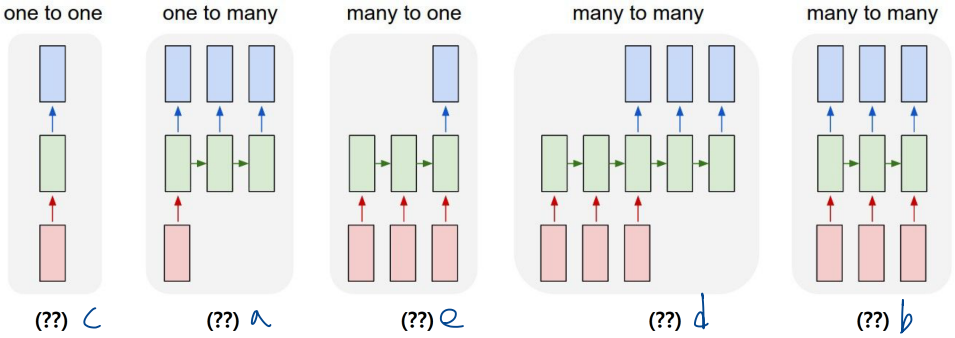
Day33: FinalTerm

1. [10 points] For each of the following statements, indicate if it is *true* or *false*. A correct answer will get 2 points, but a wrong answer will get -2 points. No answer will get 0 point.

   (a) Each element of a cell vector in LSTM is always between -1 and 1.
   (b) Gated recurrent units (GRUs) have fewer parameters than LSTMs.
   (c) Attention mechanism helps to interpret the output of the networks.
   (d) In GAN, output of the discriminator is between 0 to 1.
   (e) A self-supervised learning is under the category of the unsupervised learning.

2. [5 points] Fill in the blank (marked with ??) with one among (a)-(e)



   (a) Image captioning
   (b) Video classification on a frame level
   (c) Vanilla neural networks
   (d) Machine translation
   (e) Sentiment classification

3. [6 points] Consider the following vectors in a time step t in an LSTM cell:

$$c_{t-1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, f = \begin{bmatrix} 1 \\ 0.5 \\ 0 \end{bmatrix}, i = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, g = \begin{bmatrix} -0.7 \\ 0.4 \\ 0.8 \end{bmatrix}, \text{ and } o = \begin{bmatrix} 0.3 \\ 0.2 \\ 0.6 \end{bmatrix}.$$

Compute c_t and h_t. (You do not need to compute $\tanh(x)$. Instead, just write $\tanh(x)$ in your answer by replacing x with a particular value you computed.)

$c_t = f \cdot c_{t-1} + i \cdot g = \begin{bmatrix} 1 \\ 0.5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -0.7 \\ 0.4 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.4 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.4 \\ 0.8 \end{bmatrix}$

$h_t = o_t \cdot \tanh(c_t) = \tanh \begin{bmatrix} 1 \\ 1.4 \\ 0.8 \end{bmatrix} \cdot \begin{bmatrix} 0.3 \\ 0.2 \\ 0.6 \end{bmatrix}$

4. [5 points] Consider a variant of the original RNNs as follows:

$$h_t = \tanh(W_h(h_{t-1} + x_t))$$
$$y_t = W_y h_t.$$

$h_t = \tanh(W_{hh} h_{t-1} + W_{hx} x_t)$ (d_h, $d_h \times d_h$, d_h)

