Decision Regression Tree Continuous Multivariate

In [234]:
```
# Importing libraries

import pandas as pd
import numpy as np
import math
from sklearn.tree import DecisionTreeRegressor  # Import decision tree regressor
from sklearn.model_selection import train_test_split # Import train_test_split function
from sklearn import metrics #Import scikit-learn metrics module for accuracy calculation
from sklearn.metrics import classification_report, confusion_matrix, r2_score


df = pd.read_csv('/Users/juanrquilesjr/Downloads/UCI_MachineLearningDataSets/Bike-Sharing-Dataset/day.csv')

# Setting indexing to 1
df.index = df.index + 1
df = df.head(20)
df
```

Out[234]:

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
1	1	2011-01-01	1	0	1	0	6	0	2	0.344167	0.363625	0.805833	0.160446	331	654	985
2	2	2011-01-02	1	0	1	0	0	0	2	0.363478	0.353739	0.696087	0.248539	131	670	801
3	3	2011-01-03	1	0	1	0	1	1	1	0.196364	0.189405	0.437273	0.248309	120	1229	1349
4	4	2011-01-04	1	0	1	0	2	1	1	0.200000	0.212122	0.590435	0.160296	108	1454	1562
5	5	2011-01-05	1	0	1	0	3	1	1	0.226957	0.229270	0.436957	0.186900	82	1518	1600
6	6	2011-01-06	1	0	1	0	4	1	1	0.204348	0.233209	0.518261	0.089565	88	1518	1606
7	7	2011-01-07	1	0	1	0	5	1	2	0.196522	0.208839	0.498696	0.168726	148	1362	1510
8	8	2011-01-08	1	0	1	0	6	0	2	0.165000	0.162254	0.535833	0.266804	68	891	959
9	9	2011-01-09	1	0	1	0	0	0	1	0.138333	0.116175	0.434167	0.361950	54	768	822
10	10	2011-01-10	1	0	1	0	1	1	1	0.150833	0.150888	0.482917	0.223267	41	1280	1321
11	11	2011-01-11	1	0	1	0	2	1	2	0.169091	0.191464	0.686364	0.122132	43	1220	1263

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
12	12	2011-01-12	1	0	1	0	3	1	1	0.172727	0.160473	0.599545	0.304627	25	1137	1162
13	13	2011-01-13	1	0	1	0	4	1	1	0.165000	0.150883	0.470417	0.301000	38	1368	1406
14	14	2011-01-14	1	0	1	0	5	1	1	0.160870	0.188413	0.537826	0.126548	54	1367	1421
15	15	2011-01-15	1	0	1	0	6	0	2	0.233333	0.248112	0.498750	0.157963	222	1026	1248
16	16	2011-01-16	1	0	1	0	0	0	1	0.231667	0.234217	0.483750	0.188433	251	953	1204
17	17	2011-01-17	1	0	1	1	1	0	2	0.175833	0.176771	0.537500	0.194017	117	883	1000
18	18	2011-01-18	1	0	1	0	2	1	2	0.216667	0.232333	0.861667	0.146775	9	674	683
19	19	2011-01-19	1	0	1	0	3	1	2	0.292174	0.298422	0.741739	0.208317	78	1572	1650
20	20	2011-01-20	1	0	1	0	4	1	2	0.261667	0.255050	0.538333	0.195904	83	1844	1927

In [218]:
```
df.shape
```

Out[218]:
```
(20, 16)
```

In [338]:
```
# Creating feature and target variables

X = df[['temp','atemp', 'hum', 'windspeed']].values
y = df['cnt'].values
```

