01.11.2021 EX-1

1) i) $\sum_{i=0}^{10} i = 0+1+\cdots+10 = 55$ (Ans)

ii) $\prod_{i=1}^{5} i = 1.2.3.4.5 = 120$ (Ans)

iii) $\sum_{i=0}^{4} i^2 = 0^2+1^2+\cdots+4^2 = 30$ (Ans)

iv) $\sum_{i=1}^{3} x_i y_i$ if $\vec{x} = (1, 2, 3)^T$

$\vec{y} = (2, 1, 0)^T$

$= x_1 y_1 + x_2 y_2 + x_3 y_3$

$= 1\times 2 + 2\times 1 + 3\times 0 = 4$ (Ans)

v) $\sum_{i=1}^{3} (x_i + y_i)$ if $\vec{x} = (1, 2, 3)^T$

$\vec{y} = (2, 1, 0)^T$

$= (1+2) + (2+1) + (3+0) = 9$ (Ans)

vi) $\frac{1}{3} \sum_{i=1}^{3} x_i^2$ if $\vec{x} = (1, 2, 3)^T$

$= \frac{1}{3}(x_1^2 + x_2^2 + x_3^2)$

$= \frac{1}{3}(1^2 + 2^2 + 3^2)$

$= 14/3$ (Ans)

2) Given, $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix}$

$\vec{x} = (1, 2, 3, 0)^T$

$\vec{y} = (2, 3, 4, 5)^T$

$\vec{z} = (2, 3, 0)^T$

1) $A\vec{y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$

$= \begin{pmatrix} 2+3+4+5 \\ 4+6+8+10 \\ 6+9+12+15 \\ 8+12+16+20 \end{pmatrix} = (14 \quad 28 \quad 42 \quad 56)^T$

(Ans)

Same: ii) $A\vec{x}$, iii) $\vec{x}^T\vec{y}$ vi) $\vec{x}^T A \vec{y}$ x) $A \cdot A$
xi) $A \cdot A^T$.

iv) $\vec{x}A = (4,1), (4,4)$

As first term's column number & second term's row number is not same, matrix multiplication not possible.

Same: v) $A\vec{y}^T$ viii) $A\vec{z}$

3. Functions

Given, $f: x \in R \longrightarrow y \in R$

& $f(x) = \begin{cases} x & ; \text{ if } x \geq 0 \\ 0 & ; \text{ else} \end{cases}$

$\vec{x} = (1, 2, 3, 0)^T$

$\vec{y} = (2, 3, 4, 5)^T$

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

I) $f(\vec{x}) = f\left(\begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}$. (Ans)

II) $f(B\vec{x})$

Now, $B\vec{x} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \\ -3 \\ 0 \end{pmatrix}$

So, $f(B\vec{x}) = f\left(\begin{pmatrix} 1 \\ -4 \\ -3 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. (Ans)

III) $f(B) = \left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)$

$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. (Ans)

IV) $f(\vec{X}^T \vec{Y})$

Now,

from 2(III), we get $\vec{X}^T \vec{Y} = (20)$

So,

$f(\vec{X}^T \vec{Y}) = f((20)) = (20)$. (Ans)

4. V-S & V-V functions :

Given, $f(x) = \langle \vec{x}, (1, 1, 0, 0)^T \rangle$

$$g_i(\vec{x}) = x_i^2$$

i) $f((1, 1, 1, 1)^T) = \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle = 1\times1 + 1\times1 + 1\times0 + 1\times0$

$= 2.$ (Ans)

ii) $f(B) = f\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)$

$$= \begin{pmatrix} f\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right) \\ f\left(\begin{pmatrix} 0 \\ -2 \\ 0 \\ 0 \end{pmatrix}\right) \\ f\left(\begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}\right) \\ f\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) \end{pmatrix} = \begin{pmatrix} \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle \\ '' \\ '' \\ '' \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 0 \end{pmatrix}$$

(Ans)

IV) $g\left((1, 2, 3)^T\right) = \begin{pmatrix} 1^2 \\ 2^2 \\ 3^2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 9 \end{pmatrix}$ (Ans)

V) $g(B) = g\left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right)$

$$= \left(g\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad g\begin{pmatrix} 0 \\ -2 \\ 0 \\ 0 \end{pmatrix} \quad g\begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix} \quad g\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

(Ans)

5. Matrix Contractions & slices:

Given, $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix}$

[II, III] → Trace

I) $\sum_{i=1}^{4} A_{2i} = A_{21} + A_{22} + A_{23} + A_{24}$

$= 2 + 2 + 2 + 2 = 8.$ (Ans)

IV) $A_{1,2:3} = \begin{pmatrix} A_{12} & A_{13} \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix}$ (Ans)

V) $A_{2:3,1} = \begin{pmatrix} A_{21} \\ A_{31} \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$ (Ans)

VI) $A_{:,1} = \begin{pmatrix} A_{11} & A_{21} & A_{31} & A_{41} \end{pmatrix}^T$

$= \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}^T$ (Ans)

VII) $A_{1:2,2:3} = \begin{pmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{pmatrix}$

$= \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$ (Ans)

Scanned with CamScanner

6. Tensor Contractions & slicing:

Given, $T \in R^{3,2,4}$

$$T_{ijk} = 1 \quad , \quad \forall\, ijk$$

I) $\tilde{T}_{ab} = \sum_{i=1}^{3} T_{iab}$; Shape is $\mathbb{R}^{2,4}$

II) $\tilde{T}_{ab} = \sum_{i=1}^{2} T_{aib}$; Shape is $\mathbb{R}^{3,4}$

III) $\tilde{T}_{ab} = \sum_{i=1}^{4} T_{abi}$; Shape is $\mathbb{R}^{3,2}$

IV) $\tilde{T}_{ab} = \sum_{i=1}^{2} \sum_{j=1}^{4} T_{aij}$; Shape is $\mathbb{R}^{3}$

$\boxed{\sim \;\rightarrow\; \text{Tilda}}$

Scanned with CamScanner

Exercise 2

08.11.2021

1. 1D derivatives.

