

9. Robust regression

Least squares regression	2
Sensitivity to outliers	3
Robust regression - methods	4
L_1 regression	5
Huber regression	6
Mallows/Schweppe regression	7
Breakpoint	8
LMS regression	9
MM-estimation	10
Some closing thoughts (see Faraway Ch 13)	11

Least squares regression

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^n (y_i - x_i^T \theta)^2 = \operatorname{argmin}_{\theta} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \operatorname{argmin}_{\theta} \sum_{i=1}^n \hat{\epsilon}_i^2$$

Why least squares regression?

- Historic (used since 1800)
- The least squares estimator $\hat{\theta} = (X^T X)^{-1} X^T y$ has a closed form solution, and is simple to compute
- If $y = X\theta + \epsilon$ and $\epsilon \sim N_n(0, \sigma^2 I)$:
 - ◆ Least squares estimator = MLE
 - ◆ Least squares estimator has smallest variance among all unbiased estimators (Gauss-Markov)

2 / 11

Sensitivity to outliers

- Least squares estimators are sensitive to outliers
- Possible solutions:
 - ◆ Good diagnostics (Cook's distance) - use least squares regression with and without the influential points
 - ◆ Robust regression - move away from least squares regression

3 / 11

Robust regression - methods

- L_1 regression (=Least Absolute Deviations (LAD) regr.)
- Huber regression
- Mallows regression
- Schweppe regression
- Least Median of Squares (LMS) regression

4 / 11

L_1 regression

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^n |y_i - x_i^T \theta|$$

- Older than LS: Boscovich (1760), Laplace (1789)
- Did not become popular, since the solution cannot be written in closed form (no problem anymore with modern computers)
- In location model $y_i = \alpha + \epsilon_i$, L_1 regression gives median of the data
- Is more robust against outliers in the y -direction, but still very sensitive to outliers in the x -direction
- Is inefficient when the errors are normally distributed

5 / 11

Huber regression

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^n \rho_c(y_i - x_i^T \theta),$$

where

$$\rho_c(u) = \begin{cases} u^2/2 & \text{if } |u| \leq c \\ c(|u| - c/2) & \text{if } |u| \geq c \end{cases}$$

- Compromise between L_1 and L_2 regression:
 - ◆ $c = \infty \Rightarrow L_2$ regression (=least squares)
 - ◆ $c = 0 \Rightarrow L_1$ regression (use $\rho_c(u) = |u|$)
- Idea: penalize small residuals quadratically, and large residuals linearly
- Computation: solve $\sum_{i=1}^n \psi_c(y_i - x_i^T \theta) x_i = 0$, where $\psi_c(u) = \rho'_c(u) = \operatorname{sign}(u) \min(|u|, c)$.
- c should be chosen suitably w.r.t residuals. Estimator is computed in R using iterated weighted least squares.

6 / 11

Mallows/Schweppe regression

$$\sum_{i=1}^n \eta \left(x_i, \frac{y_i - x_i^T \hat{\theta}}{\hat{\sigma}} \right) x_i = 0$$

- Mallows:

$$\eta(x, r) = \min \left(1, \frac{a}{\|Ax\|} \right) \psi_c(r)$$

- Schweppe:

$$\eta(x, r) = \frac{1}{\|Ax\|} \psi_c(\|Ax\|r)$$

- $\|Ax\|$ is a measure of leverage of x , for example $\|Ax\|^2 = \text{const} x^T (X^T X)^{-1} x$, but then robust version
- $\psi_c = \rho'(c)$ from Huber regression
- Robust against x - and y -outliers

7 / 11

Breakpoint

The breakpoint of an estimator = the maximum proportion of observations that can be changed without changing the estimator

- Breakpoint of average: 0
- Breakpoint of median: 1/2
- Breakpoint of Least Squares regression: 0
- Breakpoint of L_1 and Huber: 0 (in x -direction)
- Breakpoint Mallows/Schweppe: $\leq 1/p$

8 / 11

LMS regression

$$\hat{\theta} = \operatorname{argmin}_{\theta} \operatorname{median}((y_i - x_i^T \theta)^2)$$

- See picture on slide
- Hampel (1975), Rousseeuw (1984)
- Breakpoint is approximately 0.5
- Difficult to compute because of many local minima
- Inefficient when statistical errors are normally distributed (convergence rate $n^{-1/3}$). This can be improved by replacing the median by an α -truncated mean that leaves out the αn observations with the largest residuals (least trimmed squares).

9 / 11

MM-estimation

- First find highly robust M-estimate of σ (first M).
- Then keep $\hat{\sigma}$ fixed and find a close by M-estimate of θ , for example using a Newton step (second M).

10 / 11

Some closing thoughts (see Faraway Ch 13)

- Robust estimators protects against long-tailed errors, but not against problems with model choice and variance structure. These latter problems are often more serious than non-normal errors.
- Inference for $\hat{\theta}$ is more difficult. One can use the bootstrap.
- Robust methods can be used in addition to least squares. There is cause to worry if the two estimators differ a lot.

11 / 11