Šta ako vrednosti atributa nisu diskretne?

- Primena stabla odlučivanja je ograničena na atribute koji uzimaju vrednosti iz diskretnog skupa.

  1. Ciljna vrednost koju stablo treba da predvidi je diskretna.
  2. Atributi koji se testiraju u čvorovima uzimaju diskretne vrednosti.

- Uslov 2 se lako može ukloniti tako da neprekidni (kontinualni) atributi takođe budu uključeni u stablo.

Šta ako vrednosti atributa nisu diskretne?

- Definisati novi atribut sa diskretnim vrednostima koji deli neprekidni opseg vrednosti kontinualnog atributa na skup intervala.

- Za atribut A koji je kontinualan, algoritam može dinamički odrediti novi bulovski atribut A_c koji ima vrednost $true$ ako je vrednost atributa A manja od c.

- Kako odabrati prag c?

Primer

- In the current example, there are two candidate thresholds, corresponding to the values of Temperature at which the value of PlayTennis changes:

  (48 + 60)/2,

  (80 + 90)/2.

- The information gain can then be computed for each of the candidate attributes, $Temperature_{>54}$, and $Temperature_{>85}$, and the best can be selected - $Temperature_{>54}$

Temperature:	40	48	60	72	80	90
PlayTennis:	No	No	Yes	Yes	Yes	No

Algoritam određivanja testa numeriškog atributa

1. For each attribute at_i:

   (i) Sort the training examples by the values of at_i;

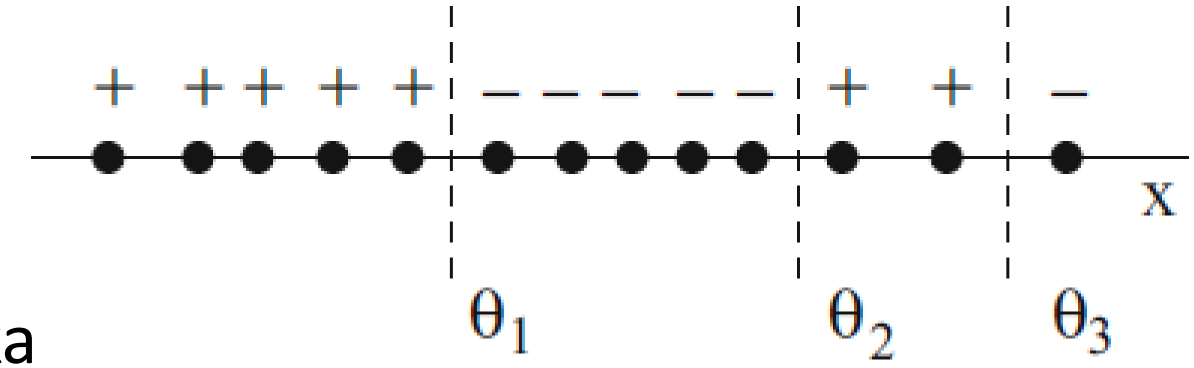
   (ii) Determine the candidate thresholds, θ_{ij}, as those lying between examples with opposite labels;

   (iii) For each θ_{ij}, determine the amount of information contributed by the boolean attribute thus created.

2. Choose the pair $[at_i, \theta_{ij}]$ that offers the highest information gain.

Primer

- Atribut x je kontinualan. Na osi su prikazane tri moguće vrednosti praga koje razdvajaju pozitivne od negativnih primera

