

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

Cleaning: This is a high dimension dataset; also, several of the variables have a significant number of NAs. When you load the dataset, be sure to convert "?" to "NA". Also, look for the variables that are mostly 'NA' and remove them – they are not useful for our purposes. The columns do not have names, and you may wish to load a file with the names (copy from the website) and use this to name the columns.

The Methods: Use Ridge, Lasso, and PCR to build models. Use CV to test them. See how accurate you can get the model. You can also try KNN.

Final Report: Write a report (no more than 8 pages, more if you need to include figures) that explains Your results for both parts of this project. A template is posted to the course homepage – do one of these for each part of the project, and combine both into one report.

xt Submission	Write Submission		
Attach Files	Browse My Computer	Browse Content Collection	
`			
COMMENTS			
COMMENTS			
COMMENTS			

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