**Week 1 Assignment**

Venkat Neelraj Nitta

College Of Professional Studies, Northeastern University

EAI6080: Advanced Analytical Utilization

Instructor: Mimoza Dimodugno PhD

Sep 26th, 2025

EAI 6080 – ASN – 1.

# I PROBABILITY

1. Given

$$\text{Mean} (\mu) = 0$$

$$\text{Variance} (\sigma^2) = 0.1$$

For a Guassian (Normal) distribution with mean $\mu$ and variance $\sigma^2$, the density at point $x$ is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Density at $x = 0$

$$p(x=0) = \frac{1}{\sqrt{2\pi(0.1)}} \, e^{-\left(\frac{0}{2(0.1)}\right)}$$

$$p(x=0) = \frac{1}{\sqrt{0.2\pi}} \approx \frac{1}{0.79} \approx 1.26$$

$$\boxed{\text{The density of } p \text{ at } x=0 \approx 1.26}$$

2.

Let $T = $ the (random) time the student finishes exam in the interval $[0,1]$.

$$\therefore \quad P(T < x) = \frac{x}{2}; \quad\quad 0 \leq x \leq 1 \quad\quad \rightarrow \text{①}$$

We want

$$P(T=1 \mid T > 0.75)$$

i.e., the probability that the student uses the full hour given that by time 0.75 they have not yet finished

We know

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$A : \{T > 1\}$

$B : \{T > 0.75\}$

If the student uses the full hour (i.e., finishes exactly at 1 or later), that means they were still working at 0.75.

$$\therefore A \cap B = A$$

$$\Rightarrow P(A|B) = \frac{P(A)}{P(B)}$$

from ①

$$P(A) = P(T \geqslant 1)$$

Since, $P(T < 1) = \frac{1}{2}$

we get $P(T \geqslant 1) = 1 - P(T < 1) = \frac{1}{2} \rightarrow ②$

$$P(B) = P(T > 0.75) = 1 - P(T \leqslant 0.75)$$

$$= 1 - \frac{0.75}{2} \rightarrow ③$$

$$\Rightarrow P(A|B) = \frac{P(A)}{P(B)} = \frac{0.5}{1 - 0.375} \quad (\text{from } ②, ③)$$

$$\boxed{P(A|B) = \frac{0.5}{0.625} = 0.8}$$

⇒ The conditional probability that the student will use the full hour, given that they are still working after 0.75 hr is __0.8 (80%)__

3.

| SICK? | Sunny | rainy | cloudy | snow |
|-------|-------|-------|--------|------|
| yes | 0.144 | 0.02 | 0.016 | 0.02 |
| no | 0.576 | 0.08 | 0.064 | 0.08 |

$$P(sick = yes \mid weather = rainy) = \, ?$$

From table,

$$P(sick = yes, \, rainy) = 0.02 \, ; \quad P(sick = \cancel{yes} \, no, \, rainy) = 0.08$$

Marginal probability of Rainy weather

$$P(rainy) = P(sick = yes, \, rainy) + P(sick = no, \, rainy)$$
$$= \quad 0.02 \quad + 0.08$$
$$= \quad 0.10$$

∴ By Bayes rule,

$$P(sick = yes \mid rainy) = \frac{P(sick = yes, \, rainy)}{P(rainy)} = \frac{0.02}{0.10} = 0.2$$

---

The probability that you are sick given the weather is rainy is 0.2 (or 20%)

## II. CALCULUS AND LINEAR ALGEBRA

1.     Given     $f(z) = \dfrac{1}{1+e^{-z}}$

$$f(z) = \left(1+e^{-z}\right)^{-1}$$

$$f'(z) = (-1) \cdot \left(1+e^{-z}\right)^{-2} \cdot \frac{d}{dz}\left(1+e^{-z}\right) \quad \left| \quad \left[\frac{d}{dz}\left(1+e^{-z}\right) = -e^{-z}\right]\right.$$

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

$$\Rightarrow \quad f'(z) = \frac{e^{-z}}{\left(1+e^{-z}\right)^2}$$

This can be written as

$$\boxed{f'(z) = \frac{1}{\left(1+e^{-z}\right)}\left[1-\frac{1}{\left(1+e^{-z}\right)}\right]}$$