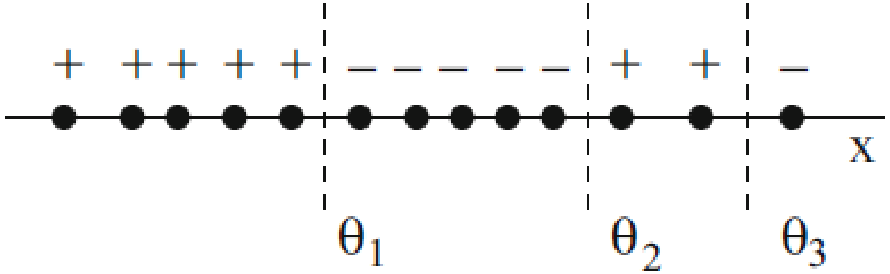


- Entropija datog skupa podataka

$$
\begin{aligned}
H(T) &= -p_+ \log p_+ - p_- \log p_- \\
&= -(7/13)\log(7/13) - (6/13)\log(6/13) = 0.9957
\end{aligned}
$$

Primer

- Entropije atributa koje definiše svaka od tri vrednosti praga



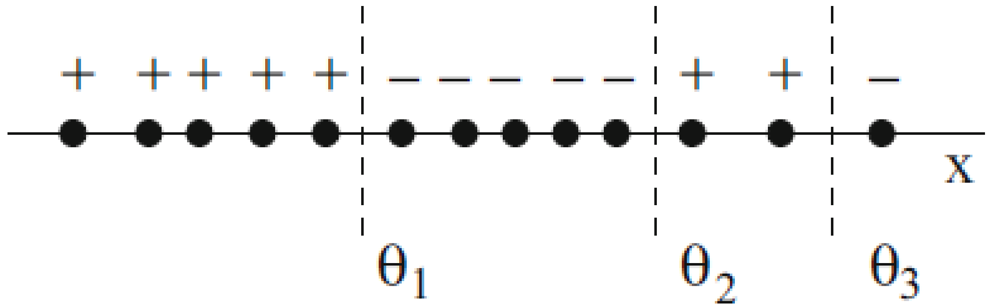
$$H(x < \theta_1) = -(5/5)\log(5/5) - (0/5)\log(0/5) = 0$$
$$H(x > \theta_1) = -(2/8)\log(2/8) - (6/8)\log(6/8) = 0.811$$

$$H(x < \theta_2) = -(5/10)\log(5/10) - (5/10)\log(5/10) = 1$$
$$H(x > \theta_2) = -(2/3)\log(2/3) - (1/3)\log(1/3) = 0.9183$$

$$H(x < \theta_3) = -(7/12)\log(7/12) - (5/12)\log(5/12) = 0.9799$$
$$H(x > \theta_3) = -(0/1)\log(0/1) - (1/1)\log(1/1) = 0$$

Primer

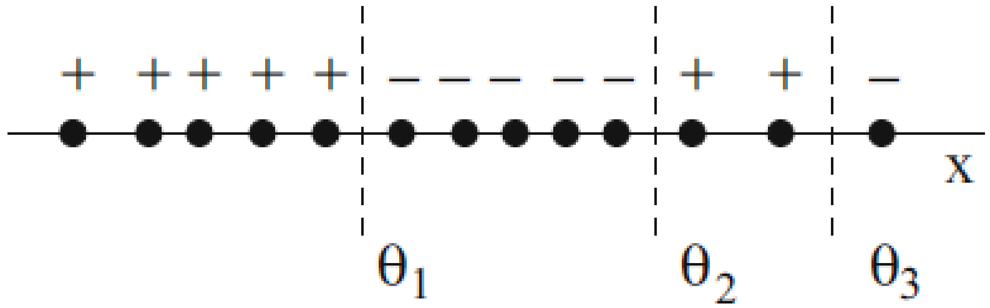
- Prosečne entropije za svaku od tri vrednosti praga



$$H(T, \theta_1) = (5/13) \cdot 0 + (8/13) \cdot 0.8113 = 0.4993$$
$$H(T, \theta_2) = (10/13) \cdot 1 + (3/13) \cdot 0.9183 = 0.9811$$
$$H(T, \theta_3) = (12/13) \cdot 0.9799 + (1/13) \cdot 0 = 0.9045$$

Primer

- Informaciona poboljšanja za svaku od tri vrednosti praga



$$I(T,\theta_1) = H(T) - H(T,\theta_1) = 0.9957 - 0.4993 = 0.4964$$
$$I(T,\theta_2) = H(T) - H(T,\theta_2) = 0.9957 - 0.9811 = 0.0146$$
$$I(T,\theta_3) = H(T) - H(T,\theta_3) = 0.9957 - 0.9045 = 0.0912$$

- Treći prag ima najveće informaciono poboljšanje.

