Tutorial 6

The “breast.txt” dataset contains benign and malignant breast tumour samples. Each sample is measured by various factors including:

1. Clump Thickness 1 - 10
2. Uniformity of Cell Size 1 - 10
3. Uniformity of Cell Shape 1 - 10
4. Marginal Adhesion 1 - 10
5. Single Epithelial Cell Size 1 - 10
6. Bare Nuclei 1 - 10
7. Bland Chromatin 1 - 10
8. Normal Nucleoli 1 - 10
9. Mitoses 1 - 10

The last column (10th) contains class labels with “1” been malignant and “0” been benign.

(1) Download the breast dataset file the course webpage.
(2) Design a 10-fold cross-validation procedure to evaluate logistic regression and *k*NN (*k*=3) classification accuracy.
(3) Calculate TP, TN, FP, FN and compute sensitivity and specificity for each classifier.
(4) Compute F1 score and compare the two classifiers.