ENGR 421/DASC 521: Introduction to Machine Learning
Homework 4: Decision Tree Regression
Deadline: April 29, 2023, 11:59 PM

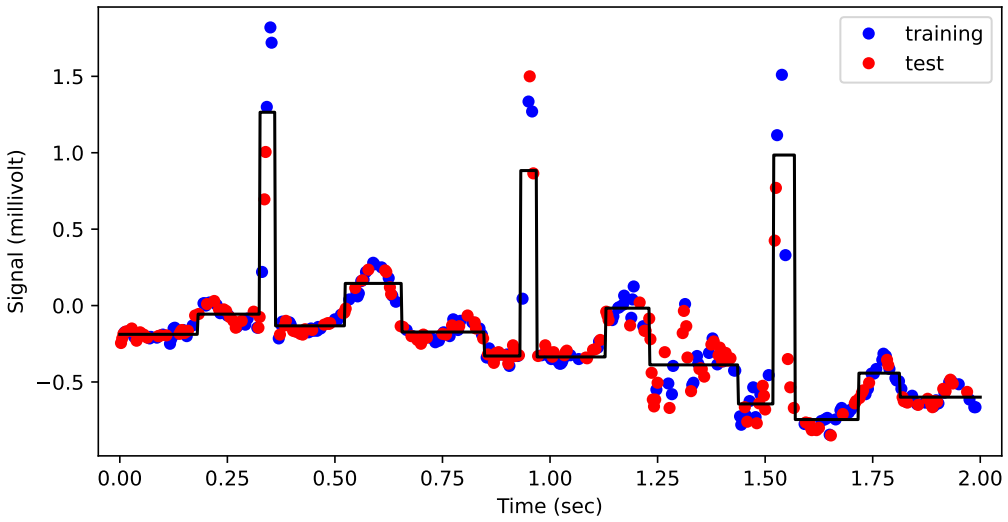
In this homework, you will implement a decision tree regression algorithms in Python. Here are the steps you need to follow:

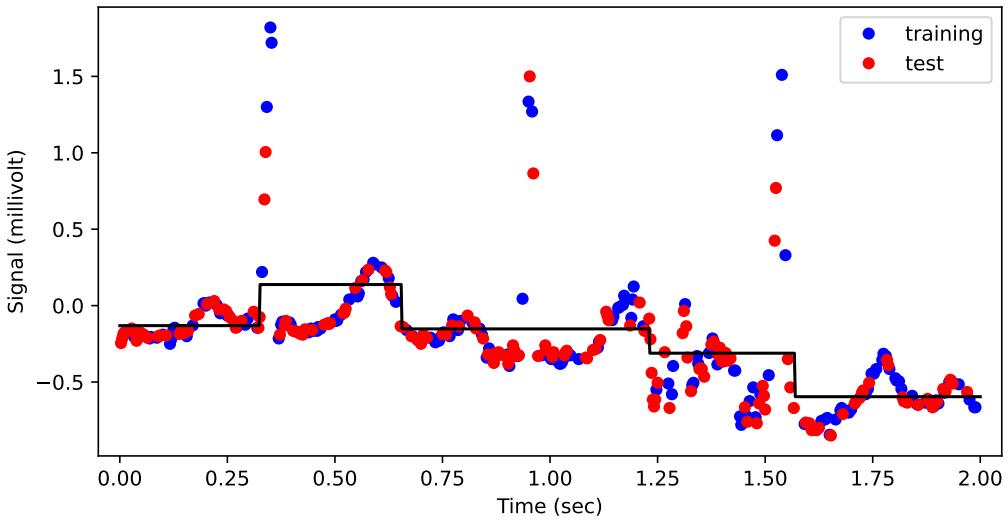
1. You are given a univariate regression data set, which contains 180 training data points in the file named `hw04_data_set_train.csv` and 180 test data points in the file named `hw04_data_set_test.csv`.

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 $P = 50$, respectively.





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 0.15366204830854638 when P is 20
RMSE on test set is 0.22615978758464964 when P is 20
RMSE on training set is 0.3451306112372657 when P is 50
RMSE on test set is 0.24805444775490876 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 04: ['x1 > 1.23' 'x1 > 1.57'] => -0.5960256410256409
Node 05: ['x1 > 1.23' 'x1 <= 1.57'] => -0.3112962962962963
Node 06: ['x1 <= 1.23' 'x1 > 0.65'] => -0.15229999999999996
Node 14: ['x1 <= 1.23' 'x1 <= 0.65' 'x1 > 0.33'] => 0.13794117647058823
Node 15: ['x1 <= 1.23' 'x1 <= 0.65' 'x1 <= 0.33'] => -0.131
```

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.
