

---

```

% Austin Welch
% EC503 HW7.1
% Exploring Boston Housing Data with Regression Trees
% https://archive.ics.uci.edu/ml/datasets/Housing

% (a) Use MATLAB's built-in functions to learn a regression tree using
the
% training data provided. Visualize and submit an image of the
regression
% tree that has a minimum of 20 observations per leaf.
% MATLAB functions: classregtree, fitrtree, eval, test, view.

% load data
clear; clc;
load('housing_data.mat');

% construct regression tree
tree = fitrtree(Xtrain, ytrain, 'PredictorNames', feature_names, ...
    'ResponseName', cell2mat(output_name), 'MinLeafSize', 20);

% display
view(tree, 'Mode', 'graph')

% (b) For the trained regression tree from part (a) that has a minimum
of
% 20 observations per leaf, what is the estimated MEDV value for the
% following test feature vector: CRIM = 5, ZN = 18, INDUS = 2.31, CHAS
= 1,
% NOX = 0.5440, RM = 2, AGE = 64, DIS = 3.7, RAD = 1, TAX = 300,
PTRATIO =
% 15, B = 390, LSTAT = 10?
testVec = [5 18 2.31 1 0.5440 2 64 3.7 1 300 15 390 10];
y_hat = predict(tree, testVec);
fprintf('Estimated MEDV value for the given feature vector is: %0.4f
\n', ...
    y_hat);

% (c) Plot the mean absolute error (MAE) of the training and testing
data
% as a function of the minimum observations per leaf ranging from 1 to
25
% (one of the many ways to change the sensitivity of a tree). What
trend do
% you notice?

% CHECK if I need train on train then test on test, or make plots for
each
% train/train, train/test

% observations per leaf range
range = 25;

