characterizinginductivesystems by theirinductivebiasallowsmodellingthembytheirequivalentdeductivesystems.Thisprovidesawaytocompareinductive systemsaccording to theirpolicies forgeneralizingbeyond the observedtrainingdata

DECISIONTREELEARNING

DECISIONTREEREPRESENTATION

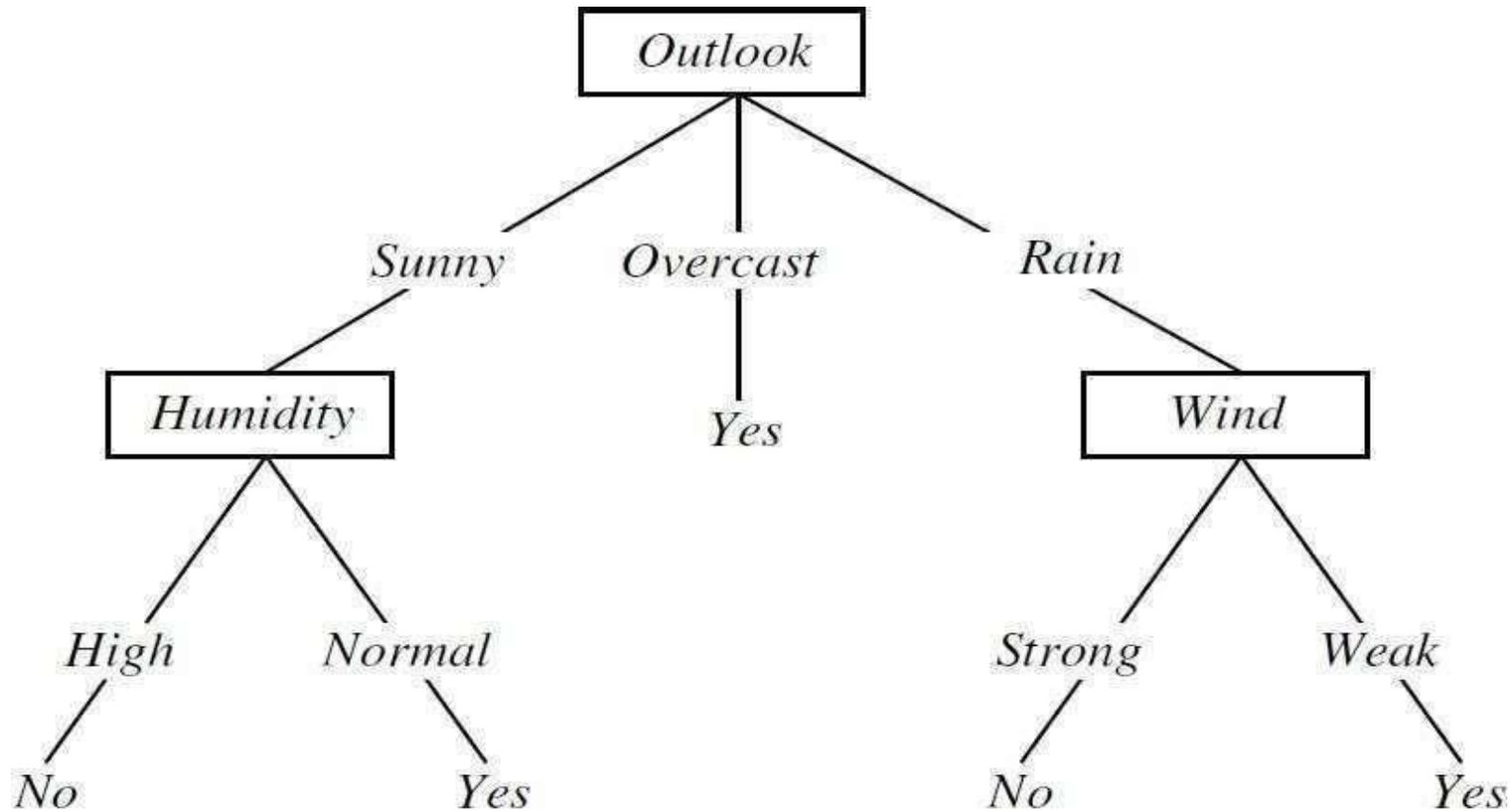


FIGURE: A decision tree for theconcept ***PlayTennis***.Anexam pleisclassified by sortingit through the tree totheappropriateleaf node, then returningtheclassific ationassociated with thisleaf

- Decision trees classify instances by sorting them down the tree from the root tosomeleaf node, which provides the classificationofthe instance.

