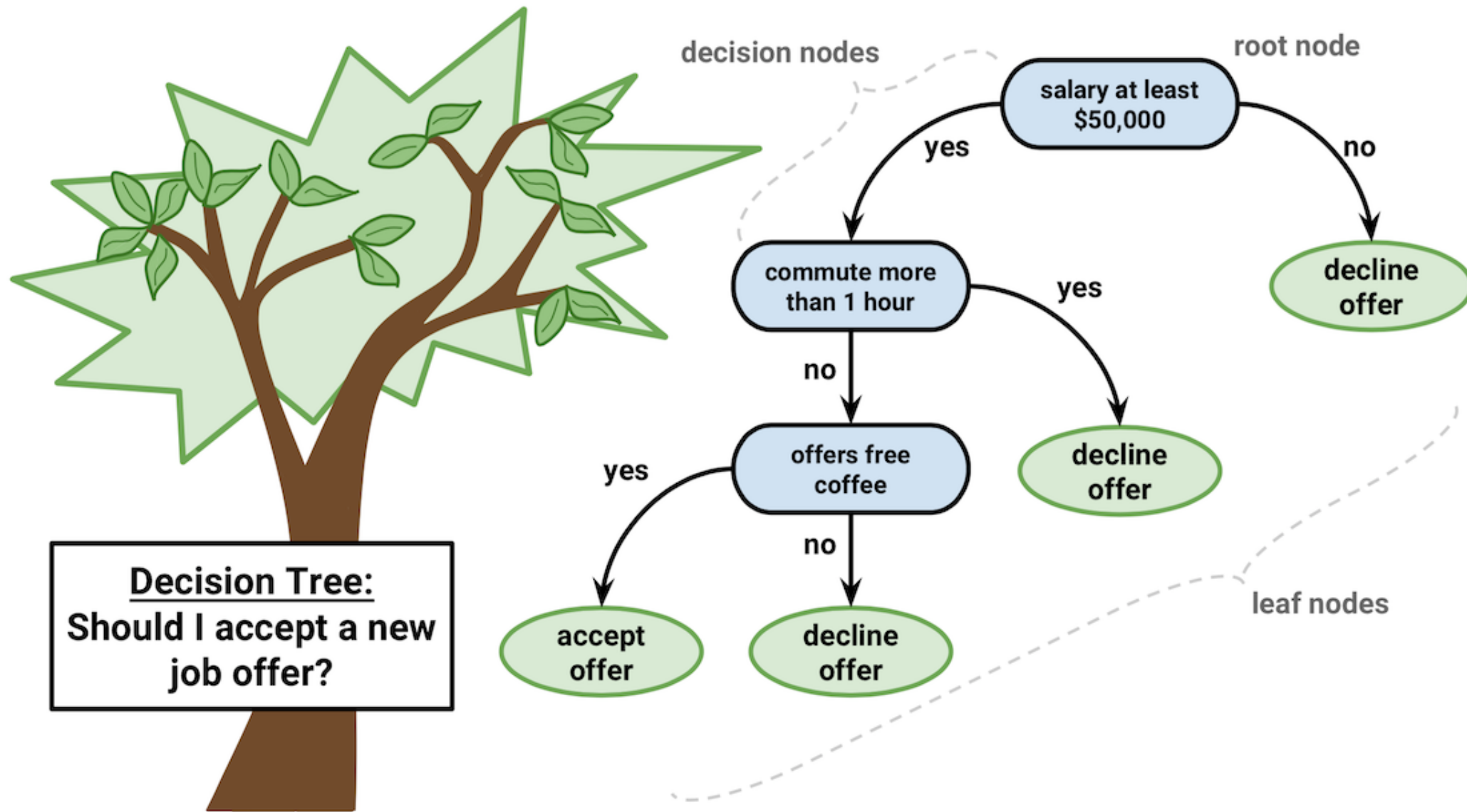# Making decisions with trees

SUPERVISED LEARNING IN R: CLASSIFICATION

Brett Lantz
Instructor

A decision tree model

Decision trees for prediction
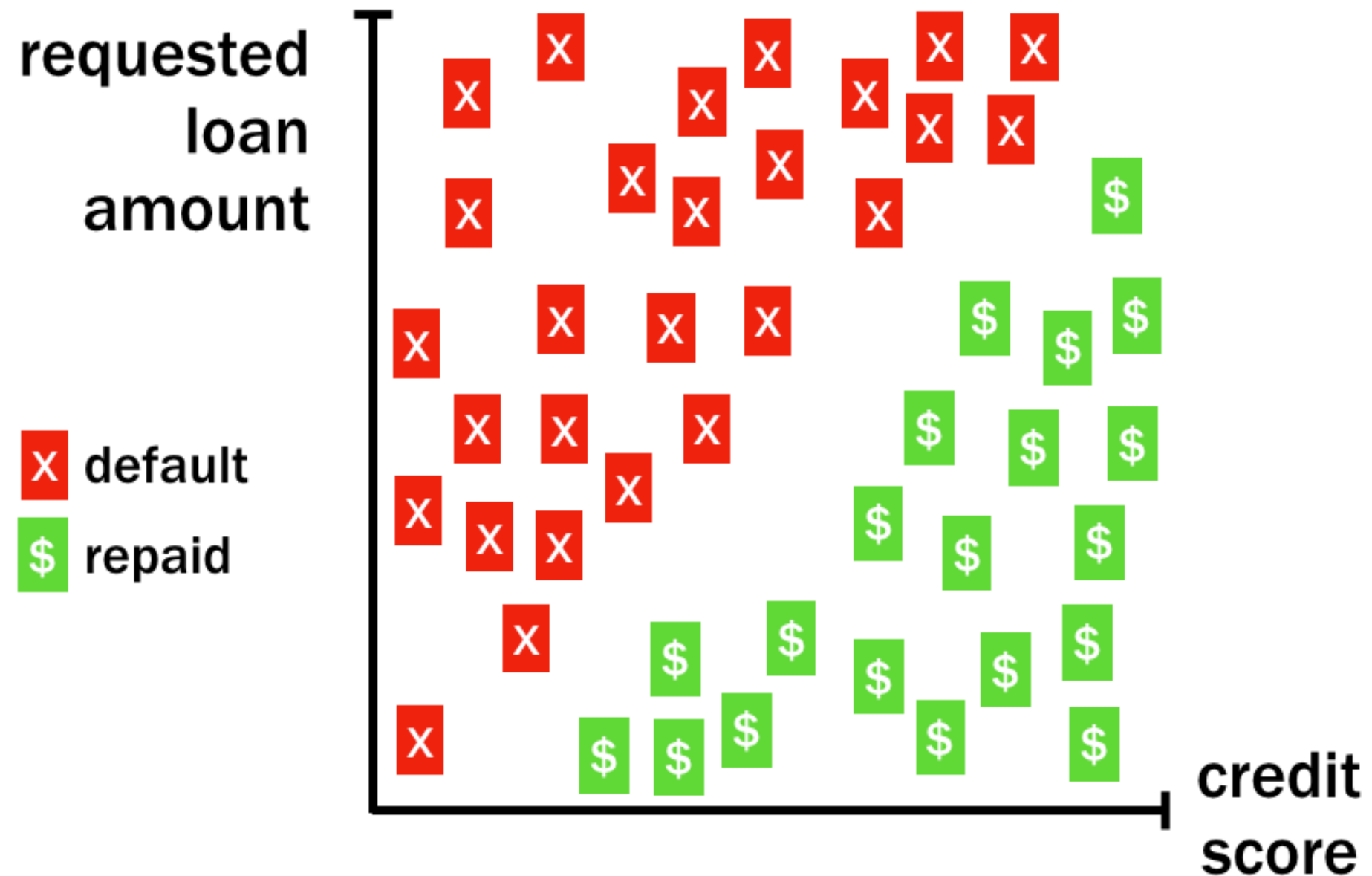
Divide-and-conquer

