



$w_{i0}$ : biases, don't forget!

$d$ : data dimension

$C$ : number of classes

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

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

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

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

$z_i = \text{sigmoid}(v_i) \in \mathbb{R}$

$0 < z_i < 1$