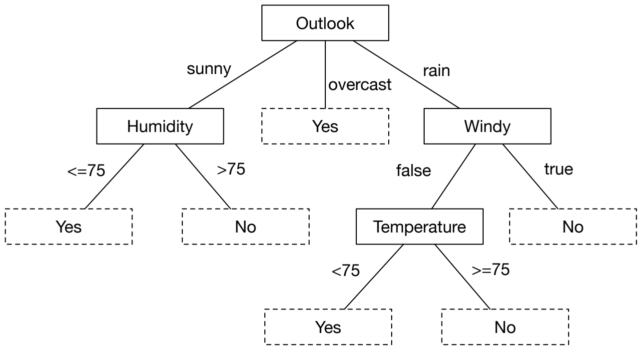
Exercises: Rule-based classifier, Naive Bayes

October 9, 2017

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 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.

rule	Coverage	Accuracy
$r_1 : (Outlook = sunny \wedge Humidity \leq 75) \rightarrow Yes$	$\frac{2}{14} = 0.143$	$\frac{2}{2} = 1$
$r_2 : (Outlook = sunny \wedge Humidity > 75) \rightarrow No$	$\frac{3}{14} = 0.214$	$\frac{3}{3} = 1$
$r_3 : (Outlook = overcast) \rightarrow Yes$	$\frac{4}{14} = 0.286$	$\frac{4}{4} = 1$
$r_4 : (Outlook = rain \wedge Windy = false \wedge Temperature \leq 75) \rightarrow Yes$	$\frac{2}{14} = 0.143$	$\frac{2}{2} = 1$
$r_5 : (Outlook = rain \wedge Windy = false \wedge Temperature > 75) \rightarrow No$	$\frac{1}{14} = 0.071$	$\frac{0}{1} = 0$
$r_6 : (Outlook = rain \wedge Windy = true) \rightarrow No$	$\frac{2}{14} = 0.143$	$\frac{2}{2} = 1$

Table 1: Rule set.

Can you simplify the rules?

r_5 has 0 accuracy, therefore it can be dropped. But, then the rule set is not exhaustive. Observe that when $Outlook = rain \wedge Windy = false$ the target class is always Yes. Therefore, r_4 may be simplified to r_4' : $(Outlook = rain \wedge Windy = false) \rightarrow Yes$. This way we the rule set becomes exhaustive.

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 new instance:

$$\mathbf{X} = \{Outlook = sunny, Temp = 70, Humidity = 65, Windy = true\} \tag{1}$$

The Naive Bayes classifier:

$$P(Y|\mathbf{X}) \propto P(Y) \prod_i P(X_i|Y) \tag{2}$$

For categorical attributes, $P(X_i = x_i|Y = y)$ is the fraction of training instances in class y that have a particular attribute value $X_i = x_i$.

a) Use discretization for the continuous attributes (e.g., split them to two categories: ≤ 75 and > 75).

$P(Yes|\mathbf{X}) \propto P(Yes)\frac{9}{14} \cdot P(Outlook = sunny|Yes)\frac{2}{9} \cdot P(Temp \leq 75|Yes)\frac{7}{9} \cdot P(Humidity \leq 75|Yes)\frac{4}{9} \cdot P(Windy = true|Yes)\frac{3}{9} = 0.0164$

$P(No|\mathbf{X}) \propto P(No)\frac{5}{14} \cdot P(Outlook = sunny|No)\frac{3}{5} \cdot P(Temp \leq 75|No)\frac{3}{5} \cdot P(Humidity \leq 75|No)\frac{1}{5} \cdot P(Windy = true|No)\frac{3}{5} = 0.0154$

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

$$\mathbf{X} = \{Outlook = rain, Temp = 87, Humidity = 90, Windy = false\} \tag{3}$$

The class-conditional probability for attribute X_i is:

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

where the sample mean μ_{ij} and variance σ_{ij} can be estimated from the training records that belong to class y_j.

Fill out the following tables and use these for looking up the probabilities/distribution parameters when classifying the new instance.

	Class prior	$P(Outlook = x_i \vert Y)$			$P(Windy = x_i \vert Y)$	
	$P(Y)$	$x_i = sunny$	$x_i = overcast$	$x_i = rain$	$x_i = false$	$x_i = true$
$Y = Yes$	$\frac{9}{14}$	$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{9}$
$Y = No$	$\frac{5}{14}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{3}{5}$

	$X_i = Humidity$		$X_i = Temperature$	
	μ	σ^2	μ	σ^2
$Y = Yes$	78.22	97.69	73	38
$Y = No$	84	92.5	74.6	62.3

$P(Yes|\mathbf{X}) \propto P(Yes)\frac{9}{14} \cdot P(Outlook = rain|Yes)\frac{3}{9} \cdot P(Temp = 87|Yes)0.0049 \cdot P(Humidity = 90|Yes)0.0198 \cdot P(Windy = false|Yes)\frac{6}{9} = 0.00001386$

$P(No|\mathbf{X}) \propto P(No)\frac{5}{14} \cdot P(Outlook = rain|No)\frac{2}{5} \cdot P(Temp = 87|No)0.0147 \cdot P(Humidity = 90|No)0.0341 \cdot P(Windy = false|No)\frac{2}{5} = 0.00002864$

c) Update the conditional probabilities using Laplace smoothing.

$$P(X_i = x_i | Y = y) = \frac{n_c + 1}{n + c} \tag{5}$$

c is the number of classes, so $c = 2$ here.

$P(X_i = x_i \vert Y)$	$X_i = Outlook$			$X_i = Windy$	
	$x_i = sunny$	$x_i = overcast$	$x_i = rain$	$x_i = false$	$x_i = true$
$Y = Yes$	$\frac{2+1}{9+2}$	$\frac{4+1}{9+2}$	$\frac{3+1}{9+2}$	$\frac{6+1}{9+2}$	$\frac{3+1}{9+2}$
$Y = No$	$\frac{3+1}{5+2}$	$\frac{0+1}{5+2}$	$\frac{2+1}{5+2}$	$\frac{2+1}{5+2}$	$\frac{3+1}{5+2}$