Part - II

1. Forward pass :-

$h_1 = 0.7 \times -0.3 + 0.5 \times 0.15 + 0.9$

$= 0.765$

$h_2 = 0.7 \times 0.8 + 0.5 \times 0.2 - 0.14 = 0.52$

$\hat{y} = 0.765 \times 0.7 + 0.52 \times 0.25 - 0.1 = 0.5655$

2. MSE $\Rightarrow \frac{1}{2}(y - \hat{y})^2 = \frac{1}{2}(0.5 - 0.5655)^2 = 0.002145$

3. Gradients calculator

$w_1 = w_1 - \alpha\left(\frac{\partial Error}{\partial w_1}\right)$

$\frac{\partial Error}{\partial w_1} = \frac{\partial error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1}$

$\Rightarrow \frac{1}{2}(y - \hat{y})^2 \cdot \frac{\partial (h_1 w_1 + h_2 w_8 + h_3 w_9)}{\partial h_1} \cdot \frac{\partial (i_1 w_1 + i_2 w_3 + i_3 w_5)}{\partial w_1}$

$\Rightarrow 2 \cdot \frac{1}{2}(y - \hat{y}) \cdot \frac{\partial (y - \hat{y})}{\partial \hat{y}} \cdot (w_7 * i_1)$

$\frac{\partial Error}{\partial w_1} = (y - \hat{y})(-1)(w_7 * i_1)$

$\rightarrow w_2 = w_2 - \alpha\left(\frac{\partial Error}{\partial w_2}\right)$

$\rightarrow \frac{\partial Error}{\partial w_2} = \frac{\partial Error}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial h_2} \times \frac{\partial h_2}{\partial w_2}$

$\Rightarrow \frac{1}{2}(y - \hat{y})^2 \cdot \frac{\partial (h_1 w_1 + h_2 w_3 + h_2 w_3)}{\partial h_2} \cdot \frac{(i_1 w_2 + i_2 w_4 + i_3 w_6)}{\partial w_2}$

$\Rightarrow 2 \cdot \frac{1}{2}(y - \hat{y}) \cdot \frac{\partial (y - \hat{y})}{\partial \hat{y}} \cdot w_8 i_2$

classmate

Date

Page

$$w_3 = w_3 - \alpha \left(\frac{\partial Error}{\partial w_3} \right)$$

$$\frac{\partial Error}{\partial w_3} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_3}$$

$$= \frac{1}{2} \frac{(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial (h_1 w_7 + h_2 w_8 + h_3 w_9)}{\partial h_1} \cdot \frac{\partial (i_1 w_1 + i_2 w_3 + i_3 w_5)}{\partial w_3}$$

$$\Rightarrow 2 \cdot \frac{1}{2} (y-\hat{y}) \frac{\partial (y-\hat{y})}{\partial \hat{y}} \cdot w_7 i_2$$

$$\Rightarrow (y-\hat{y}) \cdot (-1) (w_7 i_2)$$

$$w_4 = w_4 - \alpha \left(\frac{\partial Error}{\partial w_4} \right)$$

$$\frac{\partial Error}{\partial w_4} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_2} \cdot \frac{\partial h_2}{\partial w_4}$$

$$\frac{\partial Error}{\partial \hat{y}} \cdot \frac{1}{2} \frac{(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial (h_1 w_7 + h_2 w_8 + h_3 w_9)}{\partial h_2} \cdot \frac{\partial (i_1 w_2 + i_2 w_4 + i_3 w_6)}{\partial w_4}$$

$$\Rightarrow 2 \cdot \frac{1}{2} (y-\hat{y}) \cdot \frac{\partial (y-\hat{y})}{\partial \hat{y}} \cdot w_8 i_2$$

$$\Rightarrow (y-\hat{y})(-1) w_8 i_2 = w_4$$

$$w_5 = w_5 - \alpha \left(\frac{\partial Error}{\partial w_5} \right)$$

$$\frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_5}$$

$$\frac{\partial Error}{\partial \hat{y}\,\partial w_5} = \frac{1/2(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(h_1w_7 + h_2w_8 + h_3w_9)}{\partial h_1} \cdot \frac{\partial(i_1w_1 + i_2w_3 + i_3w_5)}{\partial w_5}$$