Divide-and-conquer
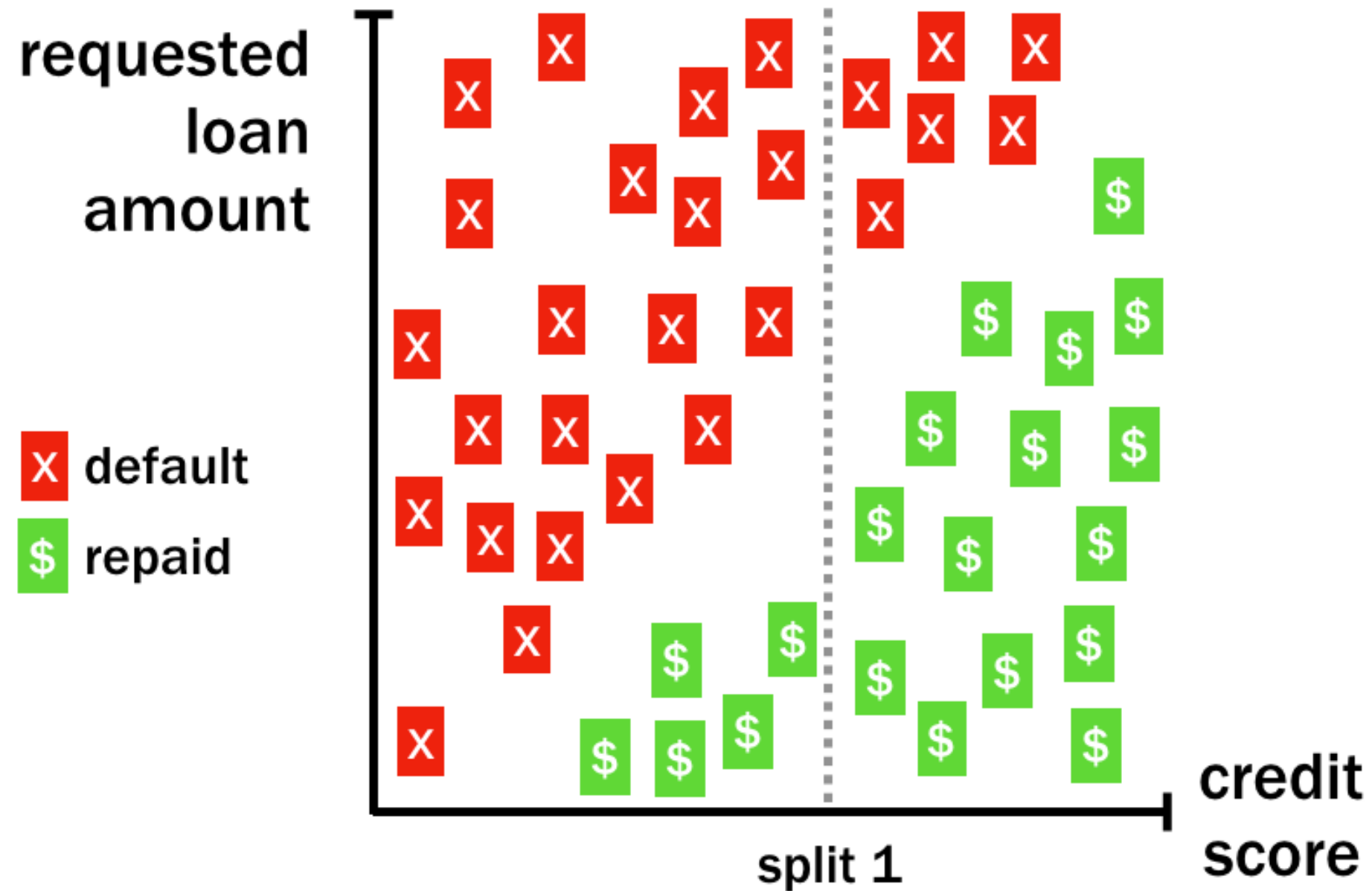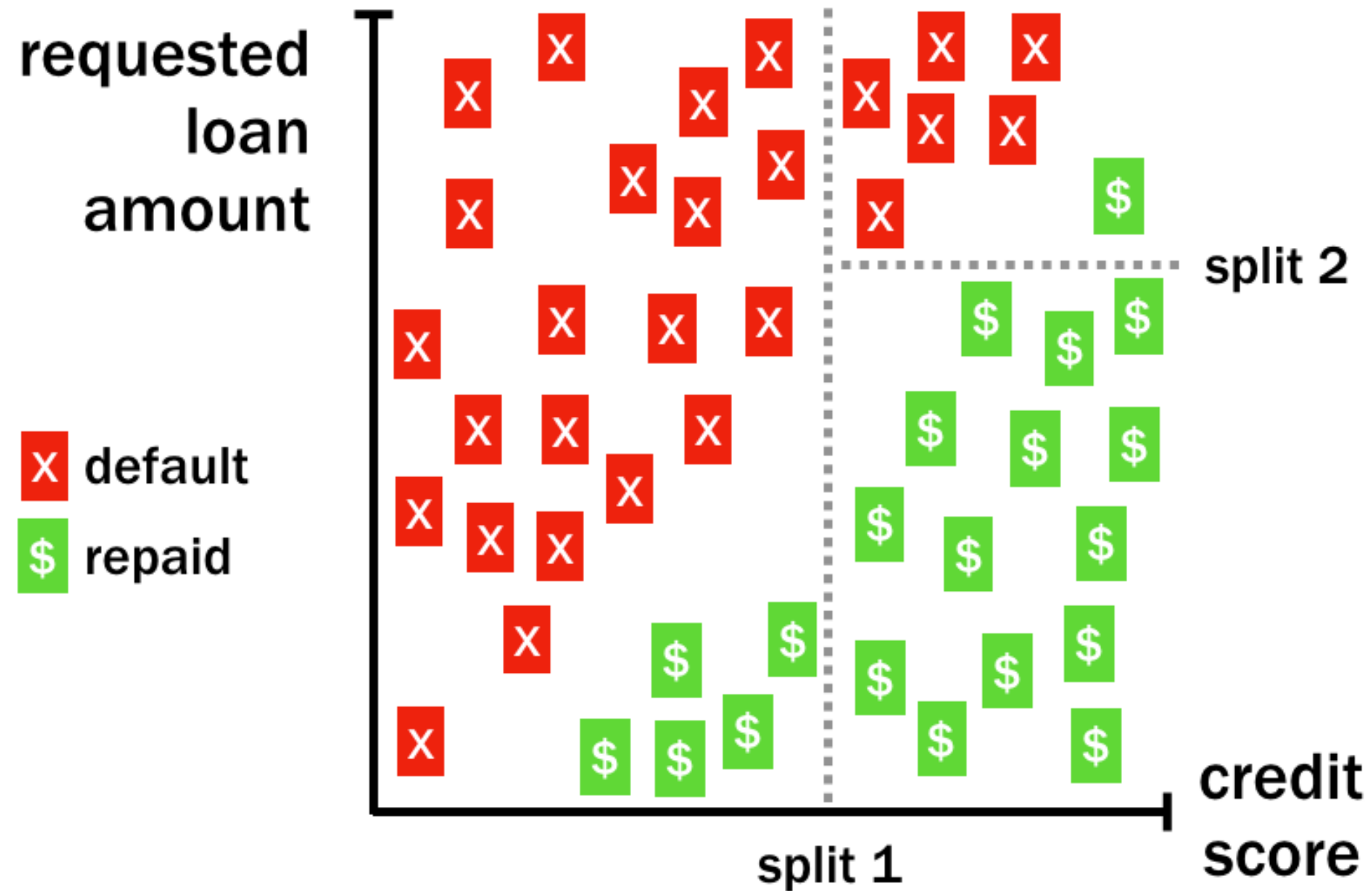