DDA3020 Machine Learning Spring 2023
Produced by TA Teams

Assignment 3

1 Problem 1

(a) Computational Graph:

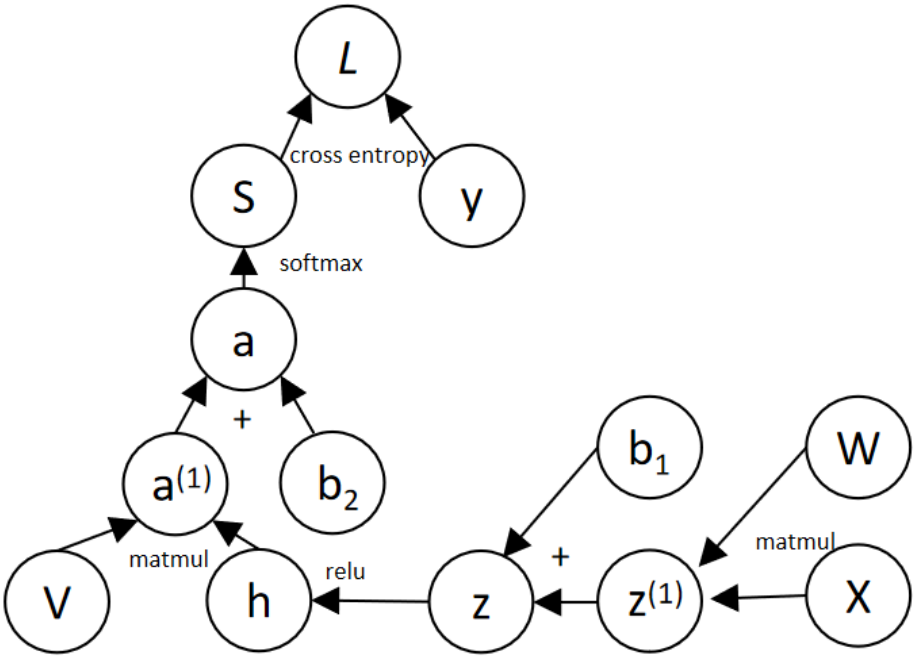

Figure 1: Computational Graph

(b) To compute $\boldsymbol{\delta}_1$, let $\mathcal{L} = \text{CrossEntropyWithLogits}\ (\boldsymbol{y}, \boldsymbol{a})$. Then

$$\delta_1 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{a}} = (\boldsymbol{p} - \boldsymbol{y})^\top$$

where $\boldsymbol{p} = Softmax(\boldsymbol{a})$. To compute δ_2, we have

$$\begin{aligned}
\boldsymbol{\delta}_2 = \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}} &= \frac{\partial \mathcal{L}}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{h}} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}} = \boldsymbol{\delta}_1 \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{h}} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}} \\
&= \boldsymbol{\delta}_1 \mathbf{U} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{z}} \text{ since } \boldsymbol{a} = \mathbf{U}\boldsymbol{h} + \boldsymbol{b}_2 \\
&= \boldsymbol{\delta}_1 \mathbf{U} \circ \text{ReLU}'(\boldsymbol{z}) \text{ since } \boldsymbol{h} = \text{ReLU}(\boldsymbol{z}) \\
&= \boldsymbol{\delta}_1 \mathbf{U} \circ H(\boldsymbol{h})
\end{aligned}$$

Now we compute the gradients wrt the parameters:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial \mathbf{U}} = \boldsymbol{\delta}_1 \boldsymbol{h}^\top$$
$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{b}_2} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{a}} \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{b}_2} = \boldsymbol{\delta}_1$$
$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}} \frac{\partial \boldsymbol{z}}{\partial \mathbf{W}} = \boldsymbol{\delta}_2 \boldsymbol{x}^\top$$
$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{b}_1} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{z}} \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{b}_1} = \boldsymbol{\delta}_2$$

2 Problem 2

(a)

Layer	Activation Volume Dimensions (memory)
INPUT	$32 \times 32 \times 1$
CONV5-10	$32 \times 32 \times 10$
POOL2	$16 \times 16 \times 10$
CONV5-10	$16 \times 16 \times 10$
POOL2	$8 \times 8 \times 10$
FC-10	10

(b) number of parameters:

Layer	Number of parameters
INPUT	0
Conv5(10)	$(5 \times 5 \times 1 + 1) \times 10$
Maxpool2	0
Conv5(10)	$(5 \times 5 \times 10 + 1) \times 10$
Maxpool2	0
FC10	$(8 \times 8 \times 10 + 1) \times 10$

3 Problem 3

(You can use both cross entropy and entropy definition to solve the problem.)

