Estimation of Parameters

- Training Data: T={(Xi, Yi)}, i=1,2, …, |T|
- Parameters: $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_M)$
- Conditional likelihood: $p(T \mid \vec{\beta}) = \prod_{i=1}^{|T|} p(Y = Y_i \mid X = X_i, \vec{\beta})$

$Y_i = 1$ $\qquad$ $Y_i = 0$

$$p(Y = 1 \mid X) = \frac{e^{\beta_0 + \sum_{i=1}^{M} x_i \beta_i}}{e^{\beta_0 + \sum_{i=1}^{M} x_i \beta_i} + 1}$$

$$p(Y = 0 \mid X) = \frac{1}{e^{\beta_0 + \sum_{i=1}^{M} x_i \beta_i} + 1}$$

- Maximum Likelihood estimate $\vec{\beta}^* = \arg\max_{\vec{\beta}} p(T \mid \vec{\beta})$

Can be computed in many ways (e.g., Newton's method)