Q1 LSTM Numerical —
(Forward Pass)

$$x_t = [0.5, -0.1]^T \simeq \mathbb{R}^2$$

$$h_{t-1} = [0.0, 0.1]^T \simeq \mathbb{R}^2$$

$$c_{t-1} = [0.2, -0.2]^T \simeq \mathbb{R}^2$$

$$W_{xi} = \begin{bmatrix} 0.5 & -0.3 \\ 0.4 & 0.1 \end{bmatrix}$$

$$W_{hi} = \begin{bmatrix} 0.1 & 0.2 \\ -0.2 & 0.05 \end{bmatrix}$$

Find out h_t and c_t.

$$W_{xf} = \begin{bmatrix} -0.4 & 0.2 \\ 0.3 & 0.3 \end{bmatrix}$$

$$W_{hf} = \begin{bmatrix} 0.05 & -0.1 \\ 0.2 & 0.1 \end{bmatrix}$$

$$W_{xo} = \begin{bmatrix} 0.3 & 0.25 \\ -0.2 & 0.2 \end{bmatrix}$$

$$W_{ho} = \begin{bmatrix} 0.15 & 0.05 \\ 0.1 & -0.2 \end{bmatrix}$$

$$W_{xg} = \begin{bmatrix} -0.5 & 0.4 \\ 0.2 & -0.3 \end{bmatrix}$$

$$W_{hg} = \begin{bmatrix} 0.2 & 0.1 \\ -0.1 & 0.05 \end{bmatrix}$$

Date

Ans:- $\boxed{i_t = \sigma(W_{hi}\, h_{t-1} + W_{xi}\, x_t)}$

$$W_{hi} \cdot h_{t-1} = \begin{bmatrix} 0.1 & 0.2 \\ -0.2 & 0.05 \end{bmatrix}_{2\times 2 \, (m\times n)} \cdot [0.0 \quad 0.1]^T_{1\times 2} \Rightarrow \begin{bmatrix} 0.0 \\ 0.1 \end{bmatrix}_{2\times 1 \, (n\times p)}$$

$$W_{hi} \cdot h_{t-1} = \begin{bmatrix} (0.1\times 0.0)+(0.1\times 0.2) \\ (-0.2\times 0.0)+(0.05\times 0.1) \end{bmatrix}_{2\times 2}$$

$$W_{hi} \cdot h_{t-1} = \begin{bmatrix} 0+0.02 \\ 0+0.005 \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.005 \end{bmatrix}_{2\times 1 \, (m\times p)}$$

$$W_{xi}\, x_t = \begin{bmatrix} 0.5 & -0.3 \\ 0.4 & 0.1 \end{bmatrix} \cdot \begin{bmatrix} 0.5 \\ -0.1 \end{bmatrix}$$

$$W_{xi}\, x_t = \begin{bmatrix} (0.5\times 0.5)+(-0.3\times -0.1) \\ (0.4\times 0.5)+(0.1\times -0.1) \end{bmatrix} = \begin{bmatrix} 0.25+0.03 \\ 0.2+(-0.01) \end{bmatrix}$$

$$W_{xi} \cdot x_t = \begin{bmatrix} 0.28 \\ 0.19 \end{bmatrix}$$

$$W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t = \begin{bmatrix} 0.02 \\ 0.005 \end{bmatrix} + \begin{bmatrix} 0.28 \\ 0.19 \end{bmatrix}$$

$$W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t = \begin{bmatrix} 0.3 \\ 0.195 \end{bmatrix}$$

$$\sigma \begin{bmatrix} 0.3 \\ 0.195 \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{-0.3}} \\ \frac{1}{1+e^{-0.195}} \end{bmatrix} = \begin{bmatrix} 0.57 \\ 0.54 \end{bmatrix}$$