(or)

$$\cancel{f'(z) = f(z)\left(1-\frac{1}{f(z)}\right)}$$

$$\boxed{f'(z) = f(z)\left(1- f(z)\right)}$$

EAI-6080-ASN-1

2.

$$f(w) = \frac{1}{1 + e^{-w^T x}}$$

Let $z = w^T x$ ; $f(w) = \sigma(z)$ ; where $\sigma(z) = \frac{1}{1 + e^{-z}}$

we know $\hspace{3cm}$ ↳ from prev problem.

$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

derivative of inner function.

$$z = w^T x = \sum_{j=1}^{D} w_j x_j \implies \frac{dz}{dw_i} = x_i$$

By chain rule

$$\frac{\partial f}{\partial w_i} = \frac{d\sigma}{dz} \frac{\partial z}{\partial w_i} = \sigma(z)(1 - \sigma(z)) x_i$$

Since $\sigma(z) = f(w)$

$$\boxed{\frac{\partial f}{\partial w_i} = f(w)(1 - f(w)) x_i}$$

3.

$$J(w) = \frac{1}{2}\left[ \sum_{i=1}^{m} (w^T x^i - y^i)^2 \right] + \cancel{\text{λ}||w||}$$

let $\; r_i(w) = w^T x^i - y^i$

for $u \neq 0$, $\frac{d}{du}|u| = \text{sign}(u)$

at $u = 0$ the subgradient is any value in $[-1, 1]$

$$\frac{\partial}{\partial w}|r_i(w)| = \text{sign}(r_i(w)) \frac{\partial r_i}{\partial w}$$

And,  $\dfrac{\partial \vartheta_i}{\partial w} = \dfrac{\partial}{\partial w}(w^T - x^{(i)} - y^{(i)}) = x^{(i)}$

$\Rightarrow$  $\boxed{\nabla_w J(w) = \dfrac{1}{2} \sum_{i=1}^{m} sign(w^T x^{(i)} - y^{(i)})\, x^{(i)}}$

Equivalently, in matrix form  with  $X = \left[ x^{(1)}, x^{(2)}, \dots x^{(m)} \right]^T$

$\&\ S_i = sign(\vartheta_i)$

$\Rightarrow$  $\boxed{\nabla_w J(w) = \dfrac{1}{2} X^T s_i}$  ;

$S_i \in \{-1, 1\}$  for  $\vartheta_i \neq 0$

$S_i \in [-1, 1]$  if  $\vartheta_i = 0$

---

4.  $J(w) = \dfrac{1}{2}\left[ \sum (w^T x^{(i)} - y^{(i)})^2 \right] + \lambda \|w\|_2^2$

$\&\ $ let  $\vartheta_i(w) = w^T x^{(i)} - y^{(i)}$

Differentiate the squared term
for each $i$,

$\dfrac{\partial}{\partial w} \dfrac{1}{2} \vartheta_i(w)^2 = \dfrac{1}{2} \cdot 2 \vartheta_i(w) \dfrac{\partial \vartheta_i}{\partial w} = \vartheta_i(w) \cdot x^{(i)}$

$\left( \text{since} \atop \dfrac{\partial \vartheta_i}{\partial w} = \dfrac{\partial (w^T x^{(i)} - y^{(i)})}{\partial w} = x^{(i)} \right)$

Summing over $i$ gives

$\dfrac{\partial}{\partial w}\left[ \dfrac{1}{2} \sum_{i=1}^{m} \vartheta_i(w)^2 \right] = \sum_{i=1}^{m} \vartheta_i(w)\, x^{(i)}$

Differentiate with $L_2$ regularizer.

