***What is decision Tree?***

**Decision Trees (DTs)** are a non-parametric supervised learning method used for [classification](https://scikit-learn.org/stable/modules/tree.html#tree-classification) and [regression](https://scikit-learn.org/stable/modules/tree.html#tree-regression). The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

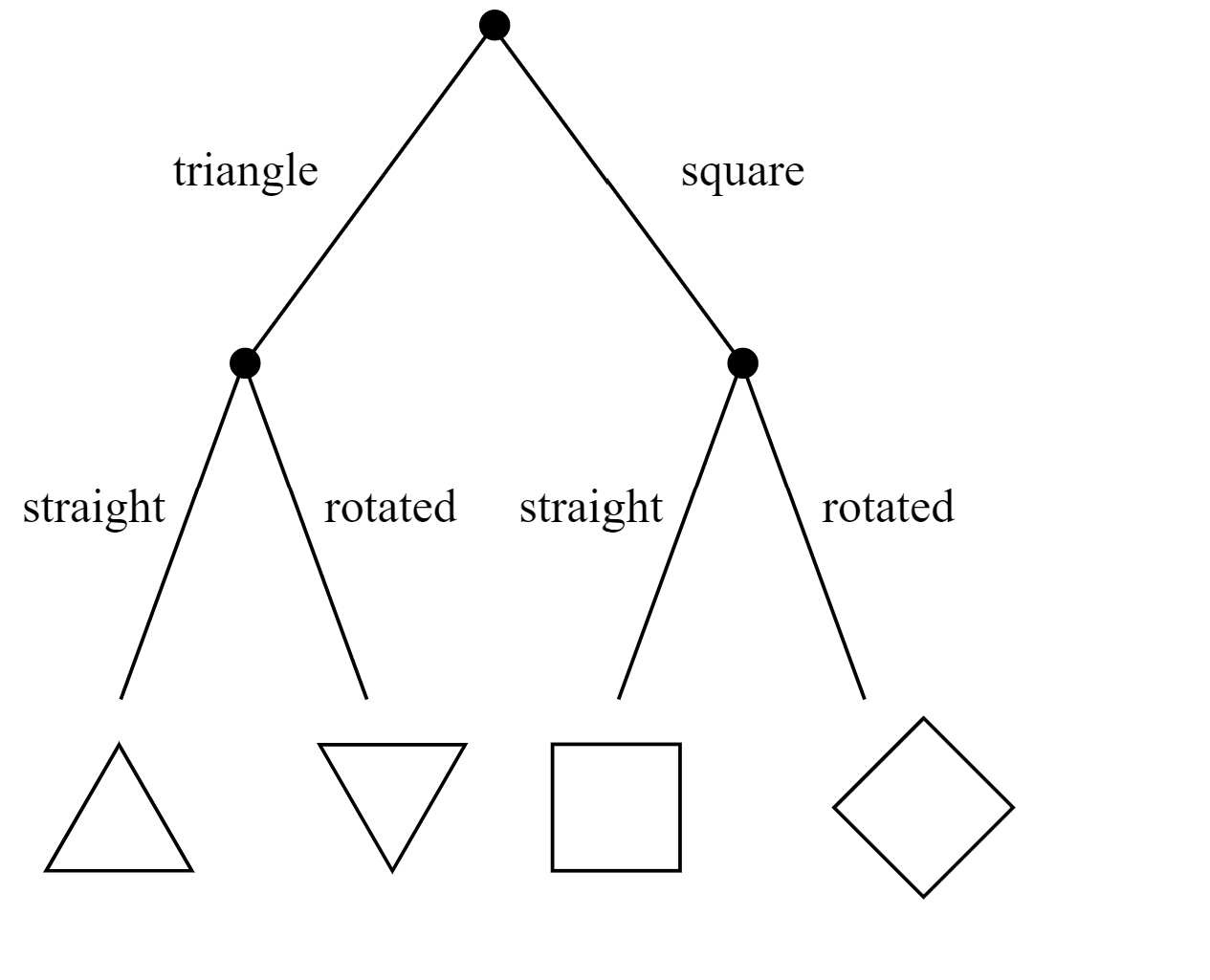
For instance, in the example below, decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model.

Some advantages of decision trees are:

* Simple to understand and to interpret. Trees can be visualised.
* Requires little data preparation. Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed. Note however that this module does not support missing values.

The disadvantages of decision trees include:

* Decision-tree learners can create over-complex trees that do not generalise the data well. This is called overfitting. Mechanisms such as pruning, setting the minimum number of samples required at a leaf node or setting the maximum depth of the tree are necessary to avoid this problem.
* Decision trees can be unstable because small variations in the data might result in a completely different tree being generated. This problem is mitigated by using decision trees within an ensemble.



*General structure of tree.*

## ***Task:***

In this notebook we implement the decision tree algorithm for indicating whether it is humid or not based on features such as ['air\_pressure\_9am', 'air\_temp\_9am', 'avg\_wind\_direction\_9am', 'avg\_wind\_speed\_9am', 'max\_wind\_direction\_9am', 'max\_wind\_speed\_9am', 'rain\_accumulation\_9am', 'rain\_duration\_9am', 'relative\_humidity\_9am']

Using these features we are predicting ***'relative\_humidity\_3pm'***

***Applications of Decision Tree:***

### Business Management

### Customer Relationship Management

### Fraudulent Statement Detection

### Energy Consumption

### *Packages used :*

### Pandas <https://pandas.pydata.org/> Matplotlib <https://matplotlib.org/> Numpy <https://numpy.org/> Seaborn <https://seaborn.pydata.org/> Scikit-learn <https://scikit-learn.org/>