Assignment 4 - Decision Tree

Keith G. Williams - 800690755

Friday, June 12, 2015

## Problem Description

A company sent out some promotion to various houses and recorded a few facts about each house and also whether the people responded or not. Please create a Decision Tree (similar to one discussed in class) for the dataset below.

## district house\_type income previous\_customer outcome  
## 1 suburban detached high no nothing  
## 2 suburban detached high yes nothing  
## 3 rural detached high no responded  
## 4 urban semi-detached high no responded  
## 5 urban semi-detached low no responded  
## 6 urban semi-detached low yes nothing  
## 7 rural semi-detached low yes responded  
## 8 suburban terrace high no nothing  
## 9 suburban semi-detached low no responded  
## 10 urban terrace low no responded  
## 11 suburban terrace low yes responded  
## 12 rural terrace high yes responded  
## 13 rural detached low no responded  
## 14 urban terrace high yes nothing

## Information Content *I(C;F)* Calculations

Each node will be split based on the information content calculation:

where C is the class (in this cases outcome), and F is the feature matrix (in this case district, house\_type, income, previous\_customer)

### Root Node

The greatest information content is , so the first internal node will split outcome three ways on district.

### Second Node Layer

* district == suburban

## district house\_type income previous\_customer outcome  
## 1 suburban detached high no nothing  
## 2 suburban detached high yes nothing  
## 3 suburban terrace high no nothing  
## 4 suburban semi-detached low no responded  
## 5 suburban terrace low yes responded

The greatest information content is , so the suburban internal node will split outcome on income. In fact, this split gives perfect separation, so this split will result in two leaf nodes.

suburban

* district == rural

## district house\_type income previous\_customer outcome  
## 1 rural detached high no responded  
## 2 rural semi-detached low yes responded  
## 3 rural terrace high yes responded  
## 4 rural detached low no responded

Since , no information can be gained by splitting this node further, so this path will terminate in a leaf node.

rural

suburban

* district == urban

## district house\_type income previous\_customer outcome  
## 1 urban semi-detached high no responded  
## 2 urban semi-detached low no responded  
## 3 urban semi-detached low yes nothing  
## 4 urban terrace low no responded  
## 5 urban terrace high yes nothing

The greatest information content is , so the urban internal node will split outcome on previous customer. In fact, this split gives perfect separation, so this split will result in the final two leaf nodes.

## Final Decision Tree

urban

rural

suburban