**MACHINE LEARNING FROM DATA**

**Fall 2018**

**Report: Lab Session 6 – Decision Trees**

**Names:**

**Group:**

Instructions

* Answer the questions
* Save the report and Matlab code in a folder, and upload the compressed folder (zip, rar).

Questions

Q1: Which are the default values of the tree depth controllers for growing the tree? (MaxNumSplits, MinLeafSize, MinParentSize).

Q2: Run the script. Copy the training, validation and test errors, and the confusion matrices. From the *Classification Tree Viewer* copy the tree graph.

Q3: Analyze the questions at each node. Which are the most relevant features for the classification task?

Q4: Include error curves for training and validation subsets (for the different values of *MinLeafSize*), and the error and confusion matrices for the best classifier.

Q5: Include the code in the report.

Q6: Include error curves for training and validation subsets (for the different values of *alpha*), and the error and confusion matrices for the best classifier.

Q7: Include the code in the report.

Q8: Compare results in Q4 and Q6. Which method is better?

Q9: Analyze the trees obtained in Q4 and Q6. Which are the most relevant features? (Analyze questions in nodes)

Q10: How many trees are trained in the ensemble? Include errors and confusion matrices for training, validation and test subsets. Compare results with those obtained with a single tree (Q4, Q6).