1. Logistic regression

1.1. **Conditional multinomial model.**

$$\log \frac{\mathbb{P}(Y=k|X)}{\mathbb{P}(Y=0|X)} = \beta_k^T X, \qquad k=1,\dots,K-1$$

Equivalently,

$$\mathbb{P}(Y=k|X) = \mathbb{P}(Y=0|X) e^{\beta_k^T X}$$

Sum the above equations for $k=1,\dots,K-1$:

$$\sum_{k=1}^{K-1} \mathbb{P}(Y=k|X) = \mathbb{P}(Y=K|X) \sum_{k=1}^{K-1} e^{\beta_k^T X}$$

But

$$\sum_{k=1}^{K-1} \mathbb{P}(Y=k|X) = 1 - \mathbb{P}(Y=0|X).$$

Therefore,

$$1 - \mathbb{P}(Y=0|X) = \mathbb{P}(Y=0|X) \sum_{k=1}^{K-1} e^{\beta_k^T X}$$

$$\mathbb{P}(Y=0|X) = \frac{1}{1+\sum_{k=1}^{K-1} e^{\beta_k^T X}}$$

Therefore,

$$\mathbb{P}(Y=k|X) = \frac{e^{\beta_k^T X}}{1+\sum_{k=1}^{K-1} e^{\beta_k^T X}}$$

When $K=2$,

$$\mathbb{P}(Y=1|X) = \frac{e^{\beta_1^T X}}{1+e^{\beta_1^T X}} = \sigma(\beta_1 X)$$

and we recover binary logistic regression.

$$p(y|\theta_1,\dots,\theta_{K-1}) = \theta_0^{y_0}\theta_1^{y_1}\cdots\theta_K^{y_k}, \quad \theta_0 = 1 - \sum_{k=1}^{K-1}\theta_k$$

$$\ell(\theta) = \sum_{k=0}^{K-1} y_k \log\theta_k$$

$$\theta_k = \sigma(\beta_k^T x)$$

$$\frac{1}{1+\sum_{k=1}^{K-1} e^{\beta_k^T x}}\left(1, e^{\beta_1^T x},\dots,e^{\beta_{K-1}^T x}\right) \in \Delta_{K-1}$$

$$\mathrm{softmax}(t_0,\dots,t_n) = \frac{1}{e^{t_0}+\cdots+e^{t_n}}\left(e^{t_0},\dots,e^{t_n}\right) \in \Delta_n$$