EEE 6002: Selected Topics in Electrical and Electronic Engineering – Privacy Preserving Machine Learning

Midterm assignment instructions

1. Choose any dataset from the UCI Machine Learning Repository.
2. Make sure the dataset satisfies the following criteria:
    a. It is a binary classification dataset.
    b. The feature dimension is at least 500.
3. Split the dataset into train, validation and test sets (e.g., using 70%, 10% and 20% splits). You can use sklearn library for splitting the data into these partitions.
4. Write the Logistic Regression code by yourself (do not use any built-in classification function).
5. Include the following plots:
    a. Training and validation loss vs iterations
    b. Norm of the gradient vs iterations
    c. Training and validation loss vs number of training samples
    d. Training and validation loss vs regularization coefficient λ
    e. Required number of iterations vs step size α
    f. Precision and Recall vs threshold
    g. True Positive Rate vs False Positive Rate (ROC curve)
6. The report must contain a brief description of the dataset and the complete code.