Rules:

I) Linearity | II) Chain rule | III) Product rule.

a) Given,

$$h(x) = \sin(\sin(x))$$

Rule:

I) Chain rule:

$$h(x) = f(g(x))$$

II) Decomposition:

$$f(x) = \sin(x) \quad ; \quad g(x) = \sin(x)$$

So, $$f'(x) = \cos(x) \quad ; \quad g'(x) = \cos(x)$$

III) Apply Rule:

$$h'(x) = f'(g(x)) \cdot g'(x)$$

$$= \cos(\sin(x)) \cdot \cos(x). \quad \text{(Ans)}$$

b) Given,

$$h(x) = 2\sin(x) + 5\cos(x)$$

Rule:

I) Linearity:

$$h(x) = af(x) + bg(x)$$

II) Decomposition: $a = 2 \quad ; \quad b = 5$

$$f(x) = \sin(x) \quad ; \quad g(x) = \cos(x)$$

So, $$f'(x) = \cos(x) \quad ; \quad g'(x) = -\sin(x)$$

iii) Apply rule:

$$h'(x) = af'(x) + bg'(x)$$

$$= 2\cos(x) - 5\sin(x) \quad \text{(Ans)}$$

c) Same as (b)

d) Given,

$$h(x) = \ln(\sin(\cos(5.2)))$$

$h'(x) = 0$; as there is no x term, the expression is considered constant. (Ans)

e) Given,

$$h(x) = \exp(x)\ln(x)$$

Rule:

i) Product rule:

$$h(x) = f(x)g(x)$$

ii) Decomposition:

$$f(x) = \exp(x) \quad ; \quad g(x) = \ln(x)$$

So,

$$f'(x) = \exp(x) \quad ; \quad g'(x) = 1/x$$

iii) Apply rule:

$$h'(x) = f'(x)g(x) + g'(x)f(x)$$

$$= \exp(x).\ln(x) + \left(\tfrac{1}{x}\right)\exp(x)$$

f) Same as (e) (Ans)

Scanned with CamScanner

g) Given, $h(x) = e^{\cos(x)}$

Rule:

I) Chain rule:

$$h(x) = \cancel{f(x)\,g(x)}\ f(g(x))$$

II) Decomposition:

$$f(x) = e^x \quad ; \quad g(x) = \cos(x)$$

so, $f'(x) = e^x \quad ; \quad g'(x) = -\sin(x)$

III) Apply rule:

$$h'(x) = f'(g(x)) \cdot g'(x)$$

$$= -e^{\cos(x)} \cdot \sin(x) \text{ . (Ans)}$$

$\boxed{\text{2. Partial Derivations}}$

a) Given, $\frac{\partial}{\partial x_3} \sum_{i=1}^{5} x_i$

$$= \frac{\partial}{\partial x_3}(x_1 + x_2 + x_3 + x_4 + x_5)$$

$$= 0 + 0 + 1 + 0 + 0 = 1 \text{ . (Ans)}$$

b) Given, $\frac{\partial}{\partial x_6} \sum_{i=1}^{5} x_i$

$$= \frac{\partial}{\partial x_6}(x_1 + \cdots + x_5)$$

$$= 0 \text{ .}$$

(Ans)

c) $\frac{\partial}{\partial x_3} \sum_{i=1}^{5} (2x_i - 5)^2$

$= \frac{\partial}{\partial x_3} (2x_3 - 5)^2$ [ignoring all x terms except x_3]

$= 2(2x_3 - 5) \cdot \frac{\partial}{\partial x_3} (2x_3 - 5)$

$= 2(2x_3 - 5) \cdot 2 = 8x_3 - 20.$ (Ans)

d) $\frac{\partial}{\partial x_3} \sum_{i=1}^{5} (x_i^2 - 5x_i)^2$

$= \frac{\partial}{\partial x_3} \left[x_3^4 - 10x_3^3 + 25x_3^2 \right]$ [ignoring all x terms except x_3 as the partial derivatives are w.r.to x_3]

$= 4x_3^3 - 30x_3^2 + 50x_3$. (Ans)

e) Given, $\frac{\partial}{\partial x_3} \cos(x_3) x_3$

Rule:

I) Product rule: $h(x_3) = f(x_3) \cdot g(x_3)$

II) Decomposition:

$f(x_3) = \cos(x_3)$; $g(x_3) = x_3$

So, $f'(x_3) = -\sin(x_3)$; $g'(x_3) = 1$

Scanned with CamScanner

iii) Apply rule:

$$h'(x_3) = f(x_3)\,g'(x_3) + f'(x_3)\,g(x_3)$$

$$= \cos x_3 + (-\sin(x_3))\,x_3 \;.\; \text{(Ans)}$$

f) $\frac{\partial}{\partial x_3} \cos(x_3)\,x_4$

$$= x_4 \cdot \frac{\partial}{\partial x_3} \cos(x_3)$$

[As partial derivative is w.r.to x_3, x_4 is considered as constant]

$$= -x_4 \sin(x_3) \;.\; \text{(Ans)}$$

g) Given, $\frac{\partial}{\partial x_3} \cos(\exp(x_3))$

Rule:

i) Chain rule: $h(x_3) = f(g(x_3))$

ii) Decomposition:

$$f(x_3) = \cos(x_3) \;;\; g(x_3) = \exp(x_3)$$

So, $f'(x_3) = -\sin(x_3) \;;\; g'(x_3) = \exp(x_3)$

iii) Apply rule:

$$h'(x_3) = f'(g(x_3)) \cdot g'(x_3)$$

$$= -\sin(\exp(x_3)) \cdot \exp(x_3).$$

(Ans)

$f \Rightarrow$ Note: if constant is with differentiable term, then it will be multiplied with the differentiated ~~part~~ part.

(3) Given,

$$g_i(\vec{x}) = x_i^2$$

$$f(\vec{x}) = \sum_i \log x_i$$

Take derivative of $f(g(\vec{x}))$ using V-V-CR.

