Dat550-exercises-3

Vinay Setty

06. Feb 2020

1 Model evaluation

Compute the Precision, Accuracy, F1-Score and ROC curve for the following classifications given by a classifier.

Table 1: Data

Predicted	**Groundtruth**
Yes	No
No	No
Yes	Yes
No	Yes
No	Yes
Yes	Yes
Yes	Yes
No	No
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	Yes
Yes	No

2 Rule-based Classifer:

Convert the decision tree into a set of classification rules $(r_i\,(Condition_i) \rightarrow y_i)$. Compute the coverage and accuracy of each rule. Compute the coverage and accuracy of each rule.

$$Coverage(r) = \frac{|A|}{|D|} Accuracy(r) = \frac{|A \cap y|}{|A|}$$

where $|A|$ is the number of rules that satisfy the rule antecendent (precondition), $|A \cap y|$ is the number of rules that satisfy both the antecedent and consequent, and $|D|$ is the total number of records.

Table 2: Data

Outlook	**Temp.**	**Humidity**	**Windy**	**Play**
sunny	85	85	false	No
sunny	80	90	true	No
overcast	83	78	false	Yes
rain	70	96	false	Yes
rain	68	80	false	Yes
rain	65	70	false	Yes
overcast	64	65	true	Yes
sunny	72	95	false	No
sunny	69	70	false	Yes
rain	75	80	false	Yes
sunny	75	70	true	Yes
overcast	72	90	true	Yes
overcast	81	75	false	Yes
rain	71	80	true	No

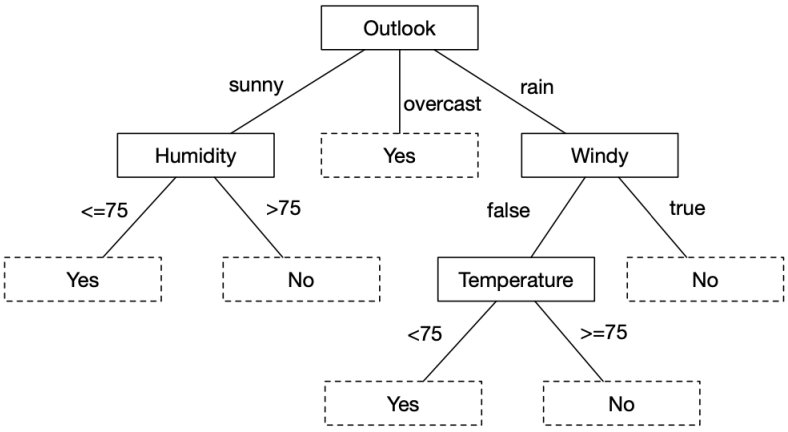

Figure 1: Decision Tree for the Data

Can you simplify the rules?

Table 3: Data

Rule	Coverage	Accuracy
$r_1 : (Outlook = sunny \wedge Humidity \leq 75) \rightarrow Yes$	$\frac{2}{14} = 0.143$	$\frac{2}{2} = 1$