$\dfrac{\partial}{\partial w} \lambda \|w\|_2^2 = \dfrac{\partial}{\partial w} \lambda w^T w = 2\lambda w.$

Combining

$$\boxed{\nabla_w J(w) = \sum (w^T x^{(i)} - y^{(i)}) x^{(i)} + 2\lambda w}$$

Matrix form (with $X \in \mathbb{R}^{m \times d}$ whose i-th row $x^{(i)T}$, $y \in \mathbb{R}^m$):

$$\boxed{\nabla_w J(w) = X^T (Xw - y) + 2\lambda w}$$

---

5.

$$J(w) = \sum_{i=1}^{m} \left[ y^{(i)} \log\left(\frac{1}{1 + e^{-w^T x^{(i)}}}\right) + (1 - y^{(i)}) \log\left(1 - \frac{1}{1 + e^{-w^T x^{(i)}}}\right) \right]$$

let $\sigma(z) = \dfrac{1}{1 + e^{-z}}$ & $z_i = w^T x^{(i)}$

$$\Rightarrow J(w) = \sum_{i=1}^{m} \left[ y^{(i)} \log \sigma(z_i) + (1 - y^{(i)}) \log (1 - \sigma(z_i)) \right]$$

for each i,

$$g_i(w) = y^{(i)} \log \sigma(z_i) + (1 - y^{(i)}) (\log 1 - \sigma(z_i))$$

using chain rule        $\sigma'(z) = \sigma(z)(1 - \sigma(z))$

$$\nabla_w \log \sigma(z_i) = \frac{\sigma'(z_i)}{\sigma(z_i)} x^{(i)} = (1 - \sigma(z_i)) x^{(i)}$$

$$\nabla_w \log (1 - \sigma(z_i)) = \frac{-\sigma'(z_i)}{1 - \sigma(z_i)} x^{(i)} = -\sigma(z_i) x^{(i)}$$

Thus

$$\nabla_w g_i(w) = y^{(i)} (1 - \sigma(z_i)) x^{(i)} + (1 - y^{(i)}) (-\sigma(z_i)) x^{(i)}$$

$$= (y^{(i)} - \sigma(z_i)) x^{(i)}$$

Sum over i

$$\nabla_w J(w) = \sum_{i=1}^{m} (y^{(i)} - \sigma(w^T x^{(i)})) x^{(i)}$$

6.

$$f(w) = \tanh [w^T x]$$

let $z = w^T x$

outer function
$$\frac{d}{dz} \tanh z = 1 - \tanh^2 z$$

inner function
$$\frac{\partial z}{\partial w} = x$$

chain rule
$$\nabla_w f(w) = \frac{d}{dz} \tanh z \cdot \frac{\partial z}{\partial w}$$

$$= (1 - \tanh^2(w^T x)) x$$

$$\Rightarrow \quad \boxed{\nabla_w f(w) = (1 - \tanh^2(w^T x)) x}$$

7. 

$$A = \begin{bmatrix} 2 & 1 & -1 \\ -3 & -1 & 2 \\ -2 & 1 & 2 \end{bmatrix} \qquad b = \begin{pmatrix} 8 \\ -11 \\ -3 \end{pmatrix}$$

$R_2 \leftarrow 2R_2 + 3R_1$

$R_3 \leftarrow R_3 + R_1$

$$[A|b] = \begin{bmatrix} 2 & 1 & -1 & | & 8 \\ 0 & -2+3 & 4-3 & | & 2 \\ 0 & 2 & 1 & | & 5 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & -1 & | & 8 \\ 0 & 1 & 1 & | & 2 \\ 0 & 2 & 1 & | & 5 \end{bmatrix}$$

$R_3 \leftarrow R_3 - 2R_2$

$$\begin{bmatrix} 2 & 1 & -1 & | & 8 \\ 0 & 1 & 1 & | & 2 \\ 0 & 0 & -1 & | & 1 \end{bmatrix}$$