$$\Rightarrow (y-\hat{y})(-1)\, w_7\, i_3$$

$$w_6 = w_6 - \alpha\left(\frac{\partial Error}{\partial w_6}\right)$$

$$\frac{\partial Error}{\partial w_6} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_2} \cdot \frac{\partial h_2}{\partial w_6}$$

$$\Rightarrow \frac{1/2(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(h_1w_7 + h_2w_8 + h_3w_9)}{\partial h_2} \cdot \frac{\partial(i_1w_2 + i_2w_4 + i_3w_6)}{\partial w_6}$$

$$\Rightarrow (y-\hat{y})(-1)(w_8\, i_3)$$

$$w_7 = w_7 - \alpha \frac{\partial Error}{\partial w_7}$$

$$\frac{\partial Error}{\partial w_7} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_7}$$

$$\Rightarrow \frac{1/2(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(h_1w_7 + h_2w_8 + h_3w_9)}{\partial w_7}$$

$$\Rightarrow (y-\hat{y})(-1)\, h_1$$

$$w_8 = w_8 - \alpha\left(\frac{\partial Error}{\partial w_8}\right) \Rightarrow \frac{\partial Error}{\partial w_8} = \frac{\partial Error}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_8}$$

$$\Rightarrow \frac{1/2(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(h_1w_7 + h_2w_8 + h_3w_9)}{\partial w_8}$$

$$\Rightarrow (y-\hat{y})(-1)\, h_2$$

$$w_9 = w_9 - \alpha \left(\frac{\partial E_{total}}{\partial w_9}\right)$$

$$\rightarrow \frac{1/2 (y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial (h_1 w_7 + h_2 w_8 + h_3 w_9)}{\partial w_9}$$

$$\rightarrow 2 \cdot 1/2 (y-\hat{y})(y-\hat{y}) \cdot \frac{h_3}{\partial \hat{y}}$$

$$\rightarrow (y-\hat{y})(-1)(h_3^1)$$

$$w_1 = w_1 - \alpha (-i_1 w_7 (y-\hat{y}))$$
$$w_2 = w_2 - \alpha (-i_1 w_8 (y-\hat{y}))$$
$$w_3 = w_3 - \alpha (-i_2 w_7 (y-\hat{y}))$$
$$w_4 = w_4 - \alpha (i_2 w_8 (y-\hat{y}))$$
$$w_5 = w_5 - \alpha (i_3 w_7 (y-\hat{y}))$$
$$w_6 = w_6 - \alpha (i_3 w_8 (y-\hat{y}))$$
$$w_7 = w_7 - \alpha (-h_1 (y-\hat{y}))$$
$$w_8 = w_8 - \alpha (-h_2 (y-\hat{y}))$$
$$w_9 = w_9 - \alpha (-h_3 (y-\hat{y}))$$

