A ^(very) brief Introduction to Decision & Regression Trees

Introduction to Data Science

Slides by: Mark Carman

What are Decision & Regression Trees?

Simple & effective Machine Learning algorithms that
- Recursively partition (divide up) the feature space into regions
- While grouping similar instances together
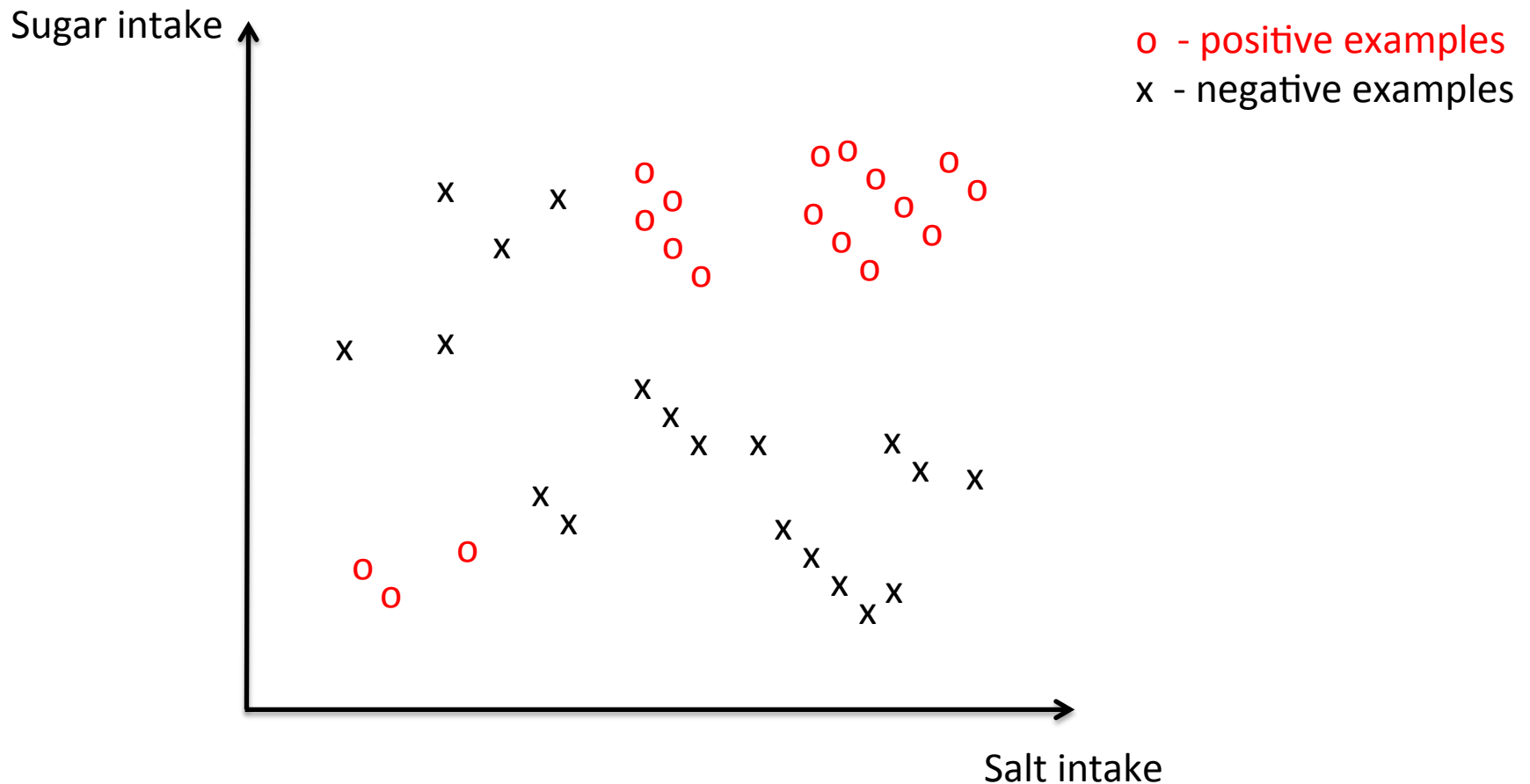
Decision Trees:
- Predict binary (or categorical) outcomes
- Prediction is the most common value in each region