Now, $$\frac{\partial f}{\partial x_1} = \sum_K \underbrace{\frac{\partial f}{\partial x_K}\Bigg|_{\vec{g}(\vec{x})}}_{(1)} \cdot \underbrace{\frac{\partial g_K}{\partial x_1}}_{(11)} \quad \text{———} \; (0)$$

(1) $\Rightarrow$ $$\frac{\partial f}{\partial x_K}\Bigg|_{\vec{g}(\vec{x})} = \frac{\partial}{\partial x_K}\left(\sum_i \log x_i\right)\Bigg|_{\vec{g}(\vec{x})}$$

$$= \sum_i \frac{1}{x_i} \cdot \delta_{iK}\Bigg|_{\vec{g}(\vec{x})}$$

$$= \sum_i \frac{1}{g_i} \delta_{iK}$$

$$= \sum_i \frac{1}{x_i^2} \cdot \delta_{iK} = \frac{1}{x_K^2}$$

(11) $\Rightarrow$ $$\frac{\partial g_K}{\partial x_1} = \frac{\partial}{\partial x_1}\left(x_K^2\right) = 2x_K \delta_{K1}$$

Re-Insert (1) & (11) into (0) $\Rightarrow$

$$\frac{\partial f}{\partial x_1} = \sum_K \cdot \frac{1}{x_K^2} \cdot 2x_K \delta_{K1} = \sum_K \frac{1}{x_K} \cdot 2 \cdot \delta_{K1}$$

$$= \frac{2}{x_1} \quad \text{(Ans)}$$

4) Given,

$$\vec{y} = \vec{f}(\vec{w}) \quad ; \quad f_i(\vec{w}) = w_i^3$$

$$\tilde{L}(\vec{y}) = \sum_i y_i \quad ; \quad \tilde{L}(\vec{w}) = \tilde{L}(f(\vec{w}))$$

Take derivative of $\tilde{L}(f(\vec{w}))$ using V-V-CR.

Now, $$\frac{\partial}{\partial w_1}(\tilde{L}) = \sum_K \underbrace{\frac{\partial \tilde{L}}{\partial y_K}\bigg|_{f(\vec{w})}}_{(i)} \cdot \underbrace{\frac{\partial y_K}{\partial w_1}}_{(ii)} \quad \text{—— } (0)$$

(i) $\Rightarrow$ $$\frac{\partial \tilde{L}}{\partial y_K} = \frac{\partial}{\partial y_K}\left(\sum_i y_i\right)\bigg|_{f(\vec{w})} = \frac{\partial y_K}{\partial y_K}\bigg|_{f(\vec{w})}$$

$$= 1.$$

(ii) $\Rightarrow$ $$\frac{\partial y_K}{\partial w_1} = \frac{\partial}{\partial w_1}\left[f_K(\vec{w})\right] = \frac{\partial}{\partial w_1}\left(w_K^3\right)$$

$$= 3w_K^2 \cdot \delta_{K1}$$

Re-insert (i) & (ii) into (0) $\Rightarrow$

$$\frac{\partial \tilde{L}}{\partial w_1} = \sum_K 1 \cdot 3w_K^2\, \delta_{K1}$$

$$= 3w_1^2. \quad \text{(Ans)}$$

5. <u>Gradient descent:</u>

Given,

$$f(\vec{x}) = x_1^2 + 2x_2^2$$

$$\vec{x}(0) = \cancel{\begin{bmatrix} 1 \\ 3 \end{bmatrix}} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \; ; \; \epsilon = 0.1$$

We know, $\because f(\vec{x}(0)) = 1^2 + 2 \cdot 3^2 = 19.$

In Gradient descent,

$$\vec{x}(i+1) = \vec{x}(i) - \epsilon \vec{\nabla} f \Big|_{\vec{x}(i)}$$

Now, $\dfrac{\partial f}{\partial x_1} = 2x_1 \quad ; \quad \dfrac{\partial f}{\partial x_2} = 4x_2$

So, $$\vec{\nabla} f = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$$

<u>Step 1:</u> $$\vec{x}(1) = \vec{x}(0) - \epsilon \vec{\nabla} f \Big|_{\vec{x}(0)}$$

$$= \begin{pmatrix} 1 \\ 3 \end{pmatrix} - 0.1 \begin{pmatrix} 2 \cdot 1 \\ 4 \cdot 3 \end{pmatrix} \Bigg|_{\begin{pmatrix} 1 \\ 3 \end{pmatrix}}$$

$$= \begin{pmatrix} 1 \\ 3 \end{pmatrix} - 0.1 \begin{pmatrix} 2 \\ 12 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 0.2 \\ 1.2 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8 \\ 1.8 \end{pmatrix}$$

So, $$f(\vec{x}(1)) = 0.8^2 + 2 \times 1.8^2 = 7.12 < f(\vec{x}(0))$$

$$= 0.64 + 6.48$$

Step 2:

$$\vec{x}(2) = \vec{x}(1) - \epsilon \vec{\nabla} f \big|_{\vec{x}(1)}$$

$$= \begin{pmatrix} 0.8 \\ 1.8 \end{pmatrix} - 0.1 \cdot \begin{pmatrix} 2 \times 0.8 \\ 4 \times 1.8 \end{pmatrix} \Bigg|_{\begin{pmatrix} 0.8 \\ 1.8 \end{pmatrix}}$$

$$= \begin{pmatrix} 0.8 \\ 1.8 \end{pmatrix} - \begin{pmatrix} 0.16 \\ 0.72 \end{pmatrix}$$

$$= \begin{pmatrix} 0.64 \\ 1.08 \end{pmatrix}$$

So, $f(\vec{x}(2)) = (0.64)^2 + 2 \times (1.08)^2$

$= 0.41 + 2.33 = 2.74 < f(\vec{x}(1))$

So, f is decreasing.

(Ans)

[In gradient ascent, equation will be:

$$\vec{x}(i+1) = \vec{x}(i) + \epsilon \vec{\nabla} f \big|_{\vec{x}(i)}$$

Exercise 3

1. Same as Ex-2 → (5)

2.

$$\boxed{\begin{array}{lcl} \partial & = & \text{Partial Derivatives} \\ \delta & = & \text{Kronecker delta} \end{array}}$$

Given,

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

a) $\frac{\partial x_i}{\partial x_j} = 1$ [when $i = j$, only then partial derivative is possible]

b) $\frac{\partial}{\partial x_j} \sum_i^I x_i$ where $j < I$ $= \sum_i^I \delta_{ij}$

$= \frac{\partial}{\partial x_j}(x_j) = 1$

c) $\frac{\partial}{\partial x_j} \sum_i^I x_i^2$ where $j < I$

$= \sum_i^I 2x_i \delta_{ij}$

$= 2x_j \delta_{jj}$

$= 2x_j \cdot 1$

$= 2x_j$. (Ans)