- Each node in the tree specifies a test of some attribute of the instance, and eachbranch descending from that node corresponds to one of the possible values forthisattribute.

- Aninstanceisclassifiedbystartingattherootnodeofthetree,testingtheattribute specified by this node, then moving down the tree branch correspondingtothevalueoftheattributeinthegivenexample.Thisprocessisthenrepea tedforthe subtree rootedat the new node.

- Decisiontreesrepresentadisjunctionofconjunctionsofconstraintsontheattributevalues of instances.
- Eachpathfromthetreeroottoaleafcorrespondstoaconjunctionofattributetests,and the treeitself toa disjunctionof these conjunctions

Forexample,
Thedecisiontreeshowninabovefigurecorrespondstotheexpression

(Outlook =Sunny AHumidity = Normal)
(Outlook=Overcast)
(Outlook=RainAWind=Weak)

APPROPRIATEPROBLEMSFOR DECISIONTREELEARNING

Decisiontreelearningisgenerallybestsuitedtoproblemswiththefollowingcharacteristics:

1. ***Instancesarerepresentedbyattribute-valuepairs***–Instancesaredescribedbyafixed set of attributes and their values
2. ***The target function has discrete output values*** – The decision tree assigns aBoolean classification (e.g., yes or no) to each example. Decision tree methodseasilyextendtolearningfunctionswithmorethantwopossibleoutputvalues.
3. ***Disjunctivedescriptionsmayberequired***

4. ***Thetrainingdatamaycontainerrors***–Decisiontreelearningmethodsarerobust to errors, both errors in classifications of the training examples and errorsinthe attributevalues that describe these examples.

5. ***Thetrainingdatamaycontainmissingattributevalues–***Decisiontreemethodscanbe used evenwhen some training exampleshaveunknownvalues

- Decision tree learning has been applied to problems such as learning to classify*medical patients by their disease*, *equipment malfunctions by their cause*, and*loan applicants by their likelihood ofdefaulting on payments*.

- Suchproblems,inwhichthetaskistoclassifyexamplesintooneofadiscrete setofpossiblecategories,areoftenreferredtoas***classificationproblems.***

THEBASICDECISIONTREE ALGORITHM

- Mostalgorithmsthathavebeendevelopedforlearningdecisiontreesarevariations on a core algorithm that employs a top-down, greedy search through thespaceofpossibledecisiontrees.Thisapproachisexemplifiedbytheid3algorithmand itssuccessor C4.5

WhatistheID3algorithm?

- ID3standsforIterativeDichotomiser3
- ID3isaprecursortotheC4.5Algorithm.
- TheID3algorithmwasinventedbyRossQuinlanin1975
- Usedtogenerateadecisiontreefromagivendatasetbyemployingatop-down,greedysearch,totesteach attributeat everynodeofthetree.
- Theresultingtreeisusedtoclassifyfuturesamples.

ID3algorithm

ID3(Examples,Target_attribute,Attributes)

Examples are the training examples. Target_attribute is the attribute whose value is tobe predictedby the tree. Attributes is a list of other attributes that may be tested by the learned decision tree.Returnsadecisiontreethatcorrectlyclassifiesthegiven Examples.

- CreateaRootnodeforthetree
- IfallExamplesarepositive,Returnthesingle-nodetreeRoot,withlabel=+
- IfallExamplesarenegative,Returnthesingle-nodetreeRoot,withlabel=-
- IfAttributesisempty,Returnthesingle-nodetreeRoot,withlabel=mostcommonvalueofTarget_attributeinExamples

- OtherwiseBegin
  - A←theattributefromAttributesthatbest*classifiesExamples
  - ThedecisionattributeforRoot←A
  - Foreachpossiblevalue,v_i,ofA,
    - Addanewtreebranchbelow*Root*,correspondingtothetestA=v_i
    - Let$Examples_{vi}$,bethesubsetofExamplesthathavevalue$v_i$for*A*
    - If$Examples_{vi}$,isempty
      - Thenbelowthisnewbranchaddaleafnodewithlabel=mostcommonvalueofTarget_attribuleinExamples
      - Elsebelowthisnewbranchaddthesubtree ID3(*Examples*$_{vi}$,Targe_tattribute,Attributes–{A}))
- End
- ReturnRoot

*Thebestattributeistheonewithhighestinformationgain

WhichAttributeIstheBestClassifier?

