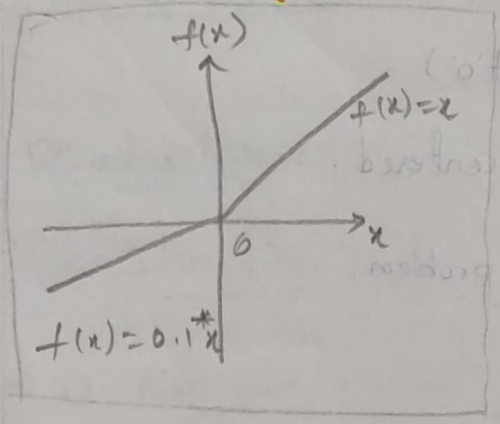
⇒ ReLU Variants → ① Linear → a) Leaky ReLU, b) Parametric "
→ ② non-Linear → a) Exp. Linear Unit (ELU), b) scaled Exp LU (selu)

① Linear:

1.a) Leaky ReLU: $f(x) = \max(0.01*z, z)$



↳ $z \geq 0 \Rightarrow f(z) = z$

↳ $z < 0 \Rightarrow f(z) = 0.01 * z$
$= z/100$

for $z < 0$; o/p is not 'zero'

↳ derivative: $f'(z) \rightarrow z \geq 0 \Rightarrow 1$

↳ $z < 0 \Rightarrow 0.01$

⇒ Adv.:

↳ ① Non-saturated

↳ ② Easily computed

↳ ③ No-dying ReLU problem

↳ ④ Close to zero-centered

1.b) Parametric ReLU: $f(x) = \max(a*x, x)$

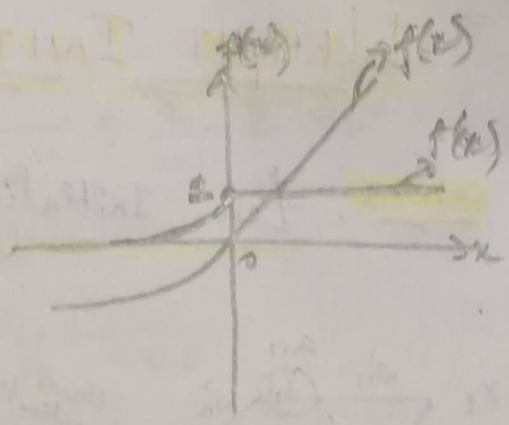
a → trainable parameter

⇒ flexibility is more

2) Non-Linear:

2.a) ELU (Exponential LU):

$$ELU(x) = \begin{cases} x & ; x \geq 0 \\ \alpha(e^x - 1) & ; x < 0 \end{cases}$$



$$ELU'(x) = \begin{cases} 1 & ; x > 0 \\ ELU(x) + \alpha & ; x < 0 \end{cases}$$

⇒ Adv:

⇒ Close to zero-centered

⇒ Better generalized

⇒ Non-Dying ReLU

⇒ Always continuous & Differentiable

Dis Adv.

⇒ Computationally expensive

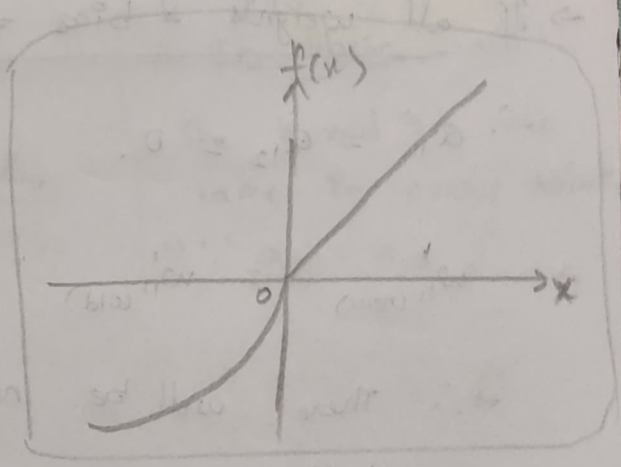
2.b) Scaled Exp. LU (SeLU):

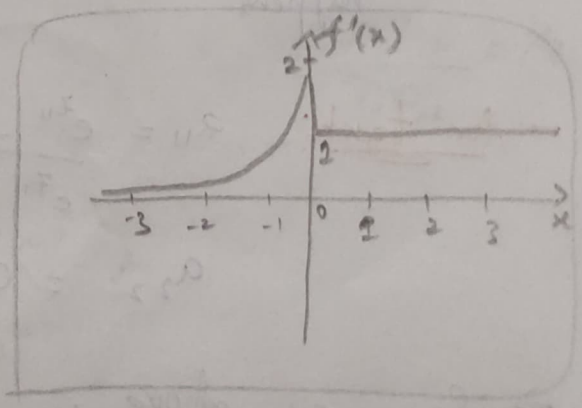
$$SELU(x) = \lambda \begin{cases} x & ; x > 0 \\ \alpha(e^x - 1) & ; x \leq 0 \end{cases}$$

⇒ λ, α are fixed parameters

↳ $\alpha \simeq 1.673$

↳ $\lambda \simeq 1.0507$





$$SELU'(x) = \lambda \begin{cases} 1 & ; x > 0 \\ \alpha e^x & ; x \leq 0 \end{cases}$$

⇒ Adv: ① Self-Normalizing ($\mu = 0$, $\sigma = 1$)

↳ ② Convergence - faster