3/21

Logistic Regression

- Needs to be negative infinity to infinity
- Take the log gives us $[0, \infty]$

Goal: fit a linear model to the log-odds of being in one of our classes

$$\log\left(\frac{P(Y=1 \mid X)}{1-P(Y=1 \mid X)}\right) = X\beta$$

- Decision boundary: ½ (probability)

  How do we get back the probability of that class? i.e. recover $P(Y=1 \mid X)$

  solution: $P(Y=1 \mid X) = \dfrac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$

- The probability is ½ only when $X\beta = 0$ which is the equation of a line

  ∴ The Boundary is a line

How to learn our model?

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

$$L(\alpha, \beta) = \prod_i P(y_i \mid x_i)$$

Log likelihood

Need to use numerical approximation, we always want maximum likelihood

* Evaluating Our Regression Model

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

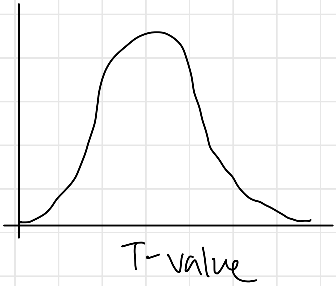
$$RSS = \sum_i^n (y_i - \hat{y}_i)^2$$

$$ESS = \sum_i^n (\hat{y}_i - \bar{y})^2$$

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

↑ measures the fraction of variance that is explained by $\hat{y}$.

Hypothesis Test:



check the T-distribution