

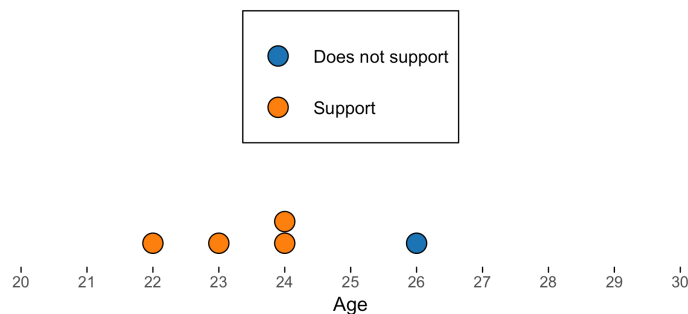
# Building Decision Trees

2018-09-21

## Building a Decision Tree from Data

In previous CART activities, you were supplied with pre-built decision trees. This time, you will be creating a tree from a small set of data. The data set (shown below) includes two variables for a sample of 24 people: the person's age and whether or not the person supports same-sex marriage. Your task is to create a classification tree using the data (not your intuition) that will predict whether or not a person supports same-sex marriage based on their age.

To build the tree you will create a set of decision rules based on age that attempt to classify people as “supports same-sex marriage” or “does not support same-sex marriage”. For example, consider the following dataset:



In these data, a good partitioning might be at 25 years of age since everyone younger than 25 supports same-sex marriage and everyone over 25 does not support same-sex marriage. Your decision rule might then be: {age < 25?} If “Yes”, they support same-sex marriage. If “No”, they do not support same-sex marriage.

To keep things consistent in each decision (and across groups), write your decision rule so that people classified as supporting same-sex marriage will end up in the right leaf. Subsequently, people classified as not supporting same-sex marriage, and will end up in the left leaf. For example, the decision rule of {age < 25?} follows these criteria, but the rule {age > 25?} reverses this.

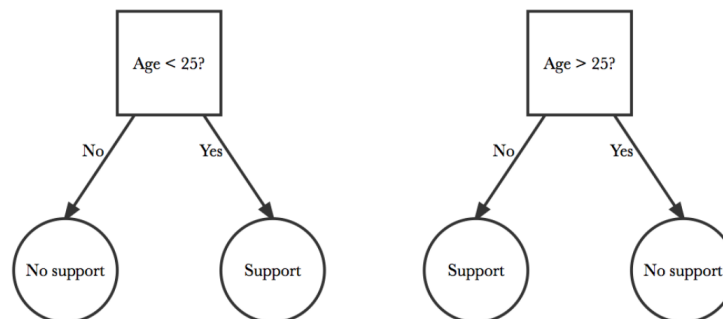


Figure 1: Left tree shows a decision rule that CORRECTLY classifies supporters into the right leaf and non-supporters into the left leaf. Right tree shows a decision rule that INCORRECTLY classifies supporters into the left leaf and non-supporters into the right leaf.

1. Create a classification tree using the training data (provided on the next page).

## Training Data

Age	Support	Age	Support
19	Support	55	Support
22	Support	58	Support
27	Support	58	Does not support
30	Support	58	Does not support
33	Does not support	61	Support
36	Support	63	Support
39	Does not support	72	Does not support
39	Does not support	74	Does not support
41	Support	80	Does not support
47	Support	80	Does not support
52	Does not support	83	Does not support
55	Support	85	Support

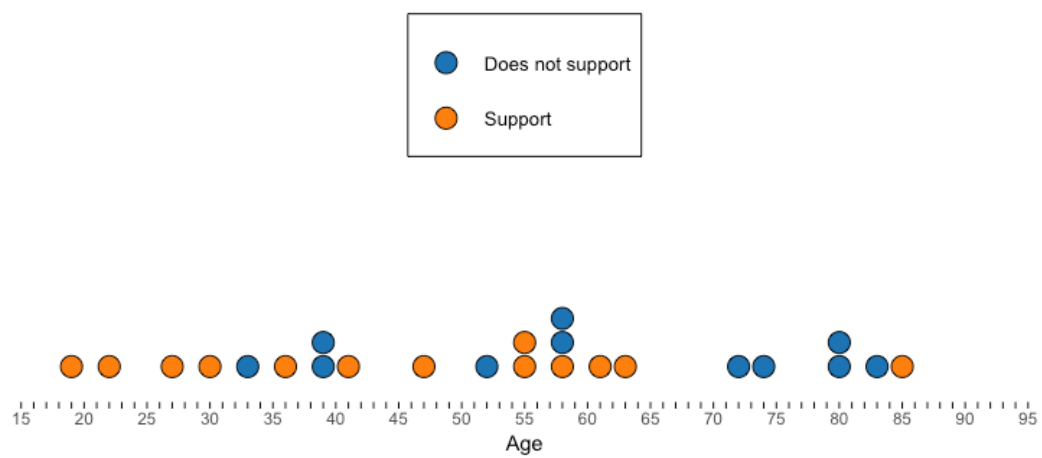


Figure 2: Dotplot of the training data.

2. Compute your classification model's overall mis-classification rate.