EE4016 AI with Deep Learning

Semester B 2025-2026

Assignment 1: Classification with Multilayer Perceptron

Due: Feb 21, 2026, at 11:00 PM

"It is not that I'm so smart. But I stay with the questions much longer." ~ Albert Einstein

Section A: Questions [40 marks] (**Submit the scanned handwritten or iPad handwritten answer sheets as PDF file.**)
Please provide brief answers to the short questions (1 – 10). Each question is worth 2 marks, and the answers should be less than 50 words.

1. What were the major limitations of early neural networks like the Perceptron and ADALINE?

2. What are the main advantages of the Multilayer Perceptron (MLP) over single-layer perceptrons, and what challenges did early MLP research face?

3. How did the backpropagation algorithm address the limitations of early neural networks?

4. What is the Universal Approximation Theorem, and why is it important in the context of neural networks?

5. Why was Mini-Batch Gradient Descent (MBGD) developed, and how does it balance the trade-offs between BGD and SGD?

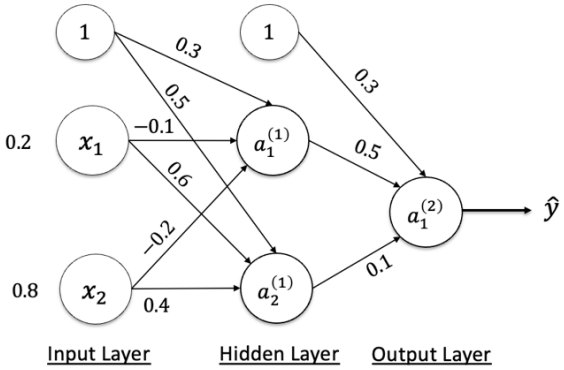
6. What are the main advantages of Adam, and how does it combine concepts from other optimization methods?

7. What is batch normalization and how does it improve neural network training?

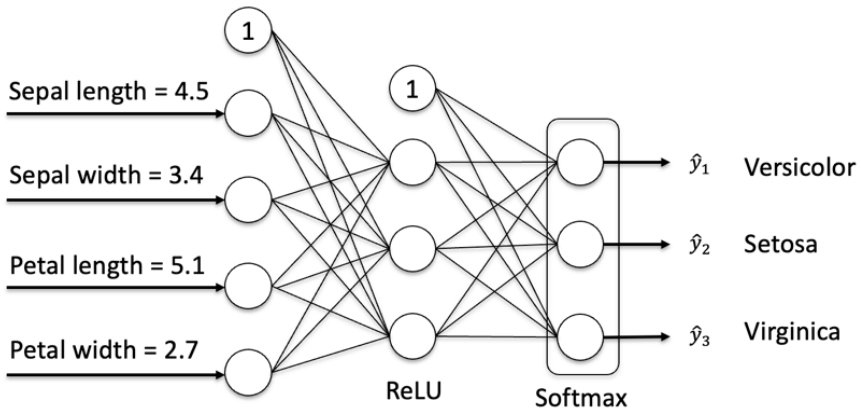
8. How does data augmentation improve neural network performance and what are some common techniques?

9. Given a two-layer MLP as shown in Figure 2 using **Sigmoid** activation function for both hidden and output layers, determine the output $\hat{y}$ by representing the network in matrix form. Determine the application of this neural network based on its architecture.

[6 marks]



10. Consider a two-layer multilayer perceptron (MLP) designed for Iris flower classification. The network consists of a hidden layer with a ReLU activation function and an output layer that uses the Softmax function for multi-class classification.