Avoiding overfitting in decision tree

- There are several approaches to avoiding overfitting in decision tree learning. These can be grouped into two classes:

- Approaches that stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data.

- Approaches that allow the tree to overfit the data, and then post-prune the tree.

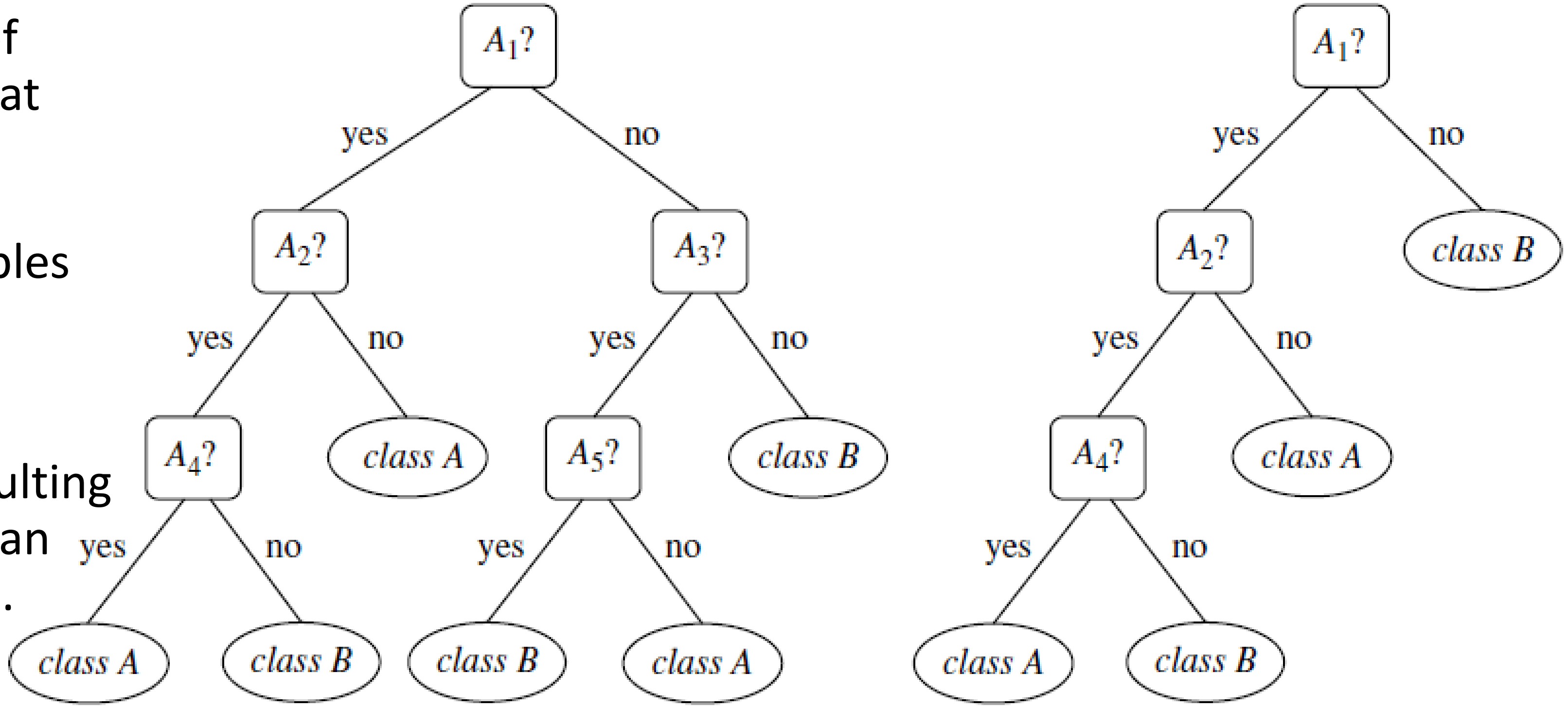
Avoiding overfitting in decision tree

- Although the first of these approaches might seem more direct, the second approach of post-pruning overfit trees has been found to be more successful in practice.

- This is due to the difficulty in the first approach of estimating precisely when to stop growing the tree.

Reduced-error pruning

