Exercise: Trace Backpropagation

Consider a neural net with one hidden layer, two inputs *a* and *b*, one hidden unit *c*, and one output unit *d*. The activation function is the sigmoid for each node. This network has five weights (w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0}), where *wx0* represents the bias or threshold weight for unit *x*. Initialize these weights to the values (.1,.1,.1,.1,.1), then give their values after each of the first two training iterations of Backpropagation algorithm. Assuming learning rate (step size) of 0.3, stochastic (incremental) gradient descent (without momentum), and the following training examples:

	a	b	d
x_1	1	0	1
x_2	0	1	0

Fill in the following tables. You can expand these to include more information (e.g. derivatives of activation functions) if you like.

Iteration	a_c	δ_c	a_d	δ_d
x_1				
x_2				

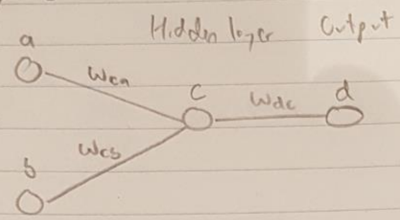
Iteration	w_{c0}	w_{ca}	w_{cb}	w_{dc}	w_{d0}
x_1					
x_2					

CMPT 880, Exercise 3

Matt Kierans

Network Architecture:

Inputs



Using notation from Bishop: Pattern Recognition and Machine Learning

Output of the final layer, z_d, uses linear activation function,

$\Longrightarrow z_d = h(a_d)$, where $h_d(x) = x$ is the activation function and a_d is the weighted sum of outputs from the previous layer;

$$a_d = w_{dc} z_c + w_{d0}$$

The output of the hidden layer unit c, z_c is given by

$\Longrightarrow z_c = h_c(a_c)$, where $h_c(x) = \frac{1}{1+e^{-x}}$, sigmoid activation

and a_c is weight sum of weights from input layer;

$$a_c = w_{ca} x_a + w_{cb} x_b + w_{c0}$$

Using squared error: $E = \frac{1}{2}(z_d - t)^2$

Define $\frac{\partial E}{\partial a_i} \equiv \delta_i$, then $\frac{\partial E}{\partial a_d} = \delta_d = (z_d - t)\frac{\partial z}{\partial a} = z_d - t$

$$\Longrightarrow \frac{\partial E}{\partial w_{dc}} = \frac{\partial E}{\partial a_d}\frac{\partial a_d}{\partial w_{dc}} = (z_d - t) z_c = \delta_d z_c$$

$$\Rightarrow \frac{\partial E}{\partial w_{d0}} = \frac{\partial E}{\partial a_d}\frac{\partial a_d}{\partial w_{d0}} = (z_d - t)\cdot 1 = z_d - t$$

$$\frac{\partial E}{\partial a_c} = \frac{\partial E}{\partial a_d}\frac{\partial a_d}{\partial a_c} = \delta_d \cdot \frac{\partial a_d}{\partial a_c}$$

where $a_d = w_{dc}\, h_c(a_c) + w_{d0} \Rightarrow \frac{\partial a_d}{\partial a_c} = w_{dc}\, h_c'(a_c)$

and $h_c(a_c) = \sigma(a_c) \Rightarrow h_c'(a_c) = \sigma(a_c)(1-\sigma(a_c))$

$$\Rightarrow \delta_c = \frac{\partial E}{\partial a_c} = \delta_d \cdot w_{dc}\cdot \sigma(a_c)(1-\sigma(a_c)) = (z_d - t)\, w_{dc}\cdot \sigma(a_c)(1-\sigma(a_c))$$
$$= \delta_d\, w_{dc}\cdot z_c(1-z_c) \quad !$$

which allows us to calculate the rest of the partial derivatives w.r.t the other weights.

ie:

$$\frac{\partial E}{\partial w_{ca}} = \frac{\partial E}{\partial a_c}\frac{\partial a_c}{\partial w_{ca}} = \delta_c \cdot x_a \quad \text{and} \quad \frac{\partial E}{\partial w_{cb}} = \delta_c \cdot x_b$$

and $\frac{\partial E}{\partial w_{c0}} = \delta_c$

Now to evaluate the SGD in the question.
Begin by forward propogating to find a_c, a_d, z_c and z_d

Initialize all weights to be 0.1, $x_1 = \{a=1, b=0, d=1\}$ and $x_2 = \{a=0, b=1, d=0\}$

First iteration over samples (x_1)

x_1: $a_c = w_{ca} x_a + w_{cb} x_b + w_{c0}$

$= 0.1 \cdot 1 + 0.1 \cdot 0 + 0.1 = 0.2$

$z_c = h_c(a_c) = \frac{1}{1+e^{-0.2}} = 0.5498$

$a_d = w_{dc} \cdot z_c + w_{d0} = 0.1 \cdot 0.5498 + 0.1$

$= 0.155$

$z_d = a_d = 0.155$

$\Rightarrow$ $\delta_d = z_d - x_d = z_d - 1 = 0.155 - 1 = -0.845$

$\delta_c = \delta_d \cdot w_{dc} \cdot z_c(1 - z_c) = -0.845 \cdot 0.1 \cdot (0.5498)(1-0.5498)$

$= -0.021$

Errors for weights:

$\frac{\partial E}{\partial w_{dc}} = \delta_d \cdot z_c = -0.4646$

$\frac{\partial E}{\partial w_{d0}} = \delta_d = -0.845$ $\qquad$ $\frac{\partial E}{\partial w_{cb}} = \delta_c \cdot x_b = 0$