from    row 3 ;    $-x_3 = 1$

$\Rightarrow x_3 = -1$

sub $x_3 = -1$ in    row 2

$x_2 + x_3 = 2 \Rightarrow x_2 - 1 = 2 \Rightarrow x_2 = 3$

Sub $x_2, x_3$ in row 1

$$2x_1 + x_2 + x_3 = 8$$

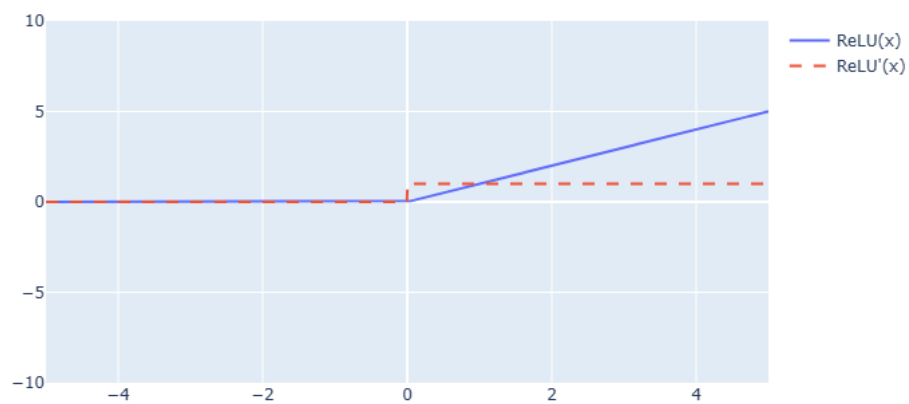$$2x_1 + 3 - (-1) \Rightarrow 8 \quad \Rightarrow \quad 2x_1 + 4 = 8$$

$$\boxed{x_1 = 2}$$

$$\Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$
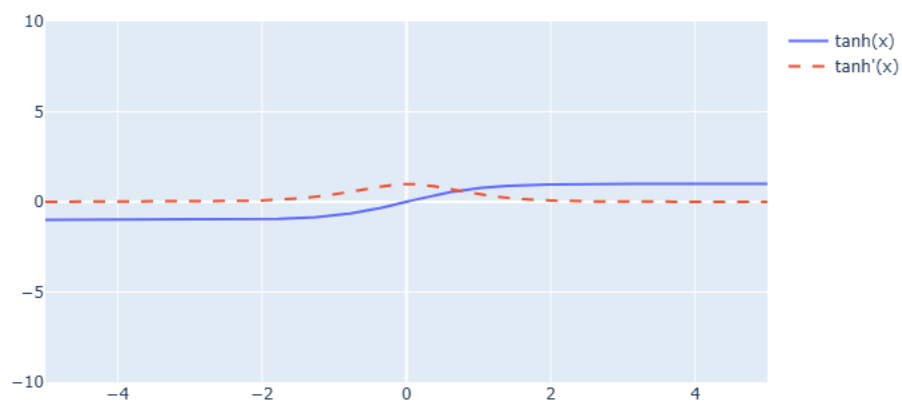
Activation Functions