Scanned with CamScanner

d) $\frac{\partial}{\partial x_j} \sum_i^I 155 x_i$ where, $j < I$

$= 155 \cdot \frac{\partial}{\partial x_j}(x_j) = 155$. (Ans)

e) $\sum_i^I \delta_{ij}$ where $j < I$

$= \delta_{jj} = 1$

f) $\sum_i^I \delta_{ij}$ where $j > I$

$= 0$. (Ans)

g) $\sum_i^I (\delta_{ij} + 5)$ where $j < I$

$= \sum_i^I \delta_{ij} + \sum_i^I 5$

$= 1 + 5I$. (Ans)

Scanned with CamScanner

3. Given,

$$\vec{f}(\vec{x}, W, \vec{b}) = \vec{s}(W\vec{x}+\vec{b})$$

a) Now,

$$\vec{f}(\vec{x}, W, \vec{b}) = \vec{f}(\vec{a})$$

$$= \vec{f}(\vec{a}(\vec{x}, W, \vec{b}))$$

$$= \vec{s}(W\vec{x}+\vec{b}).$$

So, $\vec{a}(\vec{x}, W, \vec{b}) = W\vec{x} + \vec{b}$.

(Ans)

b) Now,

$$y_i = \boxed{\cancel{\vec{f}(\vec{a})}}\ f_i(\vec{a})$$

$$= S_i(a)$$

$$= \frac{\exp(a_i)}{\sum_j \exp(a_j)}$$

&

$$a_i(\vec{x}, W, \vec{b}) = \left[\sum_k W_{ik} \cdot x_k\right] + b_i$$

(Ans)

4) We Know,

Cross-entropy for single sample $\vec{x}$ & target $\vec{t}$ is,

$$\tilde{L}(\vec{y}, \vec{t}) = -\sum_K t_K \log y_K$$

Where,
$\vec{y}$ = output vector
$\vec{t}$ = target vector.

$$= -\sum_K t_K \log y_K$$

$$= -\sum_K t_K \log S_K(\vec{a}).$$

(Ans)

[Use either log or ln]

5) We Know,

$$\tilde{L} = -\sum_K t_K \ln y_K$$

So, $$\frac{\partial \tilde{L}}{\partial b_c} = \sum_j \frac{\partial \tilde{L}}{\partial y_j} \cdot \frac{\partial y_j}{\partial b_c}$$

$$= -\sum_j \frac{\partial}{\partial y_j}\left[\sum_K t_K \ln y_K\right] \cdot \frac{\partial y_j}{\partial b_c}$$

$$= -\sum_j \frac{\partial}{\partial y_j}\left(t_j \ln y_j\right) \cdot \frac{\partial y_j}{\partial b_c}$$

$$= \cancel{-t_j \sum \frac{1}{y_j}} \cdot -t_j \cdot \frac{1}{y_j} \cdot \frac{\partial y_j}{\partial b_c}$$

$$= \left(-\frac{t_j}{y_j}\right) \cdot \frac{\partial y_j}{\partial b_c}.$$

(Done)

6) $$\frac{\partial y_j}{\partial b_c} = \sum_i \frac{\partial y_j}{\partial a_i} \cdot \frac{\partial a_i}{\partial b_c}$$

$$= \sum_i \frac{\partial S_j}{\partial a_i} \cdot \frac{\partial a_i}{\partial b_c}$$

$$= \sum_i \left(\delta_{ij} S_j - S_i S_j\right) \cdot \frac{\partial a_i}{\partial b_c}$$

V-V-CR

$$y_j = y_j(\vec{a}(\vec{x}, W, \vec{b}))$$

$$= S_j(\vec{a})$$

$$\vec{a}(\vec{x}, W, \vec{b}) = W\vec{x} + b$$

(Done)

7) We Know,

$$\vec{a}(\vec{x}, W, \vec{b}) = W\vec{x} + \vec{b}$$

Componant Notation:

$$a_i(\vec{x}, W, \vec{b}) = \left[\sum_l W_{il} x_l\right] + b_i$$

Now,

$$\frac{\partial a_i}{\partial b_c} = \frac{\partial}{\partial b_c}\left[\sum_l W_{il} x_l\right] + \frac{\partial}{\partial b_c} b_i$$

$$= 0 + \delta_{ic} = \delta_{ic}$$

$$\frac{\partial a_i}{\partial W_{ab}} = \frac{\partial}{\partial W_{ab}}\left[\sum_l W_{il} \cdot x_l\right] + \frac{\partial}{\partial W_{ab}} b_i$$

$$= \sum_l \delta_{ia} \cdot \delta_{lb} x_l + 0$$

$$= \delta_{ia} x_b$$

(Done)

Holiday Exercise

9) We Know,

$$S_i(\vec{x}) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Now,

$$\sum_i S_i(\vec{x}) = \sum_i \cancel{}$$

$$= \sum_i \cdot \frac{e^{x_i}}{\sum_j e^{x_j}} = \frac{\sum_i e^{x_i}}{\sum_j e^{x_j}} = 1,$$

(Ans)

10) We Know,

$$L(\vec{y}, \vec{t}) = -\sum_i t_i \ln(y_i)$$

$$L(\vec{y}, \vec{t}) \geq 0$$

Now, $L(\vec{y}, \vec{t}) = -\sum_i t_i \ln(y_i)$ | $y_i = S_i(\vec{x})$ & $S_i(\vec{x}) \in [0, 1]$

$$= -\sum_i \underbrace{t_i}_{>0} \underbrace{\ln(y_i)}_{<0}$$

$$= -\underbrace{\left[\sum_i t_i \ln(y_i)\right]}_{<0}$$

$$\geq 0.$$

(Proved)

16) a) We Know,

$$f'(x)\Big|_{x_0} = \lim_{\epsilon \to 0} \frac{f(x_0+\epsilon) - f(x_0)}{\epsilon}$$

here,
$$f(x) = x$$

So, $$f'(x)\Big|_{x_0} = \lim_{\epsilon \to 0} \frac{x_0 + \epsilon - \cancel{x_0}}{\epsilon}$$

$$= 1. \quad \text{(Ans)}$$

b) Local approximation of functions by linear functions.

c) We Know,

$$f(\vec{x}) \approx f(\vec{x_0}) + \vec{\nabla} f\Big|_{\vec{x_0}}^{T} \cdot (\vec{x} - \vec{x_0})$$