- ThecentralchoiceintheID3algorithmisselectingwhichattributetotestat eachnode in thetree.
- Astatisticalpropertycalled***informationgain***thatmeasureshowwellagivenattributeseparatesthetraining examplesaccordingto theirtargetclassification.
- ID3uses***informationgain***measuretoselectamongthecandidateattributesateach step while growingthe tree.

ENTROPYMEASURESHOMOGENEITYOFEXAMPLES

- Todefineinformationgain,webeginbydefiningameasurecalledentropy. *Entropymeasurestheimpurityofacollectionofexamples.*
- GivenacollectionS,containingpositiveandnegativeexamplesofsometargetconcept,theentropyof Srelative to this Booleanclassificationis

$$Entropy\,(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

Where,

p_+isthepropportionofpositiveexamplesinS

p_-isthepropportionofnegativeexamplesinS.

Example:Entropy

- Suppose S is a collection of 14 examples of some boolean concept, including 9positive and 5 negative examples. Thenthe entropy of Srelative to this booleanclassificationis

$$Entropy([9+, 5-]) = -(9/14)\log_2(9/14) - (5/14)\log_2(5/14)$$
$$= 0.940$$

- Theentropyis0ifallmembersofSbelongtothesameclass
- Theentropyis1whenthecollectioncontainsanequalnumberofpositiveandnegative examples
- Ifthecollectioncontainsunequalnumbersofpositiveandnegativeexamples,theentropyisbetween0and 1

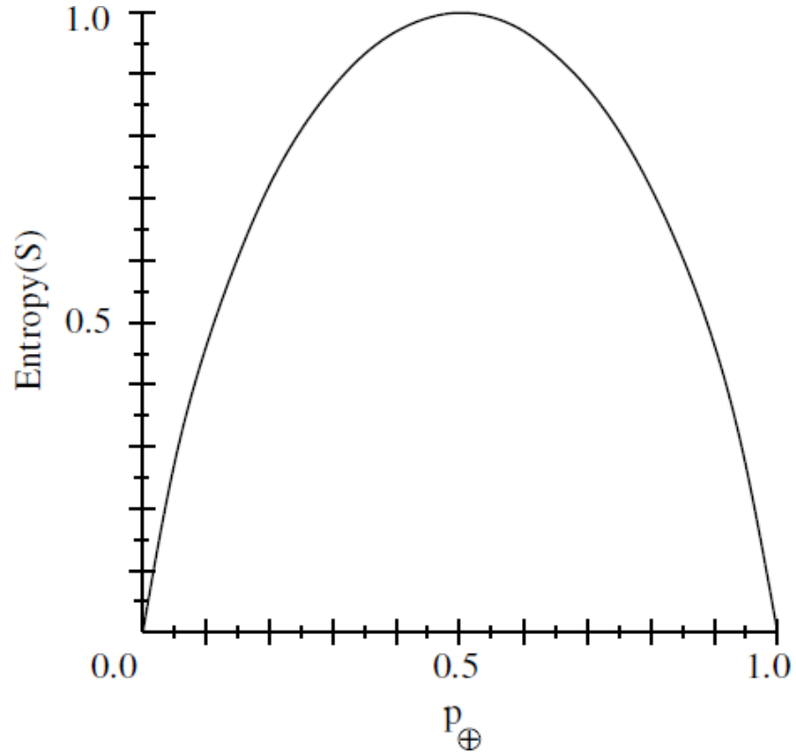


FIGURE The entropy function relative to a boolean classification, as the proportion, $p_{\oplus}$, of positive examples varies between 0 and 1.