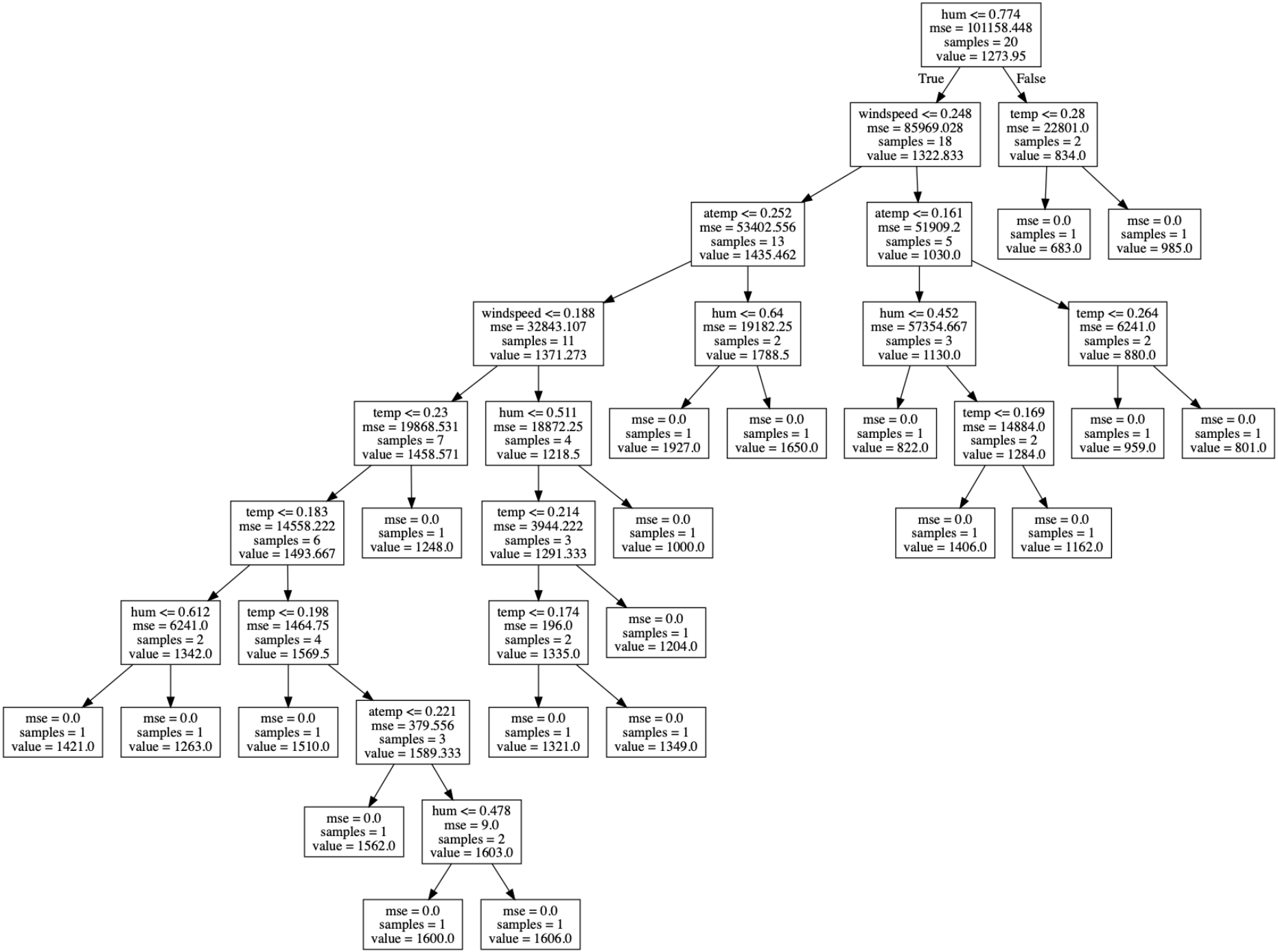
In [339]:
```
X
```

Out[339]:
```
array([[0.344167 , 0.363625 , 0.805833 , 0.160446 ],
       [0.363478 , 0.353739 , 0.696087 , 0.248539 ],
       [0.196364 , 0.189405 , 0.437273 , 0.248309 ],
       [0.2      , 0.212122 , 0.590435 , 0.160296 ],
       [0.226957 , 0.22927  , 0.436957 , 0.1869   ],
       [0.204348 , 0.233209 , 0.518261 , 0.0895652],
       [0.196522 , 0.208839 , 0.498696 , 0.168726 ],
       [0.165    , 0.162254 , 0.535833 , 0.266804 ],
       [0.138333 , 0.116175 , 0.434167 , 0.36195  ],
       [0.150833 , 0.150888 , 0.482917 , 0.223267 ],
       [0.169091 , 0.191464 , 0.686364 , 0.122132 ],
       [0.172727 , 0.160473 , 0.599545 , 0.304627 ],
       [0.165    , 0.150883 , 0.470417 , 0.301    ],
       [0.16087  , 0.188413 , 0.537826 , 0.126548 ],
       [0.233333 , 0.248112 , 0.49875  , 0.157963 ],
       [0.231667 , 0.234217 , 0.48375  , 0.188433 ],
       [0.175833 , 0.176771 , 0.5375   , 0.194017 ],
       [0.216667 , 0.232333 , 0.861667 , 0.146775 ],
       [0.292174 , 0.298422 , 0.741739 , 0.208317 ],
       [0.261667 , 0.25505  , 0.538333 , 0.195904 ]])
```

