SEOULTECH
Seoul National University of Science & Technology

Mid-term Exam

Date : 2023.11.01

<table>
<tr><th>Code</th><th colspan="2">ITM 626</th><th>Title</th><th colspan="2">Artificial Intelligence</th></tr>
<tr><td>Time for Exam</td><td>2 hours</td><td>Questions</td><td>6</td><td>Weighting</td><td>30%</td></tr>
<tr><td>Student's Number</td><td colspan="2"></td><td>Student's Name</td><td colspan="2"></td></tr>
</table>

1. (10pts) True or false questions (+2pts for correct answer, -2pt penalty for incorrect answer)

   (a) Stochastic gradient descent is always guaranteed to converge to the global optimum of a loss function. (True / False)

   (b) You train a model and you observe that the validation accuracy is much lower than the training accuracy. If you use batch normalization, the new model is likely to have a smaller gap between the train and validation accuracies (True / False)

   (c) Deep neural networks typically have many hyperparameters that can be determined based on performance on the training dataset. (True / False)

   (d) The momentum optimizer better avoids local minima by keeping running gradient statistics. (True / False)

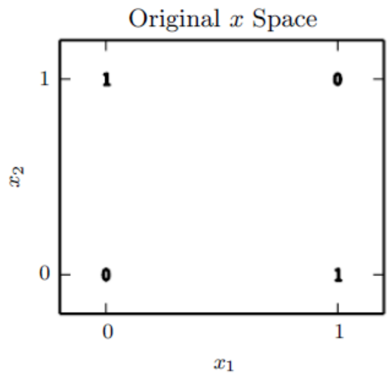
   (e) Dropout is a technique used to prevent overfitting by randomly dropping out some nodes during training. (True / False)

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**		**Title**	**Artificial Intelligence**	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

2. (10pts) Given the following XOR problem, design a neural network that can classify these points. Consider a neural network that has a single hidden layer with two nodes and ReLU as an activation function for hidden nodes.



SEOULTECH
Seoul National University of Science & Technology

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**		**Title**	**Artificial Intelligence**	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

3. (10pts) After training a neural network, you observe a large gap between the training accuracy (100%) and the test accuracy (50%). Provide at least three methods that can be used to reduce this gap, and explain why you choose them.

SEOULTECH
Seoul National University of Science & Technology

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**	**Title**	**Artificial Intelligence**

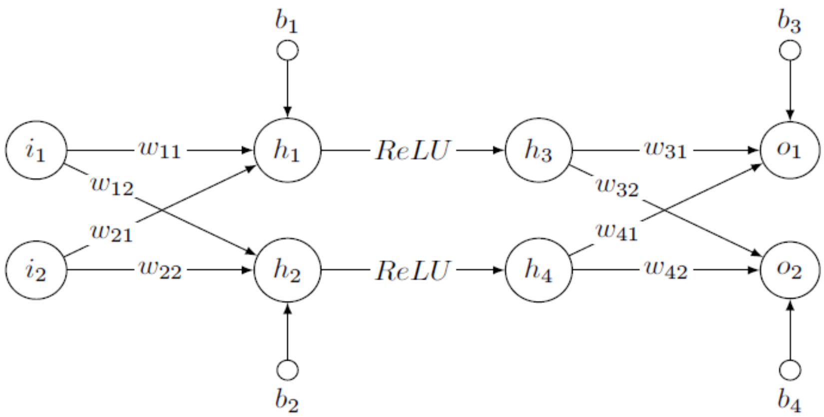
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**

Student's Number		**Student's Name**	

4. (20pts) Given the following neural network with fully connected layers and ReLU activations, including two input units (i_1, i_2), four hidden units (h_1, h_2) and (h_3, h_4). The output units are indicated as (o_1, o_2) and their targets are indicated as (t_1, t_2). The weights and bias of fully connected layer are called w and b with specific sub-descriptors.

The values of variables are given in the following table.

Var.	i_1	i_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}	w_{41}	w_{42}	b_1	b_2	b_3	b_4	t_1	t_2
Val.	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5



(a) (5pts) Compute the output (o_1, o_2) with the input (i_1, i_2) and network parameters as specified above. Write down all calculations including intermediate layer results.

SEOULTECH
Seoul National University of Science & Technology

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**		**Title**	**Artificial Intelligence**	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

(b) (5pts) Compute the mean squared error of the output (o_1, o_2) calculated above and the target (t_1, t_2).

(c) (5pts) Compute the gradient of the mean squared error with respect to w_{ij}.

(d) (5pts) Update the weights using gradient descent with learning rate 0.1.

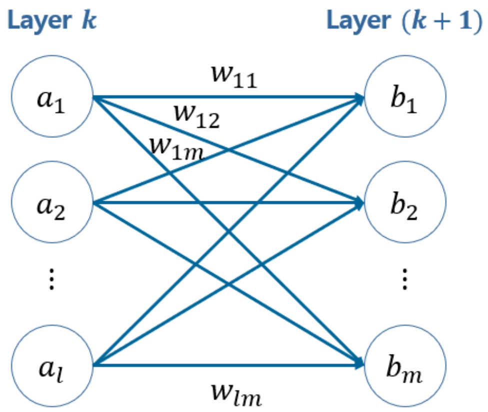
SEOULTECH
Seoul National University of Science & Technology

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**		**Title**	**Artificial Intelligence**	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

5. (20pts; each 5pts) We want to apply the dropout method to **Layer $\boldsymbol{k}$** in the following fully-connected layers. For dropout, we set the dropout rate (i.e., the probability of an element to be zeroed) to p. Here, we omit an activation function for simplicity.



(a) Describe how the dropout works during training in detail.

(b) For a specific node i in **Layer $(\boldsymbol{k}+\mathbf{1})$**, what is the expected activation value $\mathbb{E}[b_i]$?

(c) If the dropout is also applied to the test phase, we cannot expect consistent predictions due to the randomness of the dropout. Explain how we can resolve this issue with your answer in (b).

(d) Explain why dropout in a neural network provides a regularization effect.

Mid-term Exam

Date : 2023.11.01

Code	ITM 626		Title	Artificial Intelligence	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

6. (30pts; each 5pts) Write answers with precise explanations for the following questions.

(a) Draw a plot that represents a ReLU activation function and its advantages over a sigmoid activation function.

(b) You are training a neural network and notice that the validation error is significantly lower than the training error. Provide (at least) two possible reasons for this to happen.

(c) Explain why we need nonlinear activation functions.

(d) You want to solve a classification task. You first train your network on 20 samples. Training converges, but the training loss is very high. You then decide to train this network on 10,000 samples. Is your approach to fixing the problem correct? If yes, explain the most likely results of training with 10,000 samples. If not, give a solution to this problem.

Mid-term Exam

Date : 2023.11.01

Code	**ITM 626**		**Title**	**Artificial Intelligence**	
Time for Exam	**2 hours**	**Questions**	**6**	**Weighting**	**30%**
Student's Number			**Student's Name**		

(e) Given a convolutional layer with 8 filters, a filter size of 6, a stride of 2, and a padding of 1. For an input feature map of 32x32x32, what is the output shape after applying the convolutional layer to the input?

(f) Why softmax function is often used for classification problems? Explain with the formular of softmax function.