INFORMATIONGAINMEASURESTHEEXPECTEDREDUCTIONINENTROPY

- ***Informationgain,***istheexpectedreductioninentropycausedbypartitioning theexamplesaccording tothis attribute.
- Theinformationgain,Gain(S,A)ofanattributeA,relativetoacollectionofexamplesS, is defined as

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

Example:Informationgain

Let, $Values(Wind)=\{Weak, Strong\}$

$S = [9+, 5-]$

$S_{Weak} = [6+, 2-]$

$S_{Strong} = [3+, 3-]$

Informationgainofattribute*Wind*:

$$Gain(S, Wind) = Entropy(S) - 8/14\, Entropy(S_{Weak}) - 6/14\, Entropy(S_{Strong})$$

$$= 0.94 - (8/14)*0.811 - (6/14)*1.00$$

$$= 0.048$$

AnIllustrativeExample

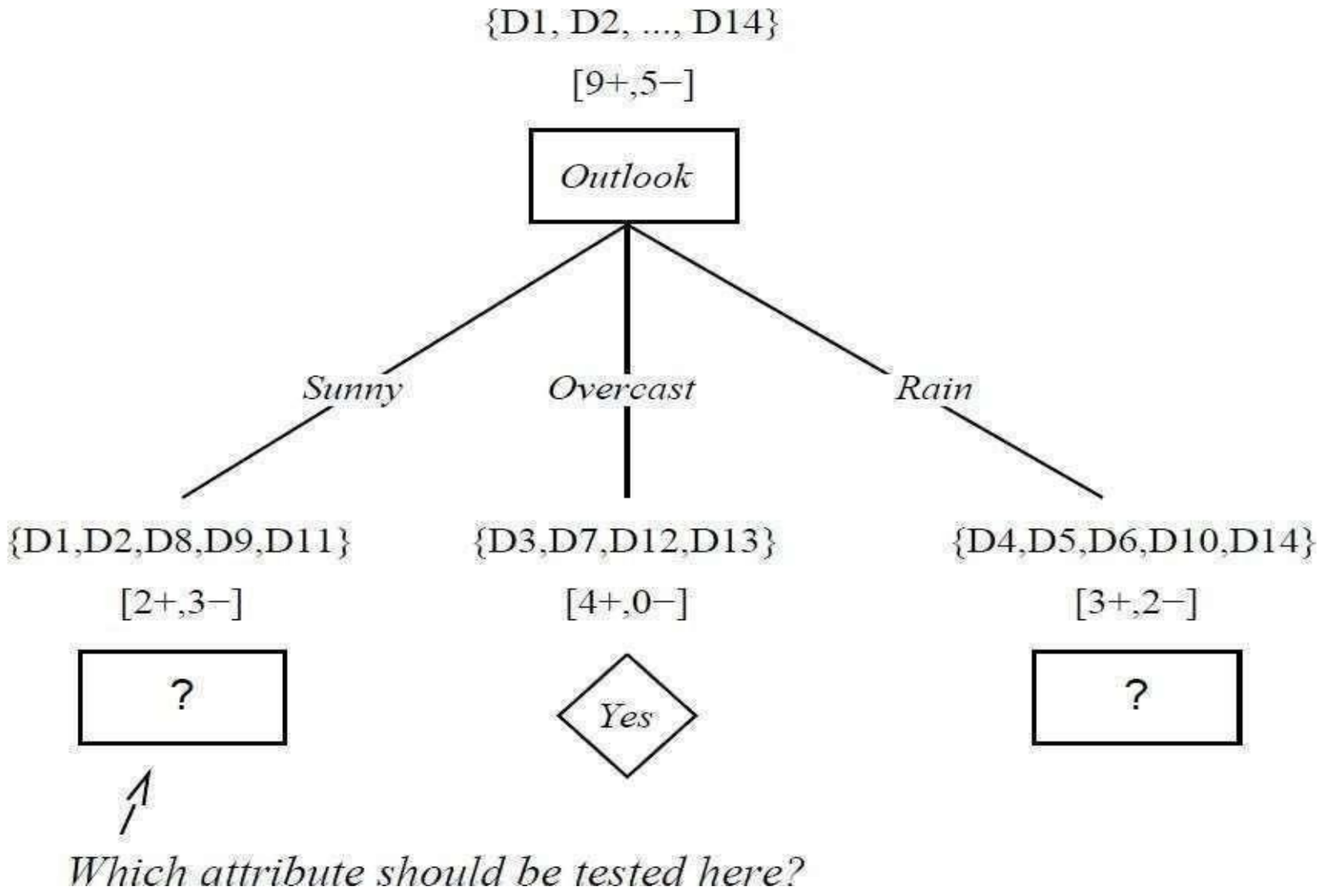
- ToillustratetheoperationofID3,considerthelearningtask representedbythetrainingexamplesofbelowtable.
- Herethetargetattribute***PlayTennis***,whichcanhavevalues***yes*** or***no***fordifferentdays.
- Considerthefirststepthroughthealgorithm,inwhichthetopmostnodeofthedecision treeiscreated.