Given, $$f(\vec{x}) = x_1^2 + x_2^2$$

$$\vec{x_0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \vec{x_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\frac{\partial f}{\partial x_1} = 2x_1 \quad ; \quad \frac{\partial f}{\partial x_2} = 2x_2$$

So, $$\vec{\nabla} f = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

For, $\vec{x}_0$

Now,
$$f(\vec{x}) = f(\vec{x}_0) + \vec{\nabla} f\Big|_{x_0}^{T} \cdot (\vec{x} - \vec{x}_0)$$

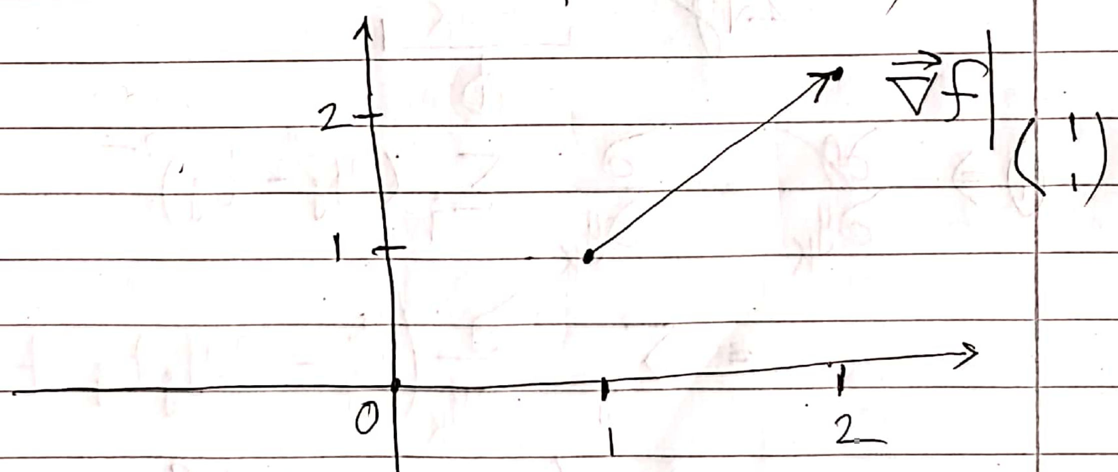
$$= f\left(\begin{pmatrix}0\\0\end{pmatrix}\right) + \begin{pmatrix}2x_1\\2x_2\end{pmatrix}\Bigg|_{\begin{pmatrix}0\\0\end{pmatrix}}^{T} \cdot \left(\vec{x} - \begin{pmatrix}0\\0\end{pmatrix}\right)$$

$$= 0 + \begin{pmatrix}0\\0\end{pmatrix} \cdot \left(\vec{x} - \begin{pmatrix}0\\0\end{pmatrix}\right) = 0$$

for, $\vec{x}_1$,
$$f(\vec{x}) = f\left(\begin{pmatrix}1\\1\end{pmatrix}\right) + \begin{pmatrix}2x_1\\2x_2\end{pmatrix}\Bigg|_{\begin{pmatrix}1\\1\end{pmatrix}}^{T} \cdot \left(\vec{x} - \begin{pmatrix}1\\1\end{pmatrix}\right)$$

$$= 1+1 + \begin{pmatrix}2\\2\end{pmatrix}^{T} \cdot \left(\vec{x} - \begin{pmatrix}1\\1\end{pmatrix}\right).$$

d)



Direction of strongest increase at point

$\vec{X}_0 = \vec{\nabla} f\Big|_{\vec{X}_0}$;

For, $\vec{x}_1 = \begin{pmatrix}1\\1\end{pmatrix}$
$\Rightarrow \vec{\nabla} f\Big|_{\vec{x}_1} = \begin{pmatrix}2\\2\end{pmatrix}$

For, $\vec{x}_0 = \begin{pmatrix}0\\0\end{pmatrix}$
$\therefore \vec{\nabla} f\Big|_{\begin{pmatrix}0\\0\end{pmatrix}} = \begin{pmatrix}0\\0\end{pmatrix}$

17) We Know,

$$\vec{y} = W\vec{x} + \vec{b}$$

$$\vec{L}(\vec{y}, \vec{t}) = \sqrt{\sum_j (y_j - t_j)^2}$$

Now,

$$\frac{\partial L}{\partial y_K} = \frac{\partial}{\partial y_K}\left[\sqrt{\sum_j (y_j - t_j)^2}\right]$$

Let, $g = \sum_j (y_j - t_j)^2$

$$= \frac{\partial \sqrt{g}}{\partial y_K}$$

$$= \frac{\partial \sqrt{g}}{\partial g} \cdot \frac{\partial g}{\partial y_K}$$

$$= \frac{1}{2\sqrt{g}} \cdot \underbrace{\frac{\partial g}{\partial y_K}}_{1}$$

$$1 \Rightarrow \frac{\partial g}{\partial y_K} = \frac{\partial}{\partial y_K} \sum_j (y_j - t_j)^2$$

$$= \sum_j \frac{\partial}{\partial y_K}\left(y_j^2 - 2y_j t_j + t_j^2\right)$$

$$= \sum_j (2y_j - 2t_j)\,\delta_{jK}$$

$$= 2y_K - 2t_K$$

Scanned with CamScanner

So, $$\frac{\partial \tilde{L}}{\partial y_K} = \frac{1}{2\sqrt{\sum_j (y_j - t_j)^2}} \cdot 2(y_K - t_K)$$

$$= \frac{y_K - t_K}{\sqrt{\sum_j (y_j - t_j)^2}} \quad \text{(Ans)}$$

18) Given,

$$a_i^{(L)} = \sin(a_i^{(L-1)})$$

So, $$\frac{\partial a_i^{(L)}}{\partial a_j^{(L-1)}} = \frac{\partial}{\partial a_j^{(L-1)}} \cdot \sin(a_i^{(L-1)})$$

$$= \cos(a_i^{(L-1)}) \cdot \delta_{ij} \quad \text{(Ans)}$$

19) We Know, $A = XW + b$

Given, $$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -1 \end{bmatrix}$$

Now, $$A^{(1)} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 1+2+3 & 1+2+3 \\ 1+0-1 & 1+0+(-1) \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 6 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ -2 & -2 \end{bmatrix}$$

$$\text{Relu} = g\left(A^{(1)}\right)$$

$$= \begin{bmatrix} 4 & 4 \\ 0 & 0 \end{bmatrix}$$

$$A^{(2)} = \begin{bmatrix} 4 & 4 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 8 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ -2 & -2 \end{bmatrix}$$

