For softmax layer

$$(z_1, \cdots, z_K) \Rightarrow (a_1, \cdots a_K) = \left(\frac{e^{z_1}}{\sum_k e^{z_k}}, \cdots \frac{e^{z_K}}{\sum_k e^{z_k}} \right)$$

where $(z_1, \cdots z_K)$ are input value to k nodes in final layer

①. for binary outcome without hidden layer, $k=2$

The probability of ith input been put in to 1st category is

$$P_{1i} = \frac{e^{z_{1i}}}{e^{z_{1i}} + e^{z_{2i}}} = \frac{1}{1+e^{z_{2i}-z_{1i}}} = \frac{1}{1+e^{-(z_{1i}-z_{2i})}}$$

$P_{2i} = 1 - P_{1i}$, which is exactly logistic regression

②. for categorical outcome ($k>2$)

consider
$$\ln P(Y_i=1) = z_{1i} - \ln(m)$$
$$\ln P(Y_i=2) = z_{2i} - \ln(m)$$
$$\vdots$$
$$\ln P(Y_i=k) = z_{ki} - \ln(m)$$

$$\Rightarrow \quad P(Y_i=1) = e^{z_{1i}} \cdot \frac{1}{m}$$
$$\vdots$$
$$P(Y_i=k) = e^{z_{ki}} \cdot \frac{1}{m}$$

$$\because \sum_{i=1}^{k} P(Y_i=i) = 1 \Rightarrow \frac{1}{m} \sum_{i=1}^{k} e^{z_{ki}} = 1 \Rightarrow m = \frac{1}{\sum_{i=1}^{k} e^{z_{ki}}}$$

$$\Rightarrow \quad P(Y_i=n) = \frac{e^{z_n}}{\sum_{i=1}^{k} e^{z_k}}$$

, which is softmax without any hidden layer