National Taiwan University of Science and Technology

Graduate Institute of Intelligent Manufacturing Technology

Ph.D. Qualifying Examination

Subject	Industrial Data Science		
Date	2025 / 05 / 05	Time	09:00 (Start) to 12:00 (End) (Total 180 Minutes)

Instructions to Candidates:

1. This exam paper consists of **TWO** pages (including this page). Please verify the number of pages.
2. Do not write any text or symbols unrelated to the answers on the exam paper.
3. Answers should be written on the answer sheet, with clear indication of the question number.
4. Upon the announcement of the exam end time, please stop writing immediately and submit both the exam paper and the answer sheet.
5. Violation of examination rules will be dealt with according to university regulations.
6. Reference materials not allowed in this exam.

Full Name: _____ Student ID: _____

◆ Question 1 (25 Points)

Classical data science emphasizes statistical applications. Industrial manufacturing has evolved from sampling inspection to full-sample analysis. If all data were to be collected, the speed of data collection would far exceed that of data analysis. Moreover, storing data incurs costs. From the perspective of industrial data science, please explain how to manage the influx of large volumes of data appropriately.

◆ Question 2 (25 Points)

The confusion matrix is used in the supervised learning of artificial intelligence. After the model is generated, there are four result, as shown in Table 1. Based on the distinction between predicted and actual states, which variables (i.e., A, B, C, D) must be strengthened to enhance accuracy and precision? Why?

Predict / al	TRUE	FALSE
E	A	B
SE	C	D

Table 1: Confusion Matrix

◆ Question 3 (25 Points)

Describe the concepts of forward and backward propagation in neural networks. Clearly explain the mathematical formulas involved in each process. Then, use a simple two-layer neural network with ReLU activation to manually demonstrate the detailed calculations for both forward and backward propagation. Consider a simple two-layer neural network with the following parameters:

- Input: $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- First layer (hidden layer):

  Weight matrix: $w_1 = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}$

  Bias vector: $b_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

  Activation function: ReLU
- Second layer (output layer):

  Weight matrix: $w_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}$

  Bias: b_2=0

  Assume the loss function is Mean Square Error (MSE):

  $L = \frac{1}{2}\left(y_{predict} - y_{true}\right)^2$ where y_{true}=5

◆ Question 4 (25 Points)

Given an image dataset, please design an analysis framework to perform image classification. For example, image preprocessing may be necessary if the figures require enhancement. In addition, please describe each step in detail and explain the reasoning behind each decision.