Lab Assignment 9

Continue your exploration of linear regression using a public dataset provided courtesy of An Introduction to Statistical Learning, called College Acceptance and Enrollment. There are four relationships we are interested in modeling:

- The amount charged for room and board, as a function of the number of: accepted students
- The number of enrolled students per college, as a function of the number of: accepted students
- The number of failed undergraduate students per college, as a function of: the number of accepted students
- The number of accepted students, as a function of: the amount charged for room and board, and the number of enrolled students.

You should be able to find these relationships given the dataset. Start by downloading the data directly from this link, or through navigating the site linked above, and then complete the following actions:

1. Read through the /Module5/**assignment9.py** starter code file. It's pretty long, but there actually isn't much in terms of coding you have to do. The hard part is in understanding the material.
2. Load up the dataset, then use indexing to slice out the features of it we're interested in examining.
3. Split your data into training and testing.
4. Use the helper drawLine() and drawPlane() methods to display the results

Lab

Lecture: Regression
Quiz

Lab: Regression
Lab

Dive Deeper

▸ 6. Data Modeling II

5. Answer the following questions.

Lab Question

(1/1 point)
Which two relationship had the worst R2 correlations?

- Accept[Room.Board,Enroll] and Accept[Enroll]
- Accept[Enroll] and Accept[F.Undergrad]
- Accept[Room.Board] and Accept[F.Undergrad] ✔
- Accept[Room.Board,Enroll] and Accept[Room.Board]

EXPLANATION

The smaller, or more negative the R2 coefficient is, the poorer the relationship is.

Accept(Room.Board) R2: -0.00266698641455

Accept(F.Undergrad) R2: 0.777991797375

Accept(Enroll) R2: 0.857820486736

Accept(Room.Board,Enroll) R2: 0.876437177421

You have used 2 of 3 submissions