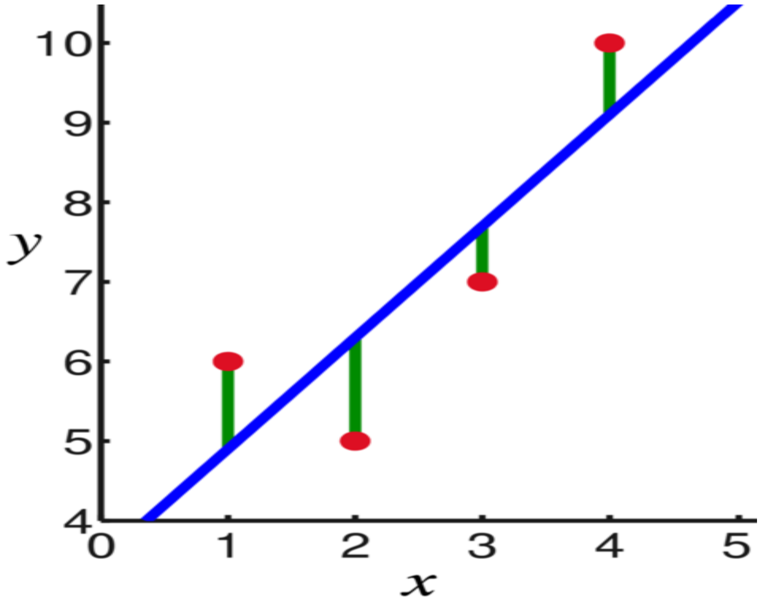
Linear Regression as Linear Optimization

Eric Olberding

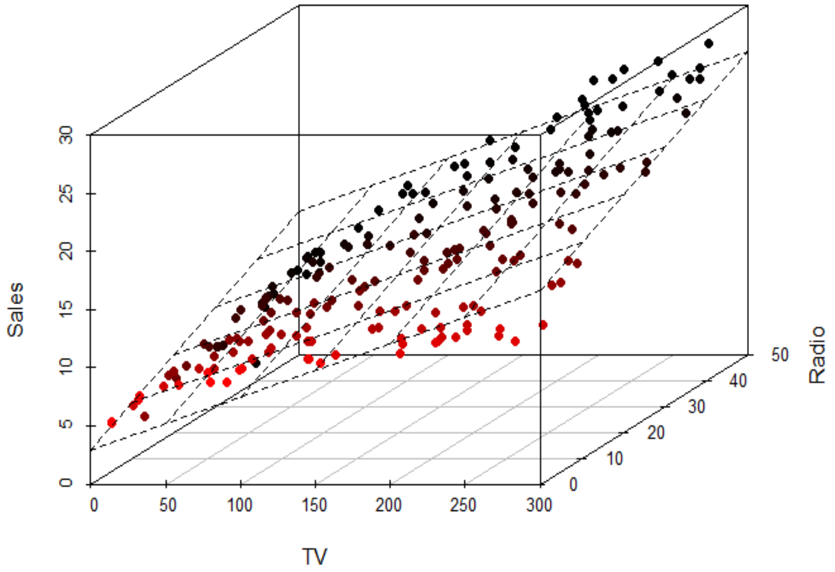
Linear Regression

- Model a dependent variable y as a linear function of some independent variables x_i. The model usually won't be perfect so there is some error term ϵ.
- Example for 1 independent variable: $y = \beta_0 + \beta x + \epsilon$



Linear Regression: General Form

- $y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \epsilon_i; \qquad i = 1, \ldots, n$
- N datapoints $(y_i, x_{i,1}, \ldots, x_{i,p})$
- 3-dimensional example (y, x_1, x_2)



Minimizing Error

- $\epsilon_i = y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j}\,;\qquad i = 1, \dots, n$
- Want to find β_j values that minimize "total error"
- Ordinary Least Squares(OLS) Regression minimizes Squared Error

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j}\right)^2 = \sum_{i=1}^{n} \epsilon_i^2 \text{ (quadratic optimization)}$$

