

## ASSIGNMENT 3

(100 POINTS) DUE BY MONDAY, APRIL 20, 2020, 4 PM

**Question 1. (IMPLEMENTING ADABOOST - 50 POINTS).** Implement the AdaBoost algorithm using a decision stump for the Breast Cancer dataset with binary labels. Report test accuracy. What are the choices of hyperparameters in your model? (*using existing machine learning libraries (example-sklearn) is not allowed*).

```
from sklearn.datasets import load_breast_cancer
X, y = load_breast_cancer(return_X_y = True)
# [0 : malignant, 1: Benign]
# Convert labels to {-1, +1}
y = 2*y - 1
train_size = 469
X_train = X[0:train_size, :]
y_train = y[0:train_size]

X_test = X[train_size: , :]
y_test = y[train_size:]
```

**Question 2. (ADABOOST REVISTED - 50 POINTS).** With the usual notation and assumptions on availability of the weak learners in the AdaBoost algorithm as in the class, show that the base learner  $h_{t+1}$  selected at time  $t + 1$  must be different from  $h_t$ .

Can this be use as a rigorous justification of the argument that Adaboost forces the weak learner to focus on the problematic samples? Justify.