Assignment 1

Analyze `weather.nominal.arff` using Weka `3.6.14`

Collaborators: None.

Exercise 1-1. Compute posterior distribution for the class

Use NaiveBayesSimple to find a Bayesian classifier (Laplace estimator = 1 is used in order to avoid frequency problems). Compute P("yes") and P("no") of the following instances.

(a) Windy = True

Solution: Given the evidence Windy = True, with independence assumption and Bayes rule, we have the probability of answer "yes" as:

$$\text{P}(\text{“yes”}|\text{Windy = True}) = \frac{\text{P}(\text{Windy = True}|\text{“yes”}) \times \text{P}(\text{“yes”})}{\text{P}(\text{Windy = True})}$$

Because of the independence assumption between attributes, we have the prior for Windy = True as follow:

$$\begin{aligned}\text{P}(\text{Windy = True}) &= \text{P}(\text{Windy = True}|\text{“yes”}) \times \text{P}(\text{“yes”}) \ + \\ &\quad\ \text{P}(\text{Windy = True}|\text{“no”}) \times \text{P}(\text{“no”}) \\ &= 0.3636 \times 0.625 + 0.5714 \times 0.375 \\ &= 0.4415\end{aligned}$$

We have the posterior distribution for the class given Windy = True:

$$\begin{aligned}\text{P}(\text{“yes”}|\text{Windy = True}) &= \frac{\text{P}(\text{Windy = True}|\text{“yes”}) \times \text{P}(\text{“yes”})}{\text{P}(\text{Windy = True})} \\ &= \frac{0.3636 \times 0.625}{0.4415} \\ &= \mathbf{0.5} \\ \text{P}(\text{“no”}|\text{Windy = True}) &= 1 - \text{P}(\text{“yes”}|\text{Windy = True}) \\ &= \mathbf{0.5}\end{aligned}$$

(b) Humidity = High, Windy = True

Solution: Similar to question 1-(a), we have the join prior probability for the case Humidity = High (abbr. H = High), Windy = True (abbr. W = True) as follow:

$$\begin{aligned}\text{P}(\text{H = High, W = True}) &= \text{P}(\text{H = High, W = True}|\text{“yes”}) \times \text{P}(\text{“yes”}) \ + \\ &\quad\ \text{P}(\text{H = High, W = True}|\text{“no”}) \times \text{P}(\text{“no”})\end{aligned}$$

Apply the attribute independence assumption, we have the following fractorization of the conditional join probability:

$$\begin{aligned} \text{P(H = High, W = True)} &= \text{P(H = High|“yes”)} \times \text{P(W = True|“yes”)} \times \text{P(“yes”)} \; + \\ &\quad \text{P(H = High|“no”)} \times \text{P(W = True|“no”)} \times \text{P(“no”)} \\ &= 0.3636 \times 0.3636 \times 0.625 + 0.5714 \times 0.7142 \times 0.375 \\ &= 0.2357 \end{aligned}$$

We have the posterior distribution for the class given Windy = True and Humidity = High:

$$\begin{aligned} \text{P(“yes”|H = High, W = True)} &= \frac{\text{P(W = True, H = High|“yes”)} \times \text{P(“yes”)}}{\text{P(H = High, W = True)}} \\ &= \frac{0.3636 \times 0.3636 \times 0.625}{0.2357} \\ &= \mathbf{0.35} \\ \text{P(“no”|H = High, W = True)} &= 1 - \text{P(“yes”|H = High, Windy = True)} \\ &= \mathbf{0.65} \end{aligned}$$

(c) Temperature = Hot, Humidity = High, Windy = True

Solution: Similar to question 1-(a,b), we have the join prior probability for the case Temperature = Hot (abbr. T = Hot) Humidity = High (abbr. H = High), Windy = True (abbr. W = True) as follow:

$$\begin{aligned} \text{P(T = Hot, H = High, W = True)} &= \text{P(T = Hot, H = High, W = True|“yes”)} \times \text{P(“yes”)} \\ &\quad + \; \text{P(T = Hot, H = High, W = True|“no”)} \times \text{P(“no”)} \end{aligned}$$

Apply the attribute independence assumption, we have the following fractorization of the conditional join probability:

$$\begin{aligned} \text{P(T = Hot, H = High, W = True)} &= \text{P(T = Hot|“yes”)} \times \text{P(H = High|“yes”)} \times \\ &\quad \text{P(W = True|“yes”)} \times \text{P(“yes”)} \; + \\ &\quad \text{P(T = Hot|“no”)} \times \text{P(H = High|“no”)} \times \\ &\quad \text{P(W = True|“no”)P(“no”)} \\ &= 0.25 \times 0.3636 \times 0.3636 \times 0.625 + 0.375 \times 0.5714 \times 0.7142 \times 0.375 \\ &= 0.078 \end{aligned}$$

