



Course Homepage

Upload Assignment: Midterm Project

Upload Assignment: Midterm Project

ASSIGNMENT INFORMATION

Due Date

Sunday, July 16, 2017

11:59 PM

Points Possible

100

Your midterm project will consist of two parts: You may work in groups of up to 4 people, or may work on your own. Please submit your final report online by midnight, July 16.

PART 1: CLUSTERING.

The Dataset: The dataset can be found here:

<https://archive.ics.uci.edu/ml/datasets/Forest+type+mapping> . It consists of image data points of Japanese forests. Your goal is to do a cluster analysis on this dataset, and then see how closely your clusters align with the forest types.

Cleaning: The dataset includes a classifier for the type of forest – you should remove this before you do your cluster analysis. The features are the columns b1 through b9; I think the other columns can be discarded, since they were used in the study for predictive purposes. Also, since this is not a supervised learning problem, you can combine the test and train datasets into one dataset.

The Methods: Use K-Means and H-Clustering as we did in class. Use the scripts as your reference.

PART 2: REGRESSION.

*When finished, make sure to click **Submit**.*

*Optionally, click **Save as Draft** to save changes and continue working later, or click **Cancel** to quit without sa*

