Backpropagation with softmax output + logistic loss fn

Softmax output is $\underbrace{z_j(t) = \frac{e^{t_j}}{\sum_{i=1}^{k} e^{t_i}}}_{\in (0,1) \quad \sum_j z_j = 1}$ $\qquad \frac{dz_j}{dt_j} = z_j(1-z_j) \qquad \underbrace{\frac{dz_j}{dt_i} = -z_i z_j}_{i \neq j}$

$t_i = W_i \cdot h$

W

$$\nabla_{W_i} L = \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} \nabla_{W_i} z_j$$

$$= \left(\frac{\partial L}{\partial z_i} - \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} z_j\right) z_i h$$

$$+ 2\lambda W_i$$

$$z_j = \frac{e^{W_j \cdot h}}{\sum_{i=1}^{k} e^{W_i \cdot h}}$$

z

$$\frac{\partial L}{\partial z_j} = \frac{1-y_j}{1-z_j} - \frac{y_j}{z_j}$$

$$- \sum_{i=1}^{k} y_i \ln z_i + (1-y_i) \ln(1-z_i)$$

$$+ \lambda |w|^2$$

L

optional L_2-regularization

[w is vector containing all the weights in matrices V & W.]

h

$$\frac{\partial L}{\partial h_i} = \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} \frac{\partial z_j}{\partial h_i}$$

$$= \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} \left(W_{ji} - \sum_{\ell=1}^{k} W_{\ell i} z_\ell\right) z_j$$

Derivatives of inputs to hidden units h are computed same way as previously.