

William Huang
CS 6140
Summer 2025

Predicting Breast Tumor Malignancy with Supervised Learning Models

Task:

- Goal: predict whether a tumor is malignant or benign based on input features (binary classification)
- Input: 30 features from digitized images of fine needle aspirate (FNA) of multiple breast masses
- Output: 1 of 2 class labels (M = malignant or B = benign)

Dataset:

- Source: Breast Cancer Wisconsin (Diagnostic)
 - <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
- Description:
 - 569 samples of breast tissue
 - Target: diagnosis (M = malignant tumor, B = benign tumor)
 - 30 continuous features (radius, texture, perimeter, area, smoothness, etc.)
 - No missing values

Prior Work:

- Multisurface Method-Tree (MSM-T) used to create separating plane:
 - Source: K. P. Bennett, "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992
- MSM-T uses robust linear programming to construct a decision tree
- Robust linear programming is a linear separator with robustness to noise/outliers
 - Source: K. P. Bennett and O. L. Mangasarian: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34

Plan:

- Preprocess data with StandardScaler
- Model Training:
 - Try logistic regression, decision tree, random forest algorithms
 - Cross-validation (10-fold)
- Evaluation:
 - Use confusion matrix, accuracy, precision, recall, F1 score
 - Plot ROC curve and compute AUC