Date

$$\boxed{g_t = \tanh(W_{hg} h_{t-1} + W_{xg} x_t)}$$

$$W_{hg} h_{t-1} = \begin{bmatrix} 0.2 & 0.1 \\ -0.1 & 0.05 \end{bmatrix} \begin{bmatrix} 0.0 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} (0.0 \times 0.2) + (0.1 \times 0.1) \\ (-0.1 \times 0.0) + (0.05 \times 0.1) \end{bmatrix}$$

$$= \begin{bmatrix} 0 + 0.01 \\ 0 + 0.005 \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.005 \end{bmatrix}$$

$$W_{xg} \cdot x_t = \begin{bmatrix} -0.5 & 0.4 \\ 0.2 & -0.3 \end{bmatrix} \begin{bmatrix} 0.5 \\ -0.1 \end{bmatrix}$$

$$= \begin{bmatrix} (0.5 \times -0.5) + (0.4 \times -0.1) \\ (0.2 \times 0.5) + (-0.3 \times -0.1) \end{bmatrix}$$

$$= \begin{bmatrix} -0.25 - 0.04 \\ 0.1 + 0.03 \end{bmatrix} = \begin{bmatrix} -0.29 \\ 0.13 \end{bmatrix}$$

$$W_{hg} h_{t-1} + W_{xg} x_t = \begin{bmatrix} 0.01 - 0.29 \\ 0.005 + 0.13 \end{bmatrix}$$

$$= \begin{bmatrix} -0.28 \\ 0.135 \end{bmatrix}$$

$$\tanh \begin{bmatrix} -0.28 \\ 0.135 \end{bmatrix} = \begin{bmatrix} -0.27 \\ 0.13 \end{bmatrix}$$

$$\boxed{f_t = \sigma\left(W_{hf}\, h_{t-1} + W_{xf}\, x_t\right)}$$

$$W_{hf}\, h_{t-1} = \begin{bmatrix} 0.05 & -0.1 \\ 0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 0.10 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 - 0.01 \\ 0 + 0.01 \end{bmatrix} = \begin{bmatrix} -0.01 \\ 0.01 \end{bmatrix}$$

$$W_{xf}\, x_t = \begin{bmatrix} -0.4 & 0.2 \\ 0.3 & 0.3 \end{bmatrix} \begin{bmatrix} 0.5 \\ -0.1 \end{bmatrix}$$

$$= \begin{bmatrix} -0.2 - 0.02 \\ 0.15 - 0.03 \end{bmatrix} = \begin{bmatrix} -0.22 \\ 0.12 \end{bmatrix}$$

$$W_{hf}\, h_{t-1} + W_{xf}\, x_t = \begin{bmatrix} -0.23 \\ 0.13 \end{bmatrix}$$

$$\sigma \begin{bmatrix} -0.23 \\ 0.13 \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{+0.23}} \\ \frac{1}{1+e^{-0.13}} \end{bmatrix} = \begin{bmatrix} 0.44 \\ 0.53 \end{bmatrix}$$

element-wise multiplication

$$\boxed{i_t \odot g_t} \Rightarrow \begin{bmatrix} 0.57 \\ 0.54 \end{bmatrix} \odot \begin{bmatrix} -0.27 \\ 0.13 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0.57 \times -0.27 \\ 0.54 \times 0.13 \end{bmatrix} = \begin{bmatrix} -0.15 \\ 0.07 \end{bmatrix}$$

$$\boxed{f_t \odot C_{t-1}} \Rightarrow \begin{bmatrix} 0.44 \\ 0.53 \end{bmatrix} \odot \begin{bmatrix} 0.2 \\ -0.2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0.088 \\ -0.106 \end{bmatrix}$$

Date

$$\boxed{c_t = f_t \odot c_{t-1} + i_t \odot g_t}$$

$$c_t = \begin{bmatrix} 0.088 \\ -0.106 \end{bmatrix} + \begin{bmatrix} -0.15 \\ 0.07 \end{bmatrix}$$

$$c_t = \begin{bmatrix} -0.06 \\ -0.03 \end{bmatrix}$$

$$\boxed{o_t = \sigma\left(W_{ho}\, h_{t-1} + W_{xo}\, x_t\right)}$$

$$W_{ho}\, h_{t-1} = \begin{bmatrix} 0.15 & 0.05 \\ 0.1 & -0.2 \end{bmatrix} \begin{bmatrix} 0.0 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 + 0.005 \\ 0 - 0.02 \end{bmatrix} = \begin{bmatrix} 0.005 \\ -0.02 \end{bmatrix}$$