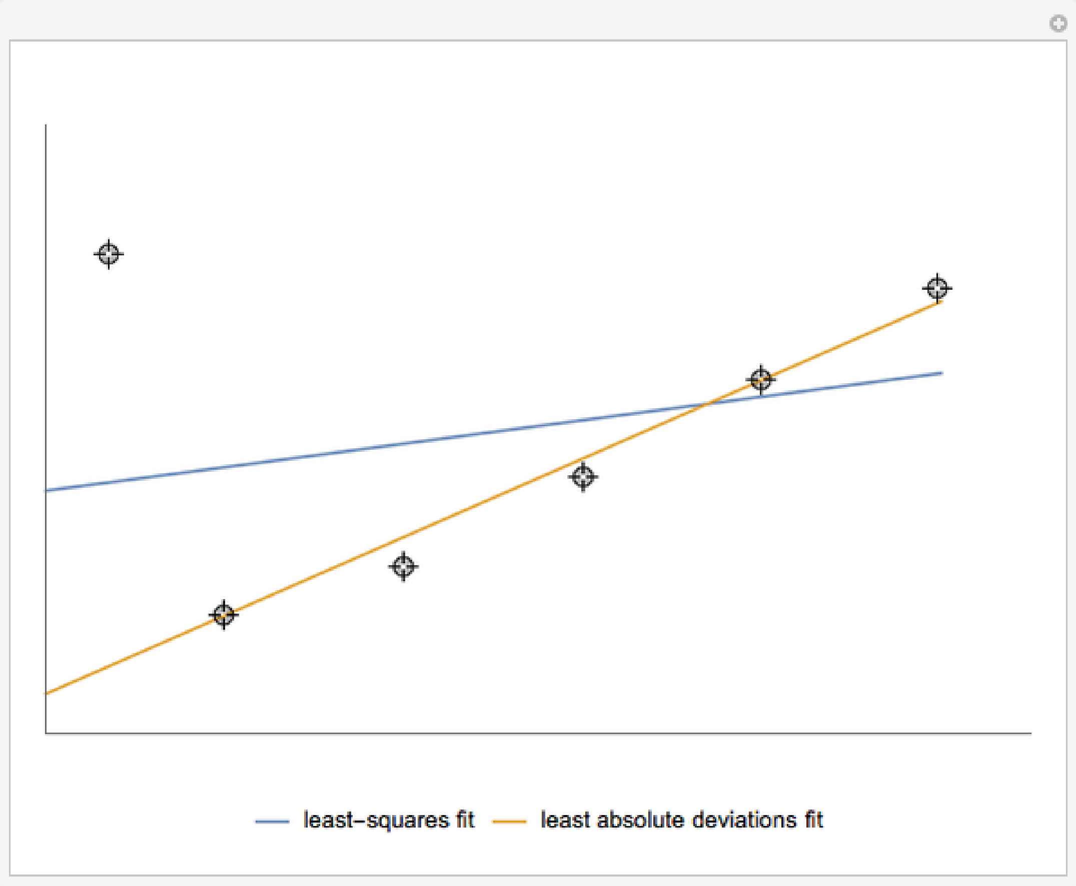
- Least Absolute Deviation(LAD) Regression minimizes Absolute Error

$$\sum_{i=1}^{n}\left|y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j}\right| = \sum_{i=1}^{n}|\epsilon_i| \quad \text{(linear optimization)}$$

