CS2008 Numerical Computing

Assignment No: 03		**CLO:**
		Semester: Spring 2024
Due date: As of GCR date		**Marks:** 50

Instructions

1. **Plagiarized work will result in zero marks.**

2. **No retake or late submission will be accepted.**

3. The submission should be a SINGLE UNZIPPED NOTEBOOK submitted on googles classroom.

4. This notebook should properly documented what you did? How you did it? And the source-code for each part as well as the generated outputs.

5. Your submission file should be according to the following **format: id_section_A2** e.g., i22123456_A_A2. (Note: A2 in the end denotes Assignment 2).

1		
2		
3		5x10=50 marks
4		
5		

The anscomb.csv dataset comprises four subdatasets (each one is plotted), each containing 11 data points. You need to perform approximation via interpolation & least square method.

subdataset I

subdataset II

subdataset III
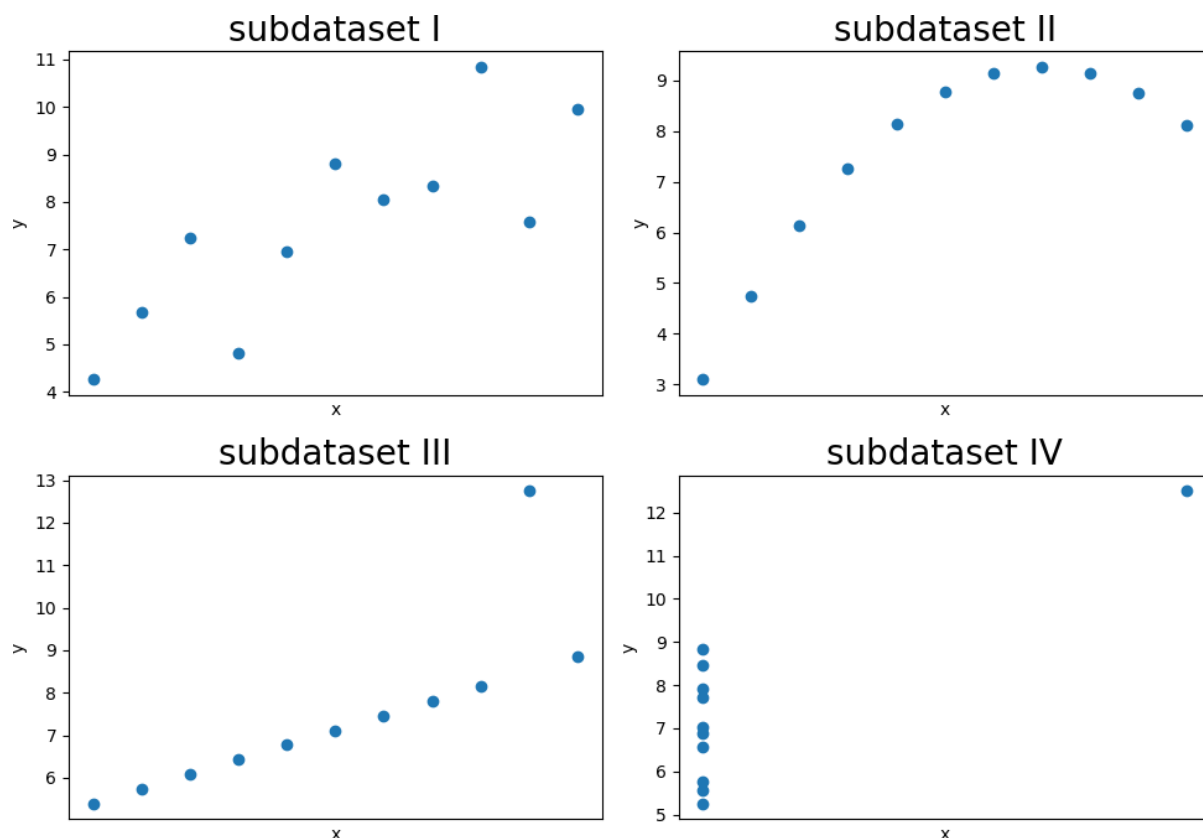
subdataset IV

Consider the subdataset II from above, you need to store the values of x from subdataset II in sorted form as mentioned in below table and also the values of y in sorted form.

Dataset 1

	X	4	5	6	7	8	9	10	11	12	13	14
Subdataset II	y	3.1	4.74	6.13	7.26	8.1	8.14	8.74	8.77	9.13	9.14	9.26

Now consider the following subset of data from above. And next perform tasks as mentioned.

Dataset 2

X1	8	9	10	11	12	13
y1	8.1	8.14	8.74	8.77	9.13	9.14

- Fit polynomial of degree 3 for **dataset 2** using least square method and test it by **dataset 1** to generate $\hat{y}lsm$
- Fit polynomial for **dataset 2** using interpolation from SciPy.Interpolate, and test p(x) on **dataset 1** to generate $\hat{y}\_lagrange$ values.
- Compare the original y values from subdataset II with $\hat{y}\_lsm$ find the error using RMSE.
- Compare the original y values from subdataset II with $\hat{y}\_lagrang$ find the error using RMSE.
- Which error is larger and why does the interpolation give larger error in this specific case? Conclude in one line.