$$W_{xo}\, x_t = \begin{bmatrix} 0.3 & 0.25 \\ -0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 0.5 \\ -0.4 \end{bmatrix}$$

$$= \begin{bmatrix} 0.15 + (-0.025) \\ -0.1 - 0.02 \end{bmatrix} = \begin{bmatrix} 0.125 \\ -0.12 \end{bmatrix}$$

$$W_{ho}\, h_{t-1} + W_{xo}\, x_t = \begin{bmatrix} 0.13 \\ -0.14 \end{bmatrix}$$

$$\sigma\begin{bmatrix}0.13\\-0.14\end{bmatrix} = \begin{bmatrix}\frac{1}{1+e^{-0.13}}\\ \frac{1}{1+e^{0.14}}\end{bmatrix} = \begin{bmatrix}0.53\\0.46\end{bmatrix}$$

$$\boxed{h_t = O_t \odot \tanh(C_t)}$$

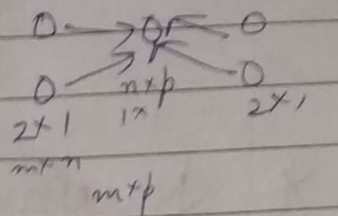
$$\tanh(C_t) = \tanh\begin{bmatrix}-0.06\\-0.03\end{bmatrix} = \begin{bmatrix}-0.05\\-0.029\end{bmatrix}$$

$$h_t = \begin{bmatrix}0.53\\0.46\end{bmatrix} \odot \begin{bmatrix}-0.05\\-0.029\end{bmatrix} = \begin{bmatrix}-0.026 \approx -0.03\\-0.013\end{bmatrix}$$

acts as PCA

Q1 $x \in \mathbb{R}^2$, $x = \begin{bmatrix} 2 \\ 0 \end{bmatrix}_{2\times1}$ **Autoencoder Numerical**

$h \in \mathbb{R}^1$
→ context vector



W_e, W_d size?

$$\left\{ \begin{array}{l} h = W_e \times x \\ 1\times1 \quad 1\times2 \quad 2\times1 \\ W_d = 2\times1 \end{array} \right.$$

Ans:- (a) $W_e = 1\times2 = \begin{bmatrix} 0.5 & -1.0 \end{bmatrix}_{1\times2}$

$W_d = 2\times1 = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}_{2\times1}$

$$\text{MSE Loss} = \frac{1}{2}\|\hat{x} - x\|_2^2$$

(b) Compute h, $\hat{x}$, L.

Ans:- $h = W_e x$

$$h = \begin{bmatrix} 0.5 & -1.0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$h = 2\times0.5 + \cancel{0\times(-1.0)} = 1.0$$

$$\hat{x} = W_d h$$

$$\hat{x} = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix} \times 1.0 = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix}$$

$L = 0.625$

$$L = \begin{bmatrix} 1.0 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$

$$L = \frac{1}{2} \times \left(\begin{bmatrix} -1 \\ 0.5 \end{bmatrix} \right)^2 = \frac{1}{2} \times \begin{bmatrix} 1 \\ 0.25 \end{bmatrix}$$

$$L = \begin{bmatrix} 0.5 \\ 0.125 \end{bmatrix}$$

$$L = 0.5 + 0.125$$

$$L = 0.625$$

(Total loss)

Attention Numerical

Page No.

Date

Q1 $h_1 = [1, 0, 1]$
$h_2 = [0, 1, 1]$
$h_3 = [1, 1, 0]$
$S_{t-1} = [1, 0, 1]$

Score function = dot product

$c_t = ?$

Ans :- $c_t = \sum_{j=1}^{T} \alpha_{t,j} \cdot h_j$

$c_t = \alpha_{t,1} \cdot h_1 + \alpha_{t,2} \cdot h_2 + \alpha_{t,3} \cdot h_3$

$$\alpha_{t,1} = \frac{\exp(\text{score}(S_{t-1}, h_1))}{(\exp(\text{score}(S_{t-1}, h_1)) + \exp(\text{score}(S_{t-1}, h_2)) + \exp(\text{score}(S_{t-1}, h_3)))}$$

