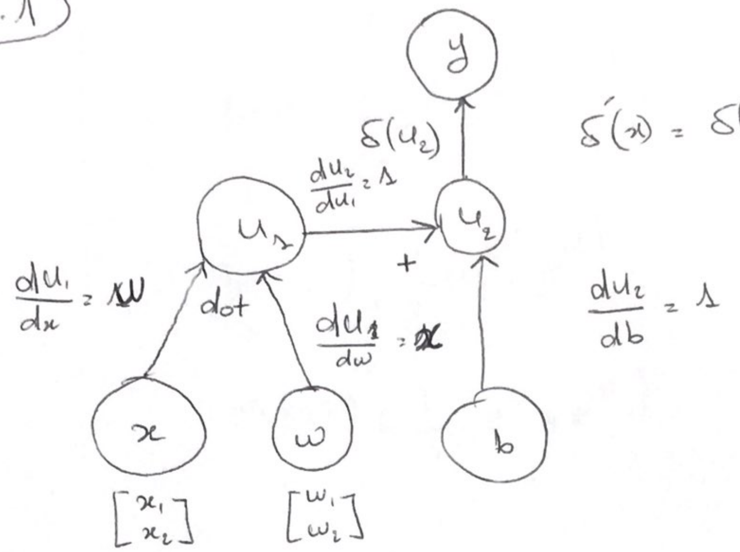
Assignment# 8

Group Members

Muhammad Hamza jamil	2572890	s8mujami@stud.uni-saarland.de
Hacane Hechehouche	2571617	S8hahech@stud.uni-Saarland.de
Shivam sharma	2576656	

