Report HW3

Simone De Carli
simone.decarli@studenti.unitn.it

Damiano Salvaterra
damiano.salvaterra@studenti.unitn.it

Exercise 1

Given a dataset of approximately $3 \cdot 10^4$ entries of the form (x, y), where x is a time value and y is its corresponding value, we are interested in modeling the statistics of the dataset. There is a clear trend in the data.

Part 1: Visualize the trend

First, we need to visualize the trend in the data. To do this, we scatter plot the data points (1a) and we plot the distribution of the data points (1b) that we cannot trust.
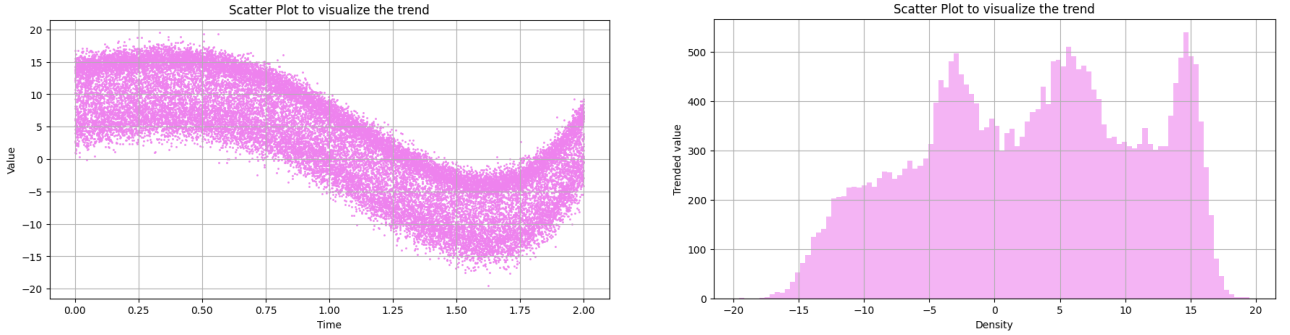
(a) Scatter plot of the dataset.

(b) Histogram of the dataset.

Figure 1: Trended dataset.

Part 2: Fit a polynomial trend

We estimate a polynomial trend using the least squares method for degrees from 1 to 8 (2).

Part 3: Find the best degree and remove the trend

To remove the trend, we need to decide which polynomial degree is the best. Then we can de-trend the data.

Find the best degree

To decide which polynomial degree is the best, we use the root mean square error (RMSE) of the polynomial fit for each degree (3a). We also use cross-validation to estimate the RMSE of the polynomial fit for each degree (3b).

Both methods in Figure 3 show that the best degree is 4.

De-trend the data

To de-trend the data, we just subtract the best polynomial fit (degree 4) from the data. The result is shown in Figure 4.

