

# Decisions based on trees

Enrique J. De La Hoz D

Data Science - UTB

# A decision tree model

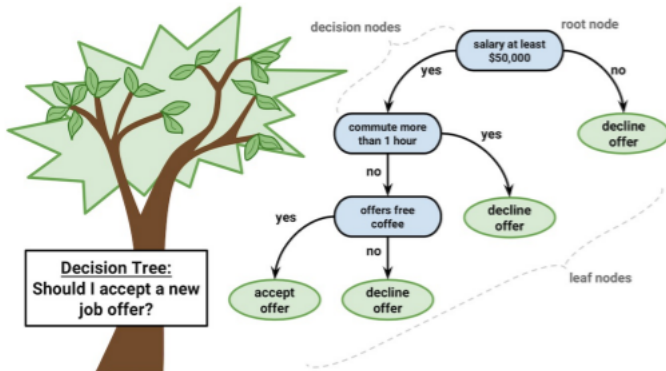


Figure 1:



# Divide and Conquer

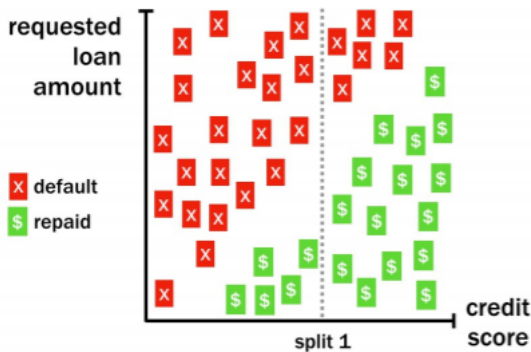


Figure 3:

# Divide and Conquer

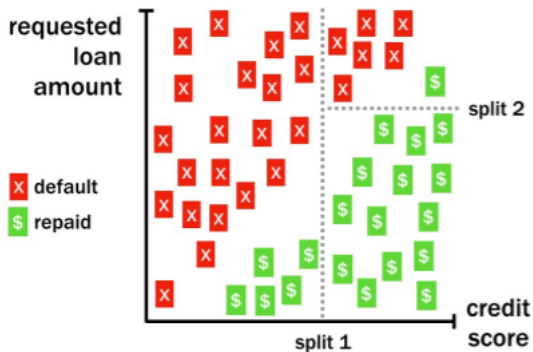


Figure 4:

# The resulting tree

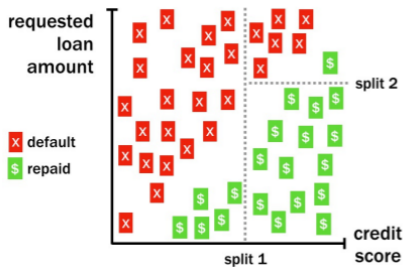


Figure 5:

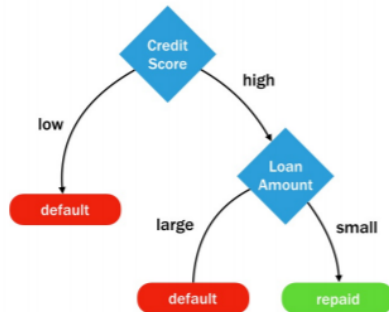


Figure 6:

# Building trees in R

```
# building a simple rpart classification tree
library(rpart)
m <- rpart(outcome ~ loan_amount + credit_score,
            data = loans,
            method = "class")
# making predictions from an rpart tree
p <- predict(m, test_data, type = "class")
```

# Larger Trees

- Choosing where to split

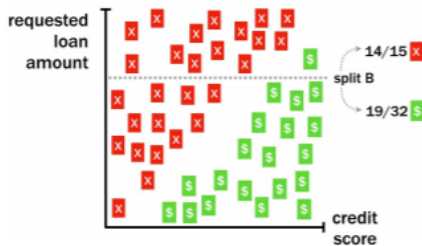
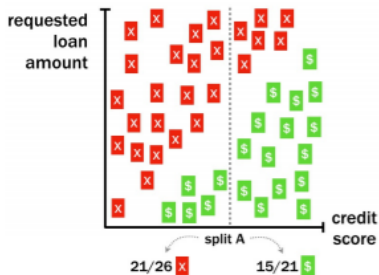