:- $\text{Score}(S_{t-1}, h_1) = S_{t-1} \cdot h_1$

$= \underset{\substack{3\times1 \\ m\times n}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}} \cdot \underset{\substack{3\times1 \\ n\times p}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}} \quad = \underset{3\times1}{m\times p}$

$= 1\times1 + 0\times0 + 1\times1 = 2$

$\exp(\text{score}(S_{t-1}, h_1)) = \exp(2)$
$= 7.389$

$\text{score}(S_{t-1}, h_2) = S_{t-1} \cdot h_2$

$= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$= 0 + 0 + 1 = 1$

$= 2.718$

$score(s_{t-1}, h_3) = \begin{bmatrix}1\\0\\1\end{bmatrix} \cdot \begin{bmatrix}1\\1\\0\end{bmatrix}$

$\underset{(score)}{exp} = 2.718$

$\alpha_{t-1} = \frac{7.389}{7.389 + 2.718 + 2.718} = \frac{7.389}{12.825}$

$\alpha_{t-1} = 0.57$

:- $\alpha_{t,2} = ?$

$score(s_{t-1}, h_2) = \begin{bmatrix}1\\0\\1\end{bmatrix} \cdot \begin{bmatrix}0\\1\\1\end{bmatrix}$

$= 2.718$

$score(s_{t-1}, h_3) = 2.718$

$score(s_{t-1}, h_1) = 7.389$

$\alpha_{t-2} = \frac{2.718}{2.718 + 2.718 + 7.389}$

$\alpha_{t-2} = 0.211 \approx 0.2$

:- $\alpha_{t,3} = ?$

$\alpha_{t,3} = \frac{2.718}{12.825} = 0.211 \approx 0.2$

$\therefore\ C_t = ?$

$$\alpha_{t,1} \cdot h_1 = 0.57 \times \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.57 \\ 0 \\ 0.57 \end{bmatrix}$$

$$\alpha_{t,2} \cdot h_2 = 0.2 \times \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.2 \\ 0.2 \end{bmatrix}$$

$$\alpha_{t,3} \cdot h_3 = 0.2 \times \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0 \end{bmatrix}$$

$$C_t = \begin{bmatrix} 0.57 \\ 0 \\ 0.57 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.2 \\ 0.2 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.2 \\ 0 \end{bmatrix}$$

$$C_t = \begin{bmatrix} 0.77 \\ 0.4 \\ 0.77 \end{bmatrix} \approx \begin{bmatrix} 0.77 & 0.4 & 0.77 \end{bmatrix}^T$$

e.g. :-

Key	Value
[1.0, 0.5, 0.2]	"Action movie Die hard"
[0.3, 1.0, 0.1]	"Romance Titanic"
[0.9, 0.4, 0.8]	"Action movie John wick"

Similarity
(dot product)
Numerical
⇐

Q = "looking for a action movies"

⇓ word2vec

[1.0, 0.3, 0.5]

Ans:- Dot product

- Q ⊙ Key-1 ⇒ [1.0 0.3 0.5] ⊙ [1.0 0.5 0.2]
  (Rank-2) ⇒ 1 + 0.15 + 0.1 = 1.25

- Q ⊙ Key-2 ⇒ [1.0 0.3 0.5] ⊙ [0.3 1.0 0.1]
  (Rank-3) ⇒ 0.3 + 0.3 + 0.05 = 0.65

- Q ⊙ Key-3 ⇒ [1.0 0.3 0.5] ⊙ [0.9 0.4 0.8]
  (Rank-1) ⇒ 0.9 + 0.12 + 0.4 = 1.42

Transformer (self-attention) — Scaled dot product Numerical

Q1 I/P ⇒ "Playing Outside"
Find z_1 and z_2?

