DDA3020 Written Homework 3

Xue Zhongkai (122090636)

April 16, 2024

Question 1.

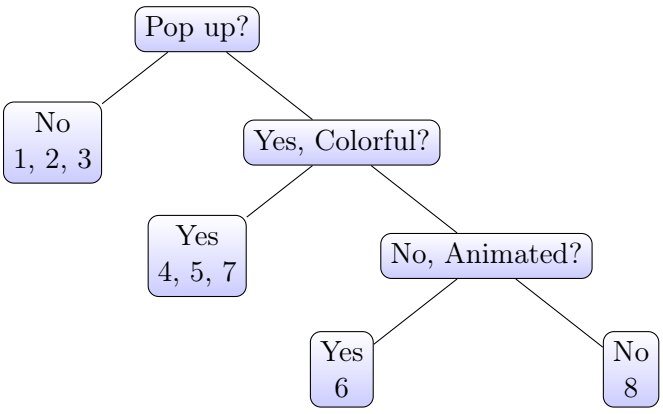
(a)

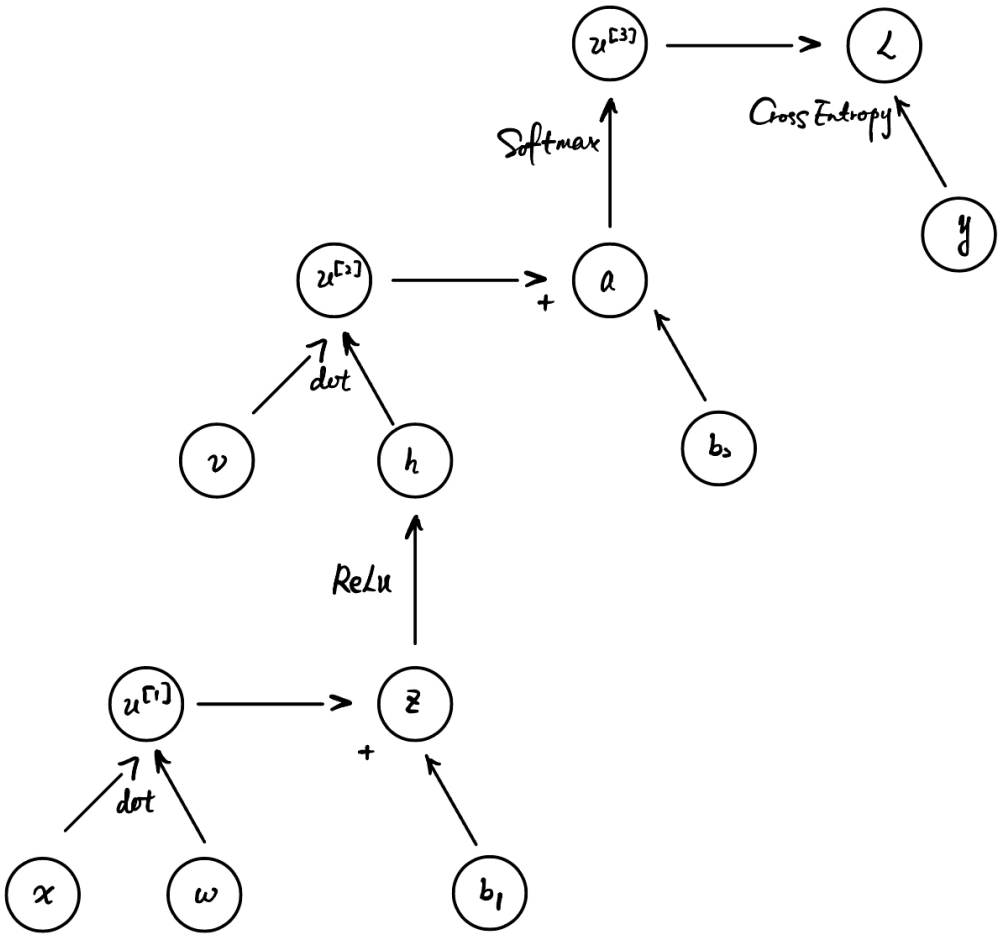
$$
\begin{aligned}
H(D) &= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \\
\text{Animated:}\quad H(D|\text{A=Y}) &= -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918 \\
H(D|\text{A=N}) &= -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971 \\
g(D) &= 1 - \frac{3}{8}\times 0.918 - \frac{5}{8}\times 0.971 = 0.049 \\
\text{Popup:}\quad H(D|\text{P=Y}) &= 0 \\
H(D|\text{P=N}) &= -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = 0.721 \\
g(D) &= 1 - \frac{5}{8}\times 0.721 = 0.549 \\
\text{Colorful:}\quad H(D|\text{C=Y}) &= -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.811 \\
H(D|\text{C=N}) &= 0.811 \\
g(D) &= 1 - 0.811 = 0.189
\end{aligned}
$$

So, **the Popup** has the largest information gain, and it should be splitted at the root.

