Section A

Answer the below Questions

Multiclass Problem

Q1) Using the given dataset carry out **Online Training** with **Multi-Class SVM Criterion**:-

x_1	0	2	1
x_2	0	0	1
y	1	2	3

$$W_0 = \begin{bmatrix} 1 & -1.5 & 2 \\ -1 & 1 & 0.9 \\ 1 & 1 & 1 \end{bmatrix}$$

What will be the delta after the first sample ? *

1 point

$$\begin{bmatrix} 0.67 & 0 & 0.33 \end{bmatrix}$$

A

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

B

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0.33 & 0.67 \end{bmatrix}$$

C

None of the above

What will be the Loss after the Second sample? *

1 point

2

11.2

8

I don't know

What will be the weights after the Second sample ? * 1 point

$$\begin{bmatrix} 0.3 & 1 & 2 \\ 0.33 & -0.33 & 0.67 \\ 0.67 & -1 & -0.33 \end{bmatrix}$$

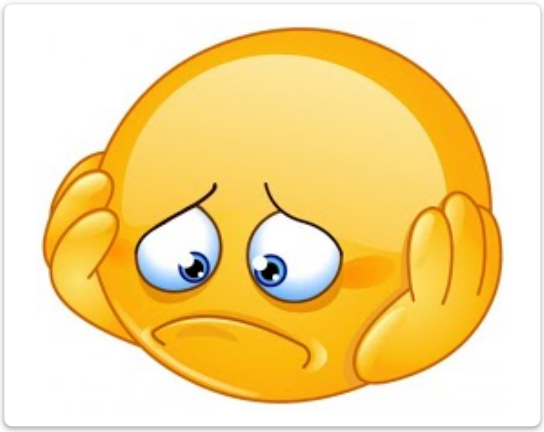
A

$$\begin{bmatrix} -0.3 & 1 & 2 \\ 0.33 & -0.33 & 0.67 \end{bmatrix}$$

B

$$\begin{bmatrix} 0.33 & -1.5 & 1.7 \\ 0.33 & 1 & 1.57 \\ 0.33 & 1 & 0.67 \end{bmatrix}$$

C



None of the above

What are the improvements that the Weston-Watkins SVM added to the Multi-Class Perceptron Loss ? (Choose all correct answers) * 1 point

- [x] Soft Margin and the ability to not care about some misclassified samples
- [ ] The ability to not use the Sign function as an activation function
- [x] Updating all the miss-classified nodes not only the most miss-classified
- [ ] Providing probability distribution to the output for better classification

Choose all correct answers * 1 point

- [x] Before Normalization the loss is very sensitive to changes in weights matrix and hard to optimize
- [x] In Batch Normalization , we use the epsilon to avoid division by zero
- [x] In Batch Normalization, Input dimension = Output dimension
- [x] To get mean and Variance, we use these equations:

$$\mu_j = \frac{1}{N} . \Sigma^N{}_{i=1} X_{i,j}$$

$$\sigma^2 = \frac{1}{N} . \Sigma^N{}_{i=1} (X_{i,j} - \mu_j)^2$$

What is the loss in the below Question? * 1 point

The following table shows the 5th example in 10 samples-dataset, the label of the example is 2. The scores of each output node is expressed in terms of x, **where x is a positive number &** $(x > 1)$

Node #	1	2	3	4
Score	x	$2x$	$x - 5$	$x^2 + x$

The loss value for this example is:

A. $-\ log(2\,x)$

B. $-\log\,(x^2\,+\,x)$

C. $-\,2x\,+\,log(e^x\,(1\,+\,e^{x^2}\,+\,e^{-5}\,+\,e^x))$

D. $-\,x^2\,-x\,+\,log(e^x\,(1\,+\,e^{x^2}\,+\,e^{-5}\,+\,e^x))$

E. $-\,log(1\,-\,1/2\,+\,2x)$

- [ ] A
- [ ] B
- [x] C
- [ ] D
- [ ] E

Section B

Answer the below Questions

Linear Regression problem

Q2) For the following dataset, Answer the following Questions:-

X	0	1	2	3
Y	0	0.81	0.95	0.31

$$W_0 = [0 \quad 0]$$

What is the delta after the first GD iteration? * 1 point

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

A

$$\begin{bmatrix} -0.91 \\ -0.51 \end{bmatrix}$$

B

$$\begin{bmatrix} -1.5 \\ 1 \end{bmatrix}$$

C

Find the Hessian matrix * 1 point

$$\begin{bmatrix} 6 & 2 \\ 10 & 2 \end{bmatrix}$$

A

$$\begin{bmatrix} 16 & 3 \\ 1 & 1 \end{bmatrix}$$

B

$$\begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}$$

C

Use the closed-form solution technique to find the weights * 1 point

$$\begin{bmatrix} 0.5 \\ -0.62 \end{bmatrix}$$

A

$$\begin{bmatrix} 0.65 \\ 0.9 \end{bmatrix}$$

B

$$\begin{bmatrix} 0.107 \\ 0.357 \end{bmatrix}$$

C

Choose all correct answers * 1 point

- [x] We use Regularization term to avoid overfitting
- [ ] We use Regularization term to avoid underfitting
- [x] $W = (D^T D + \lambda I)^{-1} D^T Y$
- [x] $L = log(1 + e^{-\hat{y}y}) + \frac{\lambda}{2}||w||^2$
- [x] Adopting the Regularization term we can compute the weights by
- [x] This is the logistic loss with Regularization