Concept Learning

Decision Trees

2 Decision Trees

• Hypotheses

- Using Trees
- Learning

3 Unpredictability

- Entropy
- Entropy for datasets
- Information Gain

Improvements

Örjan Ekeberg Brain Modeling and Machine Learning

Concepts and Hypotheses

Concepts and Hypotheses

- Definitions
- Example
- Hypotheses

- Using Trees
- Learning

- Entropy
- Entropy for datasets
- Information Gain

s and Hypotneses	Definitions
Decision Trees	
Unpredictability	Example
Improvements	Hypothese

Brain Modeling and Machine Learning

Örjan Ekeberg

Concept Learning

Concept Learning

Learning of a boolean function from examples

Conce

Categories

- "Nice weather"
- "Dog"
- "Motor vehicle"
- "Criminal offence"

Subsets of a superset X

Örjan Ekeberg Brain Modeling and Machine Learning

Terminology

c The concept to learn

Concepts and Hypotheses

Decision Trees

Improvements

$$c(x) \rightarrow \text{True/False}, x \in X$$

Definitions

h Hypothesis, Result of the learning ("guessed c")

$$h(x) \rightarrow \text{True/False}, x \in X$$

H Hypotheses space , All conceivable hypotheses (before data arrives)

 $h \in H$

D Set of available training data

Örjan Ekeberg

 $D \subseteq X$

Brain Modeling and Machine Learning

Terminology

Two kinds of training examples

Positive example:

 $x: c(x) = \text{True}, x \in D$

Negative example:

$$x: c(x) = False, x \in D$$

Örjan Ekeberg Brain Modeling and Machine Learning

Example Hypotheses

Concepts and Hypotheses Decision Trees Unpredictability Improvements

Example of a <i>concept</i>	
	" Nice Weather"

Let each "weather instance" x_i be composed of four attributes:

 $x_1 = \langle Sunny, Warm, Windy, Dry \rangle$ $x_2 = \langle Cloudy, Warm, Calm, Dry \rangle$ $x_3 = \cdots$

Generally: $Sky \times Temperature \times Wind \times Humidity$

Assume that the attributes can only take on certain discrete values:

Number of possible weathers: $|X| = 3 \cdot 2 \cdot 2 \cdot 2 = 24$

Concepts and Hypothese

Decision Trees

Concepts and Hypotheses Decision Trees Unpredictability Improvements

What does the hypotheses space H look like?

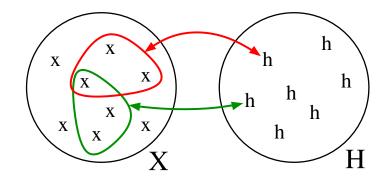


 $x_1 = \langle Sunny, Warm, Windy, Dry \rangle \rightarrow Nice$

Concepts and Hypotheses

Example Hypotheses

- $x_2 = \langle \mathsf{Sunny}, \mathsf{Warm}, \mathsf{Windy}, \mathsf{Humid} \rangle \rightarrow \mathsf{Nice}$
- $x_3 = \langle \mathsf{Rainy}, \mathsf{Cold}, \mathsf{Windy}, \mathsf{Humid} \rangle \longrightarrow \mathsf{Bad}$
- $x_4 = \langle Sunny, Warm, Calm, Humid \rangle \rightarrow Nice$



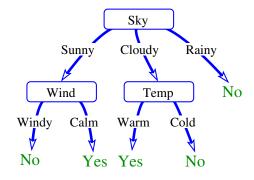
Each hypothesis h corresponds to one subset of X

Örjan Ekeberg Brain Modeling and Machine Learning	Örjan Ekeberg Brain Modeling and Machine Learning
Concepts and Hypotheses Decision Trees Unpredictability Improvements Definitions Example Hypotheses	Concepts and Hypotheses Decision Trees Unpredictability Improvements
How many hypotheses can we choose from? How many subsets does X have? $ H = 2^{ X }$ $ H = 2^{24} = 16777216$	 Concepts and Hypotheses Definitions Example Hypotheses Decision Trees Using Trees Learning Unpredictability Entropy Entropy for datasets Information Gain Improvements