CNN 2

Given,

Input Image dimension = $28 \times 28 \times 3$

Assuming it, a 5-class classifier problem. A list of Keras layers are-

i) Conv2D (32, (3, 3), activation='relu', input_shape=(28, 28, 3))
ii) MaxPooling2D ((2, 2), 2)
iii) Flatten ( )
iv) Dense (100, activation = 'relu').
v) Dense (5 , activation = 'softmax')

CNN 1 Data Sample Size = (32, 32, 3)

filter Size = 3×3 , Stride $\Delta x \times \Delta y = 1 \times 1$; $C = 32$.

i) layer 1: $H = W = 1 + \frac{H' - f_y}{\Delta y} = 1 + \frac{32 - 3}{1} = 30.$

So, dimension = (H, W, C) = (30, 30, 32)   Param = $3 \times 3 \times 3 \times 32 + 32$
$= 892$

ii) Layer 2: (Relu)   dimension = (30, 30, 32)   Param = 0

iii) Layer 3: (Maxpooling)   dimension = (15, 15, 32)   Param = 0

iv) Layer 4: (Conv2D)   dimension = (12, 12, 64)   Param = $4 \times 4 \times 32 \times 64 + 64$

v) Layer 5: (Relu)   " = (12, 12, 64)   Param = 0

vi) Layer 6: (Maxpooling)   " = (6, 6, 64)   " = 0

CNN 3

a) False b) True c) False d) False

e) False f) True. g) True.

Exercise Pool 2

Probabilities :

$X = 1 \quad 1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 2 \quad 1 \quad 1 \quad 1 \quad 1$

$Y = 2 \quad 1 \quad 1 \quad 2 \quad 1 \quad 2 \quad 3 \quad 1 \quad 3 \quad 1 \quad 1 \quad 3$

$N = 12.$

a) $\#(Y=1) = 6 \quad | \quad \#(Y=2) = 3 \quad | \quad \#(Y=3) = 3$

b) $\#(X=1, Y=1) = 4 \quad | \quad \#(X=1, Y=2) = 2$

$\#(X=1, Y=3) = 2$

c) $P(X=1, Y=1) = \dfrac{\#(X=1, Y=1)}{N} = \dfrac{4}{12} = \dfrac{1}{3}$

$P(X=1, Y=2) = \dfrac{\#(X=1, Y=2)}{N} = \dfrac{2}{12} = \dfrac{1}{6}$

$P(X=1, Y=3) = \dfrac{\#(X=1, Y=3)}{N} = \dfrac{2}{12} = \dfrac{1}{6}$

d) $P(X=1 | Y=1) = \dfrac{\#(X=1, Y=1)}{\#(Y=1)} = \dfrac{4}{6}$

$P(X=1 | Y=2) = \dfrac{\#(X=1, Y=2)}{\#(Y=2)} = \dfrac{2}{3}$

$P(X=1 | Y=3) = \dfrac{\#(X=1, Y=3)}{\#(Y=3)} = \dfrac{2}{3}$

e) $P(Y=1 | X=1) = \dfrac{\#(Y=1, X=1)}{\#(X=1)} = \dfrac{4}{8}$

f) $P(X=1 | Y \neq 1) = \dfrac{\#(X=1, Y \neq 1)}{\#(Y \neq 1)} = \dfrac{4}{6}$

Confusion Matrix:

$$\begin{pmatrix} 10 & 60 & 10 \\ 0 & 80 & 0 \\ 0 & 20 & 60 \end{pmatrix}$$

a) $N_r = \cancel{240} \sum_{i,j} C_{ij} = 240$

b) $P(\hat{y}=1 \mid \hat{t}=2) = \dfrac{\#(\hat{y}=1, \hat{t}=2)}{\#(\hat{t}=2)} = \dfrac{0}{80} = 0$

c) $P(\hat{t}=1) = \dfrac{\#(\hat{t}=1)}{N_r} = \dfrac{80}{240} = \dfrac{1}{3}$

$P(\hat{t}=2) = \dfrac{\#(\hat{t}=2)}{N_r} = \dfrac{80}{240} = \dfrac{1}{3}$

$P(\hat{y}=2) = \dfrac{\#(\hat{y}=2)}{N_r} = \dfrac{160}{240} = \dfrac{2}{3}$

d) Probability of an incorrect classification when restricting ourselves to classifier outputs of 2

$\Rightarrow P(\hat{y} \neq \hat{t} \mid \hat{y}=2) = \cancel{\dfrac{\#(\hat{y}\neq 2, \hat{t}=2)}{\#(\hat{t}=2)}} = \dfrac{\#(\hat{y} \neq \hat{t}, \hat{y}=2)}{\#(\hat{y}=2)}$

$= \cancel{\dfrac{0}{80}} = \dfrac{60+20}{160} = \dfrac{1}{2}$

e) Probability of an incorrect classification

$\Rightarrow P(\hat{y} \neq \hat{t}) =$

$= E = \dfrac{\#(\hat{y} \neq \hat{t})}{N_r}$

$= \dfrac{90}{240}$

(Ans)

Binary Classification 1

Given,

$$\#(\hat{t}=2) : \#(\hat{t}=1) = 1:9$$

So, $P_p = \frac{1}{10}$; $P_n = \frac{\#(\hat{t}=1)}{\#(\hat{t}=1)+\#(\hat{t}=2)} = \frac{9}{10}$

As, model will only output negative class, so,

$tpr = 0$; $fnr = 1 - tpr = 1$

$fpr = 0$; $tnr = 1 - fpr = 1$

& $Error = fpr \times P_n + fnr \times P_p$
$= 0 \times .9 + 1 \times .1 = 0.1$.

