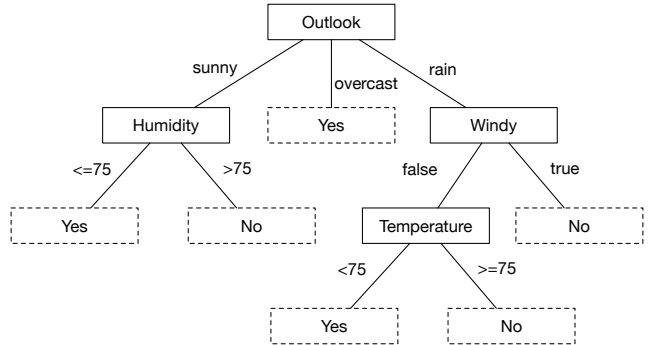


# Exercises: Rule-based classifier, Naive Bayes

September 13, 2016

## 1 Rule-based classifier

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	true	No
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



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.

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

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

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

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

Table 1: Rule set.

Can you simplify the rules?

## 2 Naive Bayes

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	true	No
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

Using the above training data set, build a Naive Bayes classifier to classify the following record. Use discretization for the continuous attributes.

$$x = \{Outlook = sunny, Temp = 70, Humidity = 65, Windy = true\} \quad (1)$$

The Naive Bayes classifier

$$P(x|y) = P(y) \prod_i P(x_i|y) \quad (2)$$

$$P(x|Yes) = P(Yes) \cdot P(Outlook = sunny|Yes) \cdot P(Temp = 70|Yes) \cdot P(Humidity = 65|Yes) \cdot P(Windy = true|Yes) =$$

$$P(x|No) = P(No) \cdot P(Outlook = sunny|No) \cdot P(Temp = 70|No) \cdot P(Humidity = 65|No) \cdot P(Windy = true|No) =$$

Classify the following record using a Gaussian distribution for continuous attributes.

$$x = \{Outlook = rain, Temp = 87, Humidity = 90, Windy = false\} \quad (3)$$

The class-conditional probability for attribute  $x_i$  is:

$$P(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}, \quad (4)$$

where the sample mean  $\mu_{ij}$  and variance  $\sigma_{ij}$  can be estimated from the training records that belong to class  $y_j$ .

$$P(x|Yes) = P(Yes) \cdot P(Outlook = rain|Yes) \cdot P(Temp = 87|Yes) \cdot P(Humidity = 90|Yes) \cdot P(Windy = false|Yes) =$$

$$P(x|No) = P(No) \cdot P(Outlook = rain|No) \cdot P(Temp = 87|No) \cdot P(Humidity = 90|No) \cdot P(Windy = false|No) =$$