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

Theinformationgainvaluesforallfourattributesare

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

- Accordingtotheinformationgainmeasure,the***Outlook***attributeprovidesthebest prediction of the target attribute, ***PlayTennis***, over the training examples.Therefore,***Outlook***isselectedasthedecisionattributefortherootnode, andbranchesarecreatedbelowtherootforeachofitspossiblevaluesi.e.,


{D1, D2, ..., D14}
[9+,5−]
Outlook
Sunny
Overcast
Rain
{D1,D2,D8,D9,D11}
[2+,3−]
{D3,D7,D12,D13}
[4+,0−]
{D4,D5,D6,D10,D14}
[3+,2−]
?
Yes
?
Which attribute should be tested here?

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

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

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

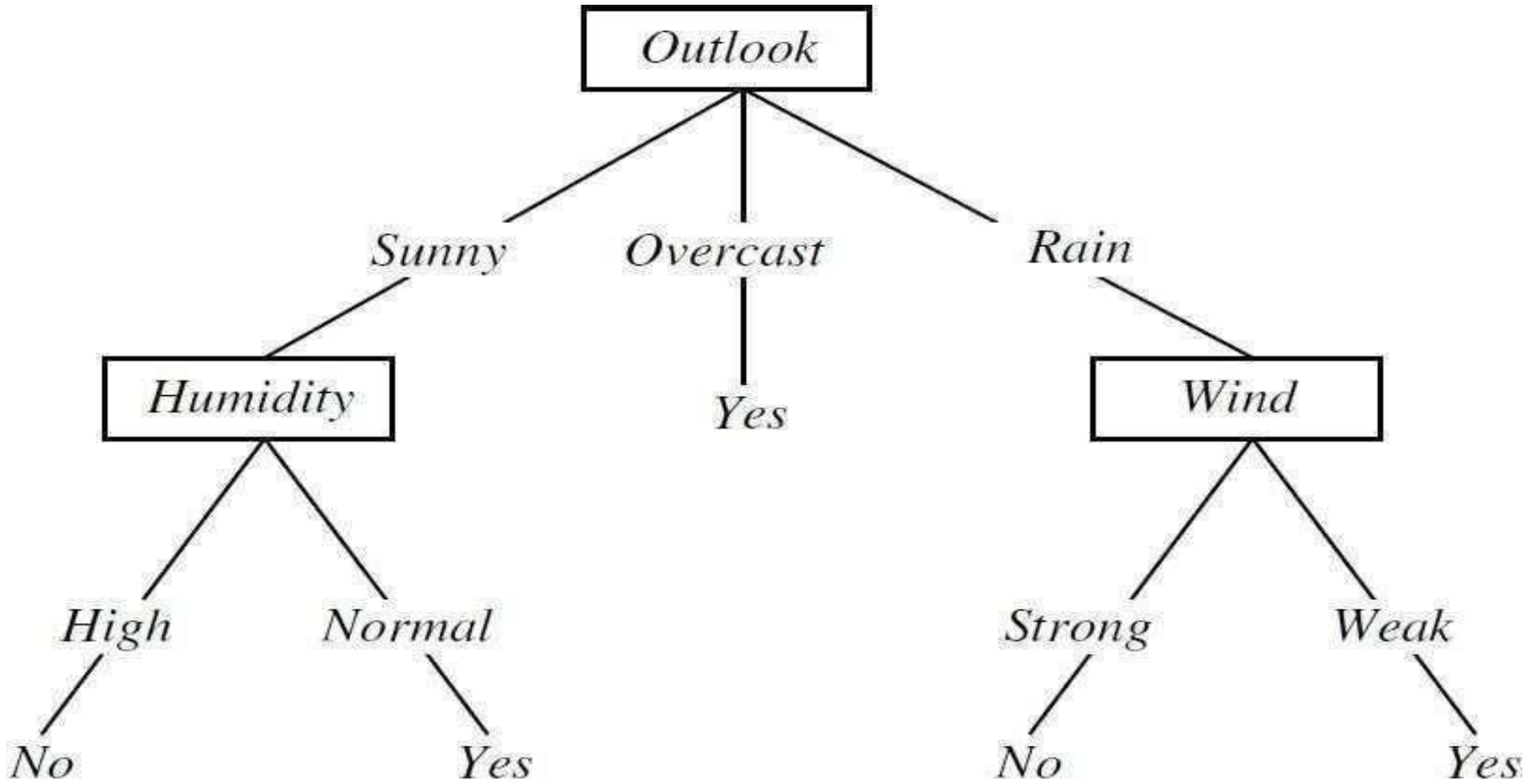
$Gain(S_{sunny}, Wind) = .970 - (2/5)\,1.0 - (3/5)\,.918 = .019$