4. Update weights and bias.

$w_1 = -0.3 - 0.03 (0.7 \times 0.7 \times 0.0655) = -0.3005$

$w_2 = 0.8 - 0.03 (0.7 \times 0.25 \times 0.0655) = 0.7996$

$w_3 = 0.15 - 0.03 \cdot (0.5 \times 0.7 \times 0.0655) = 0.1493$

$w_4 = 0.2 - 0.03 (0.5 \times 0.25 \times 0.655) = 0.1997$

$w_5 = 0.9 - 0.03 (1 \times 0.7 \times 0.0655) = 0.8981$

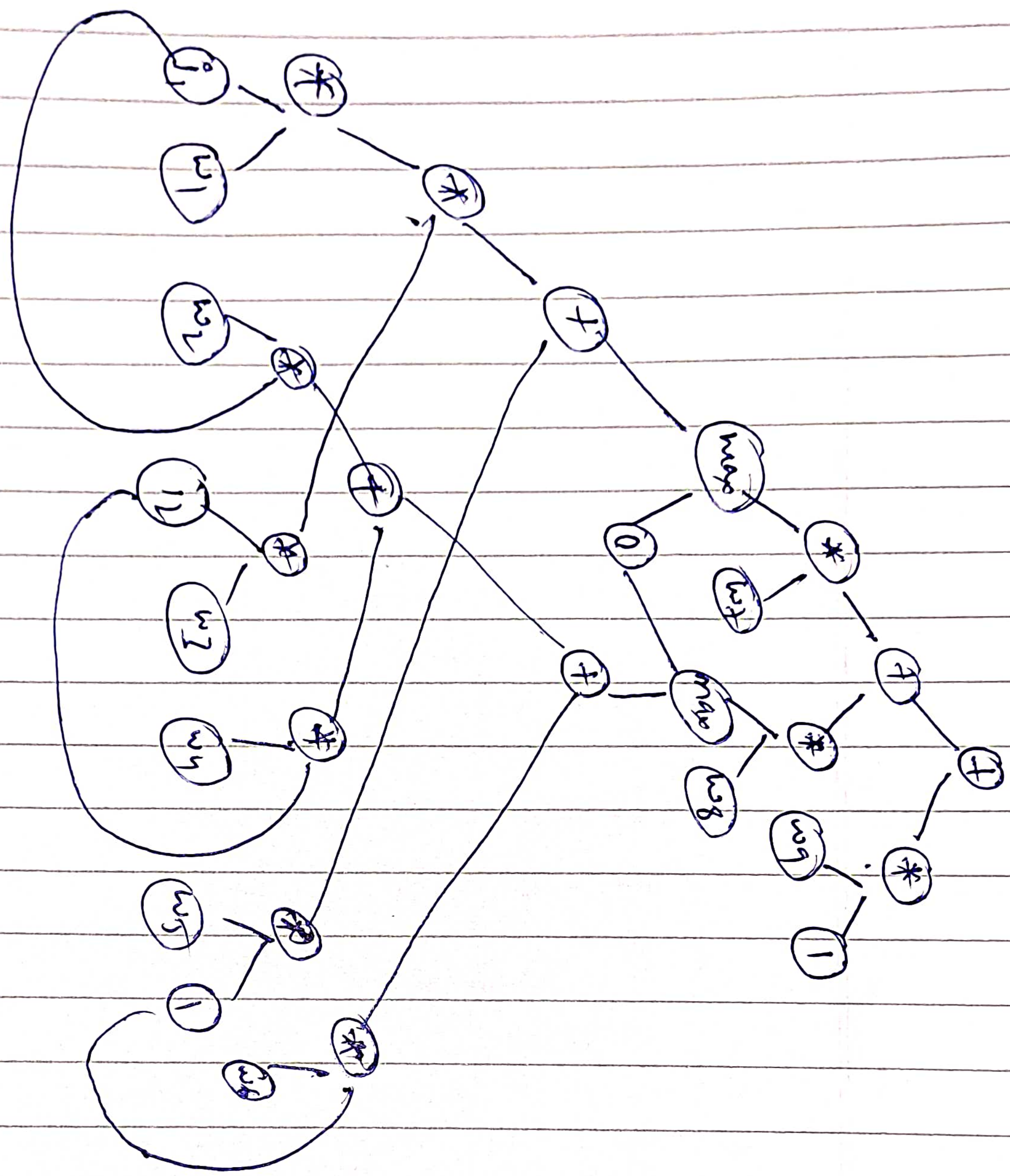
$w_6 = -0.14 - 0.03 (1 \times 0.25 \times 0.0655) = -0.1404$