In [308]:
```
# Instantiation and fitting decision tree regressor

regressor = DecisionTreeRegressor()
regressor = regressor.fit(X,y)
```

In [309]:
```
# Visualization of Results

import matplotlib.pyplot as plt
%matplotlib inline

from sklearn import tree
from sklearn.tree import export_graphviz
import pydotplus
from IPython.display import Image

# Create DOT data
dot_data = tree.export_graphviz(regressor, out_file=None,
                                feature_names= ['temp', 'atemp', 'hum', 'windspeed'],
                                class_names= ['cnt'])

# Draw graph
graph = pydotplus.graph_from_dot_data(dot_data)

# Show graph
Image(graph.create_png())
```

Out[309]:



In [310]:
```
pred = regressor.predict(np.array([0.23333, 0.187783, 0.75587, 0.13456]).reshape(1,-1))
pred
```

Out[310]:
```
array([1248.])
```

Decision Tree Regression Discrete Multivariate

In [350]:
```
# Creating feature and target variables

X = df[['season','yr','mnth', 'holiday', 'weekday', 'workingday', 'weathersit']].values
y = df['cnt'].values
```

In [351]:
```
print(X.shape)
print(y.shape)
```

```
(20, 7)
(20,)
```

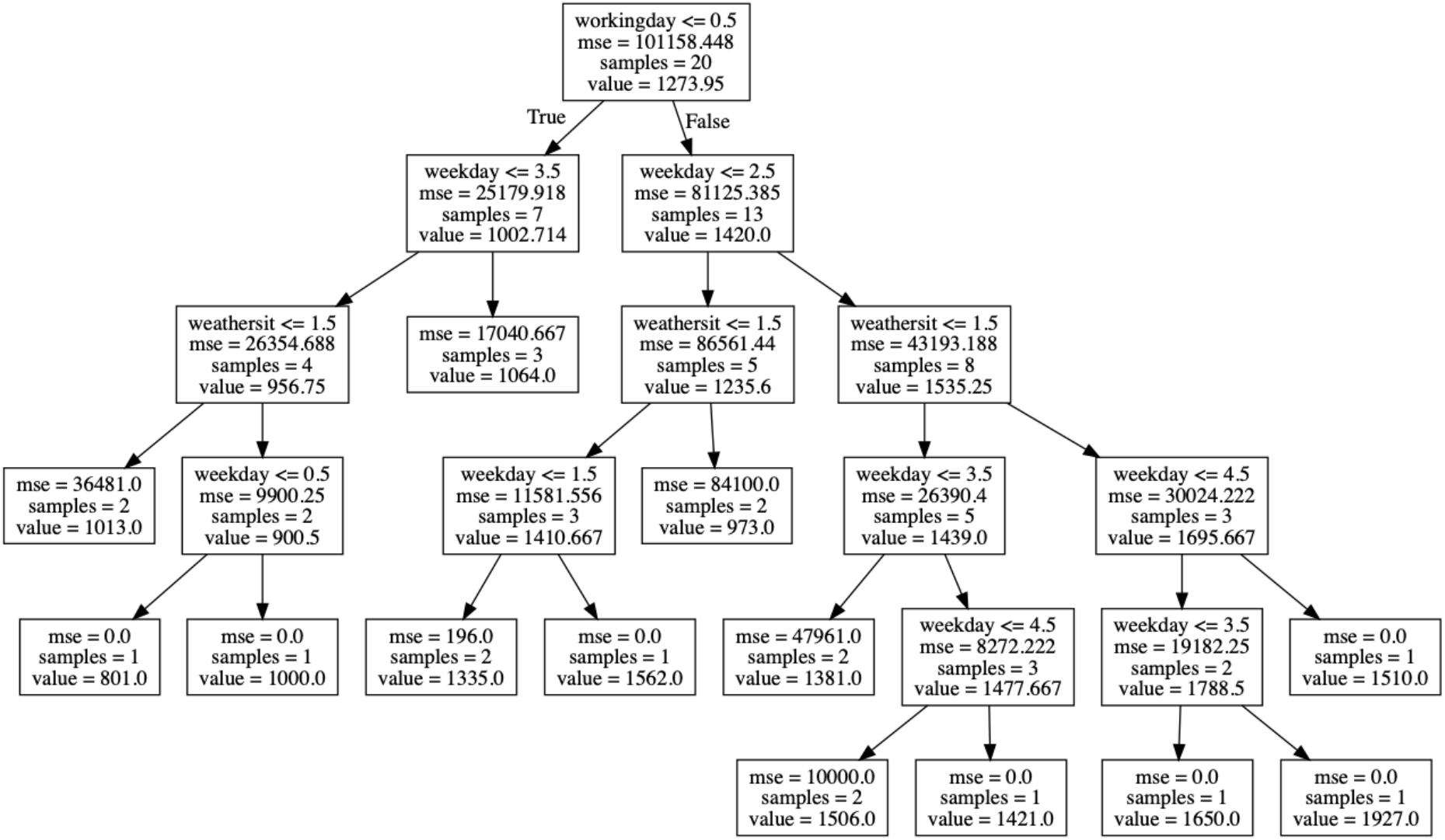
In [352]:
```
# Instantiation and fitting decision tree regressor

regressor = DecisionTreeRegressor()
regressor = regressor.fit(X,y)
```