Regression Trees:
- Predict continuous (i.e. real) values
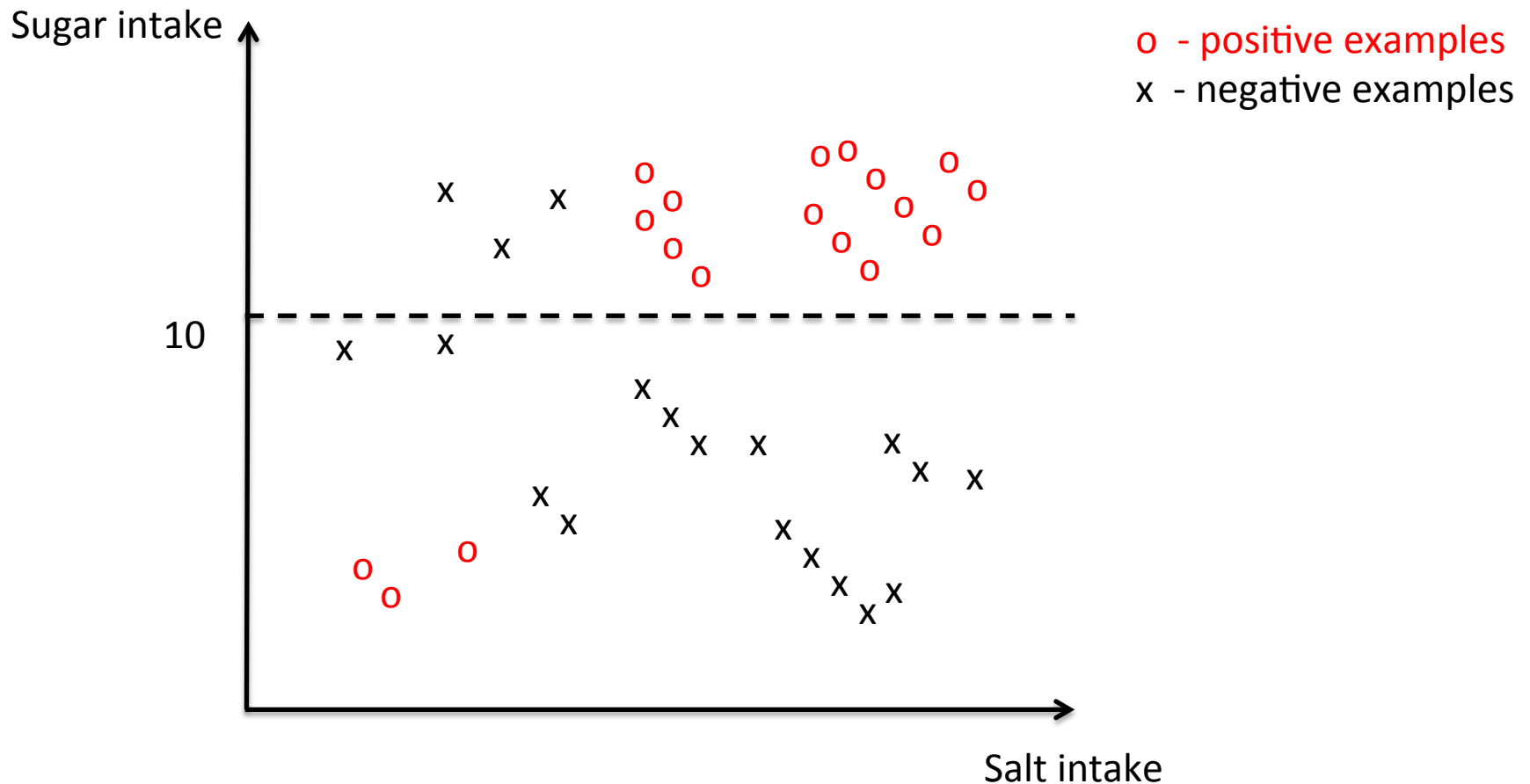