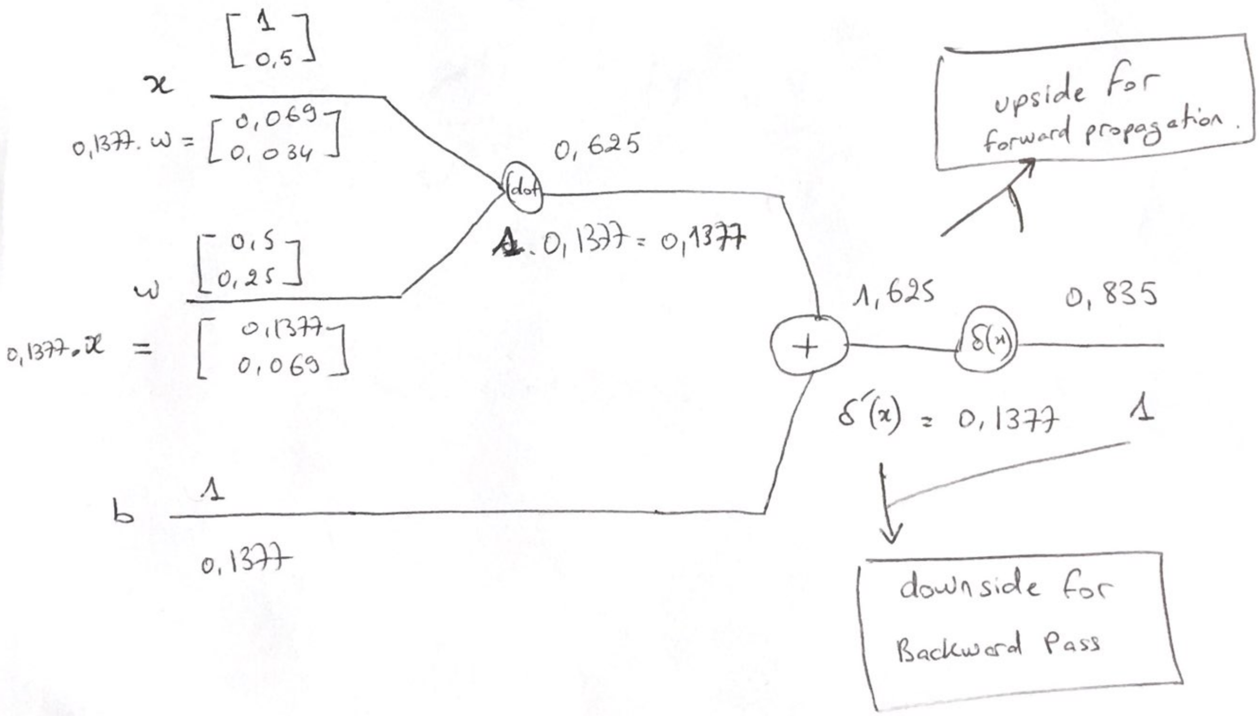
Ex 8.1

a



$\delta'(x) = \delta(x) - \delta^2(x)$

b



Ex 8.2

a) Derivative of tanh

$$\tanh = \frac{\sinh}{\cosh}$$

$$\therefore \sinh(x) = \frac{1}{2}(e^x - e^{-x})$$

$$\therefore \cosh(x) = \frac{1}{2}(e^x + e^{-x})$$

$$f'(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

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

$$= \cancel{\frac{(e^x + e^{-x})^2}{(e^x + e^{-x})^2}} - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$

$$= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2$$

$$\boxed{f'(x) = 1 - \tanh^2(x)}$$

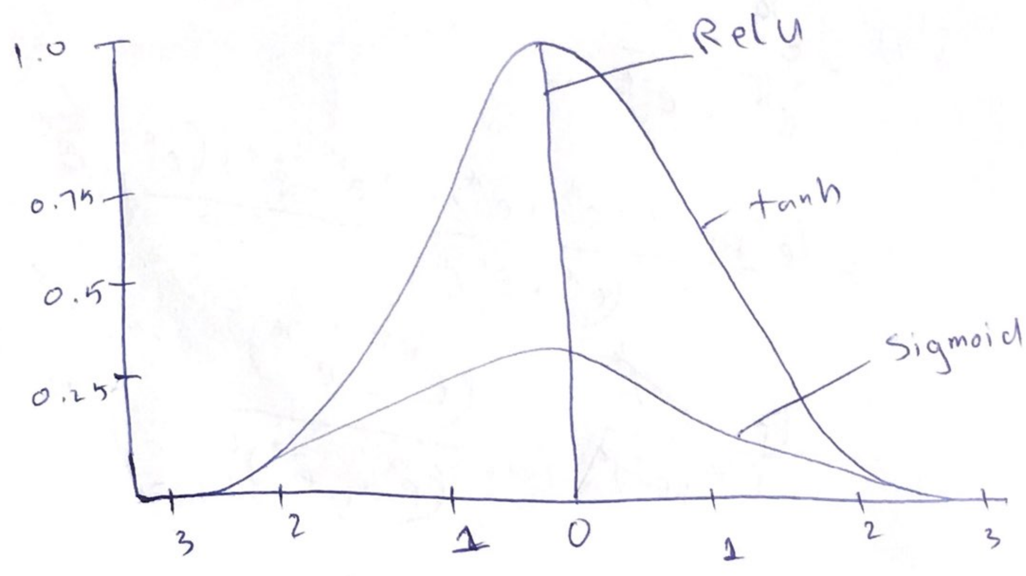
ReLU:

$$f(x) = max(0, x)$$

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

$$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$$

b)



Sigmoid function

1) It used to predict Probabilities as an output

2) Activation function that transform linear input to non linear output

3) It derivative is not monotonic unlikis its function

ReLU:

1) It avoids and Rectifies Vanishing gradient Problem

2) Derivative in Unit Step function at $x = 0$ its ignores a problem when gradient is not strictly definen

~~Hence Sig~~

Hence ReLU is better than Sigmoid.

