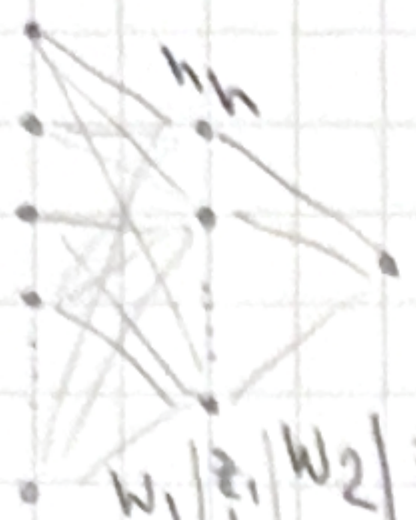
Back propagation

Binary Classification

$\hat{y}_i = a_2$



x_i | W_1, b_1 | $z_1 \downarrow a_1$ | W_2, b_2 | $z_2 \downarrow \hat{y} = a_2$

$z_1 = W_1 x_i + b_1$
$a_1 = \tanh(z_1) = \dfrac{e^{z_1} - e^{-z_1}}{e^{z_1} + e^{-z_1}}$

$z_2 = W_2 a_1 + b_2$
$a_2 = \hat{y} = \text{sigmoid}(z_2) = \dfrac{1}{1+e^{-z_2}}$

loss function: $J = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)\right]$

$\frac{\partial J}{\partial z_2} \to \frac{\partial J}{\partial W_2}, \frac{\partial J}{\partial b_2}$

$\to \frac{\partial J}{\partial z_1} \to \frac{\partial J}{\partial W_1}, \frac{\partial J}{\partial b_1}$

h_h: neurons out of first layer

h_o: output neuron

we need: $W_1 = W_1 - \eta \frac{\partial J}{\partial W_1}$
$b_1 = b_1 - \eta \frac{\partial J}{\partial b_1}$
$W_2 = W_2 - \eta \frac{\partial J}{\partial W_2}$
$b_2 = b_2 - \eta \frac{\partial J}{\partial b_2}$

Steps

1. $\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial a_2} \cdot \frac{\partial a_2}{\partial z_2}$
2. $\frac{\partial J}{\partial W_2} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial W_2}$
3. $\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2}$
4. $\frac{\partial J}{\partial z_1} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1}$
5. $\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial z_1} \cdot \frac{\partial z_1}{\partial W_1}$
6. $\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial z_1} \cdot \frac{\partial z_1}{\partial b_1}$

components:

prev gradient — activation gradient

$\frac{\partial J}{\partial z_1} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial J}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1}$

weight matrix

1. $\frac{\partial J}{\partial z_2}$

$\frac{\partial J}{\partial a_2} = -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial}{\partial a_2}[y_i \ln(a_2)] + \frac{\partial}{\partial a_2}[(1-y_i)\ln(1-a_2)]\right]$
$= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{a_2} - \frac{1-y_i}{1-a_2}\right]$
$= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i(1-a_2)}{a_2(1-a_2)} - \frac{a_2(1-y_i)}{a_2(1-a_2)}\right]$
$= -\frac{1}{n}\sum_{i=1}^{n}\frac{y_i - a_2}{a_2(1-a_2)}$

$\frac{\partial J}{\partial a_2} = \frac{1}{n}\sum_{i=1}^{n}\frac{a_2 - y_i}{a_2(1-a_2)} = \frac{A_2 - Y}{A_2(1-A_2)}$

$\frac{\partial a_2}{\partial z_2} = \frac{\partial}{\partial z_2}\left[\frac{1}{1+e^{-z_2}}\right]$ sigmoid gradient

$\frac{\partial a_2}{\partial z_2} = a_2(1-a_2) = A_2(1-A_2)$

$\frac{\partial J}{\partial z_2} = \left(\frac{A_2 - Y}{A_2(1-A_2)}\right)(A_2(1-A_2))$

$\boxed{\frac{\partial J}{\partial z_2} = A_2 - Y}$

$R^{h_o \times 1}$

2. $\frac{\partial J}{\partial W_2}$

$\frac{\partial z_2}{\partial W_2} = \frac{\partial}{\partial W_2}[W_2 a_1 + b_2]$
$= a_1$

$\boxed{\frac{\partial J}{\partial W_2} = (A_2 - Y) \cdot a_1^T}$ add transpose for dimension

$R^{h_o \times h_h}$ | $R^{h_o \times 1}$ | $R^{1 \times h_h}$

3. $\frac{\partial J}{\partial b_2}$

$\frac{\partial z_2}{\partial b_2} = \frac{\partial}{\partial b_2}[W_2 a_1 + b_2]$
$= 1$

$\frac{\partial J}{\partial b_2} = (A_2 - Y) \cdot 1$

$\boxed{\frac{\partial J}{\partial b_2} = (A_2 - Y)} = \frac{\partial J}{\partial z_2}$

$R^{h_o \times 1}$

similarity

4. $\frac{\partial J}{\partial z_1}$

$\frac{\partial z_2}{\partial a_1} = \frac{\partial}{\partial a_1}[W_2 a_1 + b_2]$
$= W_2$

$\frac{\partial a_1}{\partial z_1} = \frac{\partial}{\partial z_1}\left[\frac{e^{z_1} - e^{-z_1}}{e^{z_1} + e^{-z_1}}\right]$
$\vdots$
$\frac{\partial a_1}{\partial z_1} = 1 - a_1^2$ tanh gradient

$\frac{\partial J}{\partial z_1} = (A_2 - Y) W_2 \cdot (1 - a_1^2)$

$\boxed{\frac{\partial J}{\partial z_1} = W_2^T (a_2 - Y) * (1 - a_1^2)}$

$\in R^{h_h \times 1}$ | $R^{h_h \times h_o}$ | $R^{h_o \times 1}$ | $R^{h_h \times 1}$

5. $\frac{\partial J}{\partial W_1}$

$\frac{\partial z_1}{\partial W_1} = \frac{\partial}{\partial W_1}[W_1 x_i + b_1]$
$= x_i$

$\boxed{\frac{\partial J}{\partial W_1} = \frac{\partial J}{\partial z_1} \cdot x_i^T}$

$R^{h_h \times d}$ | $R^{h_h \times 1}$ | $R^{1 \times d}$

similarity

6. $\frac{\partial J}{\partial b_1}$

$\frac{\partial z_1}{\partial b_1} = \frac{\partial}{\partial b_1}[W_1 x_i + b_1]$
$= 1$

$\boxed{\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial z_1}}$

$R^{h_h \times 1}$