$w_7 = 0.7 - 0.03 (0.765 \times 0.0655) = 0.6984$

$w_8 = 0.25 - 0.03 (0.5 \times 0.0655) = 0.2489$

$w_9 = -0.1 - 0.03 (1 \times 0.0655) = -0.1019$

5. Computation graph for forward pass:

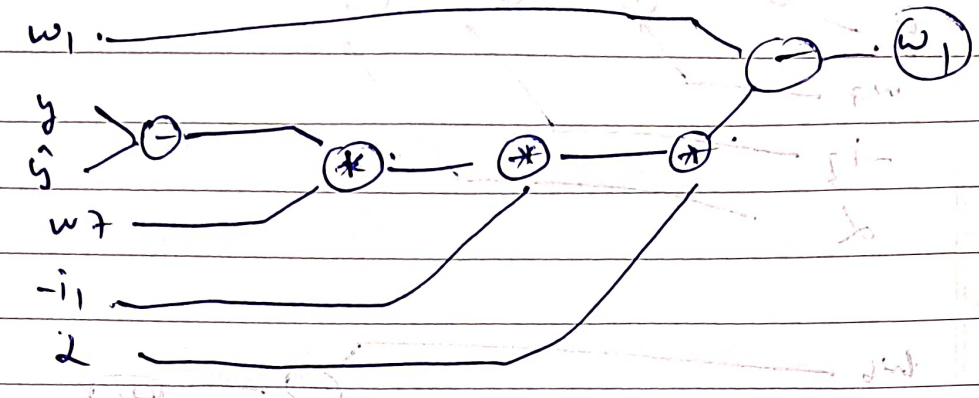