Decision Tree

- Test the attributes sequentially
- Choose attributes to test depending on earlier attribute values



Brain Modeling and Machine Learning

The results (classifications) are coded by the *leaves*

Örjan Ekeberg

	(Sky		
	_ /	_		
	Sunny	Cloudy	Rainy	
	4	J J		
Wi	nd	Ter	np No	
	$\overline{}$		$\overline{\mathbf{T}}$	
Windy	Calm	Warm	Cold	
Ļ	l l	Ļ	7	
No	V	V	No	
INU	res	Yes	INO	

What does the tree encode?

 $(Sunny \land Calm) \lor (Cloudy \land Warm)$

Works as a disjunction of conjunctions

Normal Form for boolean functions Arbitrary boolean functions can be represented!

Örjan Ekeberg Brain Modeling and Machine Learning

Using Trees Learning Decision Trees Decision Trees Unpredictability Definitions How can a decision tree be constructed automatically? • Example • Hypotheses Choose an attribute to test Pranches with a unique class become leaves • Using Trees **③** Other branches are extended recursively • Learning Remaining question: how do we choose attributes? Output Contraction Contract • Entropy Greedy approach: • Entropy for datasets Choose the attribute which tells us most about the answer • Information Gain

Entropy

Entropy — measure of unpredictability

Entropy =
$$\sum_{i} -p_i \log_2 p_i$$

 p_i probability for outcome *i*

Entropy

Entropy

Example: tossing a coin $p_{
m head} = 0.5;$ $p_{\rm tail} = 0.5$

Entropy =
$$\sum_{i} -p_{i} \log_{2} p_{i} =$$

= -0.5 log₂ 0.5 + -0.5 log₂ 0.5 = -0.5 log₂ 0.5 + -0.5 log₂ 0.5 = 1

The result of a coin-toss has 1 bit of information



Example: rolling a dice $p_1 = \frac{1}{6}; \quad p_2 = \frac{1}{6}; \dots \quad p_6 = \frac{1}{6}$

Entropy =
$$\sum_{i} -p_i \log_2 p_i =$$

= $6 \times -\frac{1}{6} \log_2 \frac{1}{6} =$
= $-\log_2 \frac{1}{6} = \log_2 6 \approx 2.58$

Example: rolling a fake dice $p_1 = 0.1; \ldots p_5 = 0.1; p_6 = 0.5$

Entropy =
$$\sum_{i} -p_i \log_2 p_i =$$

= $-5 \cdot 0.1 \log_2 0.1 - 0.5 \log_2 0.5 =$
 ≈ 2.16

A real dice is more unpredictable (2.58 bit) than a fake (2.16 bit)

The result of a dice-roll has 2.58 bit of information

Decision Trees Unpredictability

Back to the decision trees

Smart idea:

Ask about the attribute which maximizes the expected reduction of the entropy.

Information Gain

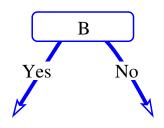
Assume that we ask about attribute A for a dataset S

$$\operatorname{Gain} = \operatorname{Ent}(S) \underbrace{\operatorname{Ent}(S)}_{\operatorname{before}} - \sum_{\nu \in \operatorname{Values}(A)} \frac{|S_{\nu}|}{|S|} \underbrace{\sum_{\nu \in \operatorname{Values}(A)} \frac{|S_{\nu}|}{|S|}}_{\underset{\operatorname{average}}{\operatorname{weighted}}} \operatorname{Ent}(S_{\nu}) \underbrace{\operatorname{Ent}(S_{\nu})}_{\operatorname{after}}$$