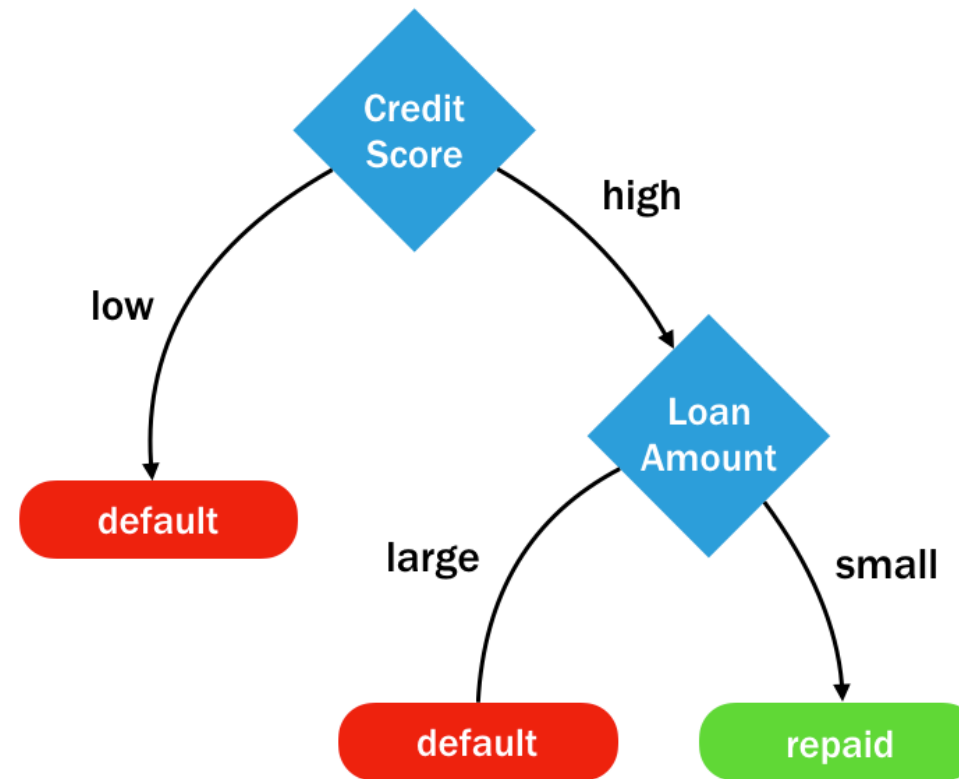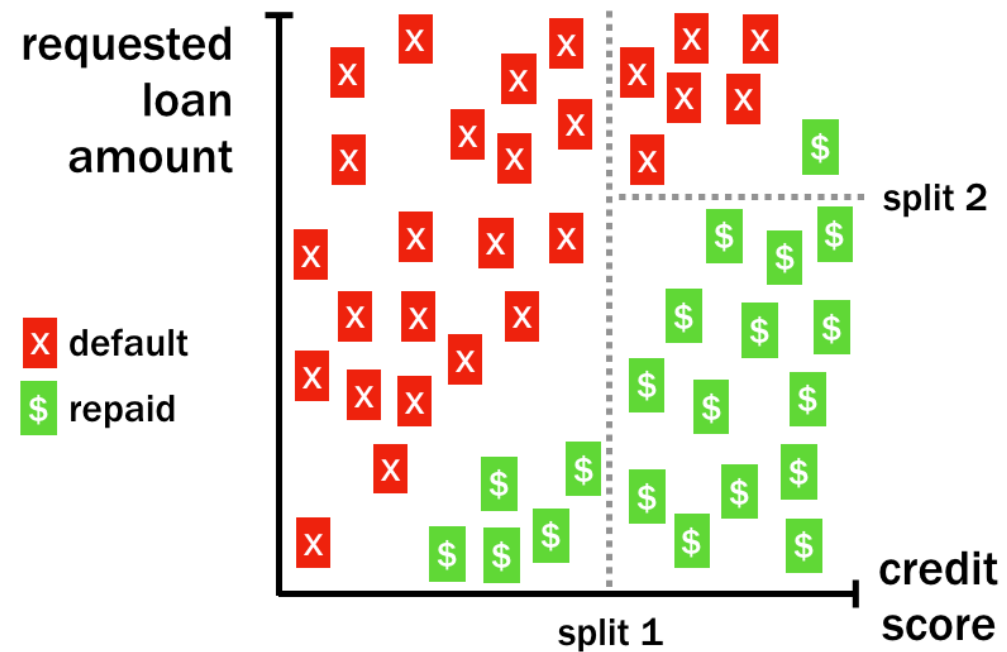
Divide-and-conquer

