INSTITUT TEKNOLOGI BANDUNG

1920

Modul : Issues in Decision Tree Learning (DTL)

Overfitting

Nur ULFA Maulidevi

KK IF - Teknik Informatika- STEI ITB

Pembelajaran Mesin
(Machine Learning)

Issues in DTL

Overfitting training data

Continuous-valued attribute

Handling attributes with differing costs

Handling missing attribute value

Alternative measures for selecting attributes

What is Overfit

H: Hypothesis space

A hypothesis: $h \in H$; Alternative hypothesis: $h' \in H$

train: training examples; D: entire distribution of data

$error_{train}(h) < error_{train}(h')$

↓

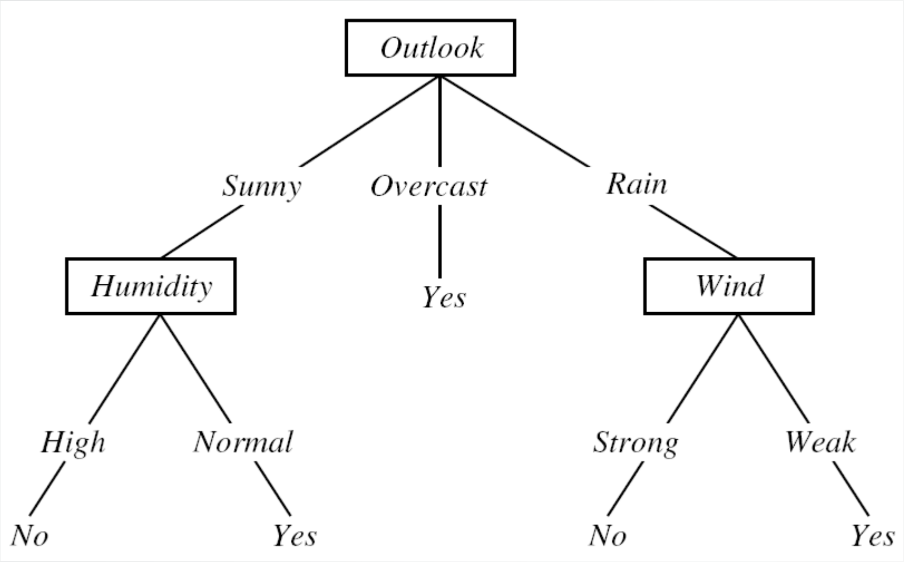
Overfit

↑

$error_D(h) > error_D(h')$

Illustration

D15 (noisy training examples):
Outlook = Sunny;
Temperature = Hot;
Humidity = Normal;
Wind = Strong;
PlayTennis = No



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

?

Overfitting can happen even training examples is noise-free (when small numbers of examples are associated with leaf Nodes) → decrese accuracy 10 – 25% on most problems

Solution Approaches

	Pros	Cons
1. Stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data	More Direct	Difficulty of estimating precisely when to stop growing the tree
2. allow the tree to overfit the data, and then post-prune the tree	More Successfull in practice	Requires more steps (grow until fit, then prune)

what criterion is to be used to determine the correct final tree size

Approaches in Determine the Correct Final Tree Size

1. Use separate examples distinct from training to evaluate the pruning tree

- 2/3 Training set
- 1/3 Validation set

2. Use all available data for training, test whether expanding (or pruning) a node will produce improvement

3. Use explicit measure of the complexity for encoding the training examples and decision tree → Minimum Description Length Principle

Source:
Machine Learning by Tom Mitchell chapter 6.6

$$h_{MDL} = \underset{h \in H}{\operatorname{argmin}} L_{C_1}(h) + L_{C_2}(D|h)$$

$L_{C_1}(h)$: *Length (number of bits) of hypothesis encoding*
$L_{C_2}(D|h)$: *Length of data D given hypothesis h encoding*

Reduced Error Pruning

Consider decision (attribute) node as candidates for pruning → assign the most common classification affiliated with that node

→ Grow until fit then prune

Split data into training and validation set
Do until further pruning is harmful:

1. Evaluate impact on validation set of pruning each possible node (plus those below it)
2. Greedily remove the one (node) that most improves validation set accuracy

an effective approach provided a large amount of data is available

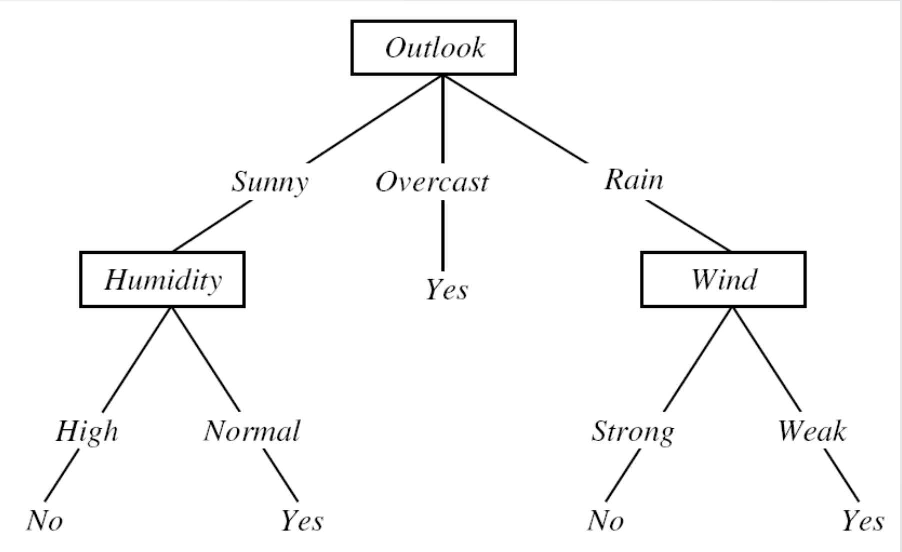
Rule Post-Pruning

Improvement of ID3 Algorithm: C4.5

Suitable for limited data

1. Growing the tree from training set, until the training data is fit as well as possible and allowing overfitting to occur.
2. Convert the learned tree into an equivalent set of rules by creating one rule for each path from the root node to a leaf node.
3. Prune (generalize) each rule by removing any preconditions that result in improving its estimated accuracy.
4. Sort the pruned rules by their estimated accuracy, and consider them in this sequence when classifying subsequent instances.

Example



Decision Rules:
If Outlook = Sunny and Humidity=High then No

Pruning:
If Outlook = Sunny then No
OR
If Humidity=High then No

↓

Increase/ Reduce Accuracy?

↓

Over validation set/ training set (C4.5)

Why Decision Tree → Decision Rule ?

1. Distinct path ~ distinct rule: independent pruning
2. No distinction between attribute tests
3. Improves readability

THANK YOU

INSTITUT TEKNOLOGI BANDUNG

1920