SGN-13006 Introduction to Pattern Recognition and Machine Learning (5 cr)

Decision Tree Learning

Joni-Kristian Kämäräinen

September 2018

Laboratory of Signal Processing
Tampere University of Technology

Material

- Lecturer's slides and blackboard notes
- T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997: Chapter 3
- (Advanced: T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2009: Chapter 15)

Contents

Beyond concept learning

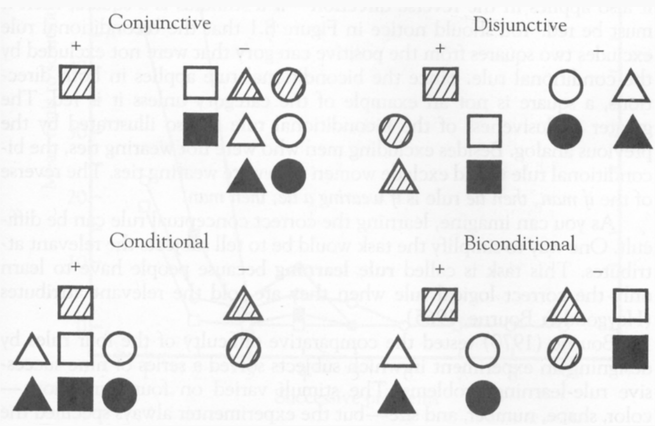


Figure 1: `http://cognitrn.psych.indiana.edu/rgoldsto/courses/concepts2.pdf`

Decision Tree Learning

Decision Tree Representation

Decision tree representation:

- Each internal node tests an attribute
- Each branch corresponds to attribute value
- Each leaf node assigns a classification

How would we represent:

- $\wedge, \vee$, XOR
- $(A \wedge B) \vee (C \wedge \neg D \wedge E)$
- M of N

Training Data (Mitchell Table 3.1)

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Top-Down Induction of Decision Trees

Main loop:

1. $A \leftarrow$ the "best" decision attribute for next *node*
2. Assign A as decision attribute for *node*
3. For each value of A, create new descendant of *node*
4. Sort training examples to leaf nodes
5. If training examples perfectly classified, Then STOP, Else iterate over new leaf nodes

Decision Tree Learning

ID3 algorithm

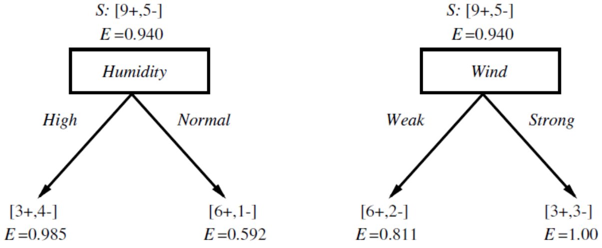
ID3 Algorithm

ID3 (Examples, Target_Attribute, Attributes)

```
Create a root node for the tree
if all examples are positive then
    Return the single-node tree Root, with label = +.
end if
if all examples are negative then
    Return the single-node tree Root, with label = -.
end if
if number of predicting attributes is empty then
    Return the single node tree Root, with label = most common value of the target attribute in the examples.
end if
Otherwise Begin
A = The Attribute that best classifies examples.
Decision Tree attribute for Root = A.
for each possible value, v_i, of A do
    Add a new tree branch below Root, corresponding to the test A = v_i.
    Let Examples(v_i) be the subset of examples that have the value v_i for A
    if Examples(v_i) is empty then
        Then below this new branch add a leaf node with label = most common target value in the examples
    else
        below this new branch add the subtree ID3 (Examples(v_i), Target_Attribute, Attributes {A})
    end if
end for
Return Root
```

Example

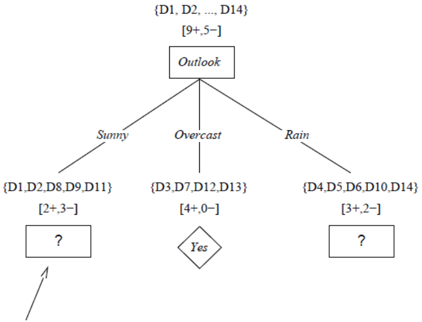
Which attribute is the best classifier?



Gain (S, Humidity)
= .940 - (7/14).985 - (7/14).592
= .151

Gain (S, Wind)
= .940 - (8/14).811 - (6/14)1.0
= .048

Example (cont.)



Which attribute should be tested here?

S_{sunny} = {D1,D2,D8,D9,D11}

$Gain(S_{sunny}, Humidity) = .970 - (3/5)\,0.0 - (2/5)\,0.0 = .970$

$Gain(S_{sunny}, Temperature) = .970 - (2/5)\,0.0 - (2/5)\,1.0 - (1/5)\,0.0 = .570$

$Gain(S_{sunny}, Wind) = .970 - (2/5)\,1.0 - (3/5)\,.918 = .019$

Application Example: C-Section Risk

```
[833+,167-] .83+ .17-
Fetal_Presentation = 1: [822+,116-] .88+ .12-
| Previous_Csection = 0: [767+,81-] .90+ .10-
| | Primiparous = 0: [399+,13-] .97+ .03-
| | Primiparous = 1: [368+,68-] .84+ .16-
| | | Fetal_Distress = 0: [334+,47-] .88+ .12-
| | | | Birth_Weight < 3349: [201+,10.6-] .95+ .05
| | | | Birth_Weight >= 3349: [133+,36.4-] .78+ .2
| | | Fetal_Distress = 1: [34+,21-] .62+ .38-
| Previous_Csection = 1: [55+,35-] .61+ .39-
Fetal_Presentation = 2: [3+,29-] .11+ .89-
Fetal_Presentation = 3: [8+,22-] .27+ .73-
```