Multinomial logistic regression

$y_i = 1, \cdots, k$

k categories.

Y	X_1	X_2	$\cdots$	X_m
y_1	x_{11}	x_{21}	$\cdots$	x_{m1}
y_2	x_{12}	x_{22}	$\cdots$	x_{m2}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_n	x_{1n}	x_{2n}	$\cdots$	x_{mn}

Take y_1, x_{11}, x_{21}, $\cdots$, x_{m1} as example

Let
$$\underline{z}_1 = \begin{bmatrix} z_{01} \\ z_{11} \\ \vdots \\ z_{k1} \end{bmatrix} = \underset{k\times m}{W} \cdot \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \end{bmatrix} + \underset{k\times 1}{\underline{\varepsilon}} \quad \swarrow \text{bias}$$

k categories

Then $\underline{z}_1 \in \mathbb{R}^k$

Soft-max : $\quad g(z_{j1}) = P(y_1 = j \mid z_{j1}) = \dfrac{e^{z_{j1}}}{\sum_{j=1}^{k} e^{z_{j1}}}$

$$L(y_1) = \prod_{j=1}^{k} P(y_1 = j \mid z_{j1})^{I(y_1 = j)}$$

$$= \prod_{j=1}^{k} \left(\frac{e^{z_{j1}}}{\sum_{l=1}^{k} e^{z_{l1}}} \right)^{I(y_1 = j)}$$, maximize this to get W_1 for y_1

$$\Rightarrow L(y) = \prod_{i=1}^{n} \left[\prod_{j=1}^{k} \left(\frac{e^{z_{ji}}}{\sum_{l=1}^{k} e^{z_{li}}} \right)^{I(y_i = j)} \right]$$