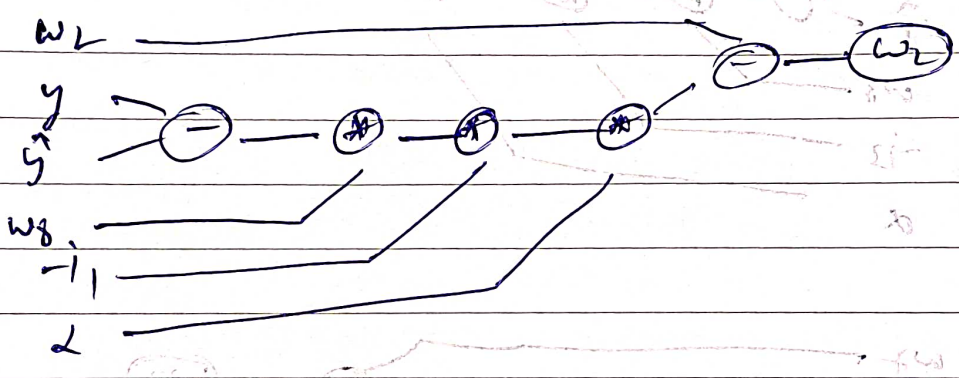
classmate

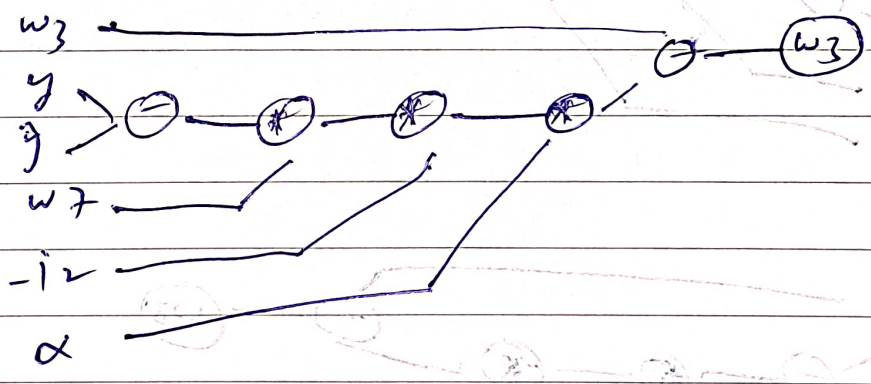
Date _____

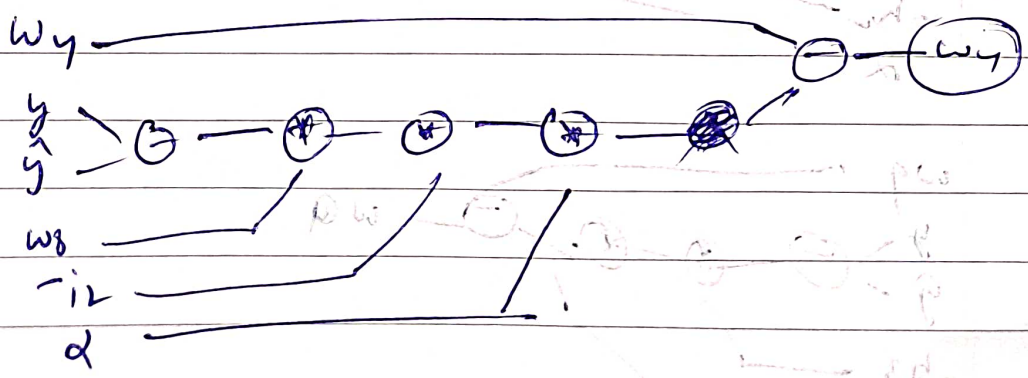
Page _____

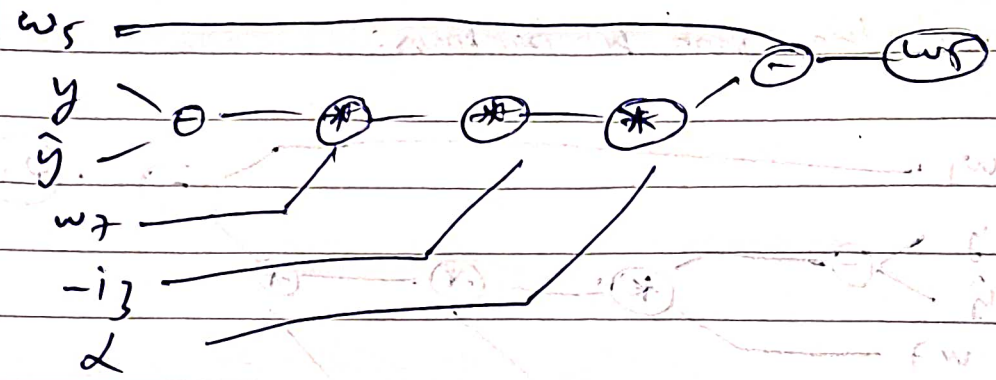
Computation graph for Backpass:

