Home Work 1

100 Possible Points

Due: Sun Jan 21, 2024 11:59pm

Add Comment

Available: Jan 10, 2024 12:00am until Jan 24, 2024 11:59pm

Details

1.Insert a column in the data set where the entries are 1 if the stock outperforms SPY in the earnings period and -1 if it underperforms or has the same return

2.Insert a column in the data set with entries: 2 if the stock return is more then 5% higher than the SPY return, 1 if it is more than 1% but less than 5% higher, 0 if it is between -1% and 1%, -2 if the stock underperforms the SPY by more than -5% and -1 if the performance is between -1% and -5%

1.Insert a column in the data set where the entries are 1 if the stock outperforms SPY in the earnings period and -1 if it underperforms or has the same return

2.Insert a column in the data set with entries: 2 if the stock return is more then 5% higher than the SPY return, 1 if it is more than 1% but less than 5% higher, 0 if it is between -1% and 1%, -2 if the stock underperforms the SPY by more than -5% and -1 if the performance is between -1% and -5%

3.A regression tree is used when the labels are real numbers instead of categories. Think of a linear regression situation when we have data points {xi} and response variables {yi} that are real numbers.
A regression tree uses the variance of the response variables instead of the Gini index. If n_j is a node with data $\{x_{ij}\}$ and responses $\{y_{ij}\}$ the variance of the response variables is

$$Var(\{y_{ij}\}) = \frac{1}{\#\{y_{ij}\}} \sum_i (y_{ij} - \overline{y}_{ij})^2$$

where $\overline{y}_{ij}$ is the average of the $\{y_{ij}\}$

To split the node into n_{j_1} and n_{j_2} we look for the split that minimizes

$$\frac{\#n_{j1}}{\#n_j} Var(\{y_{ij1}\}) + \frac{\#n_{j2}}{\#n_j} Var(\{y_{ij2}\})$$

In the notebook “Visualizing Trees” use a DecisionTreeRegressor instead of the DecisionTreeClassifier, directly on the data1 and the target (so do not transform the target into labels).

Instead of taking max in each rectangle take the average and generate the image.

Experiment with different color schemes (cm.?)