Exercise 8.3

$$f(x, y) = x^3 + x y^2$$

$$g\cancel{(x)} = 2x + y^2 = 2$$

$$L(x, y, \lambda) = x^3 + x y^2 - \lambda(2x + y^2 - 2)$$

$$\nabla L(x, y, \lambda) = 0$$

$$\begin{bmatrix} \frac{\partial L(x,y,\lambda)}{\partial x} \\ \frac{\partial L(x,y,\lambda)}{\partial y} \\ \frac{\partial L(x,y,\lambda)}{\partial \lambda} \end{bmatrix} = 0$$

$$\frac{\partial L(x,y,\lambda)}{\partial x} = 3x^2 + y^2 - \lambda(2)$$

$$\frac{\partial L(x,y,\lambda)}{\partial y} = 2xy - \lambda(2y) \Rightarrow 2xy - 2\lambda y$$

$$\frac{\partial L(x,y,\lambda)}{\partial \lambda} = -(2x + y^2 - 2)$$

$$3x^2 + y^2 - 2\lambda = 0 \quad \text{——} \quad (1)$$

$$2xy - 2\lambda y = 0 \quad \text{——} \quad (2)$$

$$-2x - y^2 + 2 = 0 \quad \text{——} \quad (3)$$

eq 2

$$2xy - 2\lambda y = 0$$

$$x = \frac{2\lambda y}{2y}$$

$$\boxed{x = \lambda}$$

put x in eq 3

$$-2\lambda - y^2 + 2 = 0$$

$$y^2 = -2\lambda + 2$$

$$\boxed{y = \sqrt{-2\lambda + 2}}$$

Put x and y in eq (1)

$$3\lambda^2 + 2 - 2\lambda - 2\lambda = 0$$

$$3\lambda^2 - 4\lambda + 2 = 0$$

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-4 \pm \sqrt{16 - 24}}{6}$$

$$= \frac{-4 \pm \sqrt{-8}}{6}$$

$$\boxed{\lambda_1 = 0.66 + 0.47i}$$

$$\boxed{\lambda_2 = 0.66 - 0.47i}$$

For x

$$\boxed{x_1 = \lambda_1}$$

$$\boxed{x_2 = \lambda_2}$$

$$y_1 = \sqrt{-2(0.66 + 0.47i) + 2}$$

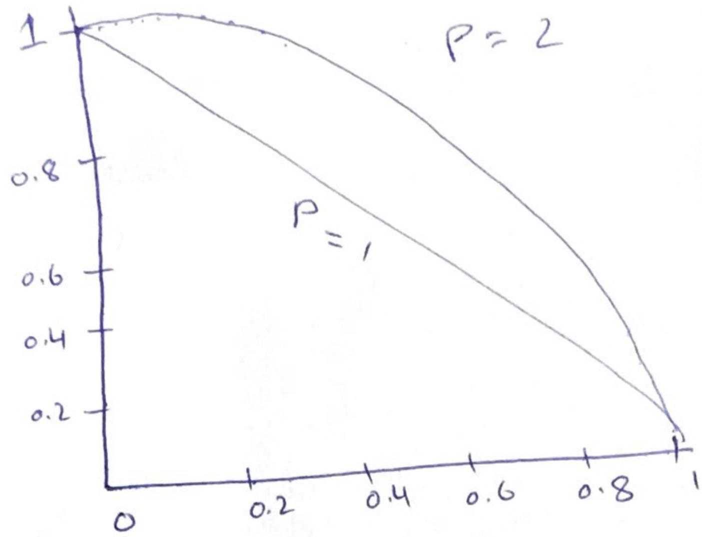
$$\boxed{y_1 = 0.95 - 0.48i}$$

$$y_2 = \sqrt{-2(0.66 - 0.47i) + 2}$$

$$\boxed{y_2 = 0.959 + 0.489i}$$

Qno 8.4

a)



b) Advantage of Lasso

1) Some of the coffients are shrunken all the way to zero

Drawback

Because Lasso penalty has the absolute value operation ~~is~~ it, the objective function is not differentiable and as a results lacks a closed form in gernal.

Limitation of Rigid Regression

1) Heavy bias toward zero for regression cofficents

2) Interpretability Unimportant coffients may be shrunken towards zero, but still they are in the model.