1) In softmax regression (Multinomial Logistic Regression), we compute the probability $P(y = k|x)$ for each value of $k = 1, 2, \dots, K$ as

$$\begin{bmatrix} P(y=1|x) \\ P(y=2|x) \\ \vdots \\ P(y=K|x) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} e^{\theta_j^T x}} \begin{bmatrix} e^{\theta_1^T x} \\ e^{\theta_2^T x} \\ \vdots \\ e^{\theta_K^T x} \end{bmatrix}$$

where θ_j are the parameters

Show that when $K = 2$, softmax regression reduces to the two-class logistic regression problem

We first note that $P(y = 1|x) = 1 - P(y = 2|x)$ since we have two distinct classes. We next solve for $P(y = 2|x)$

$$\begin{aligned} P(y=2|x) &= \frac{e^{\theta_2^T x}}{\sum_{i=1}^{2} e^{\theta_i^T x}} \\ &= \frac{e^{\theta_2^T x}}{e^{\theta_1^T x} + e^{\theta_2^T x}} \\ &= \frac{e^{\theta_2^T x}}{e^{\theta_1^T x} + e^{\theta_2^T x}} \cdot \frac{e^{-\theta_1^T x}}{e^{-\theta_1^T x}} \\ &= \frac{e^{(\theta_2 - \theta_1)^T x}}{1 + e^{(\theta_2 - \theta_1)^T x}} \\ &= \frac{e^{\theta'^T x}}{1 + e^{\theta'^T x}} \\ &= \sigma(\theta'^T x) \end{aligned}$$

Thus we have

$$\begin{bmatrix} p(y=1|x) \\ p(y=2|x) \end{bmatrix} = \begin{bmatrix} 1 - \sigma(\theta'^T x) \\ \sigma(\theta'^T x) \end{bmatrix}$$