- Consider each of the decision nodes in the tree to be candidates for pruning.
- Pruning a decision node consists of removing the subtree rooted at that node, making it a leaf node, and assigning it the most common classification of the training examples affiliated with that node.
- Nodes are removed only if the resulting pruned tree performs no worse than the original over the validation set.



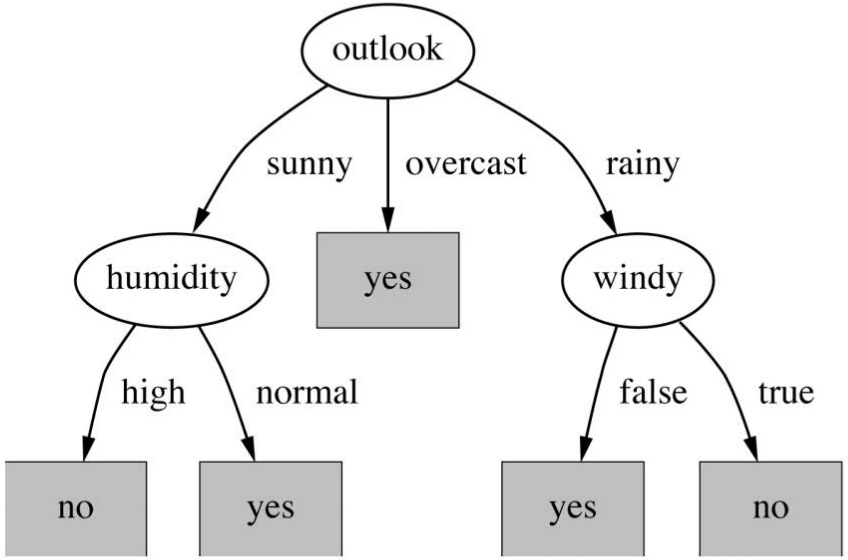
An unpruned decision tree and a pruned version of it.

Primer

(Outlook = Sunny, Temperature = Hot, Humidity = Normal, Wind = Strong, PlayTennis = No)

Kakav bi bio efekat dodavanja pozitivnog trening primera, koji je pogrešno označen kao negativan, na (inače tačne) primere iz tabele