$S_{Rain} = \{D4, D5, D6, D10, D14\}$

$$Gain(S_{Rain}, Humidity) = 0.970 - (2/5)1.0 - (3/5)0.917 = 0.019$$

$$Gain(S_{Rain}, Temperature) = 0.970 - (0/5)0.0 - (3/5)0.918 - (2/5)1.0 = 0.019$$

$$Gain(S_{Rain}, Wind) = 0.970 - (3/5)0.0 - (2/5)0.0 = 0.970$$


Outlook
Sunny
Overcast
Rain
Humidity
Yes
Wind
High
Normal
Strong
Weak
No
Yes
No
Yes

HYPOTHESISSPACESEARCHINDECISIONTREEL EARNING

- ID3canbecharacterizedassearchingaspaceofhypothesesforonethatfits thetraining examples.
- ThehypothesisspacesearchedbyID3isthesetofpossibledecisiontrees.
- ID3 performs a ***simple-to complex, hill-climbing search*** through this hypothesisspace,beginningwiththeemptytree,thenconsideringprogressivelymoreela boratehypothesesinsearchofadecisiontreethatcorrectlyclassifiesthetrainingdata

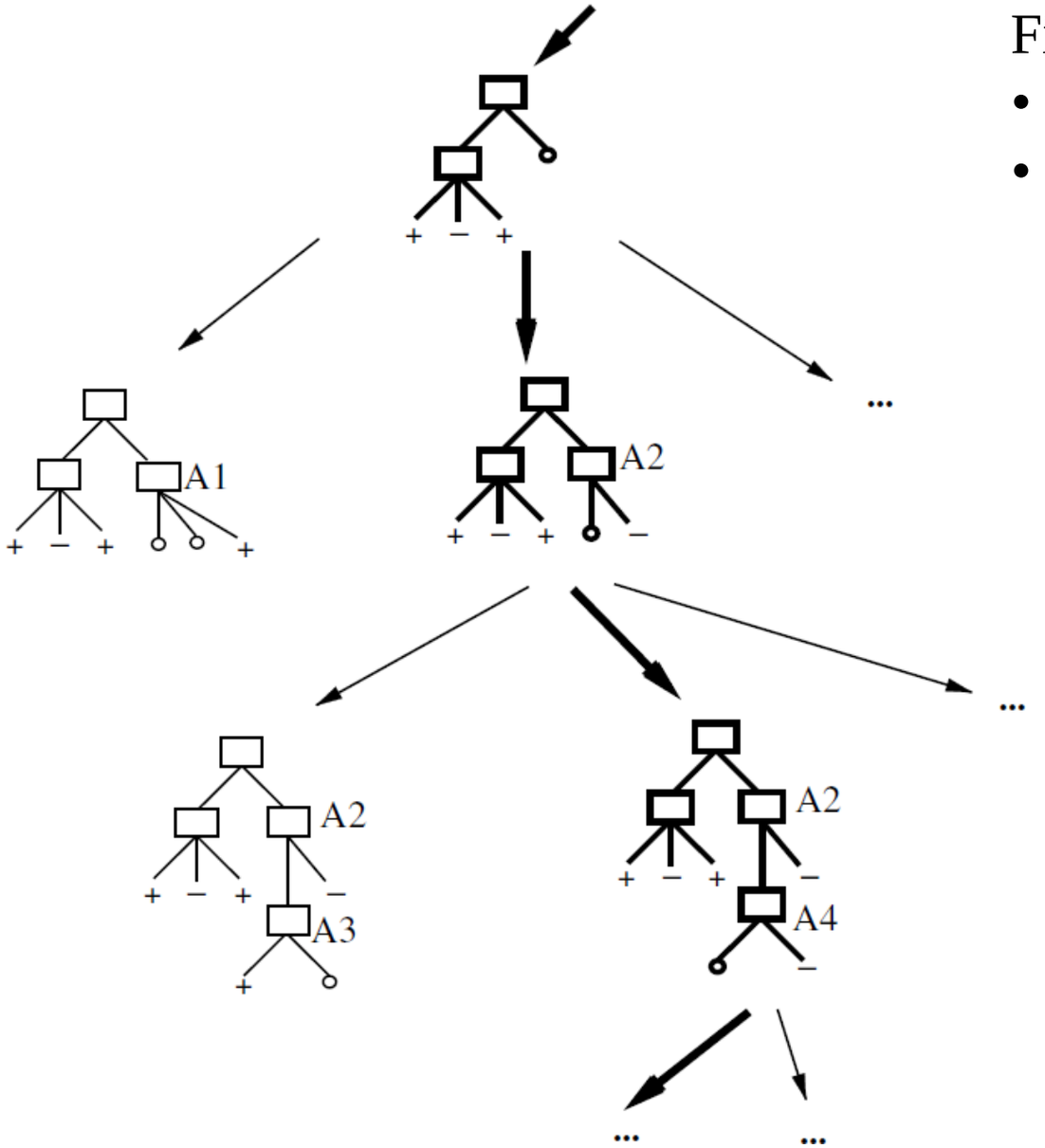


Figure:

- HypothesisspacesearchbyID3.
- ID3searchesthroughthespaceofpossible decision trees from simplest toincreasinglycomplex,guidedbytheinfo rmationgain heuristic

1. ID3's hypothesis space of all deision trees is a ***complete*** space of finite discrete-valued functions, relative to the available attributes. Because every finite discrete-valuedfunction can be represented bysome decision tree

- ID3avoidsoneofthemajorrisksofmethodsthat***searchincompletehypothesisspaces*** :thatthehypothesisspacemightnotcontainthetargetfunction.

2. ID3maintains***onlyasinglecurrenthypothesis***asitsearchesthroughthespaceof decision trees.

