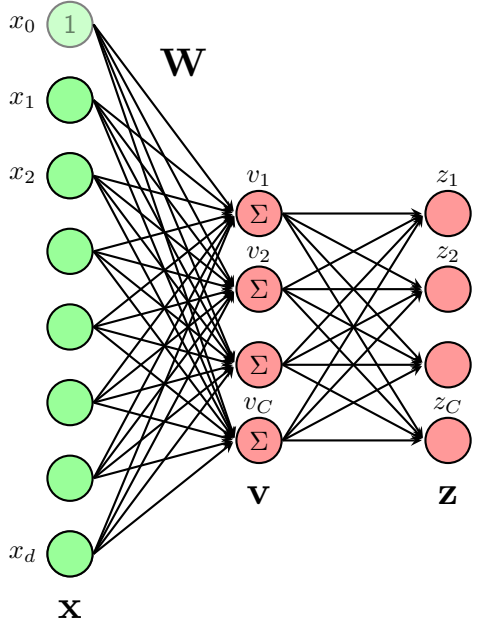

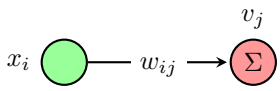

w_{i0}: biases, don't forget!

d: data dimension

C: number of classes

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

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

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

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

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

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

short form

$\mathbf{z} = \text{softmax}(\mathbf{W}^T \mathbf{x})$