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

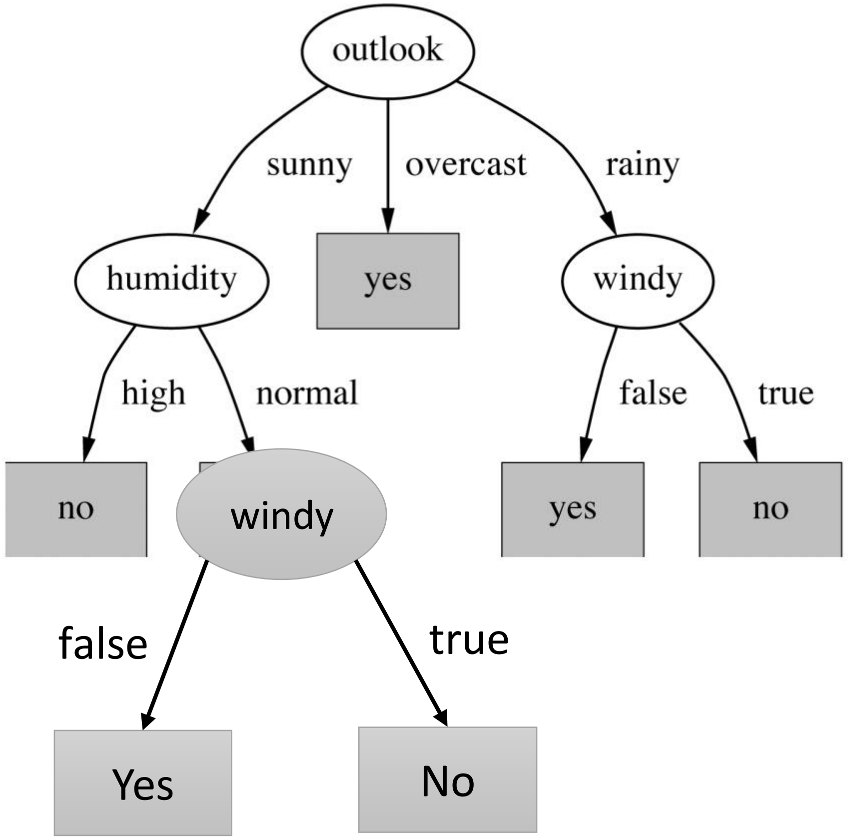


Primer

(Outlook = Sunny, Temperature = Hot, Humidity = Normal, Wind = Strong, PlayTennis = No)

Kakav bi bio efekat dodavanja pozitivnog trening primera, koji je pogrešno označen kao negativan, na (inače tačne) primere iz tabele

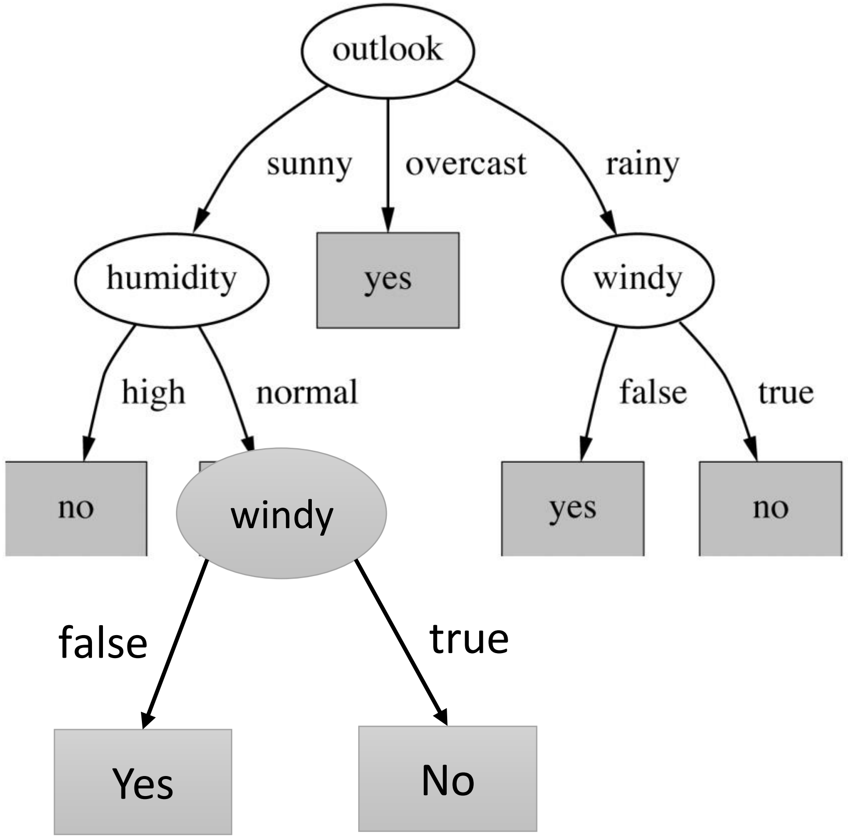
Day	*Outlook*	*Temperature*	*Humidity*	*Wind*	*PlayTennis*
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool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



Primer

- Novo stablo će se savršeno razvrstavati primere iz trening skupa.

