

ENGR 421/DASC 521: Introduction to Machine Learning

Homework 4: Decision Tree Regression

Deadline: April 28, 2025, 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 bivariate regression data set, which contains 1000 training data points in the file named `hw04_data_set_train.csv` and 1000 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 calculate the root mean squared error (RMSE) of your decision tree fit for $P = 256$ on training and test data points, you should obtain the following sentences as your outputs.

RMSE on training set is 7.675348294551647 when P is 256

RMSE on test set is 8.262182073687075 when P is 256

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 for $P = 256$, you should obtain the following sentences as your output.

Node 04: ['x2 > 0.46' 'x2 > 1.29'] => -0.37945547003366686

Node 05: ['x2 > 0.46' 'x2 <= 1.29'] => -3.0900115914710766

Node 06: ['x2 <= 0.46' 'x2 > -0.36'] => 5.570112289839696

Node 14: ['x2 <= 0.46' 'x2 <= -0.36' 'x2 > -1.29'] => -2.627269186231939

Node 15: ['x2 <= 0.46' 'x2 <= -0.36' 'x2 <= -1.29'] => -0.487968567193978

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 LearnHub 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.
