1. (a). $\text{Entropy(Animated\_Yes)} = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

$\text{Entropy(Animated\_No)} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}$

$\text{Gain(Clicked, animated)} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} + (\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3})\cdot\frac{3}{8} + (\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5})\cdot\frac{5}{8}$

$= 0.04879$

Similarly, $\text{Gain(Clicked, Popup)} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} + 0 + 0 + (\frac{4}{5}\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5})\cdot\frac{5}{8}$

$= 0.54879$

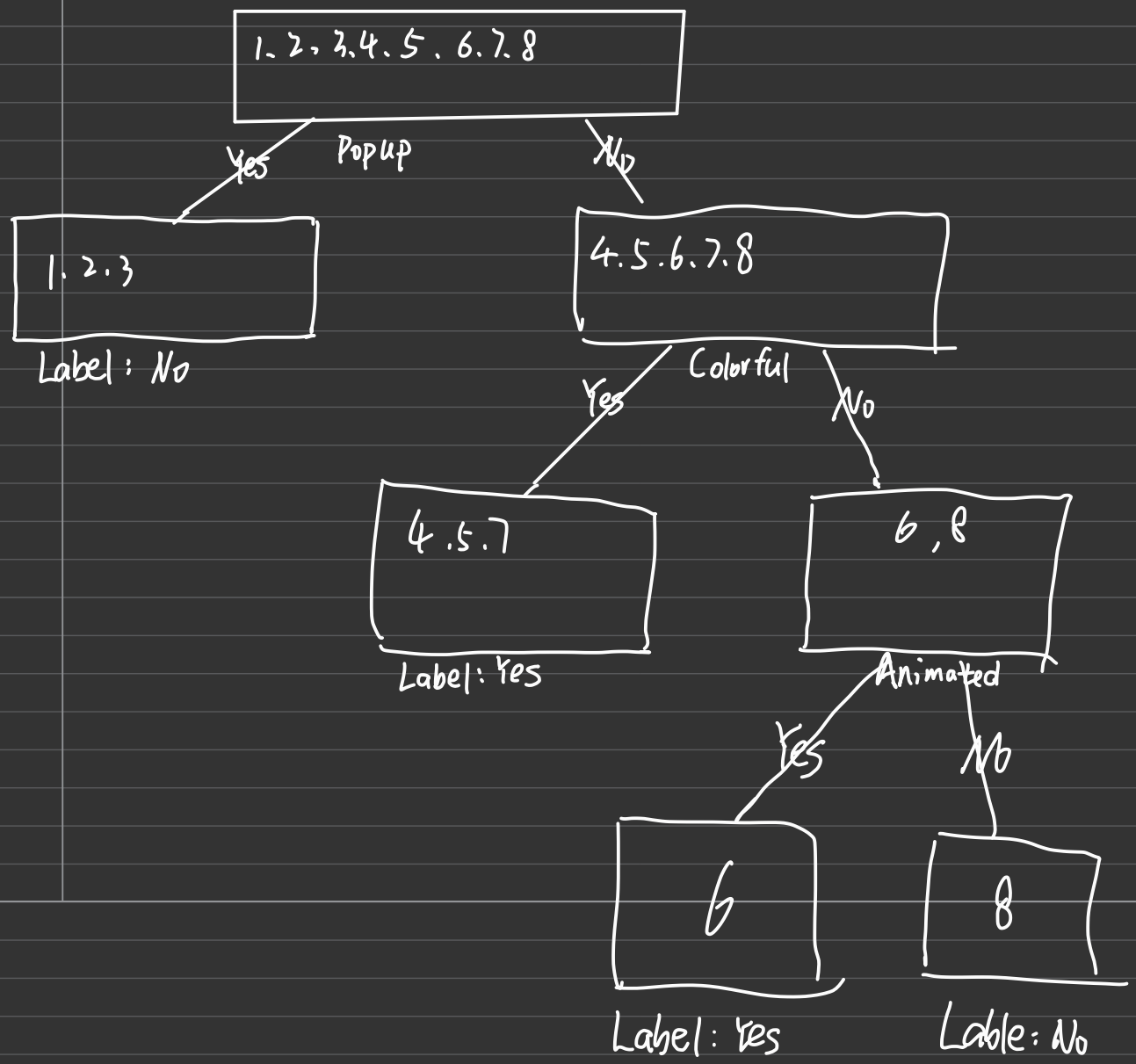
$\text{Gain(Clicked, Colorful)} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} + (\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4})\cdot\frac{1}{2} + (\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4})\cdot\frac{1}{2}$