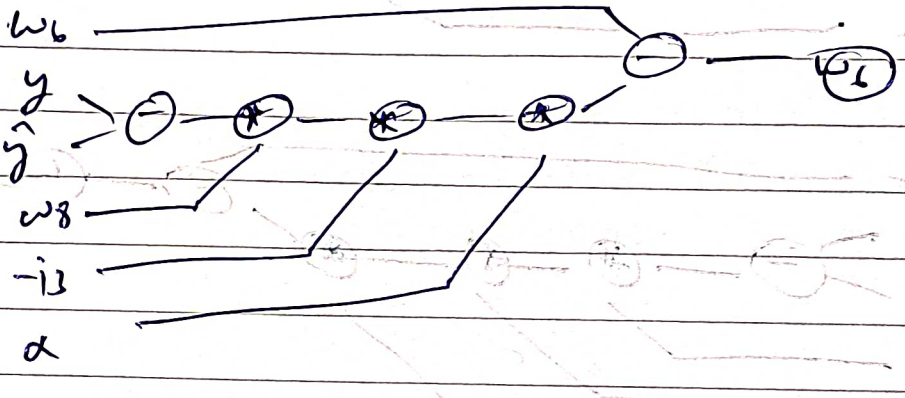


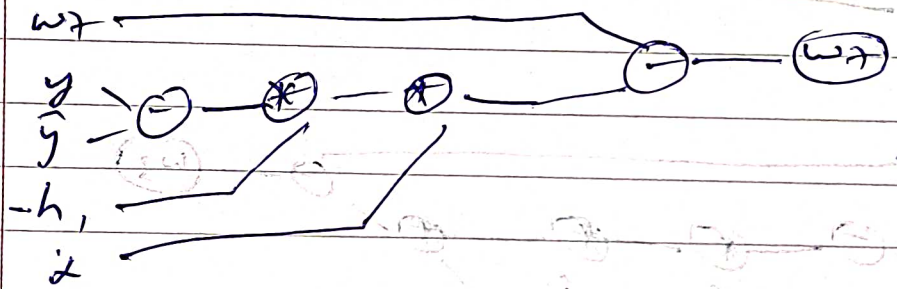


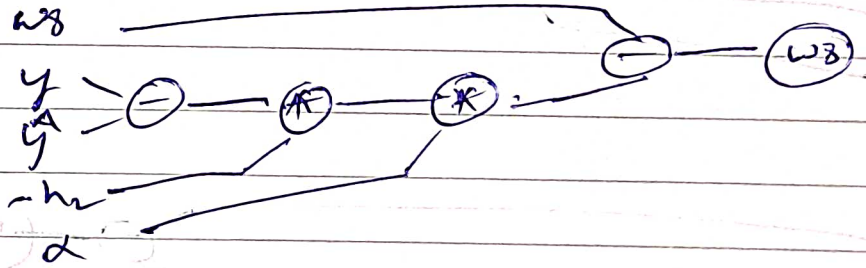


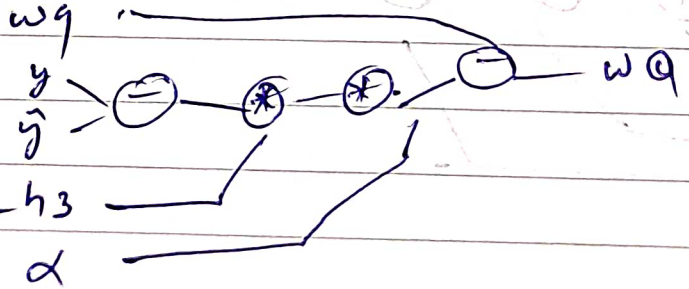












6. $h_1 = 0.7 \times -0.3009 + 0.5 \times 0.1493 + 1 \times 0.8986 = 0.7626$

$h_2 = 0.7 \times 0.7496 + 0.5 \times 0.1997 + 1 \times 0.1404)$

$= 0.5599$

7. $MSE = 1/2 (y - \hat{y})^2$

$= 1/2 (0.5 - 0.5599)^2 = 0.00179$

Mean squared error reduced from 0.00294 to 0.00179 · which explains that the model is improved in performance on training data. manipulating weights based on the performance.

Part - 2

2. Derivative of tanh

$$\tanh(x) = \left(\frac{e^x + e^{-x}}{e^x - e^{-x}}\right)$$

$$\frac{d}{dx}\tanh(x) = \frac{d}{dx}\left(\frac{e^x + e^{-x}}{e^x - e^{-x}}\right) \Rightarrow \frac{d}{dx}\tanh(x) = \frac{(e^x+e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$

$$\Rightarrow \frac{e^{2x} + e^0 + e^0 + e^{-2x} - (e^{2x} - e^0 - e^0 + e^{-2x})}{(e^x + e^{-x})^2}$$

$$\Rightarrow \frac{4}{(e^x + e^{-x})^2} \Rightarrow \frac{2}{e^x + e^{-x}} \times \frac{2}{e^x + e^{-x}}$$

We know that $\frac{2}{e^x + e^{-x}} = \frac{e^x + e^{-x}}{2} = \cosh(x)$

Therefore $\frac{d}{dx}\tanh(x) = \frac{1}{(\cosh x)^2} \Rightarrow (\operatorname{sech} x)^2$

hyperbolic property : $\tanh^2 x + \operatorname{sech}^2 x = 1$

Hence $\Rightarrow \operatorname{sech}^2 x = 1 - \tanh^2 x$

Hence, $\boxed{\frac{d}{dx}(\tanh x) = 1 - \tanh^2 x}$

if we assume $\tanh(x) = f(x)$.

then $\boxed{\frac{d}{dx} f(x) = 1 - f(x)^2}$