The resulting tree

Building trees in R

```r
# building a simple rpart classification tree
library(rpart)
m <- rpart(outcome ~ loan_amount + credit_score, data = loans,
              method = "class")
```

```r
# making predictions from an rpart tree
p <- predict(m, test_data, type = "class")
```
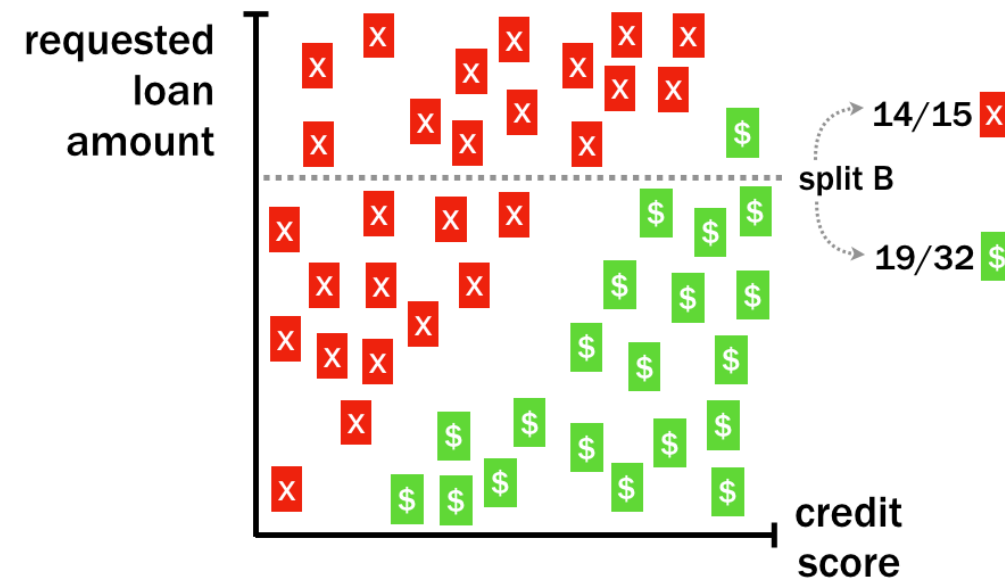
Let's practice!