In [353]:
```
pred = regressor.predict(np.array([4, 0, 3, 1, 1, 0, 1 ]).reshape(1,-1))
pred
```

Out[353]:
```
array([1013.])
```

In [354]:
```
# Visualization of Results

import matplotlib.pyplot as plt
%matplotlib inline

from sklearn import tree
from sklearn.tree import export_graphviz
import pydotplus
from IPython.display import Image

# Create DOT data
dot_data = tree.export_graphviz(regressor, out_file=None,
                                feature_names= ['season','yr','mnth', 'holiday', 'weekday', 'workingday', 'weath
                                class_names= ['cnt'])

# Draw graph
graph = pydotplus.graph_from_dot_data(dot_data)

# Show graph
Image(graph.create_png())
```

Out[354]:


workingday <= 0.5
mse = 101158.448
samples = 20
value = 1273.95
True
False
weekday <= 3.5
mse = 25179.918
samples = 7
value = 1002.714
weekday <= 2.5
mse = 81125.385
samples = 13
value = 1420.0
weathersit <= 1.5
mse = 26354.688
samples = 4
value = 956.75
mse = 17040.667
samples = 3
value = 1064.0
weathersit <= 1.5
mse = 86561.44
samples = 5
value = 1235.6
weathersit <= 1.5
mse = 43193.188
samples = 8
value = 1535.25
mse = 36481.0
samples = 2
value = 1013.0
weekday <= 0.5
mse = 9900.25
samples = 2
value = 900.5
weekday <= 1.5
mse = 11581.556
samples = 3
value = 1410.667
mse = 84100.0
samples = 2
value = 973.0
weekday <= 3.5
mse = 26390.4
samples = 5
value = 1439.0
weekday <= 4.5
mse = 30024.222
samples = 3
value = 1695.667
mse = 0.0
samples = 1
value = 801.0
mse = 0.0
samples = 1
value = 1000.0
mse = 196.0
samples = 2
value = 1335.0
mse = 0.0
samples = 1
value = 1562.0
mse = 47961.0
samples = 2
value = 1381.0
weekday <= 4.5
mse = 8272.222
samples = 3
value = 1477.667
weekday <= 3.5
mse = 19182.25
samples = 2
value = 1788.5
mse = 0.0
samples = 1
value = 1510.0
mse = 10000.0
samples = 2
value = 1506.0
mse = 0.0
samples = 1
value = 1421.0
mse = 0.0
samples = 1
value = 1650.0
mse = 0.0
samples = 1
value = 1927.0

In [ ]: