EAI6080 - ASN - 1.

I PROBABILITY

1. Given

Mean $(\mu) = 0$

Variance $(\sigma^2) = 0.1$

For a Guassian (Normal) distribution with mean μ and variance σ^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].

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

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

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

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

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

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

$$= 1 - \frac{0.75}{2} \cancel{0.375} \rightarrow ③$$

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

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

$\Rightarrow$ 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

$$P(\text{sick}=\text{yes} \mid \text{weather}=\text{rainy}) = ?$$

From table,

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

Marginal probability of Rainy weather

$$P(\text{rainy}) = P(\text{sick}=\text{yes}, \text{rainy}) + P(\text{sick}=\text{no}, \text{rainy})$$
$$= 0.02 + 0.08$$
$$= 0.10$$

∴ By Bayes rule,

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

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

II. CALCULUS AND LINEAR ALGEBRA

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

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

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

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

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

This can be written as

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

(or)

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

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

2.
$$f(w) = \frac{1}{1+e^{-w^T x}}$$

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

↳ from prev problem.

we know

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

derivative of inner function.

$$z = w^T x = \sum_{j=1}^{D} w_j x_j \Rightarrow \frac{\partial z}{\partial w_i} = x_i$$

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.
$$J(w) = \frac{1}{2}\left[\sum_{i=1}^{m}(w^T x^i - y^i)^2\right] + \cancel{\lambda \|w\|_2^2}$$

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, $\frac{\partial r_i}{\partial w} = \frac{\partial}{\partial w}(w^T x^{(i)} - y^{(i)}) = x^{(i)}$

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

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

& $s_i = \text{sign}(r_i)$

$$\Rightarrow \boxed{\nabla_w J(w) = \frac{1}{2} X^T s_i} \; ; \quad s_i \in \{-1, 1\} \text{ for } r_i \neq 0, \; s_i \in [-1, 1] \text{ if } r_i = 0$$

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

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

Differentiate the squared term for each i,

$$\frac{\partial}{\partial w} \frac{1}{2} r_i(w)^2 = \frac{1}{2} \cdot 2 r_i(w) \frac{\partial r_i}{\partial w} = r_i(w) \cdot x^{(i)}$$

(Since $\frac{\partial r_i}{\partial w} = \frac{\partial}{\partial w}(w^T x^{(i)} - y^{(i)}) = x^{(i)}$)

Summing over i gives

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

Differentiate with L_2 regularizer.

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

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

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

6.

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

let $z = w^T x$

outer function

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

inner function

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

chain rule

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

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

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

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

$R_2 \leftarrow 2R_2 + 3R_1$

$R_3 \leftarrow R_3 + R_1$

$$[A|b] = \left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & -2+3 & 4-3 & 2 \\ 0 & 2 & 1 & 5 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & 1 & 1 & 2 \\ 0 & 2 & 1 & 5 \end{array}\right]$$

$R_3 \leftarrow R_3 - 2R_2$

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 1 \end{array}\right]$$

from row3 ; $-x_3 = 1$

$\Rightarrow x_3 = -1$

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

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

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

$$\boxed{x_1 = 2}$$

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