```

---

---

```

meanAbsErr = zeros(1,range);
for opl=1:range
    % train
    tree = fitrtree(Xtrain, ytrain, 'PredictorNames',
feature_names, ...
'ResponseName',cell2mat(output_name),'MinLeafSize',opl);
    % predict
    y_hat = predict(tree,Xtest);
    %error
    err = (ytest-y_hat);
    meanAbsErr(opl) = mae(err);
end

% plot
plot(1:range,meanAbsErr);
title('Minimum observations per leaf versus MAE');
xlabel('Minimum number of observations per leaf');
ylabel('Mean absolute error');

% observations
fprintf(['\nIt looks like the plot of # of observations per leaf
\n',...
'versus MAE decreases as number of observations increases,\n',...
'reaches a local minima, increases a bit before decreasing to
\n',...
'another local minima, then continues to increase.\n\n']);

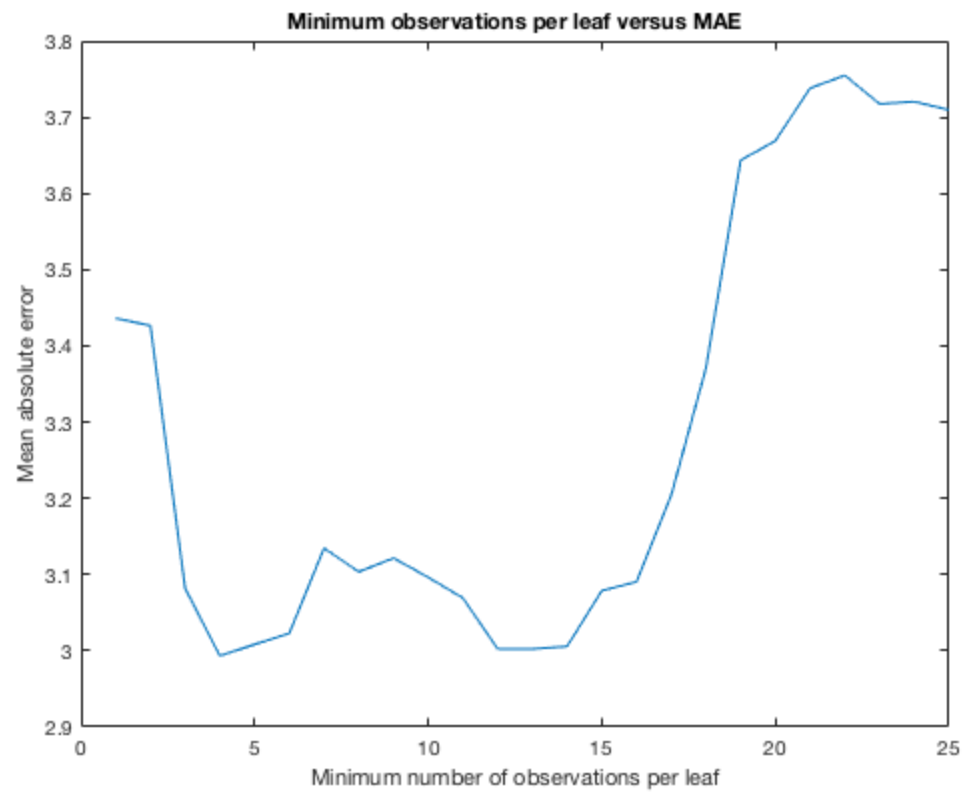
% best value for 'MinLeafSize'
[~,ind] = min(meanAbsErr);
fprintf('Best 'MinLeafSize': %d\n\n', ind);

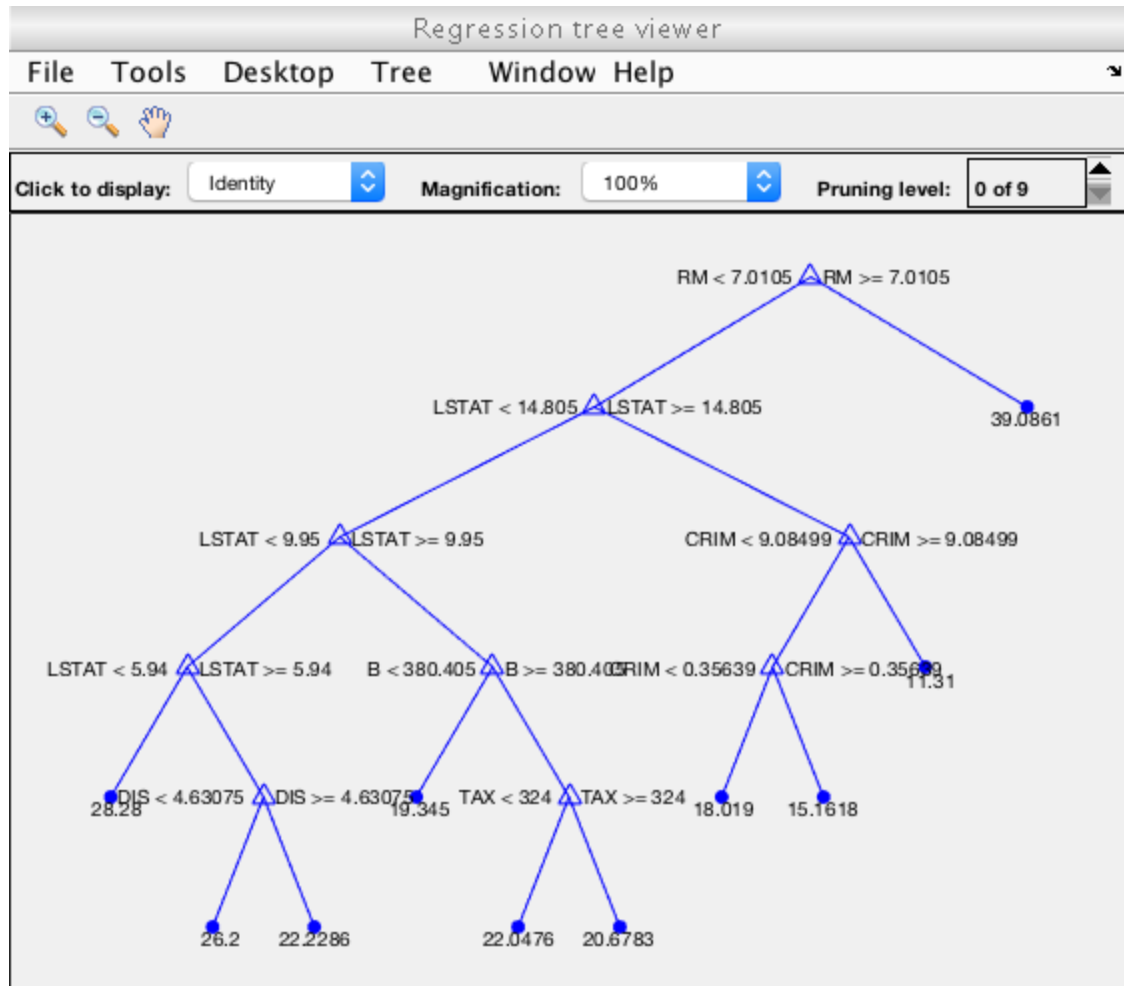
Estimated MEDV value for the given feature vector is: 22.0476

It looks like the plot of # of observations per leaf
versus MAE decreases as number of observations increases,
reaches a local minima, increases a bit before decreasing to
another local minima, then continues to increase.

Best 'MinLeafSize': 4

```





Published with MATLAB® R2017a