Hidden layers $\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(L-1)}$

$\mathbf{W}^{(1)}, \mathbf{b}^{(1)}$

$\mathbf{W}^{(2)}, \mathbf{b}^{(2)}$

$\mathbf{W}^{(L-1)}, \mathbf{b}^{(L-1)}$

$\mathbf{W}^{(L)}, \mathbf{b}^{(L)}$

$\bar{\mathbf{z}}$(mean-pooled MFCC vector)

$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\bar{\mathbf{z}} + \mathbf{b}^{(1)})$

$\mathbf{h}^{(2)} = \sigma(\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)})$

$\mathbf{h}^{(L-1)} = \sigma(\mathbf{W}^{(L-1)}\mathbf{h}^{(L-2)} + \mathbf{b}^{(L-1)})$

$\mathbf{o} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}$

$\mathbf{p} = \text{softmax}(\mathbf{o})$ $p_c = \dfrac{\exp(o_c)}{\sum_{c'} \exp(o_{c'})}$

ReLU

Dropout

ReLU

Dropout

ReLU

Dropout

$\mathcal{L} = -\displaystyle\sum_{c=1}^{C} y_c \log p_c + \lambda \|\Theta\|_2^2$

Legend / Symbol mapping

$\bar{\mathbf{z}}$: *input (mean-pooled MFCC)*

$\mathbf{h}^{(\ell)}$: *hidden vector at layer ℓ; $\sigma(\cdot)$ is elementwise nonlinearity (ReLU)*

$\mathbf{o}$: *final linear outputs; $\mathbf{p} = \text{softmax}(\mathbf{o})$ are class probabilities*

Dropout: applied between hidden layers; ℓ_2 weight decay appears in loss