Given :- For "Playing"
$q_1 = [0.212 \quad 0.04 \quad 0.63 \quad 0.36]^T$
$K_1 = [0.31 \quad 0.84 \quad 0.963 \quad 0.57]^T$
$V_1 = [0.36 \quad 0.83 \quad 0.1 \quad 0.38]^T$

:- For "Outside"
$q_2 = [0.1 \quad 0.14 \quad 0.86 \quad 0.77]^T$
$K_2 = [0.45 \quad 0.94 \quad 0.73 \quad 0.58]^T$
$V_2 = [0.31 \quad 0.36 \quad 0.19 \quad 0.72]^T$

Ans :-

Playing	Outside
Score ⇒ $q_1 \cdot K_1$	Score ⇒ $q_2 \cdot K_1$
$\Rightarrow \begin{bmatrix} 0.212 \\ 0.04 \\ 0.63 \\ 0.36 \end{bmatrix} \cdot \begin{bmatrix} 0.31 \\ 0.84 \\ 0.963 \\ 0.57 \end{bmatrix}$	$\Rightarrow \begin{bmatrix} 0.1 \\ 0.14 \\ 0.86 \\ 0.77 \end{bmatrix} \cdot \begin{bmatrix} 0.31 \\ 0.84 \\ 0.963 \\ 0.57 \end{bmatrix}$

$q_1 \cdot K_1 = 0.06572$
$+ 0.0336 + 0.60669$
$+ 0.2052$

$q_1 \cdot K_1 = 0.91121$

Also, $q_1 \cdot K_2$

$$= \begin{bmatrix} 0.212 \\ 0.04 \\ 0.63 \\ 0.36 \end{bmatrix} \cdot \begin{bmatrix} 0.45 \\ 0.94 \\ 0.73 \\ 0.58 \end{bmatrix}$$

$q_1 \cdot K_2 = 0.0954 + 0.0376$
$+ 0.4599 + 0.2088$

$q_1 \cdot K_2 = 0.8017$

Scaling $\Rightarrow \dfrac{q_1 \cdot K_1}{\sqrt{K_1}} = \dfrac{0.91121}{\sqrt{4}}$

$= \dfrac{0.91121}{2}$

$= 0.455605$

$\dfrac{q_1 \cdot K_2}{\sqrt{K_2}} = \dfrac{0.8017}{2}$

$= 0.40085$

Softmax-1 $\Rightarrow \dfrac{e^{0.455605}}{e^{0.455605} + e^{0.40085}}$

$\Rightarrow \dfrac{1.577127}{1.577127 + 1.493093}$

$q_2 \cdot K_1 = 0.031 + 0.1176 +$
$1.823 + 0.4389$

$q_2 \cdot K_1 = 2.4105$

Also, $q_2 \cdot K_2 = \begin{bmatrix} 0.1 \\ 0.14 \\ 0.86 \\ 0.77 \end{bmatrix} \cdot \begin{bmatrix} 0.45 \\ 0.94 \\ 0.73 \\ 0.58 \end{bmatrix}$

$q_2 \cdot K_2 = 0.045 + 0.1316 +$
$0.6278 + 0.4466$

$q_2 \cdot K_2 = 1.251$

Scaling $\Rightarrow \dfrac{q_2 \cdot K_1}{\sqrt{K_1}} = \dfrac{2.4105}{2}$

$= 1.20525$

$\Rightarrow \dfrac{q_2 \cdot K_2}{\sqrt{K_2}} = 0.6255$

Softmax-1 $\Rightarrow \dfrac{e^{1.20525}}{e^{1.20525} + e^{0.6255}}$

$\Rightarrow \dfrac{3.33759}{3.33759 + 1.86918}$

$\Rightarrow \dfrac{3.33759}{5.20677}$

$\Rightarrow 0.64100$

Softmax-2 $\Rightarrow \dfrac{e^{0.6255}}{e^{1.20525} + e^{0.6255}}$

$\Rightarrow 0.35899$

$$\Rightarrow \frac{1.577127}{3.07022}$$

$$\Rightarrow 0.513685$$

Also,

$$\Rightarrow \frac{e^{0.40085}}{3.07022}$$

$$\Rightarrow \frac{1.493093}{3.07022}$$

softmax-2 $\Rightarrow 0.486314$

$z_1 \Rightarrow$ softmax1 $\times v_1$ + softmax2 $\times v_2$

$$z_1 \Rightarrow 0.513685 \times \begin{bmatrix} 0.36 \\ 0.83 \\ 0.1 \\ 0.38 \end{bmatrix}$$

$$+$$

$$0.486314 \times \begin{bmatrix} 0.31 \\ 0.36 \\ 0.19 \\ 0.72 \end{bmatrix}$$

$$z_1 \Rightarrow \begin{bmatrix} 0.1836 \\ 0.4233 \\ 0.051 \\ 0.1938 \end{bmatrix} + \begin{bmatrix} 0.1488 \\ 0.1728 \\ 0.0912 \\ 0.3456 \end{bmatrix}$$

$$z_1 \Rightarrow \begin{bmatrix} 0.332 & 0.596 & 0.142 & 0.539 \end{bmatrix}^T$$