Örjan Ekeberg Brain Modeling and Machine Learning

Decision Trees Unpredictability Improvements

Gain(A) = 0.9988 - 0.9977 = **0.0011** Gain(B) = 0.9988 - 0.7210 = **0.2778** Gain(C) = 0.9988 - 0.9985 = **0.0003** Gain(D) = 0.9988 - 0.9884 = **0.0104** Attribute B gives most information



Entropy

Unpredictability of a dataset

• 100 examples, 42 positive

$$-\frac{58}{100}\log_2\frac{58}{100}-\frac{42}{100}\log_2\frac{42}{100}=0.981$$

• 100 examples, 3 positive

$$-\frac{97}{100}\log_2\frac{97}{100} - \frac{3}{100}\log_2\frac{3}{100} = 0.194$$

Örjan Ekeberg Brain Modeling and Machine Learning

Concepts and Hypotheses Decision Trees Unpredictability Improvements

Trees Entropy ability Information Gain

What is the entropy for this dataset? Ent = $-\frac{12}{25} \log_2 \frac{12}{25} - \frac{13}{25} \log_2 \frac{13}{25} \approx 0.9988$ $A = \bullet: \frac{6}{12} \text{ positive} \rightarrow 1.0$ $A = \circ: \frac{6}{13} \text{ positive} \rightarrow 0.9957$ Expected: $\frac{12}{25} \cdot 1.0 + \frac{13}{25} \cdot 0.9957 \approx 0.9977$ $B = \bullet: \frac{9}{11} \text{ positive} \rightarrow 0.684$ $B = \circ: \frac{3}{14} \text{ positive} \rightarrow 0.750$ Expected: 0.721 $C = \bullet: \frac{3}{6} \text{ positive} \rightarrow 1.0$ $C = \circ: \frac{9}{19} \text{ positive} \rightarrow 0.9980$ Expected: 0.9985

 $D = \bullet: \frac{3}{5}$ positive $\rightarrow 0.9710$ $D = o: \frac{9}{20}$ positive $\rightarrow 0.9928$ Expected: **0.9884**

А	В	l C	D	I
•	•	C 0	0	+
0	•	•	0	+
•	0	0	0	
	0	0	•	+
0	•	0	0	+
٠	0	•	0	
•	•	0	0	+
	0	0	0	
•	•	•	0	
•	•	0	0	+
0	0	0	•	+
•	0	•	0	
•	•	•	•	+
0	•	0	•	
0	0	0	0	
•	0	0	0	
•	•	0	•	
	•	0	0	+
•	0	•	0	
	0	0	0	
0	•	0	0	+
٠	•	•	•	+
0			0	+
•	0	0	0	
•	0	0	0	

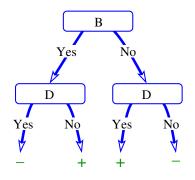
Entropy Entropy for datasets Information Gain



Entropy Entropy for datasets Information Gain



Examples where $B = \bullet$			Examples where $B = \circ$						
A	В	C	D		A	B	C	D	1
•	٠	0	0	+	0	0	0	0	
0	•	•	0	+	•	0	0	•	+
0	•	0	0	+	•	0	٠	0	
•	•	0	0	+	0	0	0	0	
•	•	0	0	+	0	0	•	0	
•	•	•	0	+	0	0	0	•	+
0	•	0	٠		•	0	0	0	
•	•	0	٠		0	0	0	0	
0	•	0	0	+	•	0	0	0	
0	•	0	0	+	0	0	•	0	
0	•	•	0	+	•	0	0	0	
					•	0	0	•	+
					0	0	0	0	
					•	0	0	0	



Örjan Ekeberg Brain Modeling and Machine Learning

Örjan Ekeberg Brain Modeling and Machine Learning



Possible ways of improving the decision trees

- Avoid overfitting
 - Limit the tree's height
 - Pruning
- Attributes with graded values
- Missing attribute values
- Variable cost for different attributes