

 Open in Colab



# Decision Trees and Random Forests

MGMT 638: Data-Driven Investments: Equity

Kerry Back, Rice University



Imports



```
In [6]: from sklearn.tree import DecisionTreeRegressor, plot_tree
        from sklearn.ensemble import RandomForestRegressor
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        sns.set_style("whitegrid")
```

Generate data



```
In [4]: np.random.seed(0)
x1 = np.random.normal(size=100)
x2 = np.random.normal(size=100)
e = np.random.normal(size=100)
y = 2*x1 + 3*x2 + e
df = pd.DataFrame(
    dict(x1=x1, x2=x2, y=y)
)
df.head()
```

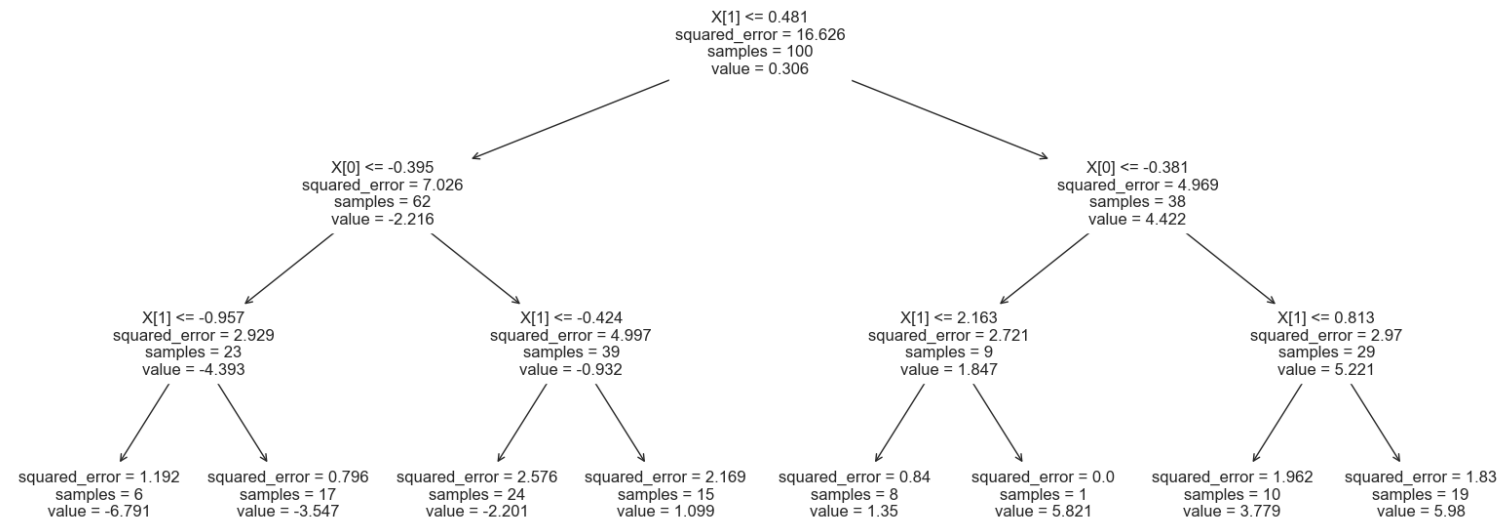
```
Out[4]:
```

	x1	x2	y
0	1.764052	1.883151	8.808375
1	0.400157	-1.347759	-3.482342
2	0.978738	-1.270485	-0.754319
3	2.240893	0.969397	8.045240
4	1.867558	-1.173123	0.855877

Fit and view a decision tree



```
In [14]: tree = DecisionTreeRegressor(max_depth=3)
tree.fit(X=df[["x1", "x2"]], y=df.y)
plt.figure(figsize=(20, 8))
plot_tree(tree, fontsize=12)
plt.show()
```





Fit a random forest and view goodness of fit



```
In [16]: forest = RandomForestRegressor(max_depth=3)
forest.fit(X=df[["x1", "x2"]], y=df.y)
predict = forest.predict(X=df[["x1", "x2"]])
sns.regplot(x=df.y, y=predict, ci=None)
plt.xlabel("Actual y")
plt.ylabel("Predicted y")
plt.show()
```

