

CSC411 Machine Learning and Data Mining
Tutorial 2 – Jan 26th, 2007

Decision Tree Example

Build an optimal decision tree based on the following data:
 (The sample data is referred to: <http://decisiontrees.net/node/25>)

District	House Type	Income	Previous Customer	Outcome
Suburban	Detached	High	No	Nothing
Suburban	Detached	High	Yes	Nothing
Rural	Detached	High	No	Responded
Urban	Semi-detached	High	No	Responded
Urban	Semi-detached	Low	No	Responded
Urban	Semi-detached	Low	Yes	Nothing
Rural	Semi-detached	Low	Yes	Responded
Suburban	Terrace	High	No	Nothing
Suburban	Semi-detached	Low	No	Responded
Urban	Terrace	Low	No	Responded
Suburban	Terrace	Low	Yes	Responded
Rural	Terrace	High	Yes	Responded
Rural	Detached	Low	No	Responded
Urban	Terrace	High	Yes	Nothing

Review:

Information Gain:

$$Gain(C, A) = I(C, A) = H(C) - \sum_{v \in V(A)} P(A = v)H(C|A = v)$$

Entropy:

$$Entropy(C) \equiv H(C) \equiv - \sum_c P(C = c) \log_2 P(C = c)$$

Tutorial Link: http://www.cs.toronto.edu/~ruiyan/index_files/page0001.htm