Binary Classification 2

Given,

Sensitivity $tpr = 0.9$ | Specificity $tnr = 0.6$

$$\#(\hat{t}=1) : \#(\hat{t}=2) = 1:1$$

So, $P_p = \frac{1}{1+1} = 0.5$; $P_n = 0.5$

$tnr + fpr = 1 \Rightarrow fpr = 1 - 0.6 = 0.4$

$tpr + fnr = 1 \Rightarrow fnr = 1 - 0.9 = 0.1$

$Error = fnr \times P_p + fpr \times P_n = 0.1 \times 0.5 + 0.4 \times 0.5$
$= 0.25$

$Accuracy = 1 - Error = 1 - .25 = 0.75$. (Ans)

Binary Classification 3

Given, $C = \begin{pmatrix} 70 & 10 \\ 50 & 30 \end{pmatrix}$

a) $tpr = P(\hat{y}=2 \mid \hat{t}=2) = \dfrac{\#(\hat{y}=2, \hat{t}=2)}{\#(\hat{t}=2)}$

$= \dfrac{30}{80}$

$tnr = P(\hat{y}=1 \mid \hat{t}=1) = \dfrac{\#(\hat{y}=1, \hat{t}=1)}{\#(\hat{t}=1)}$

$= \dfrac{70}{80}$

b) $fpr = P(\hat{y}=2, \hat{t}=1) = \underline{\hspace{3cm}}$

$= \dfrac{10}{80}$

$fnr = P(\hat{y}=1, \hat{t}=2)$

$= \dfrac{50}{80}$

c) $Precision = P(\hat{t}=2 \mid \hat{y}=2)$

$= \dfrac{30}{10+30}$

$= \dfrac{3}{4}$

Lec 10/11/12 : Classification Measures & Probability

Demonstration 1 & 2

$n = \boxed{13 \text{ Pair}}$

x	1	3	3	3	4	0	0	0	0	0	1	3	1
y	2	3	3	1	5	3	0	2	2	2	1	1	0

i) $P(X=3) = 4/13$

ii) $P(X=3, Y=3) = 2/13$

iii) $P(X=3 \mid Y=3) = 2/3$

iv) $P(Y=3) = 3/13$

v) $P(X=3, Y=3) = 2/13$

vi) $P(Y=3 \mid X=3) = 2/4$.

Scanned with CamScanner

y = Model Output

t = Target/original Label

Confusion Matrix

$$
\begin{array}{c c c c c}
 & \hat{y}=1 & \hat{y}=2 & \hat{y}=3 & \\
C = & \left(\begin{array}{cccc} C_{11} & C_{12} & C_{13} & \cdots \\ C_{21} & C_{22} & C_{23} & \cdots \\ C_{31} & C_{32} & C_{33} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array}\right) & & & \begin{array}{c} \hat{t}=1 \\ \hat{t}=2 \\ \hat{t}=3 \end{array}
\end{array}
$$

Accuracy : $\left(\sum_{i=1}^{K} C_{ii}\right)/N$

Classification Error : $1 -$ Accuracy.

Elementary Probabilities : $P(\hat{y}=K) = \frac{\sum_j C_{jK}}{N}$

$$P(\hat{t}=K) = \left(\sum_j C_{Kj}\right)/N$$

Joint Probabilities :

$$P(\hat{y}=K, \hat{t}=j) = \frac{C_{jK}}{N}$$

Conditional Probabilities :

$$P(\hat{y}=K \mid \hat{t}=j) = \frac{P(\hat{y}=K, \hat{t}=j)}{P(\hat{t}=j)}$$

Exercise :

$$C = \begin{pmatrix} 5 & 3 & 2 & 0 \\ 1 & 5 & 1 & 3 \\ 0 & 1 & 7 & 2 \\ 4 & 1 & 0 & 5 \end{pmatrix}$$

I) Samples = 40 | Classes = 4

II) Classification error $= 1 - \frac{22}{40} = \frac{18}{40}$

III) Probability of incorrectly classifying Samples from class 4 is

$$P(\hat{y} \neq 4 \mid \hat{t} = 4) = \frac{4+1+0}{4+1+0+5} = \frac{5}{10} = 0.5$$

IV) Probability of wrongly classifying a sample from class 2 as class 3.

$$P(\hat{y} = 3 \mid \hat{t} = 2) = \frac{1}{1+5+1+3} = \frac{1}{10}.$$

V) Probability of correct classifications for samples from class 1.

$$P(\hat{y} = 1 \mid \hat{t} = 1) = \frac{5}{10} = 0.5.$$

Scanned with CamScanner

Binary Classification

1 = Negative | 2 = Positive

$$\text{Confusion Matrix, } C = \begin{array}{cc} \hat{y}=1 & \hat{y}=2 \end{array} \begin{pmatrix} tn & fp \\ fn & tp \end{pmatrix} \begin{array}{c} \hat{t}=1 \\ \hat{t}=2 \end{array}$$

true negative (tn) = Correctly Predicted No Event Values $= (\hat{y}=1, \hat{t}=1)$

false negative (fn) = Incorrectly " No " " $= (\hat{y}=1, \hat{t}=2)$

true positive (tp) = Correctly " Event Values $= (\hat{y}=2, \hat{t}=2)$

false positive (fp) = Incorrectly " " " $= (\hat{y}=2, \hat{t}=1)$

negative rate $P_n = P(\hat{t}=1) = \dfrac{\text{Negative Values}}{N}$

positive rate $P_p = P(\hat{t}=2) = \dfrac{\text{Positive Values}}{N}$

TN rate $tnr = P(\hat{y}=1 \mid \hat{t}=1)$

FN rate $fnr = P(\hat{y}=1 \mid \hat{t}=2)$

TP rate $tpr = P(\hat{y}=2 \mid \hat{t}=2)$

FP rate $fpr = P(\hat{y}=2 \mid \hat{t}=1)$

$$tnr + fpr = 1$$

$$fnr + tpr = 1$$

Classification error in terms of Conditional Probabilities.