Figure 7:



# Axis-parallel splits

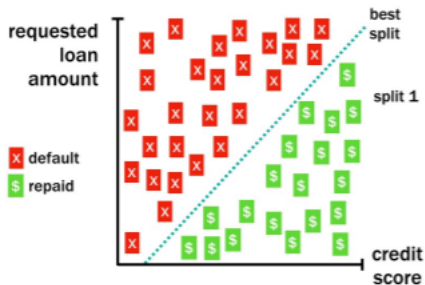
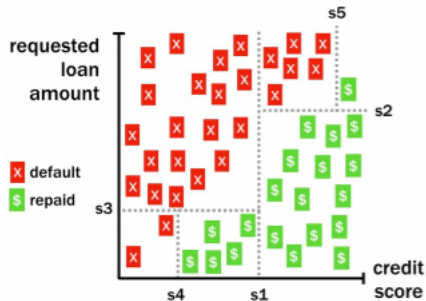


Figure 8:

# The problem of overfitting

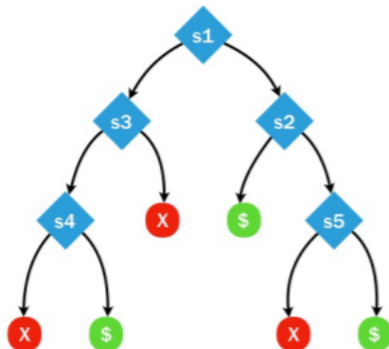
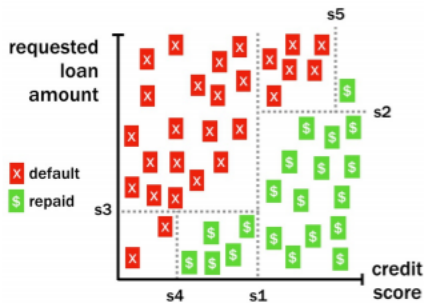


Figure 9:

# Tending to classification trees

- Pre-pruning

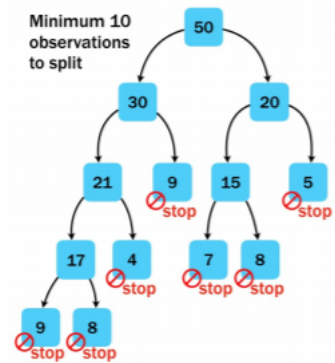
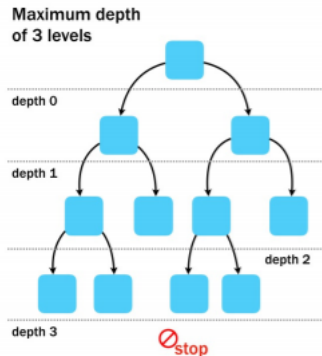


Figure 10:

# Tending to classification trees

- Post-pruning

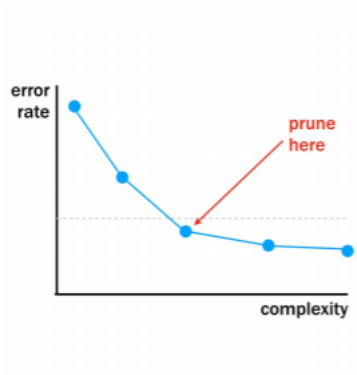
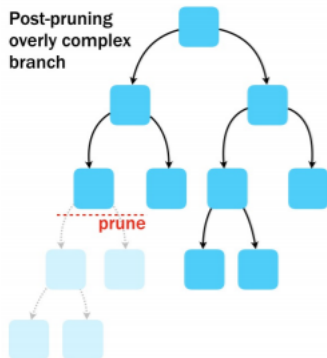


Figure 11:

## Pre- and post-pruning with R

```
# pre-pruning with rpart
library(rpart)
loans<- read.csv('loans.csv')
prune_control <- rpart.control(maxdepth = 30, minsplit = 20)
m <- rpart(repaid ~ credit_score + request_amt,
data = loans,
method = "class"
,
control = prune_control)
```

```
# post-pruning with rpart
m <- rpart(repaid ~ credit_score + request_amt,
data = loans,
method = "class")
plotcp(m)
m_pruned <- prune(m, cp = 0.20)
```