DataCamp

Growing larger classification trees

SUPERVISED LEARNING IN R: CLASSIFICATION

Brett Lantz
Instructor

Choosing where to split

requested loan amount

credit score

split A

21/26 X   15/21 $

requested loan amount

credit score

14/15 X

split B

19/32 $

Axis-parallel splits

The problem of overfitting

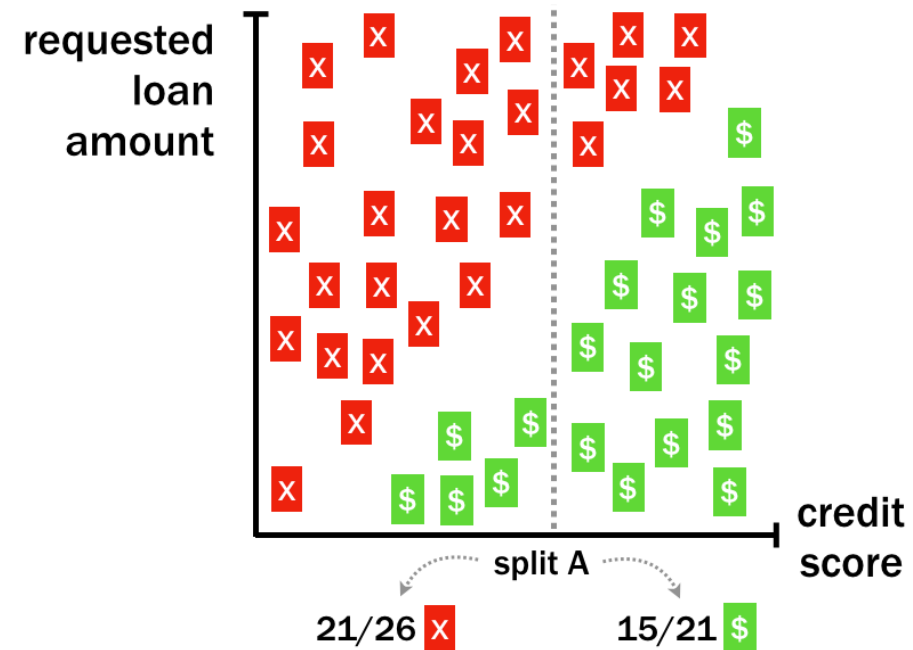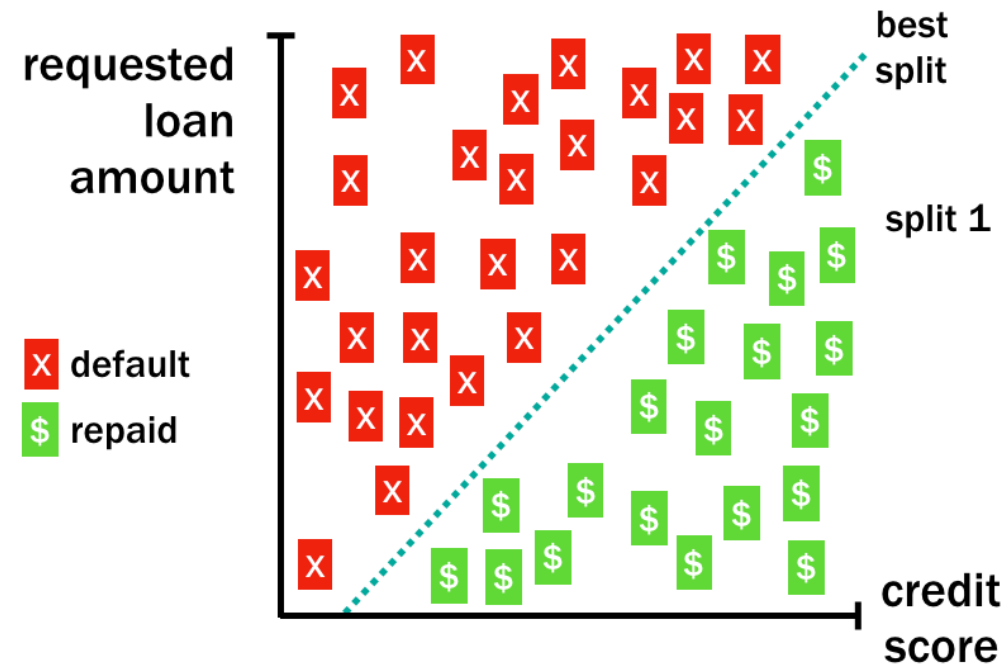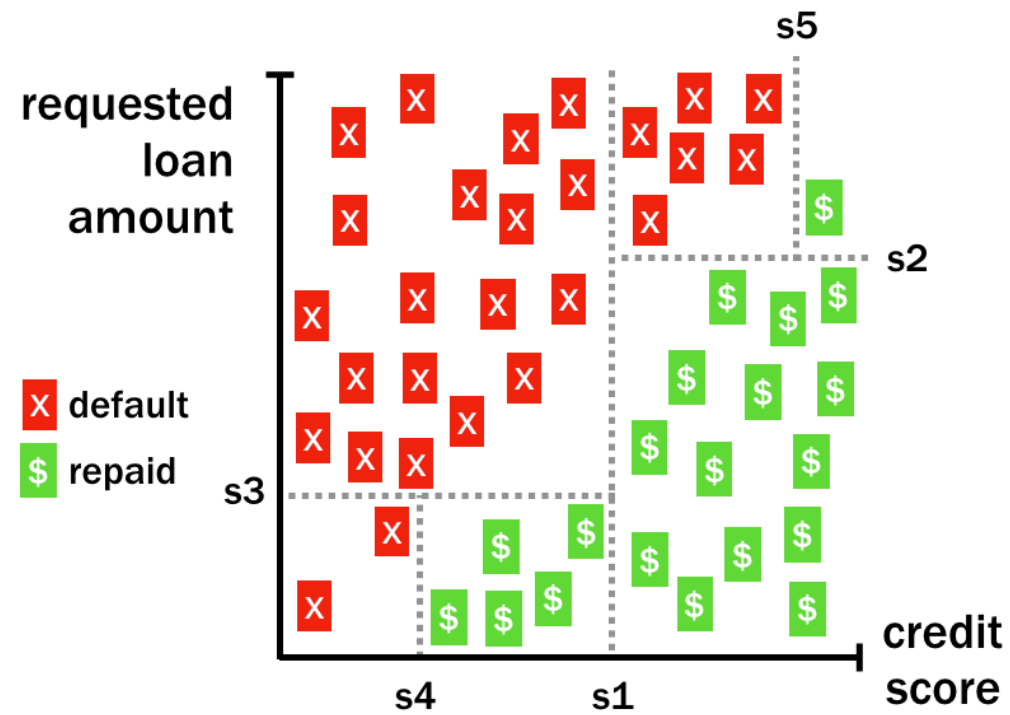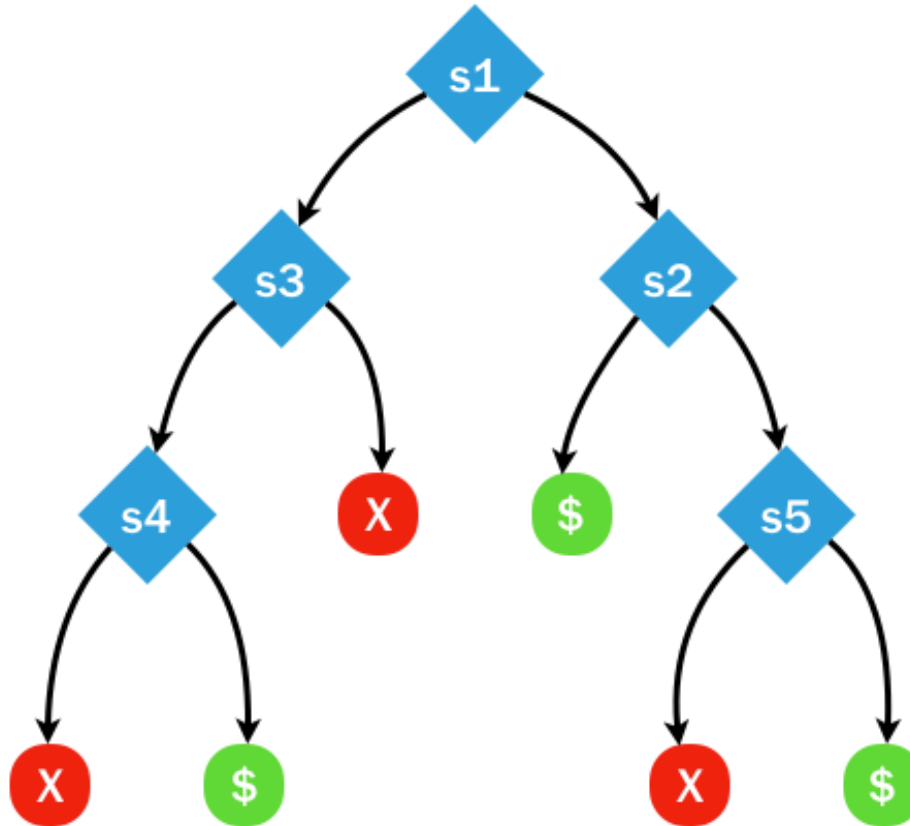# Evaluating model performance

Let's practice!