The input features $\mathbf{x}$ of the Iris flower, along with the neural network's weight and bias matrices ($\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}, \mathbf{b}^{(2)}$) are provide as follows:

$$\mathbf{x} = [4.5 \quad 3.4 \quad 5.1 \quad 2.7]^{\mathrm{T}}$$

$$\mathbf{W}^{(1)} = \begin{bmatrix} 0.5 & 0.4 & 0.3 & 0.8 \\ 0.2 & 0.4 & -0.2 & -0.5 \\ -0.9 & 0.2 & -0.5 & -0.7 \end{bmatrix} \text{ and } \mathbf{b}^{(1)} = \begin{bmatrix} 0.1 \\ 0.4 \\ 1.2 \end{bmatrix}$$

$$\mathbf{W}^{(2)} = \begin{bmatrix} 0.7 & 0.2 & 0.5 \\ 0.4 & 0.9 & 0.8 \\ 0.1 & 0.6 & 0.3 \end{bmatrix} \text{ and } \mathbf{b}^{(2)} = \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix}$$

Answer the following:

(a) The network's output classification probability distribution is denoted as $\hat{\mathbf{y}} = [\hat{y}_1 \quad \hat{y}_2 \quad \hat{y}_3]^{\mathrm{T}}$. Derive the matrix equation of $\hat{\mathbf{y}}$ in terms of $\mathbf{x}$, $\mathbf{W}^{(1)}, \mathbf{b}^{(1)}, \mathbf{W}^{(2)}$, and $\mathbf{b}^{(2)}$. [2 marks]

(b) Using the provided values for the weights and biases, compute the classification result by performing the matrix operations step-by-step. [4 marks]

(c) Assuming the true label for this sample is "Setosa," calculate the cross-entropy loss of this sample based on the predicted output $\hat{\mathbf{y}}$. [2 marks]

Section B: Programming Assignment [60 marks]

Objective:
In this assignment, students will implement a multi-layer perceptron (MLP) model for image classification on the Fashion MNIST dataset. The goal is to enhance model performance by exploring different feedforward neural network architectures, activation functions, optimization algorithms, regularization techniques, and data augmentation.

Guidelines:
A starter Jupyter notebook is provided for loading the dataset and implementing a 3-layer MLP with Sigmoid activations and Softmax output using stochastic gradient descent. Students should extend this notebook (or rewrite the whole notebook) to improve accuracy, F1 score, precision, and recall on the test set.

https://colab.research.google.com/drive/15S3-F0wCA4o3Scs6rchqLR96zxkNDoHA

The following techniques are recommended for earning marks (30 marks):

- Varying layers and neurons (5 marks)
- Trying different activation functions (5 marks)
- Using different optimizers (5 marks)
- Adding regularization such L1, L2, dropout, etc. (5 marks)
- Adding batch normalization layers (5 marks)
- Applying image data augmentation (5 marks)
    - Data augmentation requires higher computational effort to train the model. It is not recommended for students to try this technique first unless you want to achieve excellent accuracy.

Additional marks for achieving the following **test set accuracy**:

- ≥ 86% accuracy: 5 marks
- ≥ 88% accuracy: 10 marks
- ≥ 90% accuracy: 15 marks
- ≥ 92% accuracy: 20 marks
- 10 marks will be awarded for the coding style of your Jupyter notebook and the quality of your summary and analysis of the model.

Your Jupyter notebook should include a written summary and analysis of your model architecture, training procedure, experiments, results, and conclusions for assessment. Discuss what techniques worked well and how you improved upon the baseline model. Include relevant visualizations.

The **top 10** performing models of the class as measured by test set accuracy will each receive an **additional 10 marks**. In the event of a tie, preference will be given to models with fewer parameters and/or lower complexity in terms of Floating-Point Operations (FLOPs).

Please be aware that the use of MLP Mixer, Convolutional Neural Network (CNN) and Transformer architectures is not permitted for this assignment. However, students are encouraged to explore advanced techniques such as skip connections to enable deeper network architectures. Additionally, novel methods for model training that surpass standard approaches are welcomed. The objective is to encourage the maximization of accuracy while ensuring that models remain efficient in terms of size and computational demands.

Simplified models and better coding that achieve strong results through architectural innovations or training procedures will be recognized. The aim is to push beyond standard layered MLPs through creativity. Consider how techniques like residual connections, attention mechanisms, adversarial training etc. could enable better performance on this task.

Write a summary of your model and training design as well as analysis of experimental results at the end of the notebook. Submit your notebook ZIP file to the Assignment 1 on CASVAS with following file name format:

- Filename format : Assignment01_StudentName_StudentID.zip
- Filename example: Assignment01_Chen_Hoi_501234567.zip