$z_2 \Rightarrow$ softmax1 $\times v_1$ + softmax2 $\times v_2$

$$z_2 = 0.64100 \times \begin{bmatrix} 0.36 \\ 0.83 \\ 0.1 \\ 0.38 \end{bmatrix}$$

$$+$$

$$0.35899 \times \begin{bmatrix} 0.31 \\ 0.36 \\ 0.19 \\ 0.72 \end{bmatrix}$$

$$z_2 = \begin{bmatrix} 0.2307 \\ 0.5320 \\ 0.0641 \\ 0.2435 \end{bmatrix} + \begin{bmatrix} 0.1112 \\ 0.1292 \\ 0.0682 \\ 0.2584 \end{bmatrix}$$

$$z_2 = \begin{bmatrix} 0.3419 & 0.6612 & 0.1323 & 0.5019 \end{bmatrix}^T$$

Positional Embedding Numerical

Q1 I am a robot

$P_E : d = \mathbb{R}^4 \Rightarrow 4$

		$i=0$		$i=1$	
		sin	cos	sin	cos
I	0	P_{00}	P_{01}	P_{02}	P_{03}
am	1	P_{10}	P_{11}	P_{12}	P_{13}
a	2	P_{20}	P_{21}	P_{22}	P_{23}
robot	3	P_{30}	P_{31}	P_{32}	P_{33}

Notation
⇓
$P_{pos.0}$
$P_{pos.1}$
$P_{pos.2}$
$P_{pos.3}$

Find these four pos. embedding vectors for four positions?

Degree → radian

↳ $1 deg. \times \frac{\pi}{180} = 0.01745 rad.$

Date

Ans:- For pos. - 0,

- All the values currently are in degrees but later can be converted to radians.

⇓

$$P_{00} = \sin\left(\frac{pos}{100^{\wedge}\left(\frac{2i}{d}\right)}\right)$$

$$P_{00} = \sin\left(\frac{0}{100^{\wedge}\left(\frac{2(0)}{4}\right)}\right) = \sin(0) = 0$$

(in degrees)

↳ convert later into radians

$$P_{01} = \cos\left(\frac{0}{100^{\wedge}\left(\frac{2(0)}{4}\right)}\right) = \cos(0) = 1$$

$$P_{02} = \sin\left(\frac{0}{100^{\wedge}\left(\frac{2(1)}{4}\right)}\right) = 0$$

$$P_{03} = \cos\left(\frac{0}{100^{\wedge}\left(\frac{2(1)}{4}\right)}\right) = 1$$

$\boxed{0 \mid 1 \mid 0 \mid 1}$ ⇒ For pos. - 0

Now, for pos. - 1,

$$P_{10} = \sin\left(\frac{1}{100^{\wedge}\left(\frac{2(0)}{4}\right)}\right) = \sin(1) = 0.017 \approx 0$$

$$P_{11} = \cos\left(\frac{1}{100^{\wedge}\left(\frac{2(0)}{4}\right)}\right) = \cos(1) = 0.99 \approx 1$$

$$P_{12} = \sin\left(\frac{1}{100^{\wedge}\left(\frac{2(1)}{4}\right)}\right)$$

$= \cancel{\sin^{\wedge}\left(\frac{1}{50}\right)}$

$= \cancel{\sin^{\wedge}(0.02)}$

$= \cancel{0.0003}$

$$= \sin\left(\frac{1}{100^{(0.5)}}\right)$$

$$= \sin\left(\frac{1}{10}\right) = 0.0017$$

$P_{13} = \cos\left(\frac{1}{100^{(0.5)}}\right) = \cos\left(\frac{1}{10}\right) = 0.999$

Now, for pos. -2,

$P_{20} = \sin\left(\frac{2}{100^{\wedge}(2(0)/4)}\right) = \sin\left(\frac{2}{1}\right) = 0.0348$

