## Logistic regression

sigmoid is a nice, simple function that increases monotonically from $0 \to 1$ as $z$ varies from $-\infty, +\infty$. it's nice derivative properties due to exponential

- discriminative model:

$$Y_i \mid X_i = x_i \sim \mathrm{Bern}(\sigma(\beta^T x_i)) \quad ; \quad \beta, X_i \in \mathbb{R}^{n+1}$$

$$\Rightarrow P(Y_i = y_i \mid X_i = x_i) = \underbrace{[\sigma(\beta^T x_i)]^{y_i}[1 - \sigma(\beta^T x_i)]^{1-y_i}}$$

: Estimate paremeters via MLE w/ SGD

$$\Rightarrow L(\beta) = P(\vec{y} \mid X; \beta) = \prod_{i=1}^{m}$$

- Multinomial logistic regression uses a categorical (multi-category analogue of bernoulli) distribution and a softmax function instead of the sigmoid function (multi category analogue of sigmoid)

Pros

- provides probabilities for outcomes
- can penalize + use regularization
- interprettable (esp. coefficients) $\theta_i = \frac{\partial(\log \text{odds})}{\partial x_i} \approx \frac{\Delta(\log \text{odds})}{\Delta x_i}$ (holding other vars. constant)
- small # of params to fit $(n+1)$
  $\Rightarrow$ low variance, high bias
- foundational model for more complicated models like NN

↝

Cons

- possible <u>too</u> simple to model data well (a linear decision boundary)