Data Analysis (Assignment #2)

Story:

We have a data which classified if patients have heart disease or not according to some features. We will try to use this data to create a model which tries predict if a patient has this disease or not.

You need to tackle the problem using two approaches:

A. Include all the features and apply the two methods for binary classification
   1) Least square method
   2) logistic regression
B. Apply PCA and choose the number of PCA components you want (justify why you selected this number) then apply the previous two methods again for classification

Data:

"heart.csv" which contains

- age - age in years
- sex - (1 = male; 0 = female)
- cp - chest pain type
- trestbps - resting blood pressure (in mm Hg on admission to the hospital)
- chol - serum cholestoral in mg/dl
- fbs - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- restecg - resting electrocardiographic results
- thalach - maximum heart rate achieved
- exang - exercise induced angina (1 = yes; 0 = no)
- oldpeak - ST depression induced by exercise relative to rest
- slope - the slope of the peak exercise ST segment
- ca - number of major vessels (0-3) colored by flourosopy
- thal - 3 = normal; 6 = fixed defect; 7 = reversable defect
- target - have disease or not (1=yes, 0=no)

Deliverables:

1) Python code that performs each approach.
2) Charts that show the cost function and accuracy over iterations for logistic regression approach
3) PCA chart for the most prominent parameters
4) Your opinion regarding each approach and which one you recommend

Due Date: 15th of December 2019 at 11:59 PM

Delay Policy: Each day of delay cost you 2 grades out of 10 and after three days of delay you lose full assignment grade.

Submissions will be on Moodle **not by email.**

Submit your assignments individually **(no teams are permitted).**