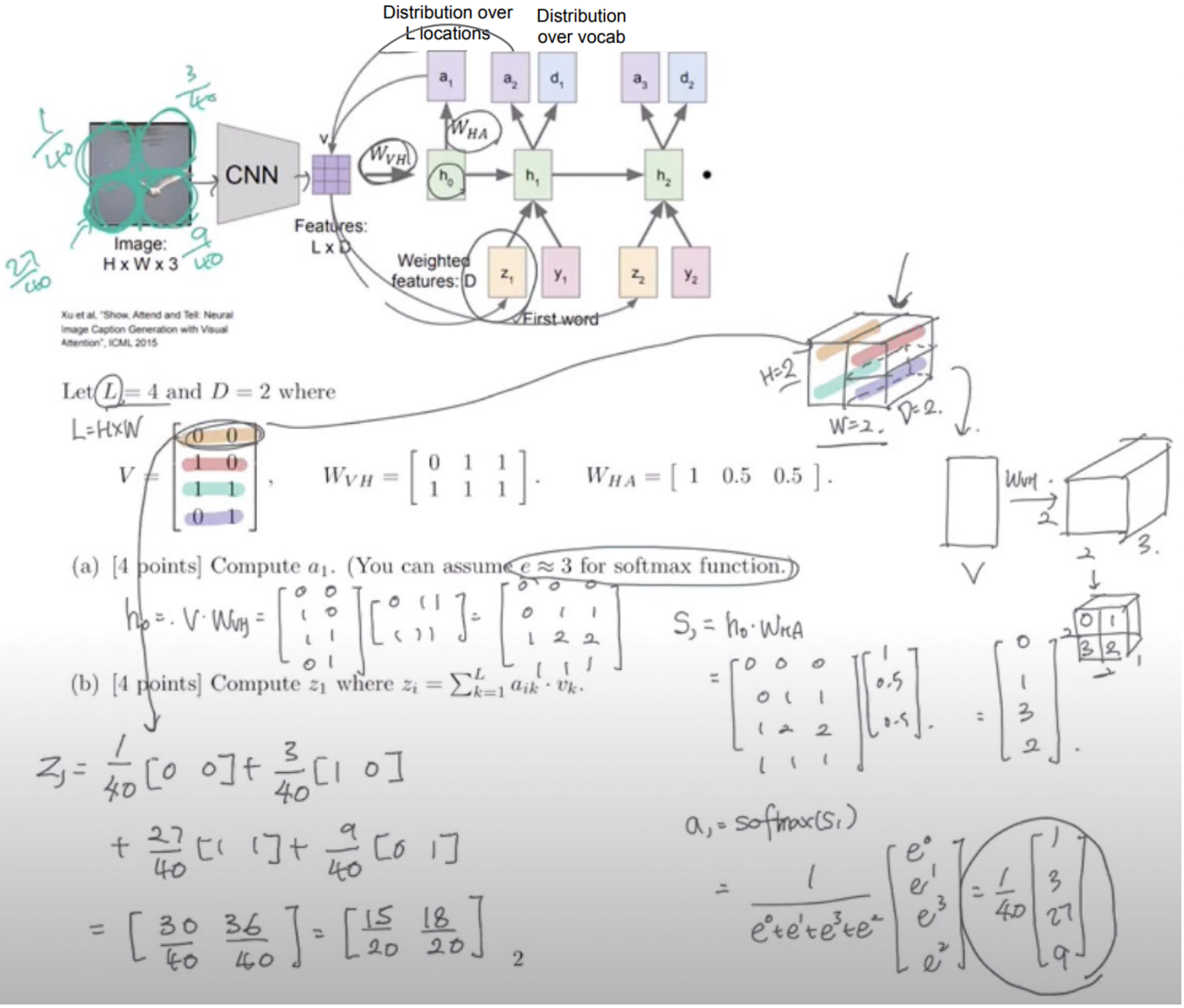
W_{hx}가 없다.

(a) [2 points] What has changed in this model, compared to the original RNNs? W_{hx}가 identity-matrix.

(b) [3 points] What is the restriction of this model in terms of an input vector dimension?

→ Input과 hidden-state의 dimension 동일 해야한다.

5. [8 points] Below is the image captioning model with attention. Please answer the following questions.



Let $L = 4$ and $D = 2$ where

$L = H \times W$

$$V = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad W_{VH} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad W_{HA} = \begin{bmatrix} 1 & 0.5 & 0.5 \end{bmatrix}.$$

(a) [4 points] Compute a_1. (You can assume $e \approx 3$ for softmax function.)

$h_0 = V \cdot W_{VH} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix}$

$S_1 = h_0 \cdot W_{HA} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 2 \end{bmatrix}$

$a_1 = \text{softmax}(S_1) = \frac{1}{e^0 + e^1 + e^3 + e^2} \begin{bmatrix} e^0 \\ e^1 \\ e^3 \\ e^2 \end{bmatrix} = \frac{1}{40} \begin{bmatrix} 1 \\ 3 \\ 27 \\ 9 \end{bmatrix}$

(b) [4 points] Compute z_1 where $z_i = \sum_{k=1}^{L} a_{ik} \cdot v_k$.

$z_1 = \frac{1}{40}[0\ 0] + \frac{3}{40}[1\ 0] + \frac{27}{40}[1\ 1] + \frac{9}{40}[0\ 1]$

$= \begin{bmatrix} \frac{30}{40} & \frac{36}{40} \end{bmatrix} = \begin{bmatrix} \frac{15}{20} & \frac{18}{20} \end{bmatrix}$

6. [6 points] Answer the following questions about the Generative Adversarial Networks.

   (a) [3 points] Which of the following loss induces the non-saturating generator for GANs (G is the generator and D is the discriminator)?

   i. $J^{(G)} = \frac{1}{m}\sum_{i=1}^{m} \log(1 - D(G(z^i)))$

   ✓ ii. $J^{(G)} = -\frac{1}{m}\sum_{i=1}^{m} \log(D(G(z^i)))$

   iii. $J^{(G)} = \frac{1}{m}\sum_{i=1}^{m} \log(1 - G(D(z^i)))$

   iv. $J^{(G)} = -\frac{1}{m}\sum_{i=1}^{m} \log(G(D(z^i)))$

7. [3 points] Training the model with the Contrastive Learning is highly unstable due to the moving targets. Fill in the blank (marked with ??) with among (a)-(c).