ENGR 421/DASC 521: Introduction to Machine Learning
**Homework 4:** Decision Tree Regression
Deadline: December 13, 2023, 11:59 PM

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

1. 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 `hw04_data_set_train.csv` and 122 test data points in the file named `hw04_data_set_test.csv`.

2. Implement a decision tree regression training function using the following pre-pruning rule: If a node has P or fewer data points, convert this node into a terminal node and do not split further, where P is a user-defined parameter. (60 points)

3. Implement a decision tree regression test function using the parameters of a trained tree to make predictions for a given set of data points. (20 points)

   When you draw training data points, test data points, and your decision tree fit in the same figure, you should obtain the following figures for $P = 20$ and $P = 50$.
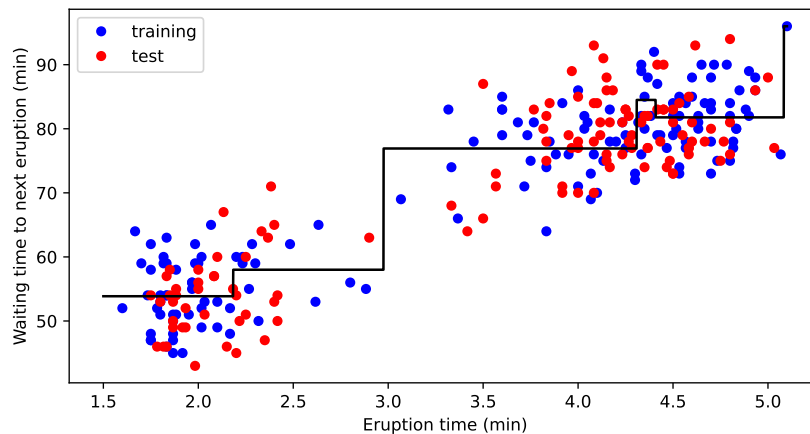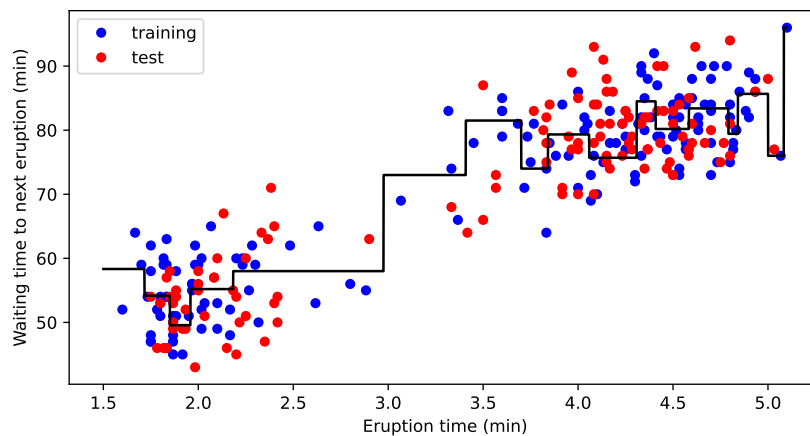
   When you calculate the root mean squared error (RMSE) of your decision tree fits for training and test data points, you should obtain the following sentences as your outputs.

```
RMSE on training set is 4.43291764487832 when P is 20
RMSE on test set is 6.714228091718778 when P is 20
RMSE on training set is 5.00414678298864 when P is 50
RMSE on test set is 6.328291469197311 when P is 50
```

4. Implement a decision tree regression rule extraction function using the parameters of a trained tree. (20 points)

   When you extract the rules of your decision tree fit, you should obtain the following sentences as your output for $P = 50$.

```
Node 05: ['x1 > 2.98' 'x1 <= 4.31'] => 76.92105263157895
Node 06: ['x1 <= 2.98' 'x1 > 2.18'] => 58.0
Node 07: ['x1 <= 2.98' 'x1 <= 2.18'] => 53.86363636363637
Node 08: ['x1 > 2.98' 'x1 > 4.31' 'x1 > 5.08'] => 96.0
Node 18: ['x1 > 2.98' 'x1 > 4.31' 'x1 <= 5.08' 'x1 > 4.41'] => 81.77777777777777
Node 19: ['x1 > 2.98' 'x1 > 4.31' 'x1 <= 5.08' 'x1 <= 4.41'] => 84.5
```

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.