Oct 25th

LINEAR REGRESSION

For the same x, y varies. y is continuous

$y = X\beta$

LEAST SQUARES

$\beta_{LS} = (X^TX)^{-1}X^Ty$ ⇝ maximize likelihood

β_{LS} is an unbiased estimator

LOGISTIC REGRESSION (classification → y - discrete)

$$\log\left(\frac{P(y=1|X)}{1-P(y=1|X)}\right) = \alpha + \beta X$$

assume 2 classes

The logit func is the log odds of being in class 1.

logit^{-1} to recover probability

How do we learn $\alpha + \beta$

$$P(y_i = 1|x_i) = \text{logit}^{-1}(\alpha+\beta x_i)^{y_i}(1-\text{logit}^{-1}(\alpha+\beta x_i))^{1-y_i}$$

$$\mathcal{L}(\alpha,\beta) = \prod_i (\text{logit}^{-1}(\alpha+\beta x_i)^{y_i}(1-\text{logit}^{-1}(\alpha+\beta x_i))^{1-y_i}$$

y_i = true value

$\hat{y}_i$ = estimate of y_i

$\bar{y}$ = sample mean of all y_i

$y_i - \hat{y}_i$ = estimates of ϵ_i (residuals)

$$TSS = \sum_i^n (y_i - \bar{y})^2$$

(total sum of squares)

$$RSS = \sum_i^n (y_i - \hat{y}_i)^2$$

$$ESS = \sum_i^n (\hat{y}_i - \bar{y})$$

epsilon

$$R^2 = \frac{ESS}{1-TSS} = 1 - \frac{RSS}{TSS}$$

R^2 measures frac of variance that is explained by $\hat{y}$

"good" value is closer to 1.