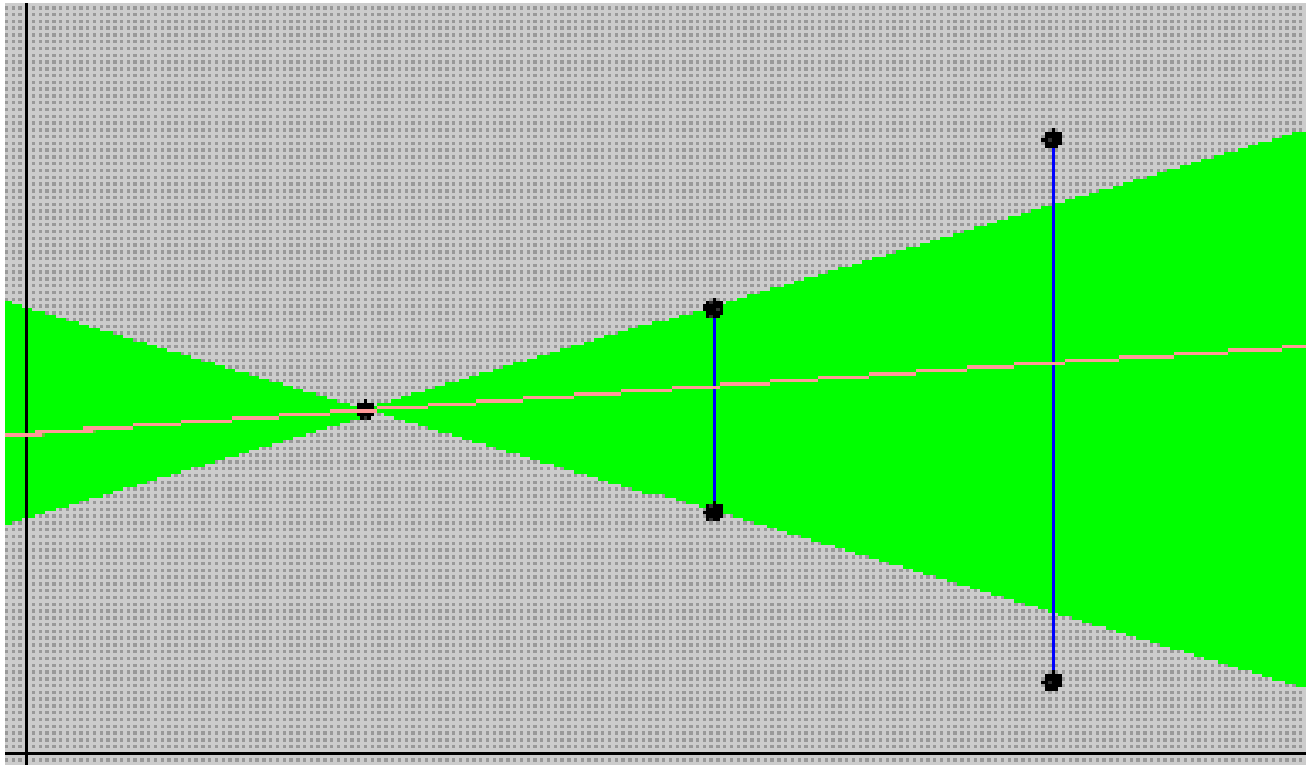
LAD vs OLS

LAD	OLS
Robust	Not very robust
Possibly many solution	Unique solution
Unstable Solution	Stable Solution