We Know,

$$P(\hat{y} | \hat{t}) = \frac{P(\hat{y}, \hat{t})}{P(\hat{t})}$$

$$\Rightarrow P(\hat{y}, \hat{t}) = P(\hat{y} | \hat{t}) . P(\hat{t})$$

Now, classification error

$$E = P(\hat{y} \neq \hat{t})$$

$$= P(\hat{y}=1, \hat{t}=2) + P(\hat{y}=2, \hat{t}=1)$$

$$= P(\hat{y}=1 | \hat{t}=2) . P(\hat{t}=2) + P(\hat{y}=2 | \hat{t}=1) . P(\hat{t}=1)$$

$$= fnr . P_p + fpr . P_n$$

Some Metrics:

Sensitivity / Recall = tpr

Specificity / Selectivity = tnr

$$\text{Precision} = \frac{tp}{fp + tp} = \frac{tpr}{\frac{P_n}{P_p} . fpr + tpr}$$

$\cancel{P(\hat{y}=2 | \hat{t}=2)}$

$$= P(\hat{t}=2 | \hat{y}=2)$$

Ex 1 : Compute tpr, fpr, tnr, fnr, E

A classifier hase,

$$\text{precision} = 0.8$$
$$\text{fpr} = 0.1$$

$$N_{\hat{t}=2} : N_{\hat{t}=1} = 0.2 : 0.8$$

So, $P_p = \dfrac{0.2}{0.2+0.8} = 0.2$

$P_n = \dfrac{0.8}{1} = 0.8$

We Know,

$$\text{precision} = \frac{\text{tpr}}{\frac{P_n}{P_p} \cdot \text{fpr} + \text{tpr}}$$

$$\Rightarrow 0.8 \times \left(\frac{0.8}{0.2} \times 0.1 + \text{tpr} \right) = \text{tpr}$$

$$\Rightarrow 0.32 = \text{tpr} - 0.8\,\text{tpr}$$

$$\Rightarrow \text{tpr} = 0.32/0.2 = 1.6$$

$$\text{tnr} + \text{fpr} = 1$$
$$\Rightarrow \text{tnr} = 1 - \text{fpr} = 1 - 0.1 = 0.9$$

$$\text{fnr} + \text{tpr} = 1$$
$$\Rightarrow \text{fnr} = 1 - \text{tpr} = (-0.6)$$

Error $E = \text{fnr} \cdot P_p + \text{fpr} \cdot P_n$

$= -0.6 \times 0.2 + 0.1 \times 0.8$

$= -0.12 + 0.08 = -0.04$ (Ans)

Ex 2:

Given,

$$\text{Precision} = 0.8$$

$$\text{Recall } tpr = 0.5$$

$$\because P_n = \frac{0.8}{0.2+0.8} = 0.8 \quad ; \quad P_p = 0.2$$

Now, $tpr + fnr = 1 \Rightarrow fnr = 0.5$

$$\text{precision} = \frac{tpr}{\frac{P_n}{P_p} fpr + tpr}$$

$$\Rightarrow 0.8 \times \left(\frac{0.8}{0.2} \times fpr + 0.5 \right) = 0.5$$

$$\Rightarrow 3.2\, fpr + 0.4 = 0.5$$

$$\Rightarrow fpr = 0.1 / 3.2 = 0.03125$$

$$fpr + tnr = 1$$

$$\Rightarrow tnr = 1 - fpr = 0.96875$$

Error : $E = fpr \times P_n + fnr \times P_p$

$$= 0.03125 \times 0.8 + 0.5 \times 0.2$$

$$= 0.125.$$

(Ans)

Example 3:

$$C = \begin{pmatrix} 30 & 1 \\ 20 & 20 \end{pmatrix}$$

Here,

$$N = 71$$

positive rate: $P_p = P(\hat{t} = 2) = \frac{40}{71}$

Negative " $P_n = P(\hat{t} = 1) = \frac{31}{71}$

Error $E = 1 - \frac{50}{71} = \frac{21}{71}$

$$fnr = P(\hat{Y} = 1 \mid \hat{t} = 2) = \frac{20}{40} = \frac{1}{2}$$

$$fpr = P(\hat{Y} = 2 \mid \hat{t} = 1) = \frac{1}{31}$$

$$Recall = tpr = P(\hat{Y} = 2 \mid \hat{t} = 2)$$

$$= \frac{20}{40} = \frac{1}{2}$$

$$Precision = P(\hat{t} = 2 \mid \hat{Y} = 2)$$

$$= \frac{20}{21}$$

Calculate ~~~~ Trainable Parameters :

Input_Image $= (h \times W \times C)$

filter $= (f_x \times f_y)$

No. of filter $= n$

Convolution Layer :

Param $= [f_x \times f_y \times \text{Input\_channel} \times n] + n$

Pooling Layer :

Reduces the Input dimension by $\left(\frac{1}{2}\right)$.

No trainable Parameters.

Fully-connected Layer :

Param $= (\text{Input\_dim} + 1) \times \text{Output\_units}$

Dimension changed to output-units.

filter_shape = 3x3

VGG-16 Architecture (224 x 224 x 3)

Stride 1
filters 64

Block 1	Output Shape	Parameter
conv2D_1	(_, 224, 224, 64)	3x3x3x64+64 = 1792
Conv2D_2	(_, 224, 224, 64)	3x3x64x64 + 64 = 36928

Block 2		
max_pooling2d_1	(_, 112, 112, 64)	0
conv2D_3	(_, 112, 112, 128)	3x3x64x128 + 128 = 73856
conv2D_4	(_, 112, 112, 128)	3x3x ~~64~~ 128x128 + 128 = 147584

Block 3		
max_pooling2D_2	(_, ~~64~~ 56, ~~64~~ 56, 128)	0
Conv_2D_5	(_, 56, 56, 256)	3x3x128x256 + 256 = 295168
conv2D_6	(_, 56, 56, 256)	3x3x256x256 + 256 = 590080
conv2D_7	(_, 56, 56, 256)	Same = 590080

Block 4

max-pooling2D-3	(-, 28, 28, ~~512~~ 256)	
conv2D-8	(-, 28, 28, 512)	$3\times3\times256\times512+512$ = 1180160
conv2D-9	(")	$3\times3\times512\times512+512$ = 2359808
conv2D-10	(")	2359808

Block 5:

max-pooling2D-4	(-, 14, 14, 512)	0
conv2D-11	(")	2359808
conv2D-12	(")	"
conv2D-13	(")	"

Block 6:

max-pooling2D-5	(-, 7, 7, 512)	0
flatten	(-, $(7\times7\times512)=25088$)	0
dense-1	(-, 4096)	$(25088+1)\times4096$ = 102764544
dropout-1	(-, 4096)	= 0
dense-2	(-, 4096)	$(4096+1) * 4096$ = 16781312
dropout-2	(-, 4096)	= 0
dense-3	(-, 1000)	$= 4097\times1000$ = 4097000