

The "breast cancer dataset" in CANVAS was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The features in the dataset, described below, have been categorized from 1 to 10. Use the knn methodology (k=3,5 and 10) to develop a classification models for the Diagnosis.

Important: make sure your categories are represented by the "factor" data type in R and delete the rows with missing value. Use 30% test 70% training data.

Features	Domain

Sample code number	id number
F1. Clump Thickness	1 - 10
F2. Uniformity of Cell Size	1 - 10
F3. Uniformity of Cell Shape	1 - 10
F4. Marginal Adhesion	1 - 10
F5. Single Epithelial Cell Size	1 - 10
F6. Bare Nuclei	1 - 10
F7. Bland Chromatin	1 - 10
F8. Normal Nucleoli	1 - 10
F9. Mitoses	1 - 10
Diagnosis Class:	(2 for benign, 4 for malignant)