$\frac{\partial E}{\partial w_{ca}} = \delta_c \cdot x_a = -0.021$ $\qquad$ $\frac{\partial E}{\partial w_{c0}} = \delta_c = -0.021$

New weights after SGD: $w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E(w^{(\tau)})$

$w_{dc} = 0.1 - 0.3(-0.4646) = 0.239$; $w_{ca} = 0.1 - 0.3 \cdot (-0.021) = 0.106$

$w_{d0} = 0.1 - 0.3(-0.845) = 0.35$; $w_{cb} = 0.1$

$w_{c0} = 0.1 - 0.3(-0.021)$ Hilroy

$= 0.106$

First iteration over Samples (X_2)

X_2: $a_c = w_{ca} X_a + w_{cb} X_b + w_{c0}$
$= 0.106 \cdot 0 + 0.1 \cdot 1 + 0.106 = 0.206$

$z_c = \sigma(a_c) = 0.55$

$a_d = w_{dc} \cdot z_c + w_{d0} = 0.239 \cdot 0.55 + 0.35 = 0.48$

$z_d = a_d = 0.48$

$\Rightarrow \delta_d = z_d - x_a = 0.48 - 0 = 0.48$

$\delta_c = \delta_d \cdot w_{dc} \cdot z_c(1 - z_c) = 0.48 \cdot 0.239 \cdot 0.55(1 - 0.55)$
$= 0.028$

Errors for weights:

$\frac{\partial E}{\partial w_{dc}} = \delta_d \cdot z_c = 0.48 \cdot 0.55 = 0.264$

$\frac{\partial E}{\partial w_{d0}} = \delta_d = 0.48$ $\qquad$ $\frac{\partial E}{\partial w_{cb}} = \delta_c \cdot X_b = 0.028 \cdot 1 = 0.028$

$\frac{\partial E}{\partial w_{ca}} = \delta_c \cdot X_a = 0$ $\qquad$ $\frac{\partial E}{\partial w_{c0}} = \delta_c = 0.028$

New weights after SGD:

$w_{dc} = 0.239 - 0.3(0.264) = 0.16$; $w_{ca} = 0.106$
$w_{d0} = 0.35 - 0.3(0.48) = 0.206$; $w_{cb} = 0.1 - 0.3(0.028)$
$w_{c0} = 0.106 - 0.3(0.028) = 0.098$ $\qquad$ $= 0.092$

Hilroy

Second Iteration over weights (x_1)

x_1: $a_c = w_{ca} x_a + w_{cb} x_b + w_{co}$

$= 0.106 \cdot 1 + 0 + 0.098 = 0.204$

$z_c = \sigma(a_c) = 0.55$

$a_d = w_{dc} z_c + w_{do} = 0.16 \cdot 0.55 + 0.206$

$= 0.294$

$z_d = a_d = 0.294$

$\Rightarrow \delta_d = z_d - x_d = 0.294 - 1 = -0.706$

$\delta_c = \delta_d \cdot w_{dc} \cdot z_c(1 - z_c) = -0.706 \cdot 0.16 \cdot 0.55(1 - 0.55)$

$= -0.028$

Errors for weights:

$\frac{\partial E}{\partial w_{dc}} = \delta_d \cdot z_c = -0.706 \cdot 0.55 = -0.39$

$\frac{\partial E}{\partial w_{do}} = \delta_d = -0.706$ $\frac{\partial E}{\partial w_{cb}} = \delta_c \cdot x_b = 0$

$\frac{\partial E}{\partial w_{ca}} = \delta_c \cdot x_a = -0.028$ $\frac{\partial E}{\partial w_{co}} = \delta_c = -0.028$

New weights after SGD:

$w_{dc} = 0.16 - 0.3(-0.39) = 0.28$; $w_{ca} = 0.106 - 0.3(-0.028) = 0.114$

$w_{do} = 0.206 - 0.3(-0.706) = 0.42$; $w_{cb} = 0.092$

$w_{co} = 0.098 - 0.3(-0.028)$

$= 0.106$

Hilroy

Second Iteration over weights (x_2)

X_2: $a_c = w_{ca} \cdot x_a + w_{cb} \cdot x_b + w_{co}$
$= 0.114 \cdot 0 + 0.092 \cdot 1 + 0.106 = 0.198$

$z_c = \sigma(a_c) = 0.55$

$a_d = w_{dc} \cdot z_c + w_{do} = 0.28 \cdot 0.55 + 0.42 = 0.574$

$z_d = a_d = 0.574$

$\Rightarrow \delta_d = z_d - x_d = 0.574 - 0 = 0.574$

$\delta_c = \delta_d \cdot w_{dc} \cdot z_c(1-z_c) = 0.574 \cdot 0.28 \cdot 0.55(1-0.55)$
$= 0.04$

Errors for weights

$$\frac{\partial E}{\partial w_{dc}} = \delta_d \cdot z_c = 0.574 \cdot 0.55 = 0.316$$

$$\frac{\partial E}{\partial w_{do}} = \delta_d = 0.574 \qquad \frac{\partial E}{\partial w_{cb}} = \delta_c \cdot x_b = 0.04 \cdot 1 = 0.04$$

$$\frac{\partial E}{\partial w_{ca}} = \delta_c \cdot x_a = 0 \qquad \frac{\partial E}{\partial w_{co}} = \delta_c = 0.04$$

New weights after SGD:

$w_{dc} = 0.28 - 0.3(0.316) = 0.19$; $w_{ca} = 0.114$
$w_{do} = 0.42 - 0.3(0.574) = 0.25$; $w_{cb} = 0.092 - 0.3(0.04)$
$= 0.08$
$w_{co} = 0.106 - 0.3(0.04)$
$= 0.094$

Hilroy