INTRODUCTION TO CLASSIFICATION MODELS

DIFFERENCE BETWEEN CLASSIFICATION & REGRESSION

FROM THE BINOMIAL DISTRIBUTION TO LOGISTIC REGRESSION

1. CALCULATING THE AVERAGE PROBABILITY OF CANCELLATION.

   CALCULATING PROBABILITIES CONDITIONED ON GROUPS.

LOGISTIC REGRESSION

HOW CAN WE CALCULATE CONDITIONING ON A CONTINOUS VARIABLE?

PROBABILITY OF EVENT VS POINT PREDICTION

MOST LIKELY OUTCOME!

THRESHOLDING: 'BIASING' OUR MODEL

MEASURING ERROR

- THE (CONFUSING) CONFUSION MATRIX
- ACCURACY
- PRECISION/RECALL
- F1-SCORE

logit function

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

S-FUNCTION! $0 < p(x) < 1$

log loss

$$\begin{cases} -\ln p_k & \text{if } y_k = \text{Success} = 1 \\ -\ln(1-p_k) & \text{if } y_k = \text{Failure} = 0 \end{cases}$$

$$y_k \cdot (-\log(p_k)) + (1-y_k) \cdot (-\log(p_k))$$

$$-y_k \log(p_k) - (1-y_k)$$