Solution

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

A

Roll No.

Total No. of Pages

Name : Date :

Quiz I Quiz II/ Mid-Sem End-Semester Make-up Part :

Semester & Degree : Course No.

Question No.	1	2	3	4	5	6	7	8	9	10
Marks										

11	12	13	14	15	16	17	18	19	20	Total

Answer on both sides of the paper including the space below

Ques01) $X = 0, \; Y = 3, \; Z = 7$

W initial layer $= p = 0.1 + 0.1 \times 0 = 0.1$

$q = 0.2 - 0.1 \times 7 = -0.5$

given $W_{ij}^{(2)} = 0.1 + 0.1Z = 0.1 + 0.1 \times 7 = 0.8$

(a) Non linearity in the final layer is softMax

(1)

(b) given $x_1 = 0.2 \quad x_2 = 0.5 \qquad y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

(3) let a_1, a_2, a_3 be activation in hidden layer

we know that.

$a_1 = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-0.2\times 0.1}} = \underline{0.51} \quad \{z = p \times x\}$

$a_2 = \frac{1}{1+e^{-(0.2\times(-0.5)+0.5\times(0.1))}} = \frac{1}{1+e^{0}} = \underline{\underline{0.5674}}$

$a_3 = \frac{1}{1+e^{0.5\times(0.5)}} = \frac{1}{\ldots 0.5\times 0.5} = 0.48\ldots$

(c) output of y. $a_1^{(2)} = \frac{e^{-z_1^{(2)}}}{\sum_{i=1}^{i=3} -z_i^{(2)}} = \frac{0.3}{0.3+0.3+0.3}$

(4)

$$\hat{y} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

(d) error function =

(2)

$$\cancel{J = -\left\{\frac{1}{3}\log 0 + \frac{1}{3}\log 1 + \frac{1}{3}\log 0\right\}}$$

$$J = -\{0 \times \ln(0.33) + 1 \times \ln(0.33) + 0 \times \ln(0.33\}$$

$$J = -\{\ln(0.33)\}$$

$$J = 1.0996$$

Ans

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

A

Roll No. | | | | | | | | |

Total No. of Pages

Name :

Quiz I | Quiz II/ Mid-Sem | End-Semester | Make-up | Date :

Semester & Degree : | Course No. | Part :

Question No.	1	2	3	4	5	6	7	8	9	10
Marks										

11	12	13	14	15	16	17	18	19	20	Total

Answer on both sides of the paper including the space below

5 Backpropagation - $x=0, y=0, z=2$

$p = 0.1 \quad q = 0.2$

$$\frac{\partial J}{\partial p} = \frac{\partial J}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial z_1^{(2)}} \cdot \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(1)}}{\partial p}$$

$$\frac{\partial J}{\partial z^{(2)}} = \bar{y} - Y = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.33 \\ -0.67 \\ 0.33 \end{bmatrix}$$

$$\frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} = W_{ij} = 0.3$$

$$\frac{\partial z_2^{(2)}}{\partial a_1^{(1)}} = W_{ij} = 0.3$$

$$\frac{\partial z_3^{(2)}}{a a_1^{(1)}} = 0.3$$

$$\frac{\partial J}{\partial a_2} \times \frac{\partial \hat{y}}{\partial z_2} \times \frac{\partial z_2}{\partial a_1} \times \frac{\partial a_1}{\partial z_1}$$

$$= \begin{bmatrix} 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 \end{bmatrix} \begin{bmatrix} 0.333 \\ -0.667 \\ 0.33 \end{bmatrix}_{3\times1} \odot \begin{bmatrix} 0.2499 \\ 0.2499 \\ 0.2498 \end{bmatrix}_{3\times1}$$

3×1

$$g'(a) = \begin{bmatrix} a_1(1-a_1) \\ a_2(1-a_2) \\ a_3(1-a_3) \end{bmatrix} = \begin{bmatrix} 0.5049(1-0.5049) \\ 0.5225(1-0.5225) \\ 0.5249(1-0.5249) \end{bmatrix}$$

$$= \begin{bmatrix} 0.2499 \\ 0.2499 \\ 0.2499 \end{bmatrix}$$

$$\frac{\partial J}{\partial z} \cdot \frac{\partial z^{(2)}}{\partial a^{(1)}} \times \frac{\partial a^{(1)}}{\partial z^{(1)}} \, ?$$

