



w_{0j} : biases, don't forget!

d : data dimension

C : number of classes

$$\mathbf{x} \in \mathbb{R}^{d+1}$$

$$\mathbf{W} \in \mathbb{R}^{(d+1) \times C}$$

$$z_i = \mathbf{w}_i^T \mathbf{x}$$

$$\mathbf{z} = \mathbf{W}^T \mathbf{x} \in \mathbb{R}^C$$

$$\mathbf{a} = \text{softmax}(\mathbf{z}) \in \mathbb{R}^C$$

$$a_i > 0, \quad \sum_{i=1}^C a_i = 1$$

short form \longrightarrow

