Ordinary least squares

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \|Ax - y\|^2 = \sum_{i=1}^{m} (a_i^T x - y_i)^2$$

- linear measurements of unknown vector $x \in \mathbb{R}^n$:

$$y_i = a_i^T x + \epsilon_i, \qquad i = 1, \ldots, m$$

  where
  - y_i is ith measurement
  - $a_i \in \mathbb{R}^n$ describes how ith measurement depends on x
  - ϵ_i is noise or error in ith measurement
- choose value of x that minimizes sum of squared errors:

$$\|\epsilon\|^2 = \sum_{i=1}^{m} \epsilon_i^2 = \sum_{i=1}^{m} (a_i^T x - y_i)^2$$

Weighted least squares

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \sum_{i=1}^{m} w_i (a_i^\mathsf{T} x - y_i)^2$$

- ordinary least squares (OLS):
  equal weight for each measurement ($w_i = 1$)
- weighted least squares (WLS):
  different weight for each measurement
- $w_i \in \mathbb{R}_+$ is weight of the ith measurement
- more accurate or reliable measurements receive larger weight

Solving weighted least-squares problems

$$\sum_{i=1}^{m} w_i(a_i^\top x - y_i)^2$$

$$= \left\| \begin{bmatrix} \sqrt{w_1}(a_1^\top x - y_1) \\ \vdots \\ \sqrt{w_m}(a_m^\top x - y_m) \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} \sqrt{w_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{w_m} \end{bmatrix} \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix} x - \begin{bmatrix} \sqrt{w_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{w_m} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \right\|^2$$

Solving weighted least-squares problems

$$\sum_{i=1}^{m} w_i(a_i^\mathsf{T} x - y_i)^2 = \|W^{\frac{1}{2}}Ax - W^{\frac{1}{2}}y\|^2$$

where

$$W = \mathrm{diag}(w_1, \dots, w_m), \qquad W^{\frac{1}{2}} = \mathrm{diag}(\sqrt{w_1}, \dots, \sqrt{w_m})$$

- ordinary least-squares problem!
  - measurement matrix: $W^{\frac{1}{2}}A$
  - observation vector: $W^{\frac{1}{2}}y$
- solution:

$$\begin{aligned} x &= ((W^{\frac{1}{2}}A)^\mathsf{T}(W^{\frac{1}{2}}A))^{-1}(W^{\frac{1}{2}}A)^\mathsf{T}(W^{\frac{1}{2}}y) \\ &= (A^\mathsf{T}WA)^{-1}A^\mathsf{T}Wy \end{aligned}$$

Example: heteroscedastic errors

- measurement model:

$$y_i = mx_i + b + \epsilon_i, \qquad i = 1, \dots, N$$

- error model:

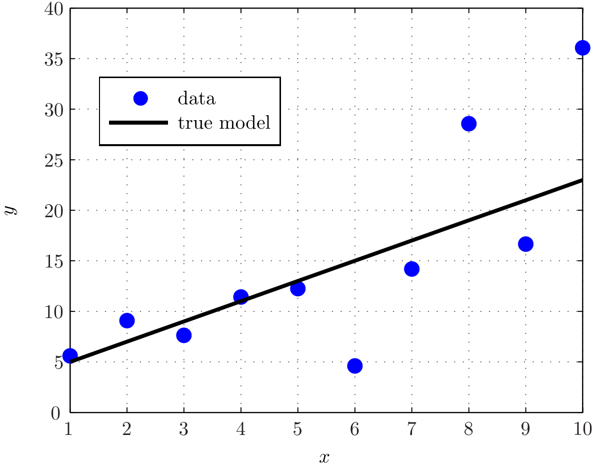
$$\epsilon_i \sim \mathcal{N}(0, i^2), \qquad i = 1, \dots, N$$

- weights:

