

Project 2 – Regression, Due September 30th before 11:00pm.

CPSC 4820/6820

Project Description

You have been hired by an ambitious entrepreneur for a new app that aims to help Clemson students make informed decisions concerning studying and drinking beer. Your job is to use meticulously gathered data to create a regression hypothesis function that will predict a student's grade point average based on how many minutes per week they study and how many ounces of beer they consume per week.

Your data is in a file named GPADData.txt that is available on Canvas (beware that Canvas might download it with a different name!). The first line in the file is an Integer value indicating how many lines of data are in the file. Each line after that contains three tab-separated real values that represent minutes studying/week, ounces of beer/week, and semester grade point average.

CPSC 4820 Assignment.

Your assignment is to create a linear regression solution ($y = w_0 + w_1x_1 + w_2x_2$). You should randomly divide your data set into a training set (70% of data) and a test set (30% of data).

Deliverables.

1. Project report that includes: (lastname_firstname_P2Report.pdf)
 - Problem description
 - Initial values that you chose for your weights, alpha, and the initial value for J .
 - Final values for alpha, your weights, how many iterations your learning algorithm went through and your final value of J on your training set.
 - Include a plot of J (vertical axis) vs. number of iterations (horizontal axis).
 - If you did feature scaling, describe what you did.
 - Value of J on your test set.
2. A python file (lastname_firstname_P2.py) that prompts the user to enter values for minutes spent studying per week and ounces of beer consumed per week, and then predicts their semester GPA (rounded to two decimal places). The program should keep prompting the user until you enter zeros for both values.

CPSC 6820 Assignment.

Your assignment is to create a polynomial regression solution ($y = w_0 + w_1x_1 + w_2x_2 + w_3x_1x_2 + w_4x_1^2 + w_5x_2^2$). You should randomly divide your data set into a training set (70% of data) and a test set (30% of data).

Deliverables.

1. Project report that includes (lastname_firstname_P2Report.pdf):
 - Problem description
 - Initial values that you chose for your weights, alpha, and the initial value for J .
 - Final values for alpha, your weights, how many iterations your learning algorithm went through and your final value of J on your training set.
 - Include a plot of J (vertical axis) vs. number of iterations (horizontal axis).
 - If you did feature scaling, describe what you did.
 - Value of J on your test set.
2. A python file (lastname_firstname_P2.py) that prompts the user to enter values for minutes spent studying per week and ounces of beer consumed per week, and then predicts their semester GPA (rounded to two decimal places). The program should keep prompting the user until you enter zeros for both values.

Special note since I really got this question! If you use a machine learning library and just call a regression library routine to do this assignment, your grade will be zero.