i - m_i)^2\right) \\
&= \underset{m}{\arg\min}\left(\text{squared error}\right)
\end{aligned}$$

*Kendrick Kay, Washington Univ. in St. Louis*

# Squared error vs. absolute error




$$\text{squared error} = \sum_{i=1}^{n} \left(d_i - m_i\right)^2$$

$$\text{absolute error} = \sum_{i=1}^{n} \left|d_i - m_i\right|$$

Kendrick Kay, Washington Univ. in St. Louis