$= 0.18872$

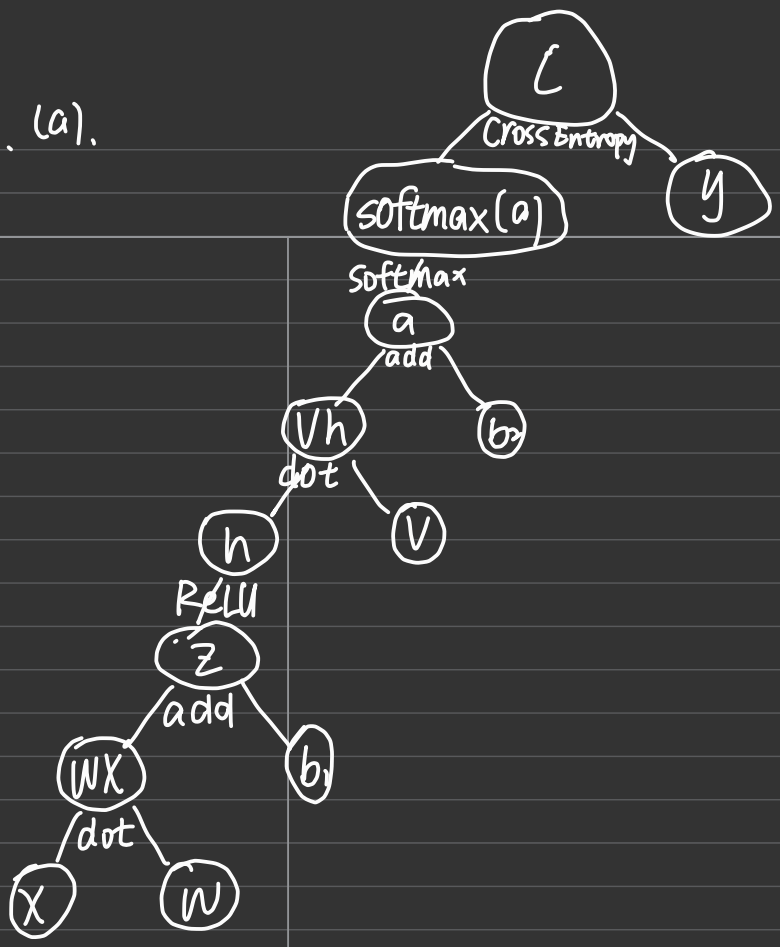
So popup is chosen. expression is

$\text{Gain(Clicked, Popup)} = -\log_2\frac{1}{2} + \frac{5}{8}\cdot(\frac{4}{5}\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5})$

(b).



2. (a).



(b). $\nabla_a L = \text{softmax}(a) - y$

$\nabla_V L = \nabla_a L \cdot \nabla_V a = (\text{softmax}(a) - y) \cdot h^T$

$\nabla_{b_2} L = \nabla_a L \cdot \nabla_{b_2} a = \nabla_a L = \text{softmax}(a) - y$

$\nabla_z L = \nabla_h L \cdot \nabla_z h$

$= \nabla_a L \cdot \nabla_h a \cdot \nabla_z h$

$= V^T (\text{softmax}(a) - y) \cdot H(z)$

$\nabla_W L = \nabla_z L \cdot \nabla_W z = V^T \text{softmax}(a) - y \circ H(z) \cdot x^T$

$\nabla_{b_1} L = \nabla_z L \cdot I = V^T \cdot (\text{softmax}(a) - y) \circ H(z)$

($\circ$ means element-wise product)

3.

(a) $32 - (a-1) = 28$

$\Rightarrow a = 5$

$32 - (b-1) = 28$

$\Rightarrow b = 5$

so the size is 5×5. (from $32 \times 32 \rightarrow 28 \times 28$)

(b). $6 \times 28 \times 28 = 4704$ neurons

(c). stride distance is 2

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

(d). Convolution layer: Extract features from input.

pooling layer: Reduce the spatial dimension of the features extracted by convolution layers and make the features more robust, which can reduce training cost and prevent overfitting.

Fully connected layers at the output end:

(i). Extract more specific features from previous layers.

(ii). Map the feature to output classification result.

4. (1). (i) Convolutional layer helps capture local patterns regardless of their location in the image. It provides local understand of the image, which reduces the risk of overfitting.

(ii). Convolutional layer requires less parameter than fully connected layer, which makes computation effective and reduce memory use.

(2). The filter size is 1D, length is 3.

assume it's $[a_1 \ a_2 \ a_3]$

so we have $\begin{cases} a_1 + 4a_2 = -2 \\ 4a_1 - 2a_3 = 2 \\ -2a_2 + 3a_3 = 11 \end{cases} \Rightarrow \begin{cases} a_1 = 2 \\ a_2 = -1 \\ a_3 = 3 \end{cases}$

so the filter is $[2 \ -1 \ 3]$

(3). input $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, filter $\begin{bmatrix} m & n \\ p & q \end{bmatrix}$

output is $\begin{bmatrix} am & an+bm & bn \\ ap+cm & aq+bp+cn+dm & bq+dn \\ cp & cq+dp & dq \end{bmatrix}$

so the output is $\begin{bmatrix} -1 & 3 & -2 \\ 3 & -3 & 1 \\ 0 & 3 & 1 \end{bmatrix}$