$$w_i = \frac{1}{i}, \qquad i = 1, \dots, N$$

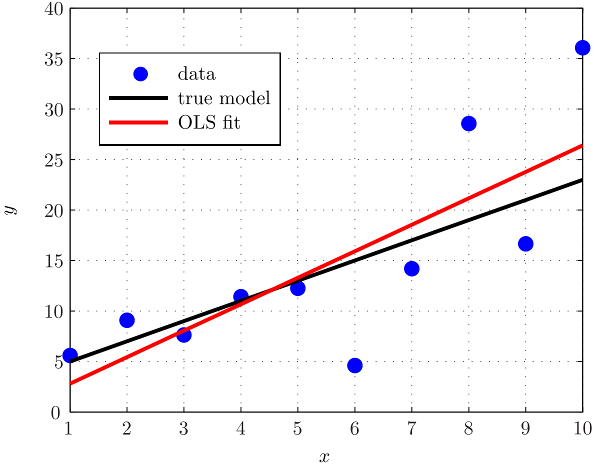
Example: heteroscedastic errors

example data: $m = 2$, $b = 3$



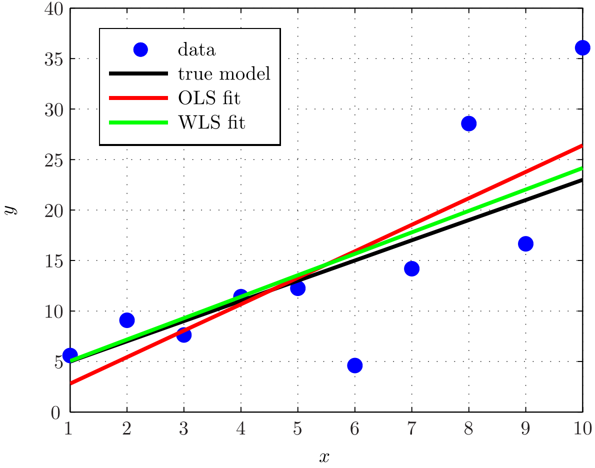
Example: heteroscedastic errors

ordinary least-squares fit: $m_{\text{ols}} = 2.6223$, $b_{\text{ols}} = 0.1845$



Example: heteroscedastic errors

weighted least-squares fit: $m_{\mathrm{wls}} = 2.1234$, $b_{\mathrm{wls}} = 2.9287$



Weighted least squares
- Weighted least squares
- Example: heteroscedastic errors

Iteratively reweighted least squares
- Iteratively reweighted least squares
- Example: ℓ_1 regression

Iteratively reweighted least squares

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \sum_{i=1}^m \phi(a_i^\mathsf{T} x - y_i)$$

- $\phi : \mathbb{R} \to \mathbb{R}_+$ is penalty function
- write as weighted least-squares problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \sum_{i=1}^m w_i(x)(a_i^\mathsf{T} x - y_i)^2$$

- where weight function is

$$w_i(x) = \frac{\phi(a_i^\mathsf{T} x - y_i)}{(a_i^\mathsf{T} x - y_i)^2}$$

Iteratively reweighted least squares

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \sum_{i=1}^m w_i(x)(a_i^\mathsf{T} x - y_i)^2$$

- the weights $w_i(x)$ depend on x
- choose initial guess $x^{(0)}$
- solve a sequence of weighted least-squares problems:
  - for $k = 0, 1, 2, \ldots$
  - compute weights using previous estimate for x:

$$W(x^{(k)}) = \text{diag}(w_1(x^{(k)}), \ldots, w_m(x^{(k)}))$$

  - solve weighted-least squares problem for next estimate of x:

$$x^{(k+1)} = (A^\mathsf{T} W(x^{(k)}) A)^{-1} A^\mathsf{T} W(x^{(k)}) y$$

Example: ℓ_1 regression

ℓ_1 regression problem:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} : \|Ax - y\| = \sum_{i=1}^{m} |a_i^\mathsf{T} x - y_i|$$

iteratively reweighted least squares:

- penalty function:

$$\phi(d) = |d|$$

- weight function:

$$w_i(x) = \frac{\phi(a_i^\mathsf{T} x - y_i)}{(a_i^\mathsf{T} x - y_i)^2} = \frac{1}{|a_i^\mathsf{T} x - y_i|}$$

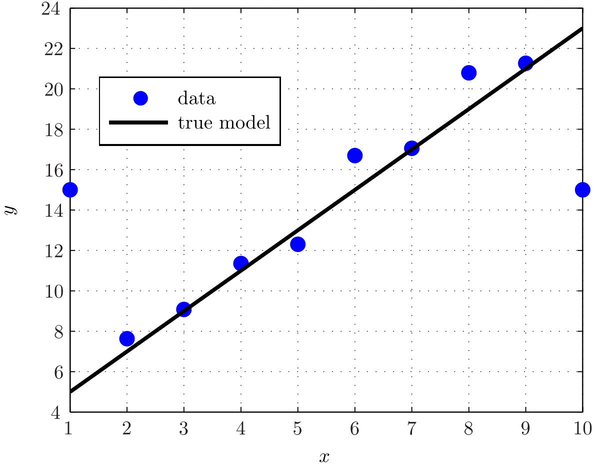
- practical weight function:

$$w_i(x) = \frac{1}{\max\{|a_i^\mathsf{T} x - y_i|, \delta\}}$$

where δ is small, positive constant

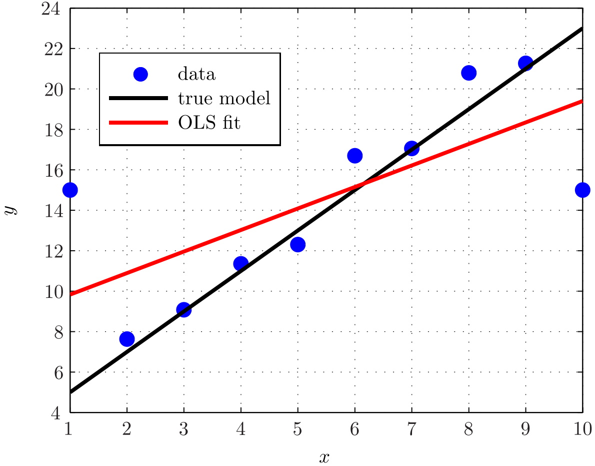
Example: ℓ_1 regression

example data: $m = 2$, $b = 3$



Example: ℓ_1 regression

ordinary least-squares fit: $m_{\text{ols}} = 1.0642$, $b_{\text{ols}} = 8.7645$



Example: ℓ_1 regression

weighted least-squares fit: $m_{\ell_1} = 1.8864$, $b_{\ell_1} = 3.8539$