Logistic Regression

Wednesday, September 8, 2021 11:09 AM

- Overview
  - ↳ Model: Probability $(Y = c_k \mid X)$ Conditional Distribution (Discriminative)

    $\{Yes, No\}$

- Decision threshold:

  $P(Y = Yes \mid X = ?)$

  $> x_1$

  $> x_2$

  ⋮

  ← Specific value $X = x_0$

  $P(Y = Yes \mid X = x_0)$

  > Threshold value

  (0.5 → 50%)

  or

  (0.1 → 10%)

  ⋮

- Linear Framework
  - General: $P(X) = \beta_0 + \beta_1 X$
  - Logistic Function:

$$\frac{P(X)}{1 - P(X)} = e^{\beta_0 + \beta_1 X}$$

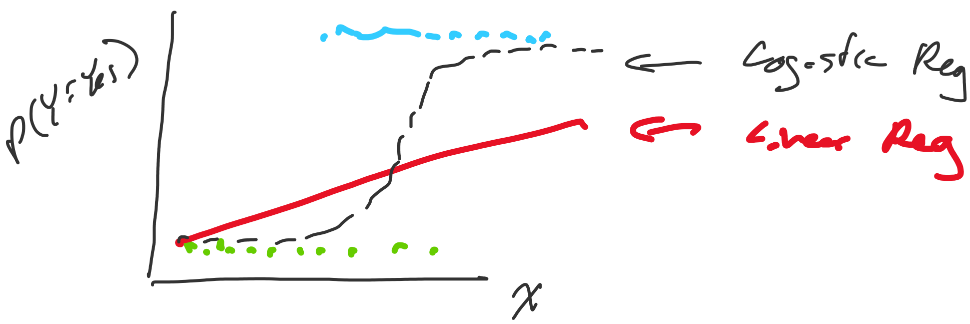
odds: [0, ∞)

Odds: $[0, \infty]$

- Logit function

$$\underbrace{\log \frac{p(x)}{1-p(x)}}_{\text{log odds}} = \beta_0 + \beta_1 x$$

*≥50% Prob ⇒ Class 1 / Yes , <50% class 0/No



- Estimation:
  - Likelihood function:

$$\ell(\beta_0, \beta_1) = \left[\prod_{i: y_i = 1} p(x_i)\right]\left[\prod_{i': y_{i'} = 0} (1 - p(x_{i'}))\right]$$

$\hat{\beta}_0, \hat{\beta}_1$ such that maximize $\ell(\beta_0, \beta_1)$

↳ Maximum Likelihood!

- Prediction
  ↳ Transform :

$f(x_0)$

$$\therefore \quad \hat{p}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_0}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_0}} = ?$$

$\geq 50\%$ $Y = T$

$< 50\%$ $Y = F$

- Multiple Logistic Regression
  - Adding terms:

$$\log \frac{p(x)}{1 - p(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

$$\hat{p}(x) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots \hat{\beta}_p x_p}}{1 - e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots \hat{\beta}_p x_p}}$$

* As x_j increases, holding all other x_j's fixed, Log odds of $Y = True$ increases/decreases by β_j factor

↳ Confounding variables!