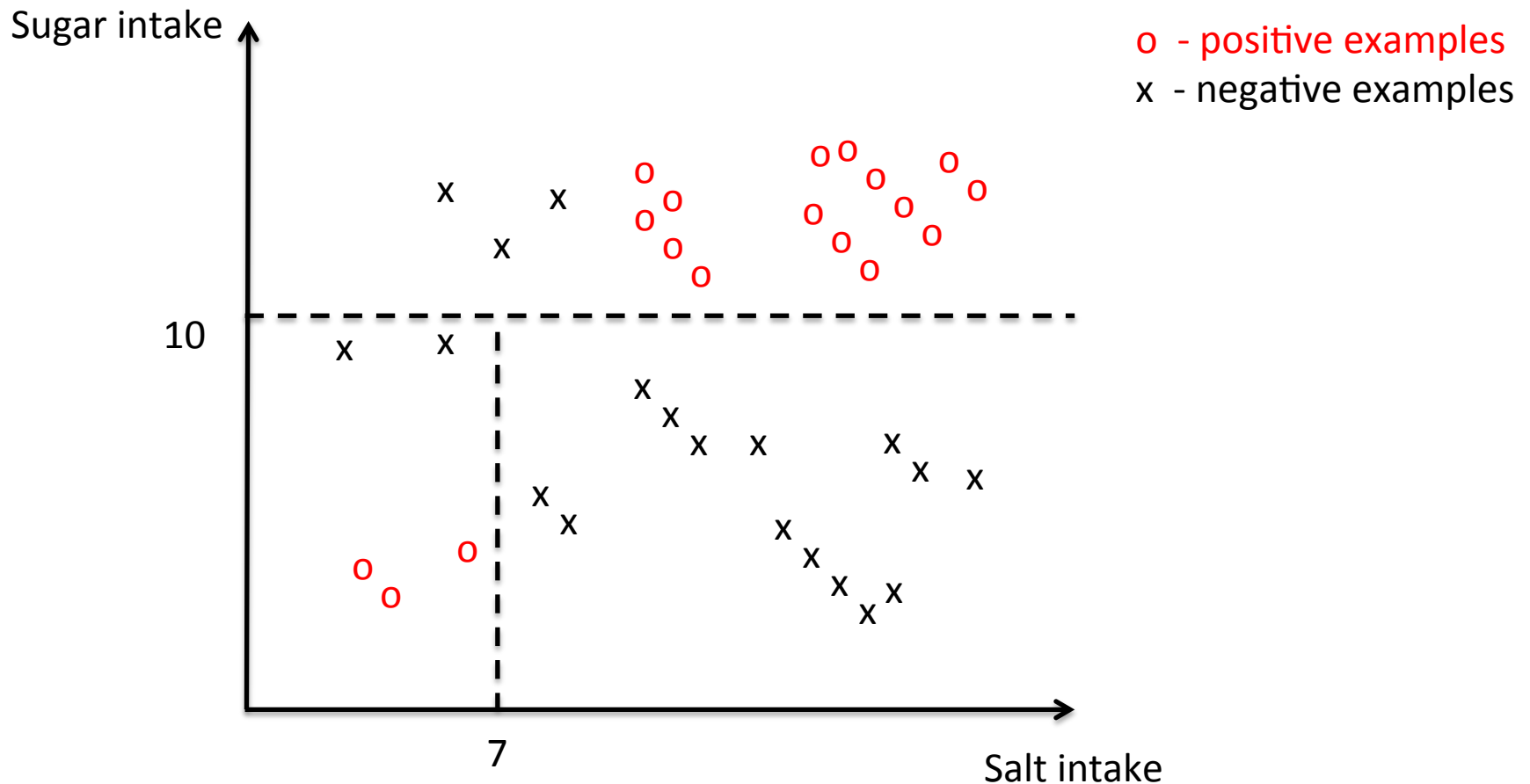- Prediction is usually the average value in each region

