

Small test running the Boosting Tree Code: IDE, PyCharm

The screenshot displays the PyCharm IDE interface. The top toolbar shows the 'Run' button (a green play icon) and a 'Readme_Test_Examples' dropdown menu. The left sidebar contains a project tree for 'Project2-Panama' with subfolders 'boosting_tree' and 'data_generators'. The 'Readme_Test_Examples.py' file is selected in the tree and open in the editor. The editor shows the following Python code:

```
2 from boosting_tree.BoostingTreeModel import BoostingTreeModel, MyRSquared, MyMSE
3
4 # Sample Data
5 X = np.random.rand(100, 3)
6 y = 2 * X[:, 0] + X[:, 1] - 1.5 * X[:, 2] + np.random.normal(loc=0, scale=0.05, size=100)
7
8 # Model Initialization
9 model = BoostingTreeModel(num_trees=50, learning_rate=0.05, max_depth=4, tol=1e-4, subsample=0.7)
10
11 # Train the Model
12 results = model.fit(X, y)
13
14 # Predict
15 y_pred = results.predict(X)
16
17 # Calculate Mean Squared Error
18 mse = MyMSE.calculate(y, y_pred)
19 # Calculate R-squared
20 r_squared = MyRSquared.calculate(y, y_pred)
21
22 print(f"Mean Squared Error: {mse:.4f}")
23 print(f"R-squared: {r_squared:.4f}")
24
```

Below the editor, the 'Run' tab is active, showing the execution output:

```
/usr/local/bin/python3.12 /Users/harleeliz/CS584-Project2-Panama/Project2-Panama/Readme_Test_Examples.py
Mean Squared Error: 0.2965
R-squared: 0.4600

Process finished with exit code 0
```

Test 1: Test1_BTModel_Tested_8_Datagenerators

Project2-Panama

boosting_tree

__init__.py

BoostingTreeModel.py

data_generators

__init__.py

data_generators.py

main.py

Readme_Test_Examples.py

small_test.csv

Test1_BTModel_Tested_8_Datagenerators.py

Test2_BTModel_SmallTest.py

Test3_BTModel_CaliforniaHousing.py

External Libraries

Scratches and Consoles

Readme_Test_Examples.py

Test1_BTModel_Tested_8_Datagenerators.py

```
1 import numpy as np
2 from matplotlib import pyplot as plt
3 from sklearn.ensemble import GradientBoostingRegressor
4 from sklearn.metrics import r2_score, mean_squared_error
5 from scipy import stats
6 from boosting_tree.BoostingTreeModel import BoostingTreeModel, MyRSquared, MyMSE
7 from data_generators.data_generators import (
8     linear_data_generator1,
9     linear_data_generator2,
10     nonlinear_data_generator1,
11     generate_collinear_data,
12     generate_periodic_data,
13     generate_higher_dim_data,
14     generate_high_collinear_data,
15     generate_horrible_data,
16 )
```

Run

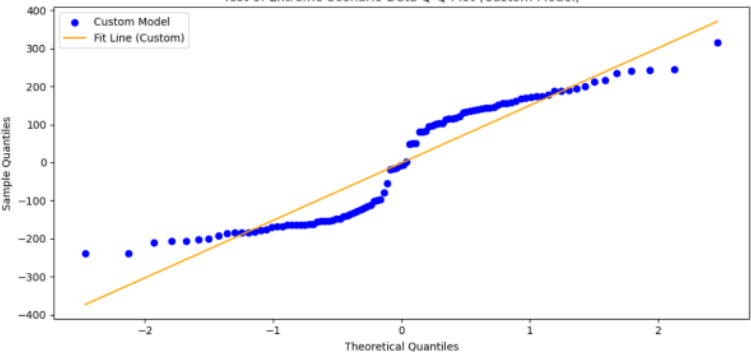
Test1_BTModel_Tested_8_Datagenerators

```
/usr/local/bin/python3.12 /Users/harleeliz/CS584-Project2-Panama/Project2-Panama
Running Test 1: Linear Data (Single Feature)...
Custom Model - R²: 0.7128, MSE: 11.1915, Time: 0.0653s
Sklearn Model - R²: 0.9124, MSE: 3.4143, Time: 0.0077s
Running Test 2: Linear Data (Multiple Features)...
Custom Model - R²: 0.1090, MSE: 2239.6081, Time: 0.1284s
Sklearn Model - R²: 0.6631, MSE: 846.7990, Time: 0.0072s
Running Test 3: Nonlinear Data...
Custom Model - R²: 0.9591, MSE: 0.2591, Time: 0.0836s
Sklearn Model - R²: 0.9859, MSE: 0.0895, Time: 0.0063s
Running Test 4: Collinear Data...
Custom Model - R²: 0.9718, MSE: 10.4852, Time: 0.4062s
Sklearn Model - R²: 0.9989, MSE: 0.4168, Time: 0.0090s
Running Test 5: Periodic Data...
Custom Model - R²: 0.3202, MSE: 22.5624, Time: 0.0538s
Sklearn Model - R²: 0.9932, MSE: 0.2269, Time: 0.0063s
Running Test 6: Higher Dimensional Data...
Custom Model - R²: 0.9934, MSE: 17.8907, Time: 0.3099s
Sklearn Model - R²: 0.9990, MSE: 2.6319, Time: 0.0079s
Running Test 7: High Collinearitv Data...
```

