ENGR 421/DASC 521: Introduction to Machine Learning
Homework 3: Nonparametric Regression
Deadline: December 4, 2023, 11:59 PM

In this homework, you will implement three nonparametric regression algorithms in Python. Here are the steps you need to follow:

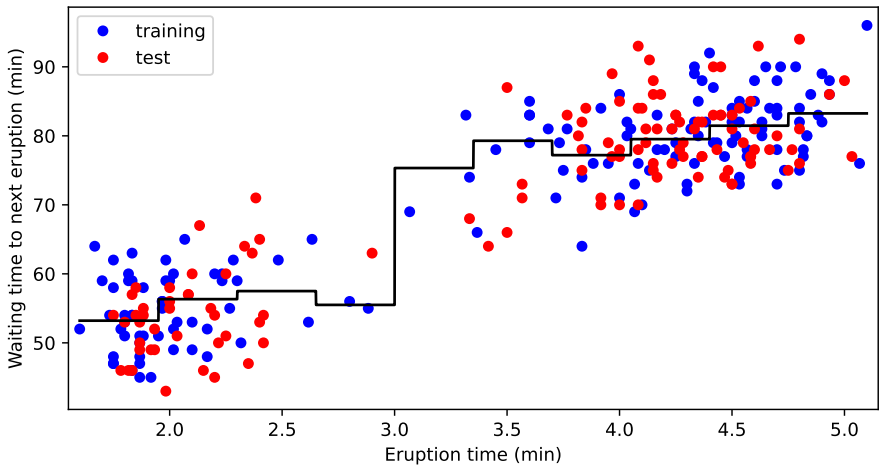
1. Read Section 8.8 from the textbook.

2. You are given a univariate regression data set about the duration of the eruption and waiting time between eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA (`https://www.yellowstonepark.com/things-to-do/about-old-faithful`), which contains 150 training data points in the file named `hw03_data_set_train.csv` and 122 test data points in the file named `hw03_data_set_test.csv`.

3. Learn a regressogram by setting the bin width parameter to 0.35 and the origin parameter to 1.6. (30 points)

$$g(x) = \frac{\sum_{i=1}^{N_{train}} b(x, x_i) y_i}{\sum_{i=1}^{N_{train}} b(x, x_i)} \qquad \text{where } b(x, x_i) = \begin{cases} 1 & \text{if } x_i \text{ is in the same bin with } x \\ 0 & \text{otherwise} \end{cases}$$

When you draw training data points, test data points, and your regressogram in the same figure, you should obtain the following figure.



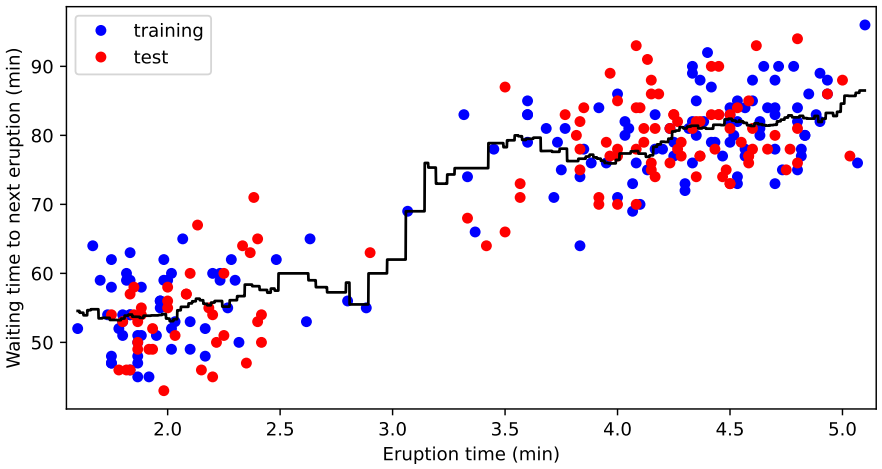
When you calculate the root mean squared error (RMSE) of your regressogram for test data points, you should obtain the following sentence as your output.

```
Regressogram => RMSE is 5.96665297472927 when h is 0.35
```

4. Learn a running mean smoother by setting the bin width parameter to 0.35. (35 points)

$$g(x) = \frac{\sum_{i=1}^{N_{train}} w\left(\frac{x - x_i}{h}\right) y_i}{\sum_{i=1}^{N_{train}} w\left(\frac{x - x_i}{h}\right)} \qquad \text{where } w(u) = \begin{cases} 1 & \text{if } -1/2 \leq u < 1/2 \\ 0 & \text{otherwise} \end{cases}$$

When you draw training data points, test data points, and your running mean smoother in the same figure, you should obtain the following figure.



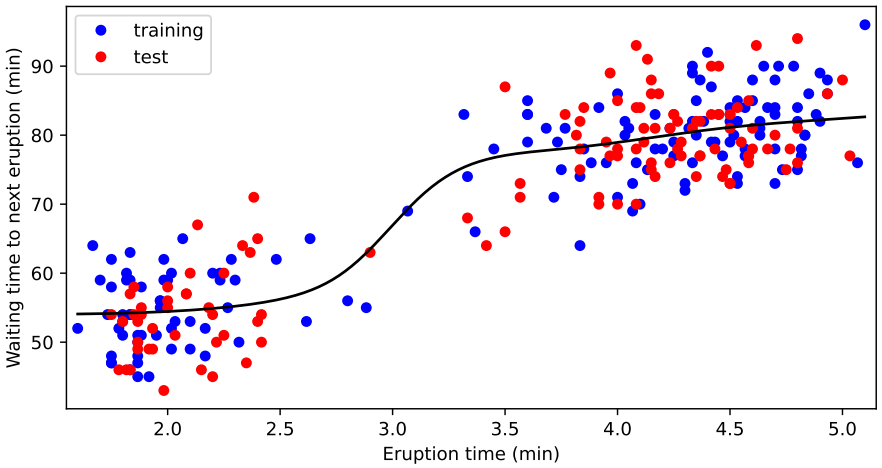
When you calculate the RMSE of your running mean smoother for test data points, you should obtain the following sentence as your output.

```
Running Mean Smoother => RMSE is 6.093996956866121 when h is 0.35
```

5. Learn a kernel smoother by setting the bin width parameter to 0.35. (35 points)

$$g(x) = \frac{\sum_{i=1}^{N_{train}} K\left(\frac{x - x_i}{h}\right) y_i}{\sum_{i=1}^{N_{train}} K\left(\frac{x - x_i}{h}\right)} \qquad \text{where } K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

When you draw training data points, test data points, and your kernel smoother in the same figure, you should obtain the following figure.



When you calculate the RMSE of your kernel smoother for test data points, you should obtain the following sentence as your output.

```
Kernel Smoother => RMSE is 5.8781517265154095 when h is 0.35
```

What to submit: You need to submit your source code in a single file (`.py` file). You are provided with a template file named as `0099999.py`, where `99999` should be replaced with your 5-digit student number. You are allowed to change the template file between the following lines.

```
# your implementation starts below

# your implementation ends above
```

How to submit: Submit the file you edited to Blackboard by following the exact style mentioned. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.