COL865: Special Topics in Computer Application

SIL8123: Artificial Intelligence for Cybersecurity

Semester I, 2025-2026

Assignment 3

November 09, 2025

Note

- The code (with comments) and the report written for the assignment should be submitted on Gradescope.
- Be careful about plagiarism.
- The assignment will be graded in a demo.

Assignment 3

To enhance the robustness of a CNN model trained with the CIFAR-10 dataset, implement a defense against each of the following types: adversarial attack, training set poisoning attack, membership inference attack, and model inversion attack. In the report, describe the methodology and demonstrate the results in terms of images and appropriate metrics.