Plots

1,000x500 PNG (32-bit color) 37.01 kB

Test 8: Extreme Scenario Data Q-Q Plot (Custom Model)



18

19

20

21

22

23

6:1 (80 chars) UTF-8 4 spaces Python 3.12

Test 2: Test2_BTModel_SmallTest

Project2-Panama main

Project

- Project2-Panama ~/CS584-Project2-Panama/Proje
 - boosting_tree
 - __init__.py
 - BoostingTreeModel.py
 - data_generators
 - __init__.py
 - data_generators.py
 - main.py
 - Readme_Test_Examples.py
 - small_test.csv
 - Test1_BTModel_Tested_8_Datagenerators.py
 - Test2_BTModel_SmallTest.py
 - Test3_BTModel_CaliforniaHousing.py
- External Libraries
- Scratches and Consoles

Run Test2_BTModel_SmallTest

```
/usr/local/bin/python3.12 /Users/harleeLiz/CS584-Project2-Panama/Project2-Panama/Test2_BTModel_SmallTest.py
```

Test - Dataset from small_test.csv

Custom Boosting Tree - R²: 0.8908, MSE: 54.6706, Time: 0.1904s

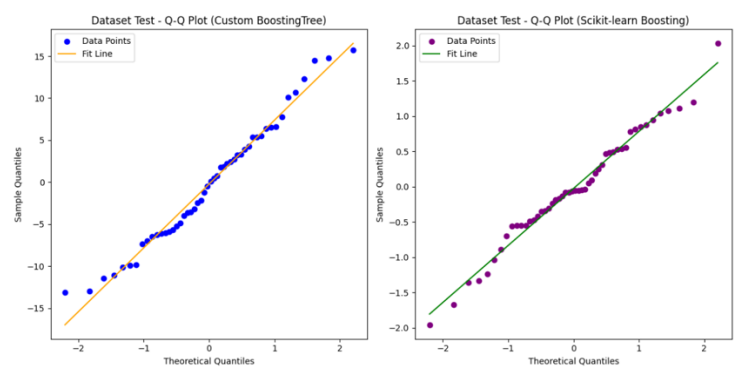
scikit-learn Boosting Tree - R²: 0.9988, MSE: 0.6190, Time: 0.0088s

Process finished with exit code 0

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 from sklearn.ensemble import GradientBoostingRegressor
5 from sklearn.metrics import r2_score, mean_squared_error
6 from scipy import stats
7 from time import time
8 from boosting_tree.BoostingTreeModel import BoostingTreeModel, MyRSquared, MyMSE
9
10 # Load the dataset
11 dataset_path = 'small_test.csv'
12 data = pd.read_csv(dataset_path)
13
14 # Split into features (X) and target (y)
15 X = data.iloc[:, :-1].values # All columns except the last one
16 y = data.iloc[:, -1].values # Last column
```

Plots

1,200x600 PNG (32-bit color) 54.63 kB



Test 3 : Test3_BTModel_CaliforniaHousing

Project2-Panama

main

Project2-Panama

boosting_tree

data_generators

External Libraries

Scratches and Consoles

Test3_BTModel_CaliforniaHousing.py

Test2_BTModel_SmallTest.py

Test1_BTModel_Testesd_8_Datagenerators.py

Readme_Test_Examples.py

main.py

data_generators.py

__init__.py

BoostingTreeModel.py

__init__.py

```
46 california_data.data
47 california_data.target
48 feature_names = california_data.feature_names
49
50 # Parameters for the model
51 num_trees = 20
52 learning_rate = 0.1
53 depth = 3
54 subsample = 0.5
55
56 # Custom Boosting Tree Model
57 custom_model = BoostingTreeModel(num_trees, learning_rate, depth, subsample)
58 start_time = time()
59 custom_results = custom_model.fit(X, y)
60 sklearn_results = custom_results.predict(X)
61 end_time = time() - start_time
62
63 # Scikit-learn Gradient Boosting Tree Model
64 sklearn_model = GradientBoostingRegressor()
65 start_time = time()
66 sklearn_model.fit(X, y)
67 sklearn_results = sklearn_model.predict(X)
68 end_time = time() - start_time
69
```

Run

Test3_BTModel_CaliforniaHousing

```
/usr/local/bin/python3.12 /Users/harleeliz/CS584-Project2-Panama/Project2-Panama/Test3
Test - California Housing Dataset
Custom Boosting Tree - R²: 0.0187, MSE: 1.3067, Time: 303.2805s
Scikit-learn Boosting Tree - R²: 0.6458, MSE: 0.4717, Time: 0.6525s

Process finished with exit code 0
```

Plots

1,000x500 PNG (32-bit color) 41.29 kB

California Housing Dataset Test - Q-Q Plot (Custom BoostingTree)

California Housing Dataset Test - Q-Q Plot (Scikit-learn Boosting)