Part 4: Fit a mixture of Gaussians
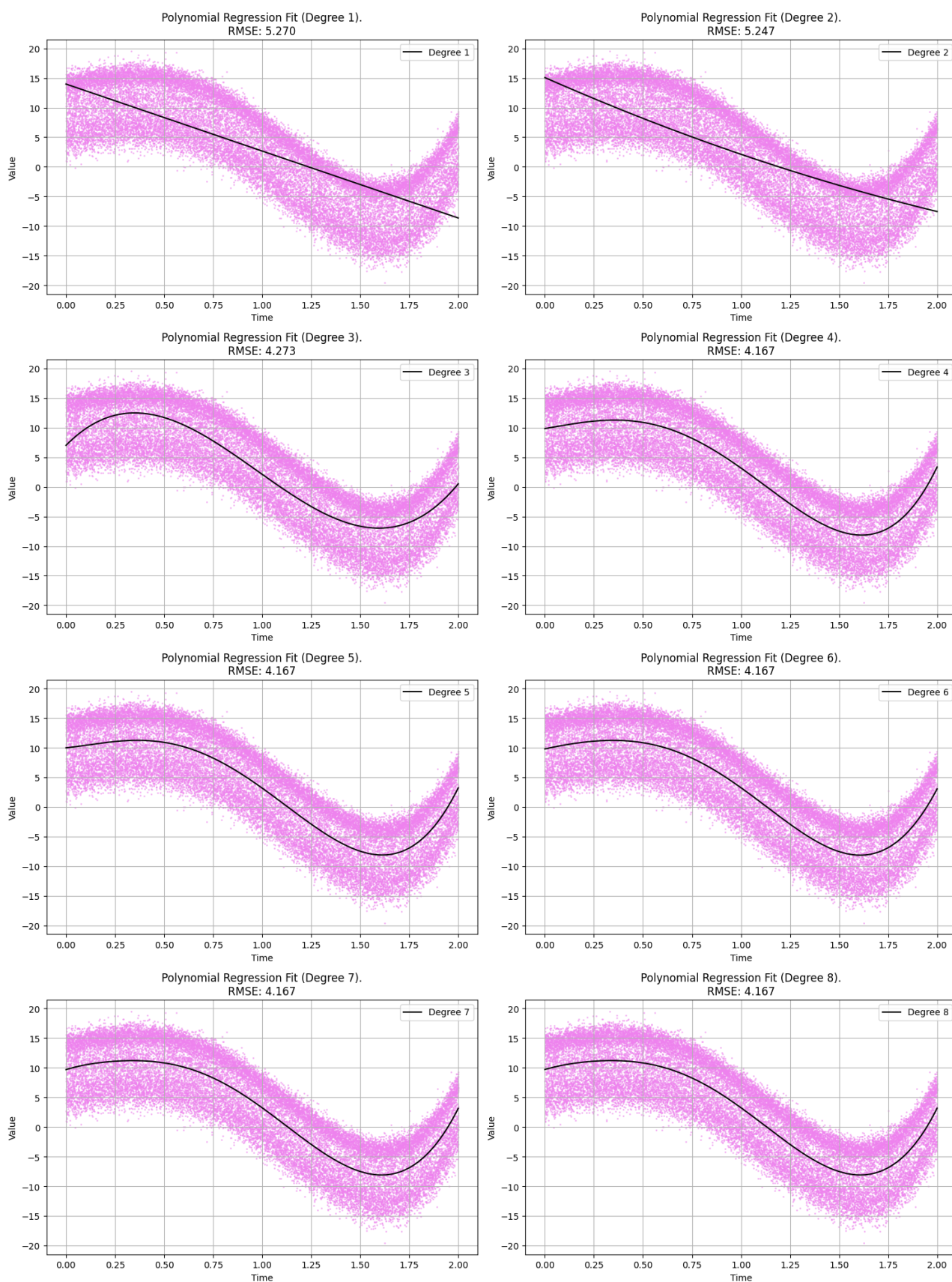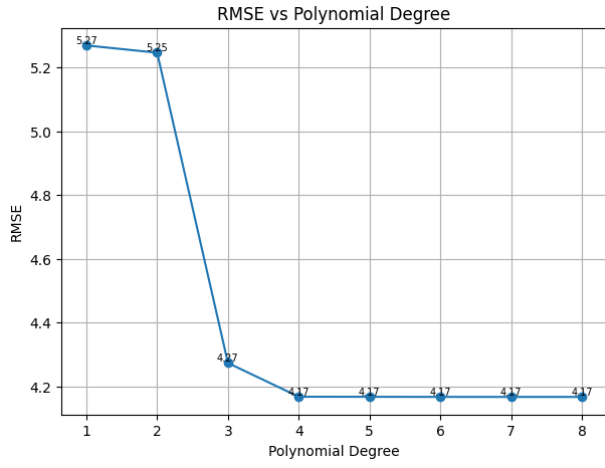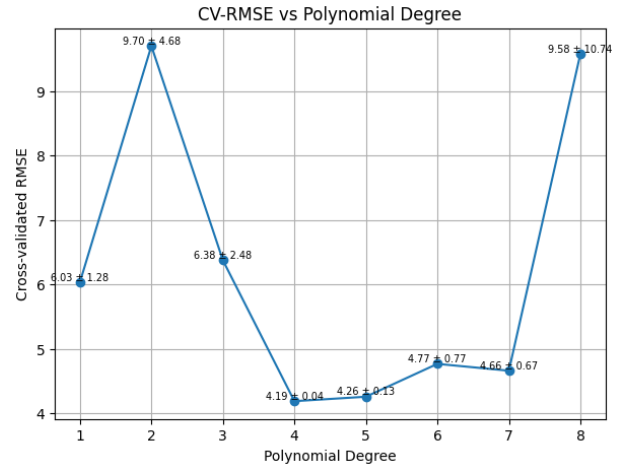
We use ....

Figure 2: Polynomial trends fitted to the dataset.
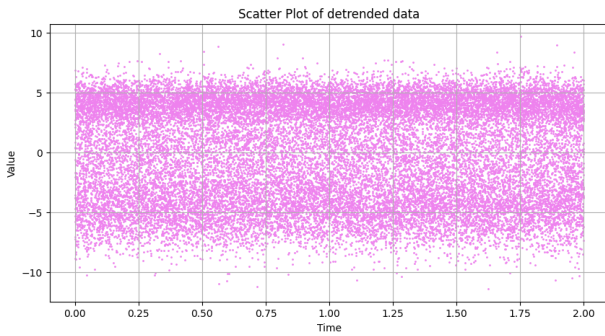
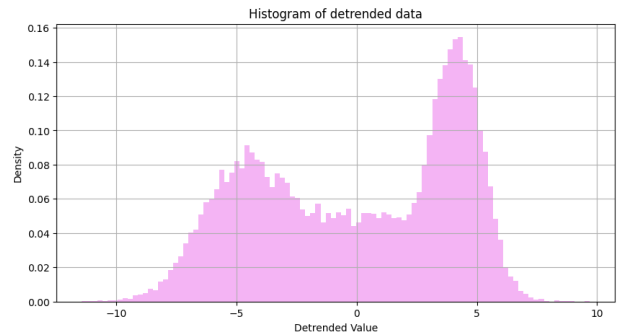(a) RMSE of the polynomial for each degree.

(b) Mean RMSE from the cross-validation for each degree.

Figure 3: Evaluation of the polynomial degree.

(a) Scatter plot of the de-trended dataset.

(b) Histogram of the de-trended dataset.

Figure 4: De-trended dataset.