Recursive Partitioning

- At each iteration, we divide the data to group similar instances together

Recursive Partitioning

- At each iteration, we divide the data to group similar instances together
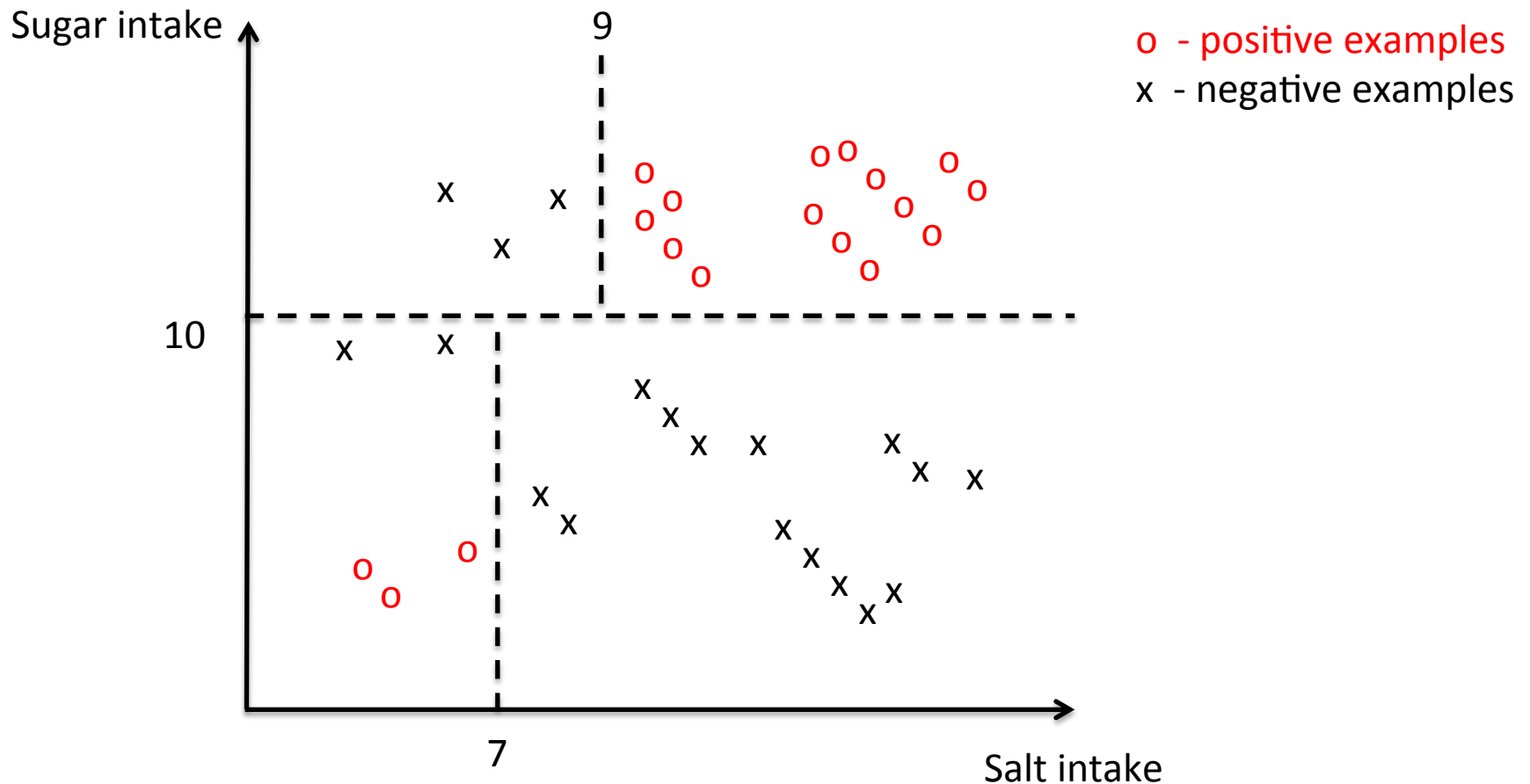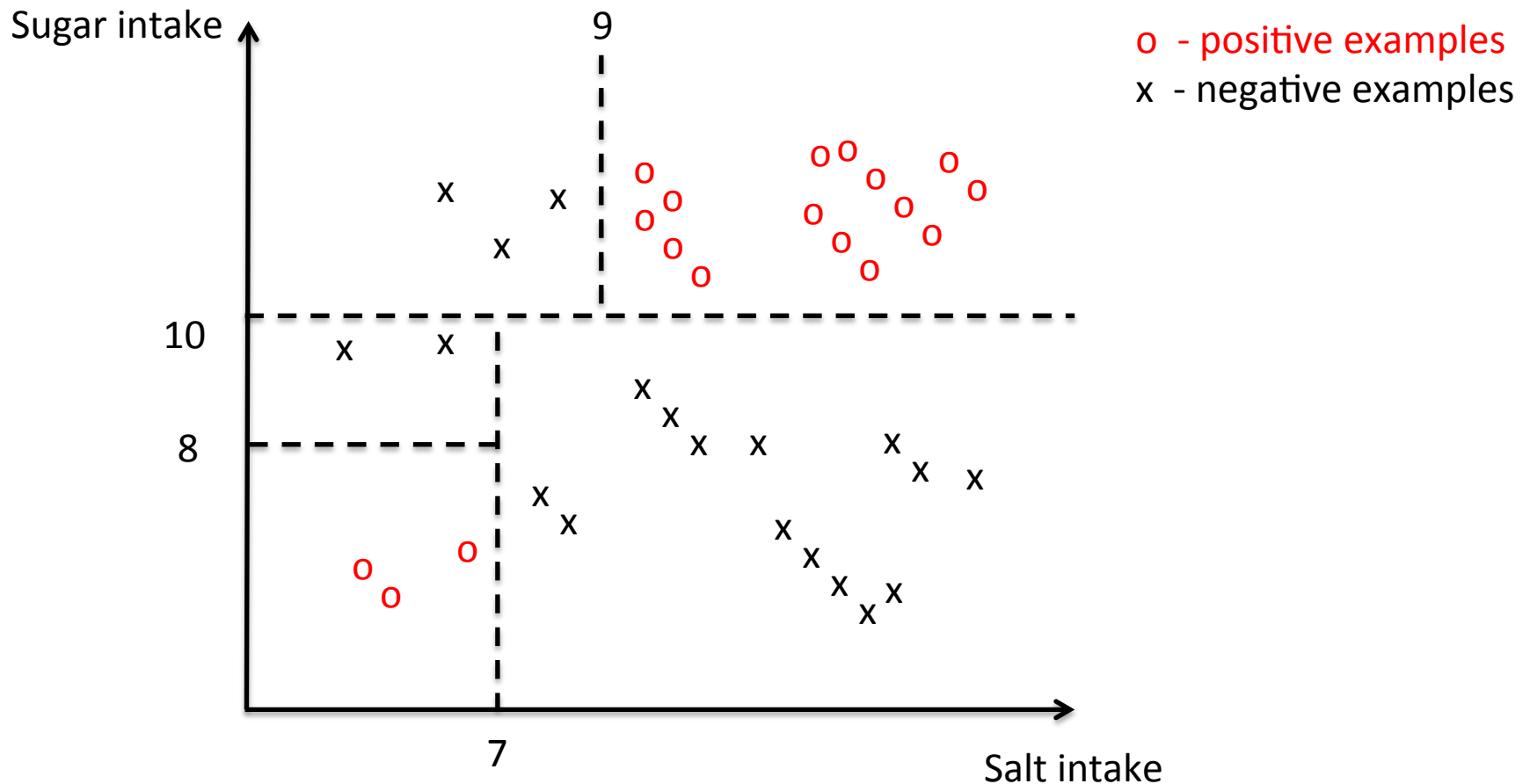
Recursive Partitioning