    1.  ReLU

ReLU and derivative
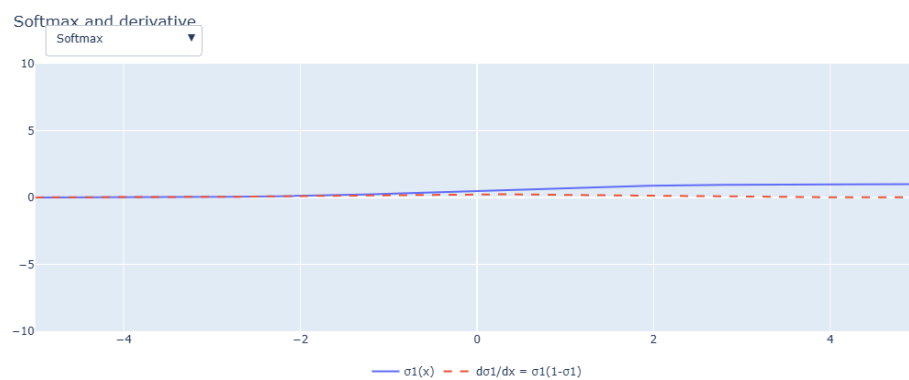


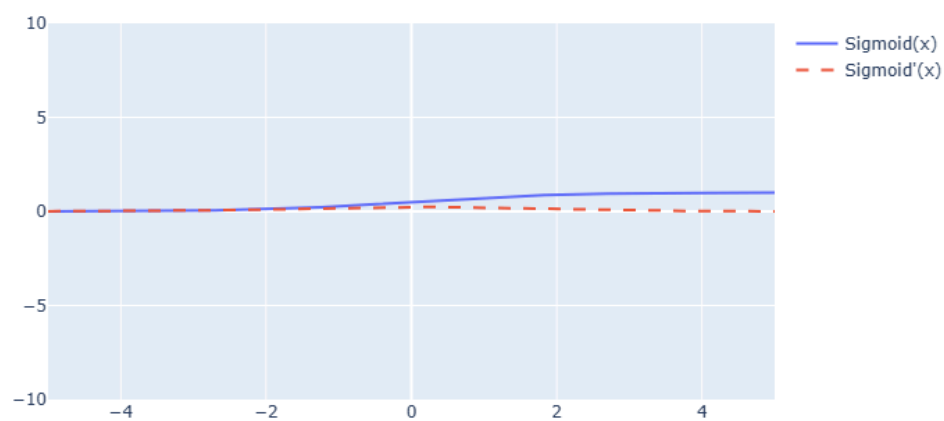    2.  Tanh

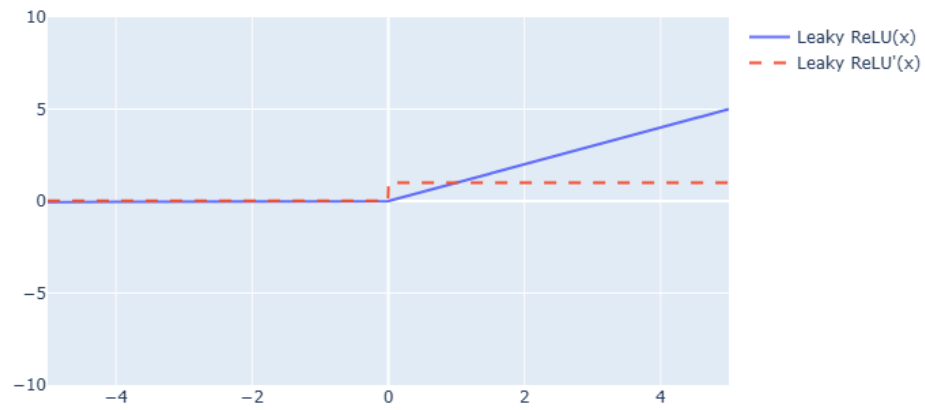Tanh and derivative


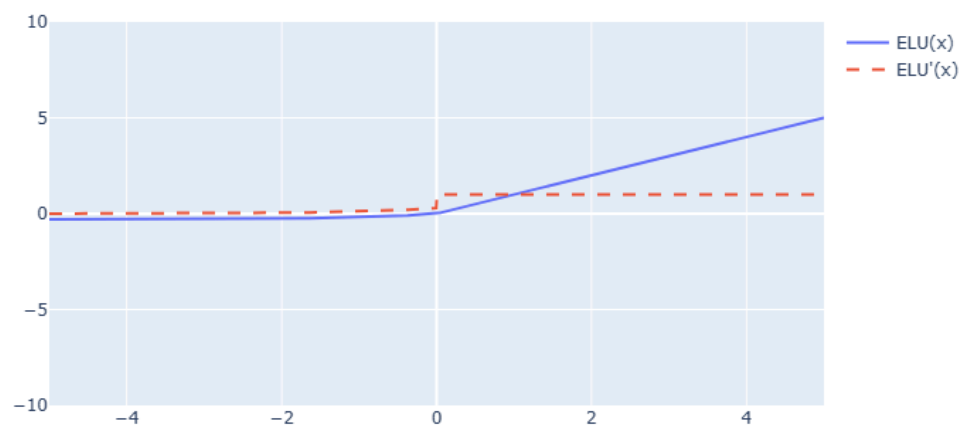
    3.  Softmax

Softmax and derivative

Softmax ▼

—— σ1(x)  – – dσ1/dx = σ1(1-σ1)

4. Sigmoid



Sigmoid and derivative

—— Sigmoid(x)
– – Sigmoid'(x)

5. Leaky ReLU

Leaky ReLU and derivative



6. ELU ( (plot with $\alpha$=0.3)

ELU and derivative (a=0.3)



7. Sinc

Sinc and derivative