Decision Tree Regression:

* The tree is built *greedily* from top to bottom
* Each split is selected to maximize info gain (IG)

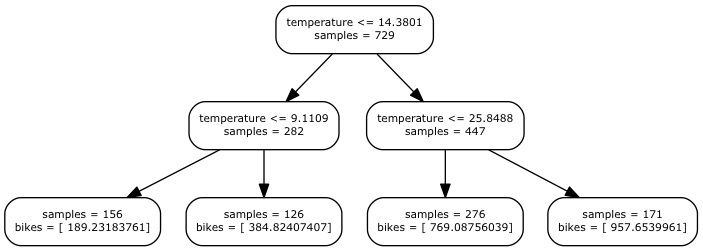
Here you can note that the prediction increases when the temperature increases. You can also note that the prediction is a stepwise function of the temperature. This is strictly related to how Decision Trees work. A decision tree splits the input features (only temperature in this case) in several regions and assigns a prediction value to each region. The selection of the regions and the predicted value within a region are chosen in order to produce the prediction which best fits the data. Where for best fit we mean that it minimizes the distance of the observations from the prediction.

You can inspect the set of rules created during the training process by exporting the tree:

from sklearn.tree import export\_graphviz

export\_graphviz(regressor, out\_file='tree.dot', feature\_names=['temperature'])

Here we exported the tree in dot format. Visualising it with Graphviz (http://www.graphviz.org/) we have the following result:



Decision Tree

The rules are organised in a binary tree: each time you ask to estimate the number of bikes hired the temperature is checked with the rules starting from the root of tree till the bottom following the path dictated form the outcome of the rules. For example, if the input temperature is 16.5, the first checked (temperature <= 14.3) will give a negative outcome leading us to its right child node in the next level of the tree (temperature <= 25.8), this time you will have a positive response and you will end in its left child. This node is a leaf, which means that it doesn't contain a rule but the value that you want to predict. In this case, it predicts 769 bikes.

Usually decision trees can be much deeper, and the deeper they are, the more complexity they are able to explain. In this case, a two level tree was configured using the parameter max\_depth during the instantiation of the model.

After splitting the data into different groups, the averages of each group is calculated. Then if we want to predict value y for a particular value X, the regressor looks for the group where X lies. Then it predicts the value of y by putting the average of the group as y.