- At each iteration, we divide the data to group similar instances together

Recursive Partitioning

- At each iteration, we divide the data to group similar instances together
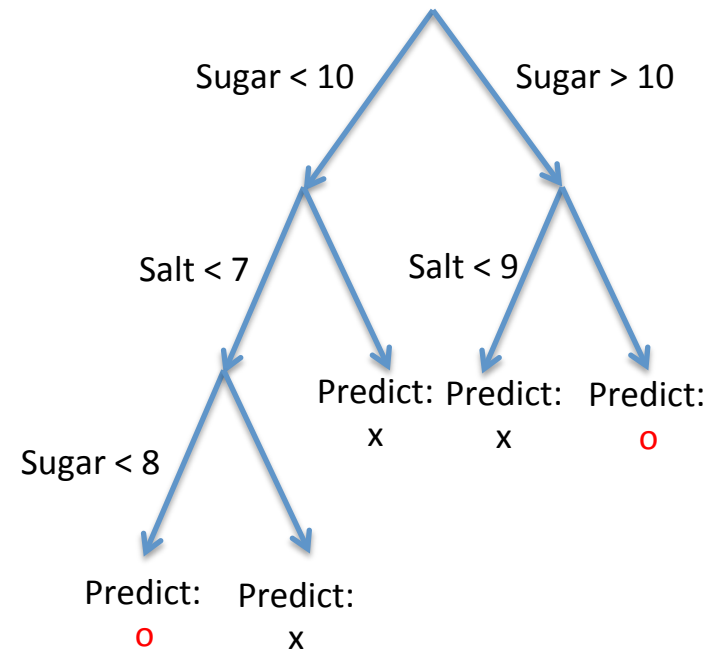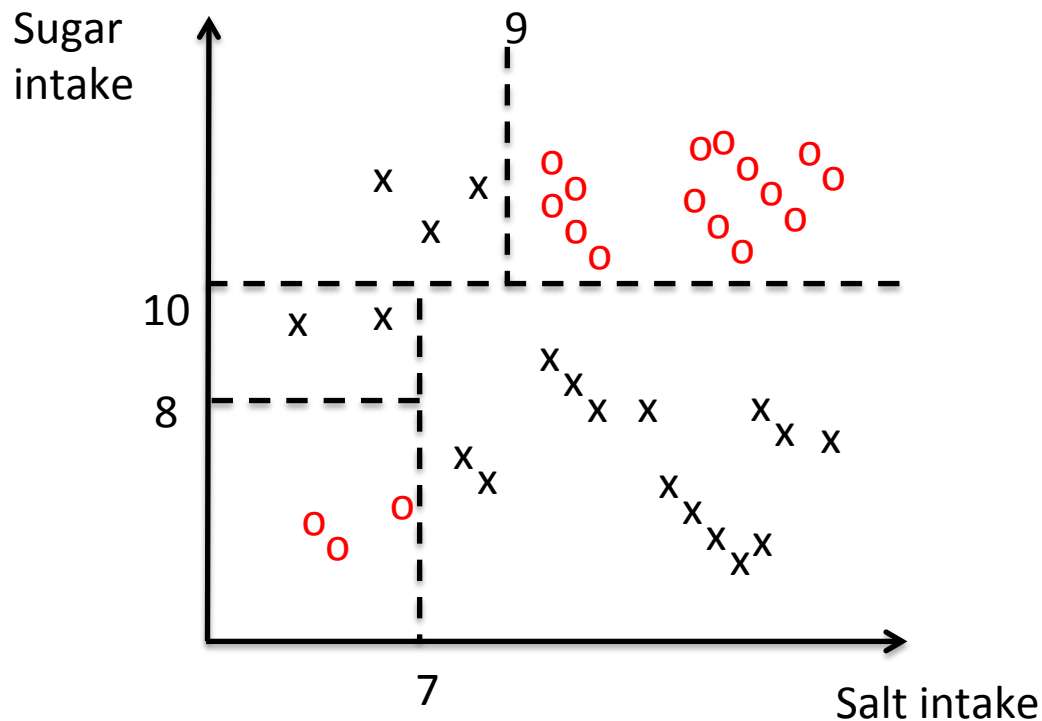
Recursive Partitioning

- At each iteration, we divide the data to group similar instances together

Sugar intake

9

o - positive examples
x - negative examples

10

8

7

Salt intake

Prediction Model is a Tree

- The model learnt can be represented as a tree with predictions at the leaves:

Using Decision/Regression Trees

- We'll train Decision and Regression Trees in the tutorial using BigML

- Algorithms for building Decision & Regression trees differ on the criteria used to:
  - decide on which feature to split on in each iteration
  - decide when to stop splitting

- More information on Decision & Regression trees available at:
  - https://en.wikipedia.org/wiki/Decision_tree_learning