

Decision Tree Learning

Richard McAllister

February 4, 2008

1 Overview

2 Tree Construction

3 Case Study: Determinants of House Price

Decision Tree Learning

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Widely Used
- Used for approximating discrete-valued functions
- Robust to noisy data
- Capable of learning disjunctive expressions

Algorithms

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- ID3
- Assistant
- C4.5
 - Uses information entropy / information gain

Inductive Bias

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Preference for small trees over large
- **Occam's Razor:** Prefer the simplest hypothesis that fits the data.

Tree

Decision Tree
Learning

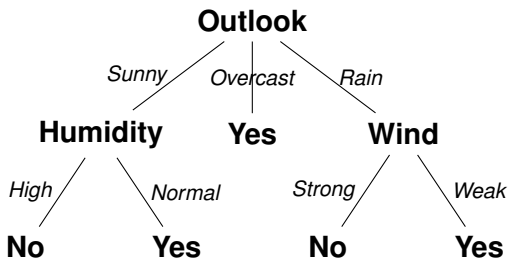
Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price



Leftmost Branch:

$\langle \text{Outlook} = \text{Sunny}, \text{Temperature} = \text{Hot}, \text{Humidity} = \text{High}, \text{Wind} = \text{Strong} \rangle$

Disjunction of Conjunctions

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

$(\textit{Outlook} = \textit{Sunny} \wedge \textit{Humidity} = \textit{Normal}) \vee (\textit{Outlook} = \textit{Overcast}) \vee (\textit{Outlook} = \textit{Rain} \wedge \textit{Wind} = \textit{Weak})$

Appropriate Problems

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Instances represented by attribute-value pairs
- Target function has discrete output values (e.g. yes or no)
- Disjunctive descriptions may be required
- Training data may contain errors
- Training data may contain missing attribute values

For Example...

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Classifying medical patients by disease
- Loan applications
- Equipment malfunctions
- Classification problems

- Learns by constructing decision trees top-down
 - "What should be tested first?"
 - Greedy
- Descendants are then chosen from possible values resulting from a decision option

How do you choose what's next?

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

Information Gain

- Measures how well a given attribute separates the training examples according to their target classifications.

■ Entropy

- Characterizes the (im)purity of an arbitrary collection of examples
- Measures expected reduction in entropy caused by knowing the value of A
- $Entropy(s) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$
 - S : the collection of positive and negative examples of some target concept
 - p_{\oplus} : the proportion of positive examples in S
 - p_{\ominus} : the proportion of negative examples in S

- incidentally - $GeneralEntropy(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i$

Information Gain (Cont'd)

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

■ Information Gain

$$\blacksquare \text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

- S : the collection of positive and negative examples of some target concept
- A : a particular attribute
- S_v : the collection of positive and negative examples of the same target concept after A has partitioned S .
- $S_v = \{s \in S \mid A(s) = v\}$

Use of Information Gain

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

Information Gain is calculated for each attribute:

- $Gain(S, Outlook) = 0.246$
- $Gain(S, Humidity) = 0.151$
- $Gain(S, Wind) = 0.048$
- $Gain(S, Temperature) = 0.029$

Maximum is chosen at each level

Continues until...

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- 1 Every attribute has already been included along the tree's path.
- 2 Training examples associated with leaf nodes all have the same target attribute value (i.e. *Entropy* is zero).

Hypothesis Space Search in Decision Tree Learning

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Search proceeds by hill-climbing.
- Information gain measure guides hill-climbing.

Capabilities and Limitations

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- ID3's hypothesis space is a complete space of finite, discrete-valued functions.
- Maintains only one current hypothesis as it searches through the tree.
- No backtracking - can reach a local optimum that is not global.
- Uses all training examples at each step in its statistics-based decisions.

