COMP 6721 - Artificial Intelligence
Introduction to Deep Learning

Question 1 Assume that the 3×3 matrix below represents a gray scale image. Further assume that we will use X to train an autoencoder.

$$X = \begin{array}{|c|c|c|} \hline 0.2 & 0.3 & 0.2 \\ \hline 0.4 & 0.1 & 0.3 \\ \hline 0.1 & 0.9 & 0.5 \\ \hline \end{array}$$

(a) What will be the input and the output of an autoencoder that processes X.

(b) What will be the size of the input and the output layers of such an autoencoder?

(c) What would be a suitable size for the number of hidden units?

(d) What would happen if the size of the hidden layer is equal to the size of the input? How would you achieve the desired feature in this case?

(e) Assume that we use an autoencoder with the following hyperparameters: the activation function is *sigmoid*, the loss/error function is the mean squared error (MSE), the hidden layer has a size of 5, and the learning rate η is 1. Perform a single iteration of forward pass and backward pass through the autoencoder. You can assume an input value of 1 for the biases, and all the weights (including the biases) are initialized to 0.5.

Question 2 Assume the following matrix that represents an image. This image will be fed to a convolutional neural network.

1	1	2	2	2	0	0
2	0	1	1	2	1	2
0	1	0	0	1	1	2
0	2	1	2	0	2	2
1	2	0	0	1	0	1
0	0	0	0	1	2	1
2	0	0	0	2	1	1

(a) Assume that we use the following convolution filter with a stride of 2.

0	1	1
0	1	0
0	-1	-1

What will be the size of the activation map? What will be the activation map?

(b) What will be the output of the pooling layer with, a size of 2*2 and a stride of 1, on the activation map of part (a) if we use the following strategies?

i. Average pooling

ii. Max pooling

Question 3 Consider the following sentence:

"the cat drinks the milk"

We will use this sentence to train a CBOW Word2Vec model. Assume that:

- you want to produce word embeddings of dimension 2,
- you use a context window of size 2 (1 word before and 1 word after the target word), and
- your vocabulary only contains the words in the sentence above

(a) Using only the sentence above, how many instances will be generated as training set?

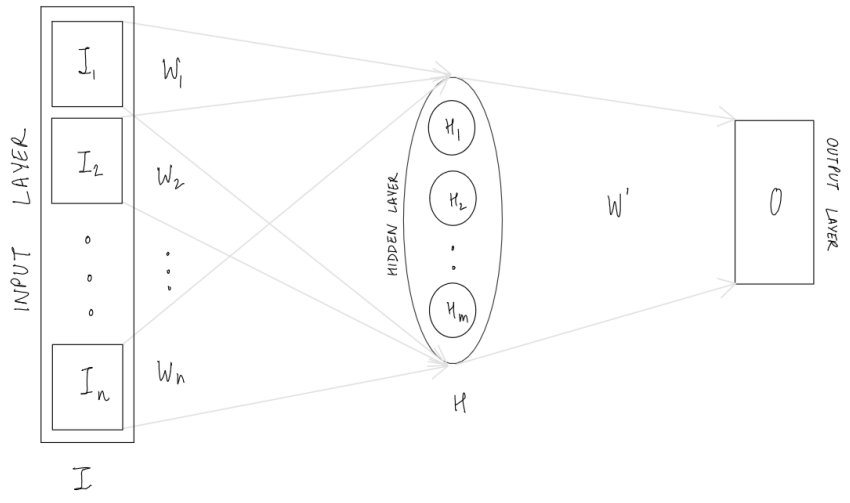
(b) List the one-hot vectors that correspond to each word in the vocabulary. (Assume alphabetical ordering)

(c) List the one-hot vectors that correspond to each training instance in the input layer.

(d) How many nodes will the hidden layer contain?

(e) What is the target hot vector for each training instance?

(f) Assume that the Word2Vec model is trained with the standard network depicted below:



i. What will be the values of n and m?

ii. What will be the sizes of I, W_i (for each $1 \leq i \leq n$), W' and O?

(g) Assume that we have these weight vectors:

$$W = \begin{bmatrix} 2 & 6 \\ 4 & 3 \\ 1 & 4 \\ 5 & 2 \end{bmatrix}$$

$$W' = \begin{bmatrix} 6 & 2 & 8 & 3 \\ 4 & 5 & 9 & 7 \end{bmatrix}$$

To compute the final probabilities at the output layer, we use the softmax function as shown in class. Recall that for a given vector of size k, the softmax function is defined as:

$$p_i = \frac{e^{x_i}}{\sum_{i=1}^{k} e^{x_i}}, \text{ where } 1 \leq i \leq k$$

i. Trace the first feed forward pass in the network and show the values propagated all the way to the output layer.

ii. What is the error after the first pass?