DAT18 SF: HOMEWORK 4 ASSIGNMENT

Assigned: Thursday, November 19, 2015
Due: Tuesday, November 24, 2015, before class
Review Due: Thursday, November 26, 2015, before midnight (no class)

The purpose of this homework is to review what we've learned about regression, regularization and expand our knowledge of Generalized Linear Models (GLMs). It consists of a set of small problems

HOMEWORK QUESTIONS

DUE MONDAY:

1. A projectile is fired upwards from the ground. The height of the projectile above the ground is shown in the following table:

Time (s)	0	0.5	1	1.5	2	2.5
Height (f)	0	20.5	31.36	36.25	30.41	28.23

    a) Find a good model to fit this data.
    b) Find the time at which the projectile hit the ground.

2. Healthcare costs have been increasing over the years. The following data shows the average cost of healthcare per person from 1976 to 1998: Year 1976 1980 1987 1993 1998 Cost (per person) 618 860 1324 1865 2256 a) Find a model that fits the data well. b) Find the time the average healthcare cost will reach $2800 per person.

Year	1976	1980	1987	1993	1998
Cost	618	860	1324	1865	2256

    a) Find a good model to fit this data.
    b) Find the time the average healthcare cost will reach $2800 per person.

3. Here's another dataset:

X	-1	0	1	2	3	5	7	9
Y	-1	3	2.5	5	4	2	5	4

    a) Find a good model to fit this data.
    b) Play around with different forms of regularization to see the effect and choose one.

4. Download the red wine quality data from here:
        https://archive.ics.uci.edu/ml/datasets/Wine+Quality

5. Import the dataset to a Pandas dataframe and explore the data using a scatter matrix plot.
    a. Do you notice any correlation?
    b. What can you say about the quality score?

    c.   How many features are there? Are they normalized?
6. Use a generalized linear model to predict the quality score of the wine.
    a.   What precautions will you take to account for the imbalance between different the amount of data in each class?
    b.   How will you treat the fact that the score is discrete in value?
7. How will you assess the goodness of your model? (hint: has to do with errors)
8. Check if regularization improves the outcome of your predictions

BONUS POINTS:

Read about Generalized Linear Models implemented in Scikit Learn here:
http://scikit-learn.org/stable/modules/linear_model.html

DUE THURSDAY:
1. Go to your new assigned review-buddy's repo
2. Read through your buddy's ipython notebook and make sure you understand what he/she is doing.
3. Open an issue in his/her repo and write comments on the things you don't understand and on the things you like in his/her code.
4. Quote the instructors in the comments so that we get notified about the open issue