Introduction to Machine Learning and Artificial Intelligence, Summer 2023 (ET1550)

Project 2

In this project, a neural network model for a classification application will be investigated. According to the lectures, a neural network contains several learning layers, each containing several nodes that would map the input features to the output. After training, this relationship (model) will be used to predict the output, given any new input. Neural networks are often employed to model complex and nonlinear systems, particularly if a large dataset is available. On the other hand, logistic regression models (for binary classification) or softmax regression models (for multiclass classification) can be used as the baseline for predictions. They are especially effective if the system is not complex or the dataset is relatively small.

The Dataset

In this project, the diagnosis of breast cancer into malignant and benign tumours will be performed. To do so, a dataset is employed with several numeric features and attributes of the tissue. These features are computed from a digitized medical image of a breast mass. The dataset used in this project is from UCI ML Breast Cancer Wisconsin (Diagnostic) and is available through the University of Wisconsin-Madison FTP server.

The Code

The project can be implemented in either Google Colab or Jupyter Notebook (please refer to the Project IDE document).

The Report

For the project report, you only need to provide answers to the questions. In addition, include your written codes (from the places that are specified in the code) for each question (if required). The project report should be a PDF file containing your answers, codes and possible outputs from the specified parts.

The deadline for submission is **30 July 2023**.