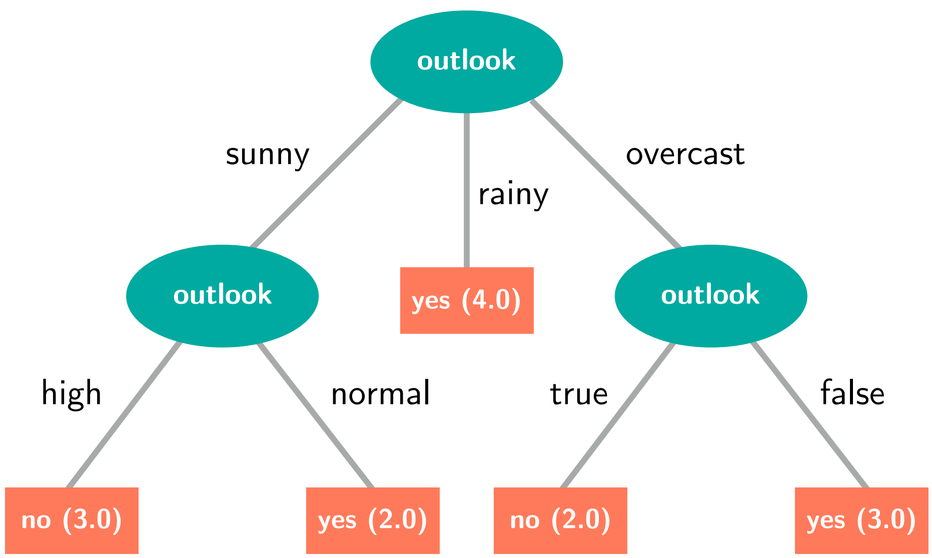
We have the posterior distribution for the class given Temperature = Hot, Windy =

True, and Humidity = High:

$$\begin{aligned}
\text{P(“yes”|T = Hot, H = High, W = True)} &= \frac{\text{P(T = Hot, W = True, H = High|“yes”)} \times \text{P(“yes”)}}{\text{P(T = Hot, H = High, W = True)}} \\
&= \frac{0.25 \times 0.3636 \times 0.3636 \times 0.625}{0.078} \\
&= \mathbf{0.26} \\
\text{P(“no”|T = Hot, H = High, W = True)} &= 1 - \text{P(“yes”|T = Hot, H = High, Windy = True)} \\
&= \mathbf{0.74}
\end{aligned}$$

Exercise 1-2. J48 decision tree

(a) Use J48 to find a decision tree.
Solution: The decision tree found by J48 with defaut setting is:



(b) Check its accuracy to the original data by hand.
Solution: The accuracy of the tree over the training data is **100%**. This result is due to the tree is created by using full dataset. However, the accuracy of the tree is only 50% if checked with 10-fold cross-validation.

(c) Why is “`outlook`” selected as its root attribute? Compare with other attribute (temperature, humidity and windy) and answer quantitatively.

Solution: "`outlook`" is selected as its root attribute because its information gain is the largest out of all attributes. We can compute the expected information of each attribute as follow:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{I}(\text{Outlook})\right] &= \text{info}([2,3],[4,0],[3,2]) = 0.693 \\
\mathbb{E}\left[\mathcal{I}(\text{Temperature})\right] &= \text{info}([3,1],[2,2],[4,2]) = 0.911 \\
\mathbb{E}\left[\mathcal{I}(\text{Humidity})\right] &= \text{info}([6,1],[3,4]) = 0.788 \\
\mathbb{E}\left[\mathcal{I}(\text{Windy})\right] &= \text{info}([6,2],[3,3]) = 0.892
\end{aligned}$$

Infomation gain for each attribute:

$$\begin{aligned}
\mathbb{G}(\text{Outlook}) &= \text{info}([9,5]) - 0.693 = 0.247 \\
\mathbb{G}(\text{Temperature}) &= \text{info}([9,5]) - 0.911 = 0.029 \\
\mathbb{G}(\text{Humidity}) &= \text{info}([9,5]) - 0.788 = 0.152 \\
\mathbb{G}(\text{Windy}) &= \text{info}([9,5]) - 0.892 = 0.048
\end{aligned}$$

Attribute **outlook** gives the largest information gain. Therefore it is selected as the root attribute.

Exercise 1-3. Use Prism to find rules

(a) Show all rules.

Solution: The rules derived from Prism algorithm:

```
=== Classifier model (full training set)
Prism rules
-----------
If outlook = overcast then yes
If humidity = normal
and windy = FALSE then yes
If temperature = mild
and humidity = normal then yes
If outlook = rainy
and windy = FALSE then yes
If outlook = sunny
and humidity = high then no
if outlook = rainy
and windy = TRUE then no
Time taken to build model:  0.01 seconds
```

(b) Check the accuracy (confidence) and the converage (support) of each rule to the original data by hand.

Solution:

- `If outlook = overcast then yes`: Support = 4; accuracy = 100%.

- `If humidity = normal and windy = FALSE then yes`: Support = 4; accuracy = 100%.
- `If temperature = mild and humidity = normal then yes`: Support = 2; accuracy = 100%.
- `If outlook = rainy and windy = FALSE then yes`: Support = 3; accuracy = 100%.
- `If outlook = sunny and humidity = high then o`: Support = 3; accuracy = 100%.
- `If outlook = rainy and windy = TRUE then no`: Support = 2; accuracy = 100%.

(c) Compare (i) the above rules and (ii) rules coverted from the above decision tree. Is there any instance that is not covered by the rules (i) or rules (ii)?

Solution: There is a slight difference between (i) and (ii) is there is two rules in set (i) are replaced by a single rule in set (i). `If humidity = normal and windy = FALSE then yes`
`If temperature = mild and humidity = normal then yes`
is replaced by: `If outlook = sunny and humidity = normal then yes`

Checking by hand we can see that both rule sets covered the dataset. There is no instance that is not covered.