DataCamp

Tending to classification trees

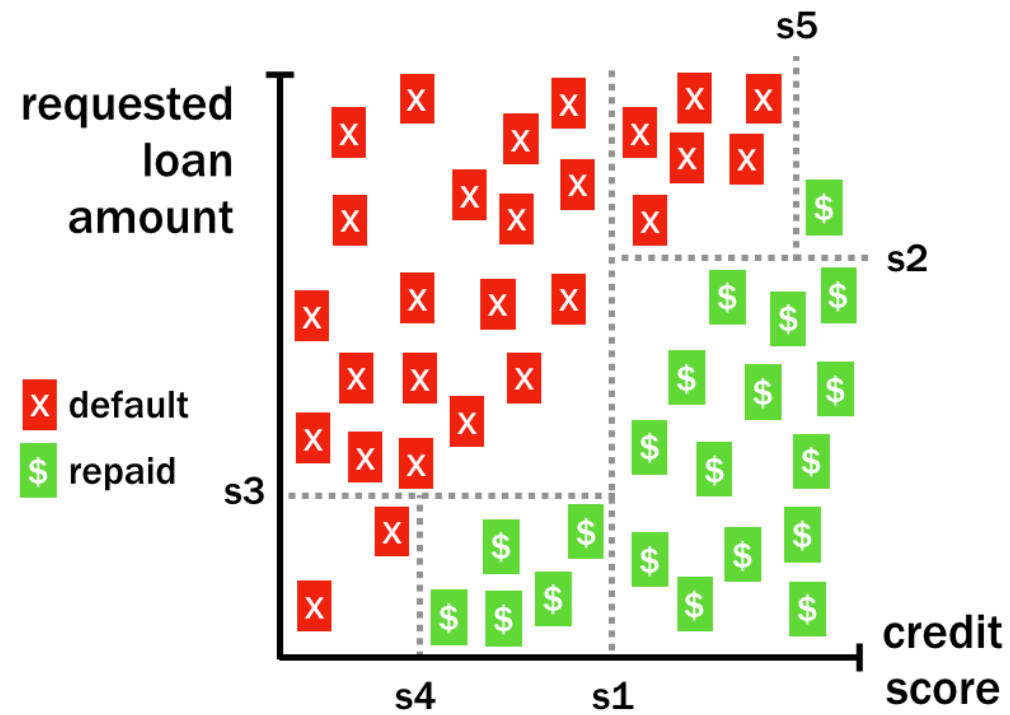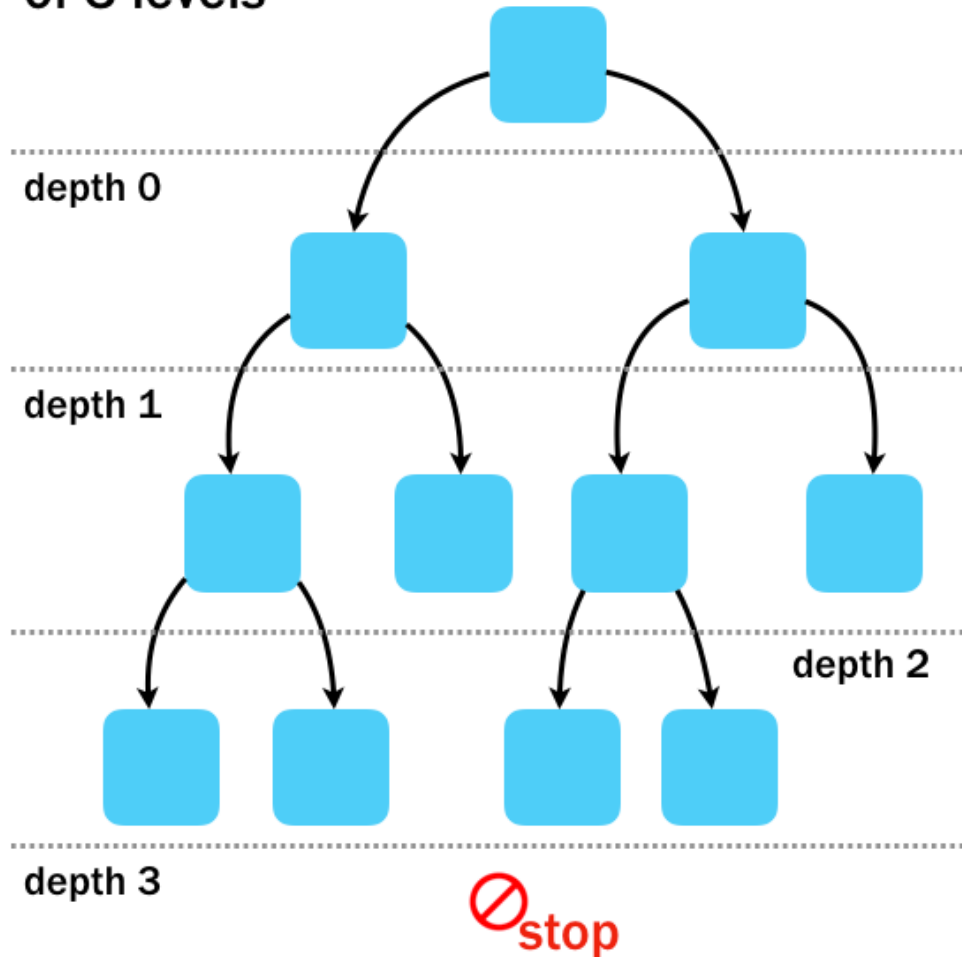### SUPERVISED LEARNING IN R: CLASSIFICATION

