

CSci 343 Fundamentals of Data Science Challenge 6

Late Penalties Start On:
November 17, 2016

Points Available:
200 XP for successful demonstration
50XP for readable & understandable code

Objectives:

- Learn the basics of Polynomial Regression
- Have fun!

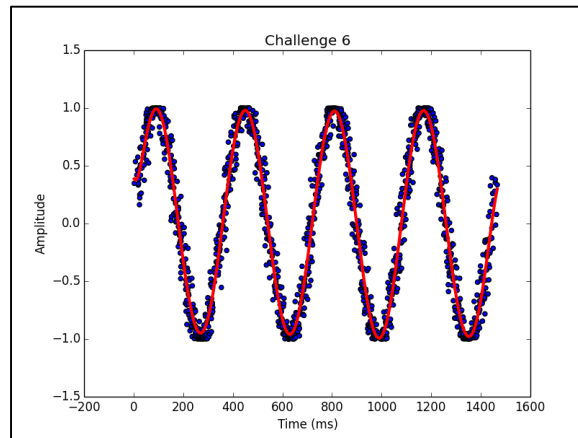
Assignment:

You're a data scientist for the National Weather Service. Usually, you are analyzing weather data, but today something has gone wrong with one of two remote weather stations that transmit their data over radio frequencies to the headquarters. One of the stations has stopped responding! What happened? Zombie apocalypse? Alien invaders? Sentient robot unicorns? Maybe, but it's probably just a blown power relay. These stations are pretty far away, so your boss doesn't want to waste time sending a repair crew to both stations. Before your boss has time to brief the repair team on which station went silent, she is called away to an urgent meeting! Oh, no! The repair crew doesn't know which station to go to. Being a clever data scientist, you suggest that we look at the radio transmission data and see which radio frequency isn't transmitting a signal and which one is. This is your task!

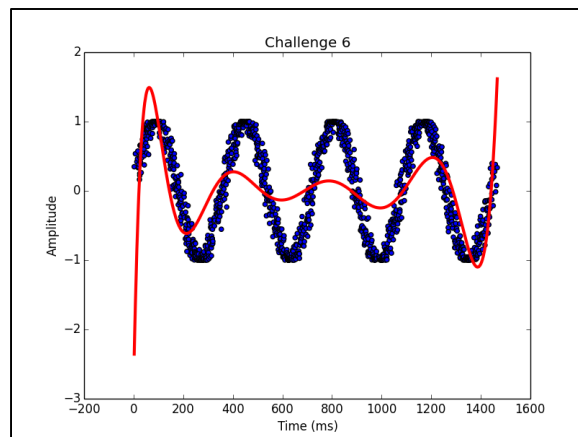
On the class website, you'll find a CSV file for Challenge 6 that has three columns: transmission time in milliseconds, signal amplitude, and FM radio frequency in kHz. You must first find out which radio frequency is generating a structured signal and which is just random noise. Once you've found the correct frequency, then do a scatter plot of the data that was transmitted. Finally, you'll need to do a polynomial regression (using the built-in NumPy functions) of this raw data and plot a regression line on top of the scatter plot. Be sure to try multiple degrees for your polynomial and pick the one that best fits your specific data (everyone in the class has a different dataset).

Goals:

1. Read in your data file
2. Find the two groups based on third column (you'll have to plot your data and give it a look over to figure this out)
3. Plot only the structured data as a scatter plot
4. Perform a polynomial regression (the exact procedure is on the Course Wiki)
http://www.cs.olemiss.edu/~jones/doku.php?id=csci343_polynomial_regression
5. Plot the regression line on top of the scatter plot (as shown below)



The above is an acceptable regression



The above is NOT an acceptable regression

Submission Instructions:

1. Demo your working code to the class TA before uploading it to Blackboard. You cannot proceed to step 2 before doing this.
2. Once your code is working and you've demoed it to the TA, upload all your code to Blackboard as a single ZIP file. Name your ZIP file *spiritAnimal.zip*, where *spiritAnimal* is your class user ID (not your

webID or ID number). Be sure to name your main source file "main.py". In a comment at the top of the file, include the following information. Spirit Animal User ID, Date the file was last edited, Challenge Number, and cite any sources that you used as a reference for code, data, and content (including title and URL).