Robustness of LAD

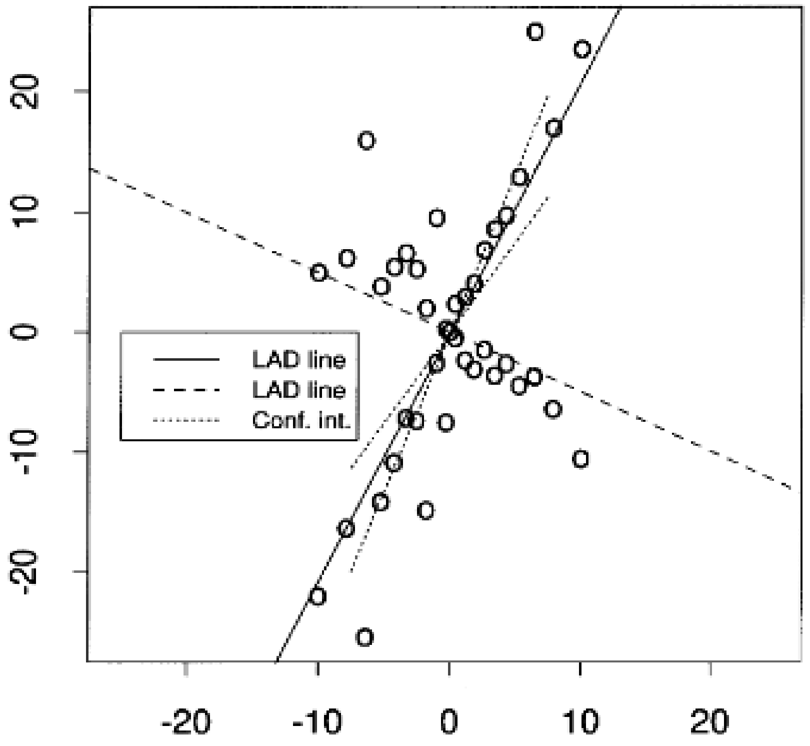


Multiple Solutions using LAD



Instability of LAD

- Instability: small changes to data lead to large changes in the fit line
- Singularity: x_0 is a singularity if limit of a statistic $\delta(x)$ does not exist as $x \to x_0$ (x is a matrix)



Deriving OLS β

- Loss function is quadratic in β with positive definite Hessian matrix so a unique global minimum exists
- $\hat{\beta} = (X^T X)^{-1} X^T y$
- jth row of X is $(1, x_{j,1}, x_{j,2}, \ldots, x_{j,p})$

Least Absolute Deviations as a Linear Program

- LP formulation:

$$\min \sum_{i=1}^{n} |\epsilon_i|$$

$$y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j} = \epsilon_i; \qquad i = 1, \dots, n$$

- x, y are data (constraints). βs are the decision variables.

LAD in form for AMPL

- Need fact that $|a| \leq b \Leftrightarrow -b \leq a \leq b$
- $t_i \geq 0 \quad i = 1, \ldots, n$ such that $|\epsilon_i| \leq t_i \Leftrightarrow -t_i \leq \epsilon_i \leq t_i$
- $\sum_{i=1}^{n} |\epsilon_i| \leq \sum_{i=1}^{n} t_i$

- LP formulation:

$$\min \sum_{i=1}^{n} t_i$$

$$-t_i \leq y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j} \leq t_i; \qquad i = 1, \ldots, n$$

Examples: mtcars data

- Variables in handout
- $mpg = 7.62 + 0.65cyl + 0.02disp - 0.03hp + 0.78drat - 4.56wt + 0.58qsec + 1.59vs + 1.33am + 2.37gear - 0.33carb + \epsilon$
- R^2 for LAD: 0.838
- R^2 for OLS: 0.869
- $R^2 = 1 - \frac{\sum_{i=1}^{n}(\epsilon_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Examples: iris data

- $petallength = -2.636 + 1.72 sepallength - 1.22 sepalwidth + \epsilon$
- R^2 for LAD: 0.864
- R^2 for OLS: 0.867

Different LP formulation

- Call the previous formulation the Bounding Form
- Splitting Variables LP formulation:

$$\min \sum_{i=1}^{n} r_i^+ + r_i^-$$

$$r_i^+ - r_i^- = y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j}\,; \qquad i = 1, \dots, n$$

$$r_i^+, r_i^- \geq 0; \qquad i = 1, \dots, n$$

- For iris data, splitting variables formulation took 119 iterations, bounding formulation 90 iterations

Maximum Absolute Deviation

- Slight modification of LAD bounded formulation
- Only minimize bound on largest error

- LP formulation:

$$\min t$$

$$-t \leq y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{i,j} \leq t; \quad i = 1, \ldots, n$$

- Iris dataset: $petallength = -2.636 + 1.84 sepallength - 1.35 sepalwidth + \epsilon$

References

- https://en.wikipedia.org/wiki/File:Linear_regression.svg
- https://stackoverflow.com/questions/26431800/plot-linear-model-in-3d-with-matplotlib
- https://demonstrations.wolfram.com/ComparingLeastSquaresFitAndLeastAbsoluteDeviationsFit/
- https://en.wikipedia.org/wiki/Linear_regression
- https://vanderbei.princeton.edu/542/lectures/lec9.pdf
- https://en.wikipedia.org/wiki/Least_absolute_deviations
- Ellis, S. (1998) *Instability of Least Squares, Least Absolute Deviation and Least Median of Squares Linear Regression*