

# STAT3612 Homework 3:

Date: November 28, 2019

Submit (in the ipynb format) via Moodle before 11:59pm December 12, 2019.

(Breast Cancer Classification Tasks). Use the `load_breast_cancer()` from `sklearn.datasets` to get a copy of the breast cancer (diagnostic) dataset with 569 samples with 212 Malignant and 357 Benign cases. Consider only the first 10 attributes (mean features) as the predictor variables and perform the following tasks.

**Step 1.** (20%) Fit a decision tree classifier with `max_depth = 3`. Visualize the fitted tree by `export_graphviz`. Report the training accuracy.

**Step 2.** (20%) Fit the random forests and gradient boosting machines. Report the training accuracy for both models.

**Step 3.** (20%) Fit support vector classifiers with linear and RBF kernels. Report the training accuracy for both models.

**Step 4.** (20%) Fit a multi-layer perceptron (MLP) classifier. Report the training accuracy.

**Step 5.** (20%) Pick the most-accurate model (likely a black-box model) from the above model fits. Run the post-hoc interpretability analysis.