Forexample,withtheearlierversionspacecandidateeliminationmethod,whichmaint ains theset of ***all*** hypotheses consistentwiththeavailable trainingexamples.

Bydeterminingonlyasinglehypothesis,ID3losesthecapabilitiesthatfollowfromexplici tlyrepresenting allconsistenthypotheses.

For example, it does not have the ability to determine how many alternativedecisiontreesareconsistentwiththeavailabletrainingdata,orto pose newinstancequeriesthat optimallyresolveamongthesecompetinghypotheses

3. ID3 in its pure form performs ***no backtracking in its search***. Once it selects anattribute to test at a particular level in the tree, it never backtracks to reconsider thischoice.

- In the case of **ID3,** a locally optimal solution corresponds to the decision tree itselects along the single search path it explores. However, this locally optimalsolution may be less desirable than trees that would have been encountered along adifferentbranch of the search.

4. ID3*usesalltrainingexamplesateachstep*inthesearchtomakestatisticallybaseddecisions regarding how to refineits current hypothesis.

- Oneadvantageofusingstatisticalpropertiesofalltheexamplesisthattheresultingsearch ismuch ***less sensitivetoerrors***inindividualtrainingexamples.
- **ID3**canbeeasilyextendedtohandlenoisytrainingdataby modifyingitsterminationcriteriontoacceptthypotheses thatimperfectlyfitthetrainingdata.