Linear Regression

OLS

SLR

Least Squares

L2 regression

Linear regression w/ Least Squares

given

response

(x_i, y_i) pairs

$i = 1, \ldots, n$

covariate

Setup: Assume have a relationship

$$y_i \approx \beta_0 + \beta_1 x_i$$

fixed parameters

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$ ← noise

$$\varepsilon_i \sim [0, \sigma_\varepsilon^2] \quad i=1,\dots,n$$

ε_i uncorrelated.

How do we find

$\hat{\beta}_0, \hat{\beta}_1$ (estimates of the coefficients)

our line estimate

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

"fitted values"

we can minimize how bad we do.

$|y_i - \hat{y}_i|^p$ is how bad we do.

We generally call

$$y_i - \hat{y}_i = \hat{\varepsilon}_i$$ "residual"

A simple choice $p=2$

Least squares

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{b_0, b_1}{\operatorname{argmin}} \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2$$

$$= \underset{b_0, b_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \underset{b_0, b_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

$$\Downarrow$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Our past typical setup

$X_i \sim P_\theta(x) \quad i = 1, \ldots, n$

we estimate θ via MLE usually

Before I said

$$\varepsilon_i \overset{iid}{\sim} [0, \sigma_\varepsilon^2]$$

I will additionally assume

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$y_i \overset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma_\varepsilon^2)$$

$$i = 1, \dots, n$$

we can do MLE

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(\frac{-1}{2\sigma_\varepsilon^2}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n \exp\left(\frac{-1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right)$$

$$\ell(\beta_0, \beta_1) = \log\left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n$$

$$-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

maximizing $\ell(\beta_0, \beta_1)$

equivalent to max

$$-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

equivalent to min

$$\sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

we've recovered Least squares from Normal MLE.

Correlation

For population

$$\rho = cor(X,Y)$$

$$= \frac{cov(X,Y)}{sd(X)\, sd(Y)}$$

$$= \frac{E\big((X-E(X))(Y-E(Y))\big)}{sd(X)\, sd(Y)}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\Rightarrow \hat{\beta}_1 = r_{xy} \frac{sdy}{sdx}.$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}$$

why not absolute deviation

(L1)

$$\underset{b_0, b_1}{\operatorname{argmin}} \sum_{i=1}^{n} |y_i - (b_0 + b_1 x_i)|$$

vs.

$$\underset{b_0, b_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

(L2)

L2 is sensitive to outliers more than L1.

Ridge

$$\underset{\beta}{\text{argmin}} \ \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$$

Lasso

$$\underset{\beta}{\text{argmin}} \ \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

$$\|\beta\|_1 = \sum |\beta_i|$$