$P_{21} = \cos\left(\frac{2}{100^{\wedge}(2(0)/4)}\right) = \cos(2) =$
0.99

$P_{22} = \sin\left(\frac{2}{100^{\wedge}(2(1)/4)}\right) = \sin(0.2)$
$= 3.49 \times 10^{-3}$
$= \frac{3.49}{100\,000}$
$= 0.00349$

$P_{23} = \cos\left(\frac{2}{100^{\wedge}(2(1)/4)}\right) = \cos(0.2)$
$= 0.99$

For pos. $-2 \Rightarrow$ $\boxed{0.0348 \mid 0.99 \mid 0.00349 \mid 0.99}$

$P_{30} = \sin\left(\frac{3}{100^{\wedge}(2(0)/4)}\right) = \sin(3)$
$= 0.05$

$P_{31} = \cos\left(\frac{3}{100^{\wedge}(2(0)/4)}\right) = \cos(3)$
$= 0.99$

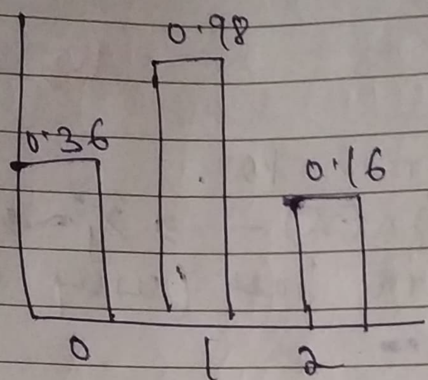
$P_{32} = \sin\left(\frac{3}{100^{\wedge}(2(1)/4)}\right) = \sin(0.3)$
$= 5.23 \times 10^{-3}$
$= 0.00523$

$P_{33} = \cos\left(\frac{3}{100^{\wedge}(2(1)/4)}\right) = \cos(0.3)$
$= 0.99$

For pos. $3 \Rightarrow$ $\boxed{0.05 \mid 0.99 \mid 0.005 \mid 0.99}$

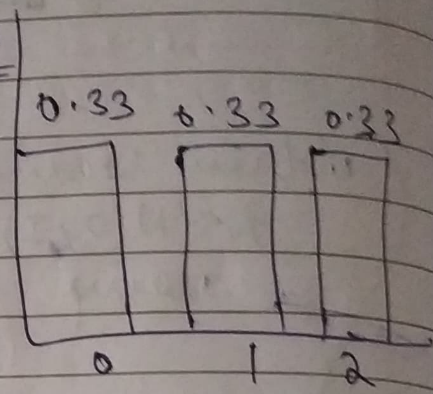
KL-example —

Q1 P



Binomial X distri.
(P)

Q



Uniform distri.
(Q)

(X ⇒ random variable
⇒ takes values 0, 1, 2

In P distri., takes 0 = prob. = 0.36
1 = " = 0.98
2 = " = 0.16)

Distri. X	0	1	2
$P(X)$	$\frac{9}{25}$	$\frac{12}{25}$	$\frac{4}{25}$
$Q(X)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Ans:- $D_{KL}(P\|Q) = \sum_{x\in X} P(X=x) \ln \frac{P(X=x)}{Q(X=x)}$

$$D_{KL}(P\|Q) = \frac{9}{25}\ln\left(\frac{9/25}{1/3}\right) + \frac{12}{25}\ln\left(\frac{12/25}{1/3}\right) + \frac{4}{25}\ln\left(\frac{4/25}{1/3}\right)$$

If DKL equal to 0, then P = Q (we don't want)

$$D_{KL}(P\|Q) = 0.0852996$$

more closer DKL to zero, more similar P & Q are.

Date

$$D_{KL}(Q \| P) = \frac{1}{3} \ln\left(\frac{1/3}{9/25}\right) + \frac{1}{3} \ln\left(\frac{1/3}{12/25}\right) + \frac{1}{3} \ln\left(\frac{1/3}{4/25}\right)$$

$$D_{KL}(Q \| P) \approx 0.09745$$

$$\therefore D_{KL}(P \| Q) \neq D_{KL}(Q \| P)$$

↳ KL-divergence is not symmetric (always).