※ 考生請注意：本試題不可使用計算機。　請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

(10%) 1: Linear perceptron is one of the earliest neural network models. However, it was found out that the model cannot even perform the XOR (exclusive OR) function. Please describe why such a model cannot perform XOR function.

(25%) 2: (a) Please describe the supervised learning and the unsupervised learning.

(b) Classification and clustering are two data analysis approaches. Please describe the differences between the classification and clustering methods. Are they supervised or unsupervised learning? Why?

(15%) 3: During the training of an AI model, we usually would divide the data into training set, validation set and test set. Please explain why we need to divide the data into three data sets, and the purpose of each of the data sets in the training.

(20%)4: Gradient descent has been commonly adopted in the training of a deep neural network. Please describe the basic operation theory why gradient descent can be used to tune a network. Please also give the update equation using the gradient descent for updating a parameter w in a deep neural network.

(10%)5: A convolutional neural network often adopts several convolutional layers of different kernel sizes. What are the advantages and disadvantages of using a smaller kernel size convolution mask compared to using a large kernel size convolution mask?

(20%)6: Deep convolutional neural networks often adopt the mini-batch training with output normalization (mentioned as batch normalization) in certain layers. Please describe the batch normalization and the training procedure adopting batch normalization.