

Problem Set 2

Fall 19

November 16, 2019

1 Problem Definition

For numerous purposes, different time series are recorded and analyzed to understand phenomena and behaviors of variables, to try to predict future values. Unfortunately, and for several reasons, there are gaps in data, irregular time steps of recordings, or removed data points that often need to be filled for data analysis. In order to fill gaps in time series, numerous methods are available, and the choice of the method(s) to apply is not easy for non-mathematicians among data users. One of such problems is the prediction of fine-grained air quality.

The information about air quality, such as the concentration of $PM_{2.5}$ is important to urban environmental governance and the health of residents. The specific monitoring station in different regions collects the air quality information. But the number of monitoring stations in cities is limited due to the expensive cost of building and maintaining, and there is no air quality monitoring station in most urban regions. Considering the equipment maintenance, re-calibration, and many other problems, the data of air quality collected by the monitoring stations may also be inaccurate or missing. Moreover, *the air quality data is not distributed equally rather quite abruptly*. Table 1 shows such a data set where the concentration of $PM_{2.5}$ are shown at different months of year 2016 in Bangladesh.

Table 1: Concentration of $PM_{2.5}$ at different months of year 2016

Month	Concentration(ug/m^3)
March	105.8104839
April	48.33055556
June	30.36527778
July	27.98252688
September	25.25833333
October	57.68145161
December	164.9502688

Your job is to find the concentrations of $PM_{2.5}$ of the missing months: January, February, May, August, November.

Tasks

1. Estimate the concentrations of $PM_{2.5}$ of the missing months by using different interpolating polynomials.
2. Plot the concentrations of $PM_{2.5}$ throughout the year including the missing months in graphs using any Python library(e.g. Matplotlib) you feel comfortable with.
3. Prepare a report justifying the decision which interpolating polynomial you are going to use and why.

Mark Distribution

Algorithm implementation(Code): 8

Plotting graphs: 1

Reasoning: 1

Rules

- At the time of evaluation, a group has to bring a printed copy of the report and a laptop for showing the code. All members of a group should be able to explain the code.
- Any type of plagiarism is strongly forbidden. **NO/NEGATIVE** marks will be given to the students who will be found to be involved in plagiarism (from internet/senior/class- mates code etc.).
It does not matter who is the server and who is the client.
- It is expected that all the members of a group will have equal participation in completing the project. If any certain member is not found to be involved in the project, **NO** marks will be provided for that particular member regardless of the group's performance.

Deadline

Deadline is set at **5 December**, 2019.