Brett Lantz
Instructor

Pre-pruning

Post-pruning

Post-pruning overly complex branch

prune

error rate

prune here

complexity
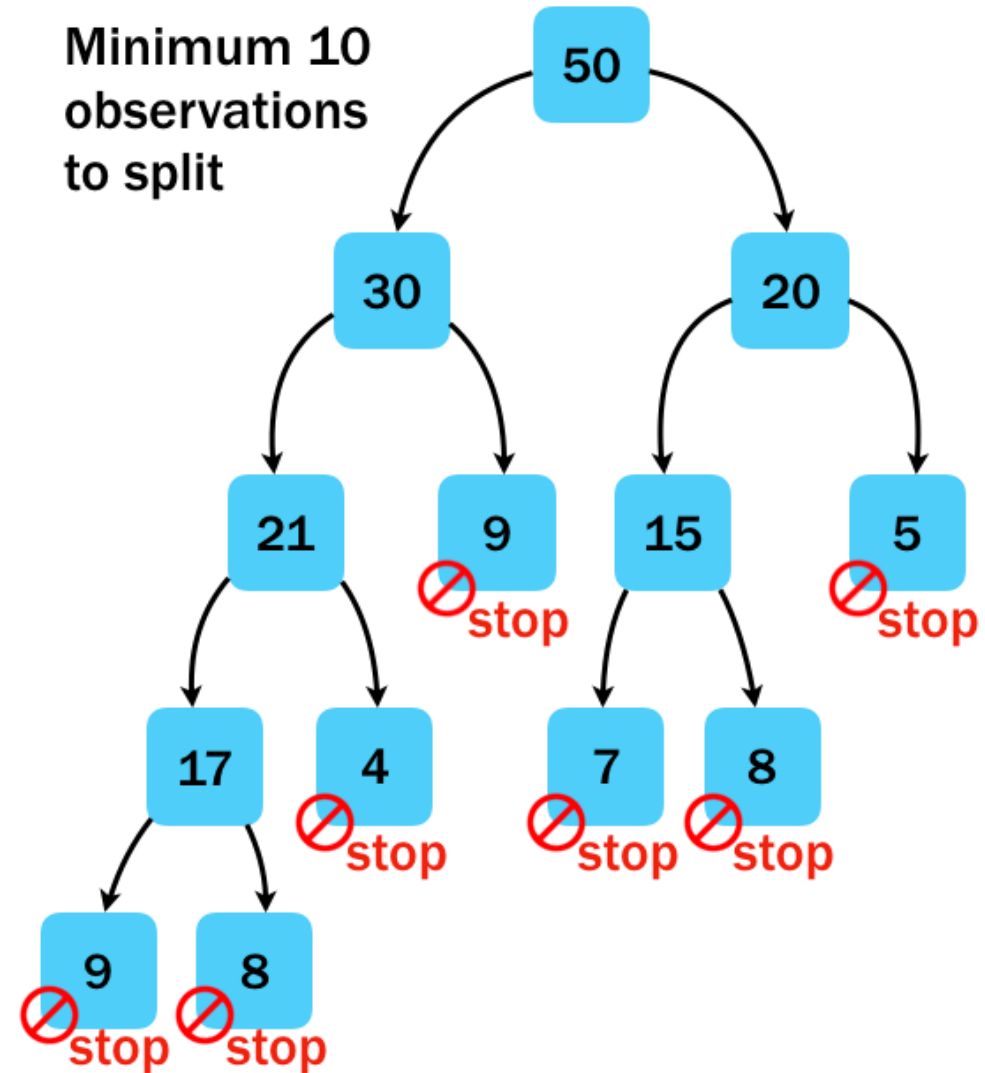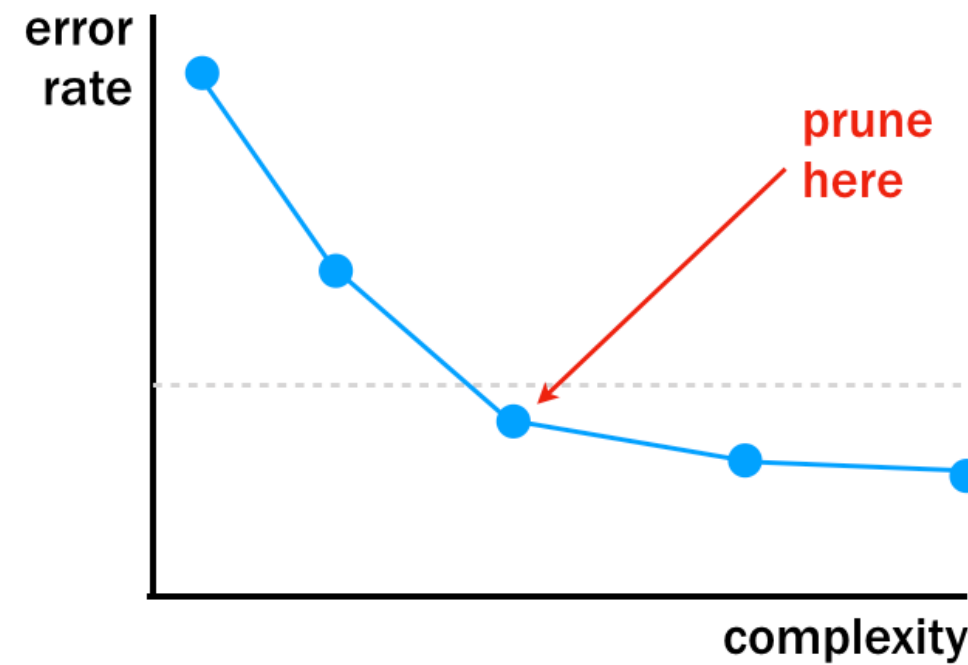
Pre- and post-pruning with R

```r
# pre-pruning with rpart
library(rpart)
prune_control <- rpart.control(maxdepth = 30, minsplit = 20)

m <- rpart(repaid ~ credit_score + request_amt,
           data = loans,
           method = "class",
           control = prune_control)
```

```r
# post-pruning with rpart
m <- rpart(repaid ~ credit_score + request_amt,
           data = loans,
           method = "class")

plotcp(m)

m_pruned <- prune(m, cp = 0.20)
```

Let's practice!