$$\frac{\partial J}{\partial z_{10}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \frac{\partial J}{\partial p} = \frac{\partial J}{\partial z_{11}} \times \frac{\partial z_{11}}{\partial p} + \frac{\partial J}{\partial z_{12}} \times \frac{\partial z_{12}}{\partial p}$$

this is zero = 0

$$\frac{\partial J}{\partial p} = 0$$

similarly

$$\frac{\partial J}{\partial q} = \frac{\partial J}{\partial z_{12}} \times \frac{\partial z_{12}}{\partial q} + \frac{\partial J}{\partial z_{13}} \times \frac{\partial z_{13}}{\partial q}$$

$$\frac{\partial J}{\partial z_1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\frac{\partial J}{\partial q} = 0$$

for p

update $\Rightarrow$ $p_{new} = 0.1 - 0.01(0)$

$$= \underline{\underline{0.1}}$$

update for - q -

$$q_{new} = 0.2 - 0.01(0)$$

$$= 0.2$$

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

A

Roll No.

Name : | Total No. of Pages

Quiz I | Quiz II/ Mid-Sem | End-Semester | Make-up | Date :

Semester & Degree : | Course No. | Part :

Question No.	1	2	3	4	5	6	7	8	9	10
Marks										

11	12	13	14	15	16	17	18	19	20	Total

Answer on both sides of the paper including the space below

QUIZ-2 Q2:

$$P(W/\text{Data Points}) = \prod_{i=1}^{N \to \text{No. of Data points.}} \frac{P(\text{DataPoint}/W)\,P(W)}{\boxed{P(\text{Datapoints})}}$$

↳ Can be ignored because it is same for all cases/terms.

$$P(W/D) \simeq \prod_{i=1}^{N}\left[\frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{(y_i-\hat{y}_i)^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi\varepsilon^2}}\cdot e^{-\frac{(W-\mu)^T\varepsilon^{-1}(W-\mu)}{2}}\right]$$

Now we need to maximize $P(W/D)$, which is equivalent to minimizing $-\ln[P(W/D)]$ i.e. Negative Log Likelihood.

$$\Rightarrow \prod_{i=1}^{N} e^{-\frac{(y_i-\hat{y}_i)^2}{2\sigma^2}}\cdot e^{-\frac{(W-\mu)^T\varepsilon^{-1}(W-\mu)}{2}} \Rightarrow \text{Maximize or}$$

$$\Rightarrow \sum_{i=1}^{N}\left[\underbrace{\frac{(y_i-\hat{y}_i)^2}{2\sigma^2}}_{\text{Original Loss term}} + \underbrace{\frac{(W-\mu)^T\varepsilon^{-1}(W-\mu)}{2}}_{\text{Additional term}}\right] \Rightarrow \text{Minimize.}$$

$$\Rightarrow Loss = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_i-\hat{y}_i)^2}{2\sigma^2} + \frac{\cancel{N}}{\cancel{N}}\frac{(W-\mu)^T\Sigma^{-1}(W-\mu)}{2}$$

$$\Rightarrow Loss = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(y_i-\hat{y}_i)^2}{2\sigma^2}\right] + \boxed{\frac{(W-\mu)^T\Sigma^{-1}(W-\mu)}{2}} \Rightarrow \text{Additional term}$$

$$\frac{\partial(Loss)}{\partial W} = \underbrace{\frac{1}{N}\frac{\partial}{\partial W}\left[\frac{(y_i-\hat{y}_i)^2}{2\sigma^2}\right]}_{\text{Original update term}} + \underbrace{\frac{\partial}{\partial W}\left[\frac{(W-\mu)^T\Sigma^{-1}(W-\mu)}{2}\right]}_{\text{Correction term}}$$

$$\Rightarrow \text{Correction Term} : \frac{\cancel{2}\,\Sigma^{-1}(W-\mu)}{\cancel{2}}$$

$$\Rightarrow \boxed{\text{Correction term} = \Sigma^{-1}(W-\mu)}$$

$$\text{Where} : \Sigma = \begin{bmatrix} Var(\alpha) & Cov(\alpha,\beta) \\ Cov(\alpha,\beta) & Var(\beta) \end{bmatrix}$$

$$Cov(\alpha,\beta) = Corr(\alpha,\beta)\cdot\sqrt{Var(\alpha)\,Var(\beta)}$$

$$W = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} ; \quad \mu = \begin{bmatrix} \alpha_{mean/expected} \\ \beta_{mean/expected} \end{bmatrix}$$