- $H(overweight) = 1$
- $H(overweight|gender) = \frac{4}{6}(-\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2}) + \frac{2}{6} \times 0 = 0.67$

- $H(overweight|Hyperlipidemia) = \frac{3}{6} \times 0 + \frac{3}{6}(-\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3}) = 0.46$
- $H(overweight|unhealthydiet)\frac{4}{6}(-\frac{3}{4}log\frac{3}{4} - \frac{1}{4}log\frac{1}{4}) + \frac{2}{6}(-\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2}) = 0.87$
- $H(overweight|exercise) = \frac{4}{6}(-\frac{1}{2}log\frac{1}{2} - \frac{1}{2}log\frac{1}{2}) + \frac{2}{6} \times 0 = 0.67$

So we choose hyperlipidemia as first classification choice since its maximize the mutual information. Then we classify the students as yes{1,2,5} and no{3,4,6}. For yes part, we have done. For no part,we can directly choose exercise/gender to classify them into {3} and {4,6} two class since their conditional/cross entropy now goes to 0.

4 Problem 4

- (a) All Predictions

  By the statement, when we define the first data sample as $(x_1^{(1)}, x_2^{(2)}) = (5, 5)$, we can get the predictions as:

$$\begin{aligned} f(x^{(1)}) &= \frac{1}{1 + \exp\left(-w_1 x_1^{(1)} - w_2 x_2^{(2)}\right)} \\ &= \frac{1}{\exp\left(-0.5 * 5 + 0.1 * 5\right)} \\ &= 0.8807970779778823 \end{aligned}$$

  Similarly, we get the all predictions:

  - Positive Samples:
    - Sample 0: 0.8808
    - Sample 1: 0.6682
    - Sample 2: 0.2142
    - Sample 3: 0.8909
    - Sample 4: 0.4256
  - Negative Samples:
    - Sample 5: 0.0691
    - Sample 6: 0.7685
    - Sample 7: 0.3543
    - Sample 8: 0.1824
    - Sample 9: 0.0163

- (b) Basic Evaluation Scores

  The threshold is 0.5 so the model predicts the samples as:

  - Positive Samples:
    - Sample 0: +
    - Sample 1: +
    - Sample 2: -
    - Sample 3: +
    - Sample 4: -
  - Negative Samples:
    - Sample 5: -
    - Sample 6: +

* Sample 7: -
* Sample 8: -
* Sample 9: -

$$\text{Accuracy} = \frac{3+4}{5+5} = 0.7$$
$$\text{Precision} = \frac{3}{3+1} = 0.75$$
$$\text{Recall} = \frac{3}{5} = 0.6$$
$$\text{F1} = 2\ \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 0.6667$$

Then, the confusion matrix is:

Thres= 0.5		Prediction	
		Positive	Negative
True Label	Positive	3	2
	Negative	1	4

- (c) ROC curve

  True Positive Rate (TPR) is the ratio of the positive samples that are predicted as positive. And, False Positive Rate (FPR) is the ratio of the negative samples that are predicted as positive. (TPR, FPR) become the following with different thresholds:

$$\text{thres} = 0.0 : (1.0, 1.0)$$
$$\text{thres} = 0.2 : (1.0, 0.4)$$
$$\text{thres} = 0.4 : (0.8, 0.2)$$
$$\text{thres} = 0.6 : (0.6, 0.2)$$
$$\text{thres} = 0.8 : (0.4, 0.0)$$
$$\text{thres} = 1.0 : (0.0, 0.0)$$

  By the statement, when we use the step-wise curve, we get the following ROC curve (Fig.2).

- (d) AUC

  We just need to compute the blue area of Fig.2. Then, AUC = 0.84.

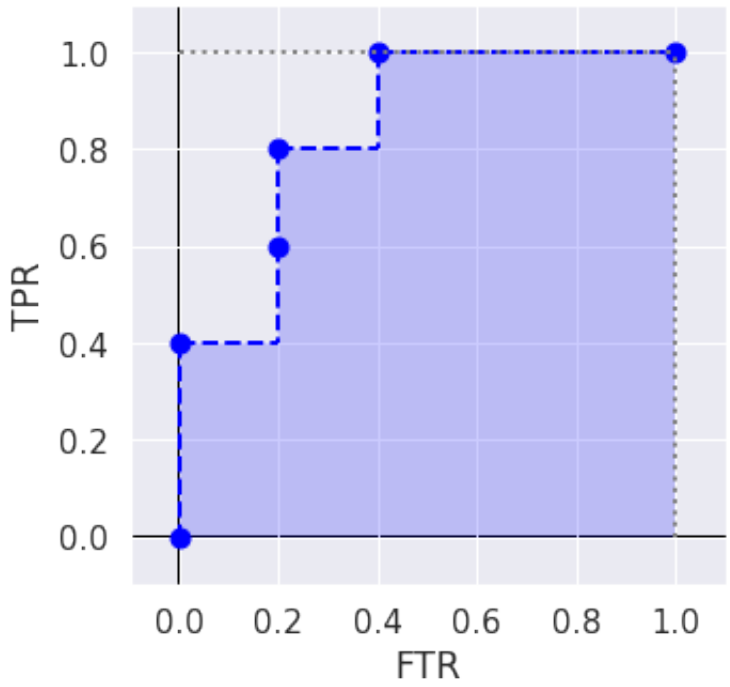

Figure 2: ROC curve