DataCamp

Seeing the forest from the trees

SUPERVISED LEARNING IN R: CLASSIFICATION
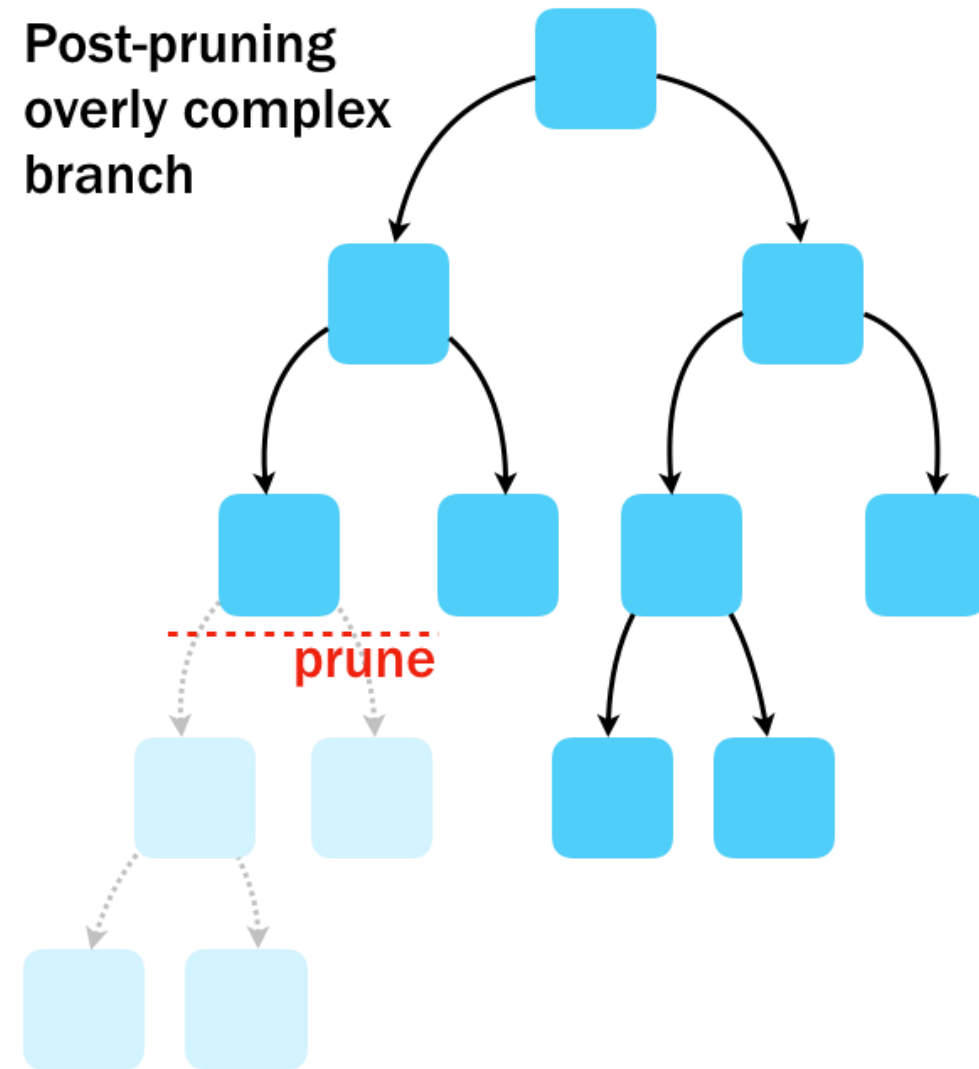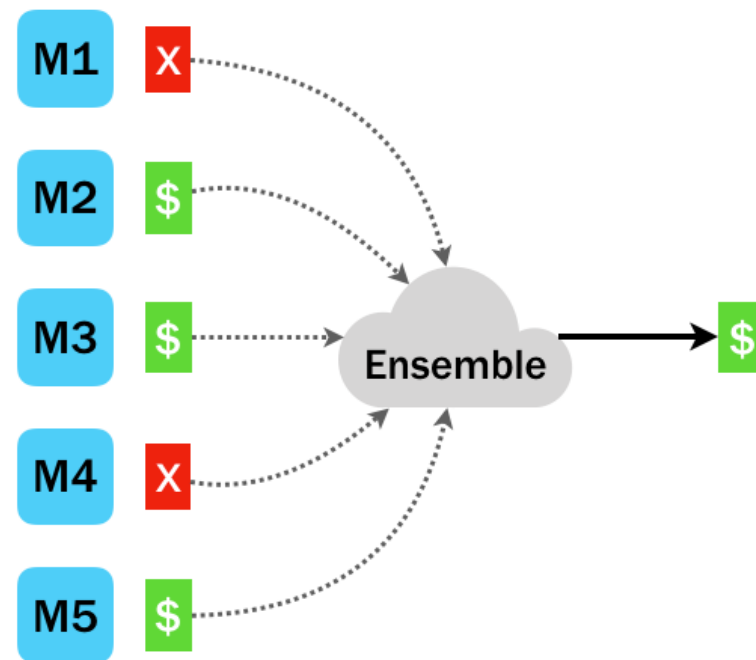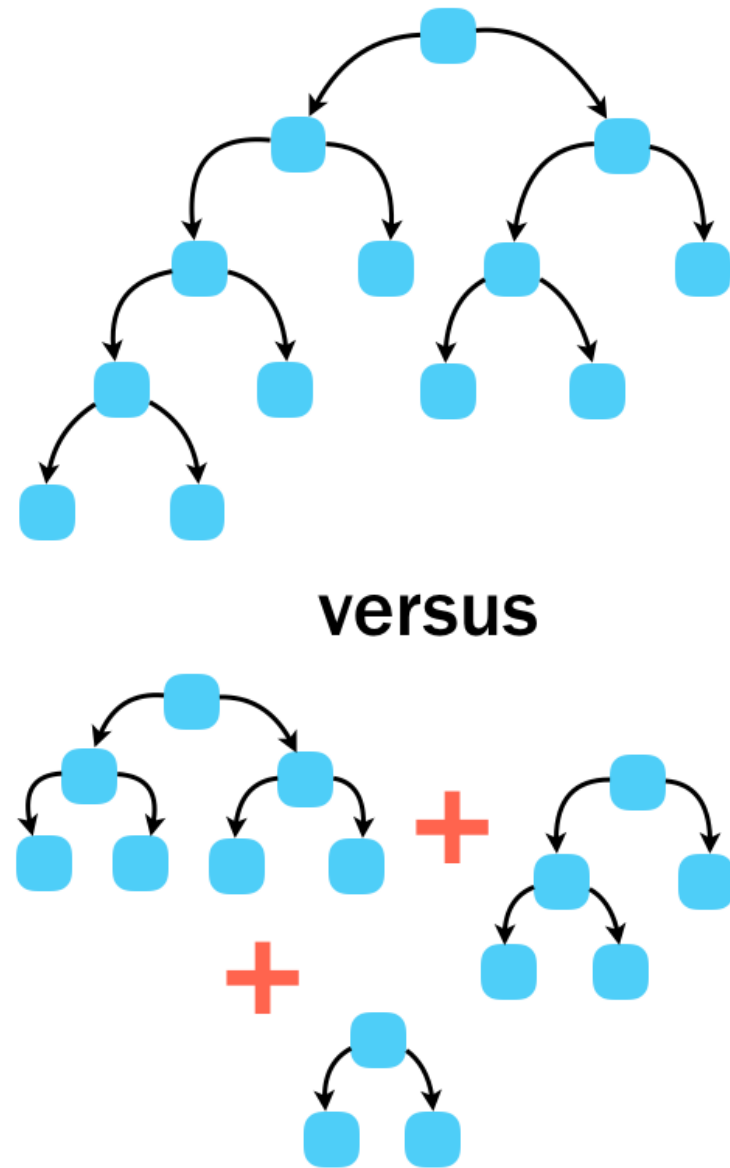
Brett Lantz

Instructor

Understanding random forests

Making decisions as an ensemble

versus

Random forests in R

```r
# building a simple random forest
library(randomForest)
m <- randomForest(repaid ~ credit_score + request_amt, data = loans,
                  ntree = 500,     # number of trees in the forest
                  mtry = sqrt(p)) # number of predictors (p) per tree
```

```r
# making predictions from a random forest
p <- predict(m, test_data)
```

Let's practice!

SUPERVISED LEARNING IN R: CLASSIFICATION