Classification trees

Instructions

In this (unassessed) mini-workshop you will draw the classification tree based on a given visual splitting rule and make predictions based on the result. We will discuss the solution after the workshop.

I suggest you go over the exercise by yourself and then compare your answers with your groupmates.

You have 20 minutes to complete the workshop.

**Learning objectives**

* Understand the how tree algorithm determines splits (i.e. rules)
* Translate splits to a tree diagram
* Classify new observations based on a tree

1. Example

Classification and regression trees represent rule-based decision using a tree structure. For example consider the following tree we saw earlier in the class.

Diagram

Description automatically generated

The first split is for whether number of shows watched in January is more or less than 4. The second split on the left is again on “no of shows”. This particular second split is only applied to the subset of the loans that fall to this branch; i.e. those users who have viewed videos on iPlayer for less than 4 shows. **See next page!**

2. Exercise 1

Consider the following data points.

Chart, scatter chart

Description automatically generated

Consider splitting at percent almost watched =0.5

1. Find the gini index of the split.
2. Draw the resulting tree.
   1. What is the root node?
   2. What are the next nodes?

**See next page!**

3. Exercise 2

Consider the splits depicted below.

Chart, scatter chart

Description automatically generated

First split is at percent\_almost\_watched=0.25 and the second split is at log(total\_time)=2.5 (for those percent\_almost\_watched>=0.25).

1. Draw the resulting tree.
2. What is the probability that the following user will watch in February according to this tree?