- **Overfitting the Data**
- Incorporating Continuous-Valued Attributes
- Alternative Measures for Selecting Attributes
- Handling Training Examples with Missing Attribute Values
- Handling Attributes with Differing Costs

Overfitting the Data

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Overfitting: when some other hypothesis that fits the training examples less well actually performs better over the entire distribution of instances.
- Addition of incorrect training data will cause ID3 to create a more complex tree.
- Addition of noise can also cause overfitting.
- Prevention:
 - Reduced error pruning: making a node out of a subtree, using most common classification of the training examples to characterize it.
 - Rule post-pruning: convert the tree to rules, then prune rules created by the decision tree by removing preconditions that result in improving the rule's estimated accuracy.

- Overfitting the Data
- **Incorporating Continuous-Valued Attributes**
- Alternative Measures for Selecting Attributes
- Handling Training Examples with Missing Attribute Values
- Handling Attributes with Differing Costs

Incorporating Continuous-Valued Attributes

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Discretizing: dynamically defining new discrete-valued attributes that partition the continuous attribute value into a discrete set of intervals.

- Overfitting the Data
- Incorporating Continuous-Valued Attributes
- **Alternative Measures for Selecting Attributes**
- Handling Training Examples with Missing Attribute Values
- Handling Attributes with Differing Costs

Alternative Measures for Selecting Attributes

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- *SplitInformation*(S, A): how broadly and uniformly the attribute A splits the data.
- Distance-based measures

- Overfitting the Data
- Incorporating Continuous-Valued Attributes
- Alternative Measures for Selecting Attributes
- **Handling Training Examples with Missing Attribute Values**
- Handling Attributes with Differing Costs

Handling Training Examples with Missing Attribute Values

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Assign the missing attribute the most common among the training examples.
- Assign a probability to each of the possible values. Estimate the value of the missing attribute based on the observed frequencies of the values among the training examples.

- Overfitting the Data
- Incorporating Continuous-Valued Attributes
- Alternative Measures for Selecting Attributes
- Handling Training Examples with Missing Attribute Values
- Handling Attributes with Differing Costs

Handling Attributes with Differing Costs

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- Introduce a cost term into the attribute selection measure.
- Example:
 - $\frac{Gain^2(S,A)}{Cost(A)}$

Case Study: Determinants of House Price

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

The housing purchasing decision depends on household preference or priority orderings for dwelling attributes or services and that, therefore, the attributes or services probably have different importance in the determination of housing prices.

Decision Tree Application

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

- CART Algorithm
 - Makes use of an impurity measure
 - Impurity:
 - low value: indicates the predomination of a single class within a set of cases
 - high value: indicates the even distribution of classes

Gini Impurity Measure

Decision Tree
Learning

Richard
McAllister

Outline

Overview

Tree
Construction

Case Study:
Determinants
of House
Price

$$\blacksquare I_G(i) = 1 - \sum_{j=1}^m f(i,j)^2 = \sum_{j \neq k} f(i,j)f(i,k)$$

- $f(i,j)$: the probability of getting value j in node i
- Gini impurity is based on squared probabilities of membership for each target category in the node. It reaches its minimum (zero) when all cases in the node fall into a single target category.

DECISION TREE FOR HDB FLATS

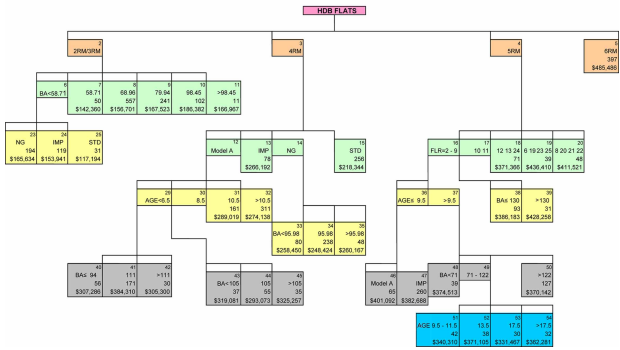


Figure 1. Decision tree results for the HDB flats

Decision Tree Learning

Richard McAllister

- Outline
- Overview
- Tree Construction

Case Study: Determinants of House Price