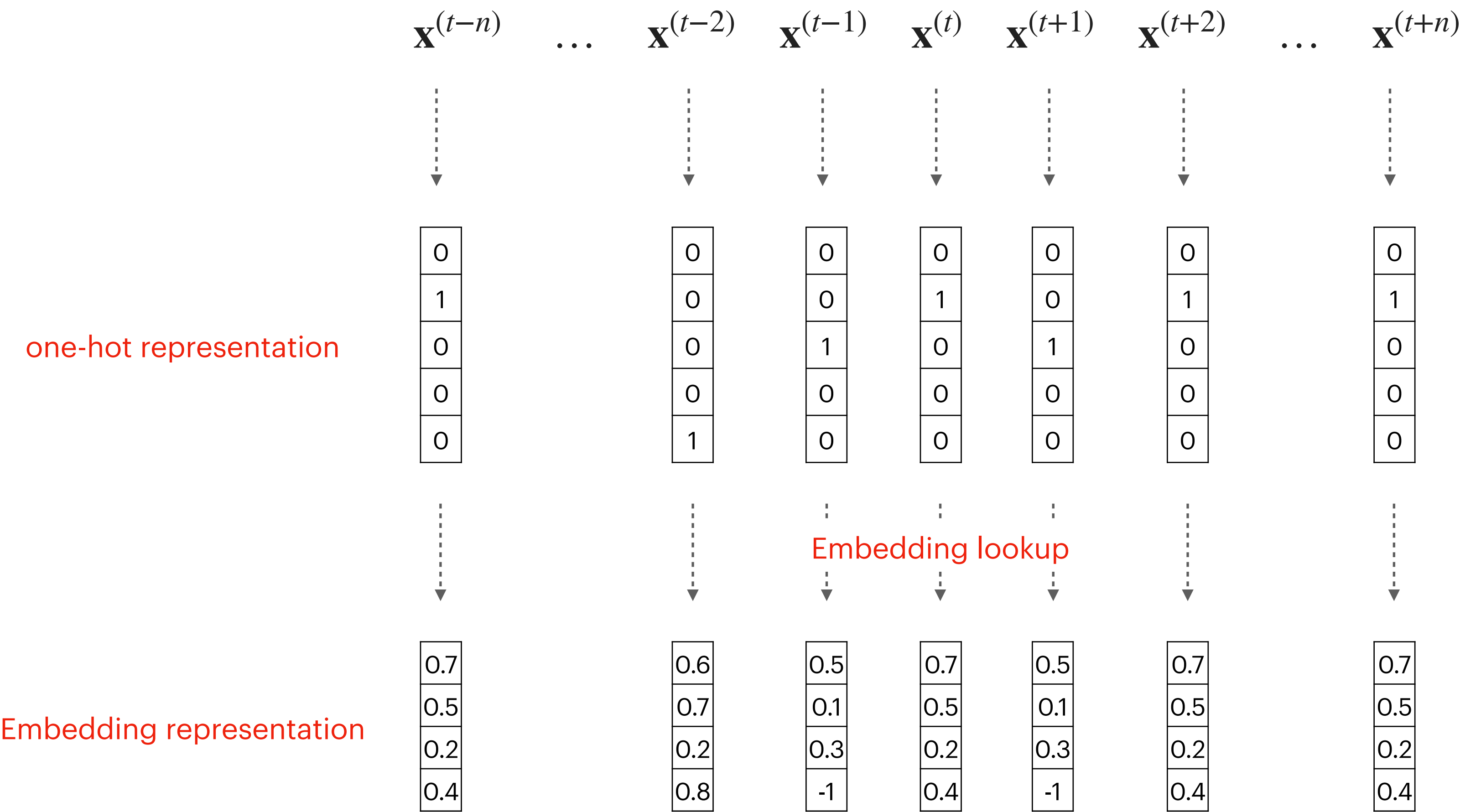
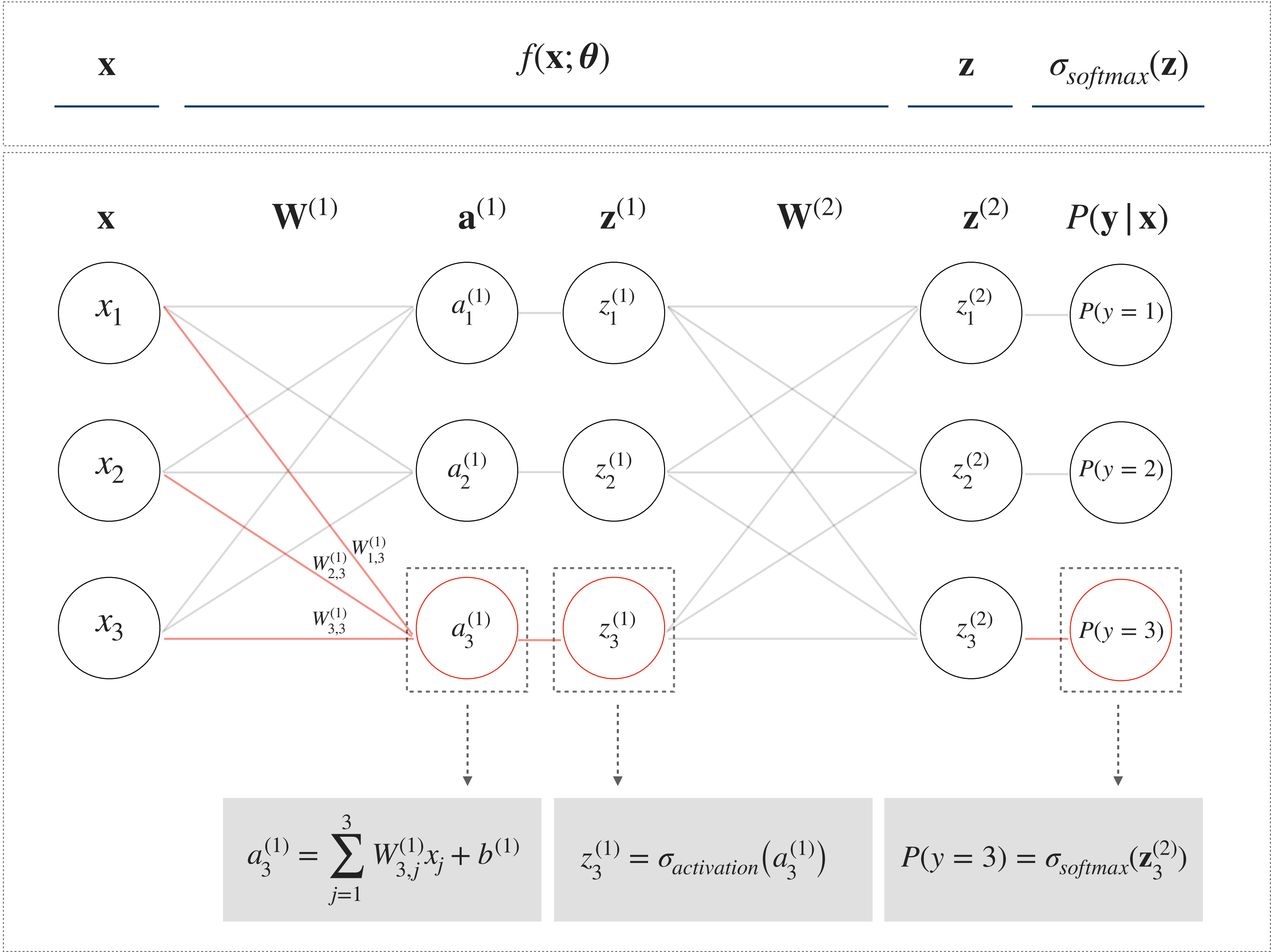


Neural language modeling: feature representation



Feed forward neural network



Common activation functions

$\sigma_{activation}$

$$\sigma_{sigmoid}(z_k) = \frac{1}{1 + e^{-z_k}}$$
$$ReLU(z_k) = \max(0, z_k)$$
$$LReLU(z_k) = \max(\alpha z_k, z_k) \quad (0 < \alpha < 1)$$
$$\tanh(z_k) = \frac{2}{1 + e^{-2z_k}} - 1$$

Softmax output

$$\sigma_{softmax}(z_k) = \frac{e^{z_k}}{\sum_{k'=1}^K e^{z_{k'}}$$