- Međutim, s obzirom na to da je novi čvor posledica uklapanja netačnog (*noisy*) trening primera, staro stablo će nadmašiti novo kada se primene za klasifikaciju testnih podataka.



Reduced-error pruning

- Nodes are pruned iteratively, always choosing the node whose removal most increases the decision tree accuracy over the validation set.
- Pruning of nodes continues until further pruning is harmful (i.e., decreases accuracy of the tree over the test set).

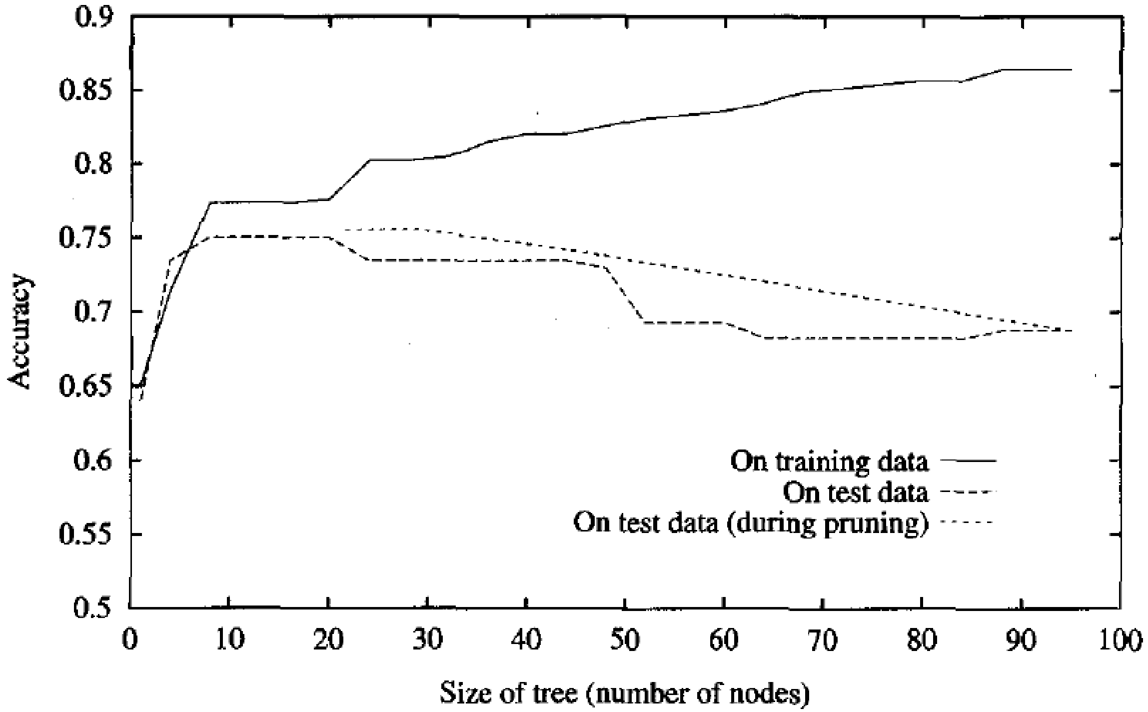


FIGURE 3.7
Effect of reduced-error pruning in decision tree learning. This plot shows the same curves of training and test set accuracy as in Figure 3.6. In addition, it shows the impact of reduced error pruning of the tree produced by ID3. Notice the increase in accuracy over the test set as nodes are pruned from the tree. Here, the validation set used for pruning is distinct from both the training and test sets.