(b)

The decision tree is shown as below:





Question 2.

(a)

The computational graph is shown above.

(b)

Every time we first calculate the single case, and generize to that of the matrix. Denote j as the labelled category.

$$\nabla_a L = \frac{\partial L}{\partial u^{[3]}} \frac{\partial u^{[3]}}{\partial a} = -y_j(1 - u_j^{[3]}) + u_j^{[3]} \sum_{i \neq j} y_i = u_j^{[3]} - y_j = \frac{\exp(a_j)}{\sum_i \exp(a_i)} - y_j \to \text{Softmax}(a) - y$$

The derivatives of L with respect to the weights and biases are:

$$\nabla_V L = \frac{\partial L}{\partial u^{[3]}} \frac{\partial u^{[3]}}{\partial a} \frac{\partial a}{\partial u^{[2]}} \frac{\partial u^{[2]}}{\partial v} = \frac{\partial J}{\partial a} \cdot \frac{\partial a}{\partial w} = h \frac{\exp(a_{j)}}{\sum_i \exp(a_i)} - h y_j \to (\text{Softmax}(a) - y)\, h^T$$

$$\nabla_{b_2} L = \frac{\partial L}{\partial u^{[3]}} \frac{\partial u^{[3]}}{\partial a} \cdot \frac{\partial a}{\partial b_2} = \frac{\exp(a_{j)}}{\sum_i \exp(a_i)} - y_j \to \text{Softmax}(a) - y$$

$$\nabla_w L = \frac{\partial L}{\partial u^{[3]}} \frac{\partial u^{[3]}}{\partial a} \frac{\partial a}{\partial u^{[2]}} \frac{\partial u^{[2]}}{\partial h} \frac{\partial h}{\partial z} \frac{\partial z}{\partial u^{[1]}} \frac{\partial u^{[1]}}{\partial w} = x \left(\frac{v \exp(a_j)}{\sum_i \exp(a_i)} - v y_j \right) I(z > 0) \to v^T (\text{Softmax}(a) - y)\, H(a) x^T$$

$$\nabla_{b_1} L = \frac{\partial L}{\partial u^{[3]}} \frac{\partial u^{[3]}}{\partial a} \frac{\partial a}{\partial u^{[2]}} \frac{\partial u^{[2]}}{\partial h} \frac{\partial h}{\partial z} \frac{\partial z}{\partial b_1} = \left(\frac{v \exp(a_j)}{\sum_i \exp(a_i)} - v y_j \right) I(z > 0) \to v^T (\text{Softmax}(a) - y) H(a)$$

Question 3.

(a)

$$L = I - F + S = 32 - 28 + 1 = 5$$

The convolutional filter size 5x5, is determined by the size of input image and the current-layer feature maps. It also depends on the stride and padding.

(b)

$$6 \times 28 \times 28 = 4704$$

Hence 4704 neurons are needed in C_1.

(c)

The stride distance in a max-pooling layer is typically set to be equal to the filter size, hence the stride distance required for filter of S_2 is 2.

$$6 \times 14 \times 14 = 1176$$

Hence 1176 neurons are required in S_2.

(d)

The purpose of the convolution layers is to extract local features from the input data, and that of the pooling layers is to reduce the spatial size of the feature maps.

The purpose of using the fully connected layers at the end is to combining the learned features in a non-linear way approach, performing the final classification or regression tasks.

Question 4.

(a)

Convolutional layers applies the same set of trainable weights across the entire image, allowing to detect local features **regardless of their position** in the image. Also, it allows **faster computation** and is more efficient compared to fully connected layers.

(b)

Given the matrix A and vector b, solve for vector w

$$A = \begin{bmatrix} 1 & 4 & 0 \\ 4 & 0 & -2 \\ 0 & -2 & 3 \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \quad b = \begin{bmatrix} -2 \\ 2 \\ 11 \end{bmatrix}$$

By solving $Aw = b$, we have the filter

$$w = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

(c)

Given matrices F, I, and O

$$F = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad I = \begin{bmatrix} -1 & 2 \\ 3 & 1 \end{bmatrix}, \quad O = \begin{bmatrix} O_{11} & O_{12} & O_{13} \\ O_{21} & O_{22} & O_{23} \\ O_{31} & O_{32} & O_{33} \end{bmatrix}$$

We have the output

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

So, the output of transpose convolution is

$$\begin{bmatrix} -1 & 3 & -2 \\ 3 & -3 & 1 \\ 0 & 3 & 1 \end{bmatrix}$$