

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 $P(\text{"yes"})$ and $P(\text{"no"})$ 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:

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

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

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

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

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

(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:

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

Apply the attribute independence assumption, we have the following factorization of the conditional joint probability:

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

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

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

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

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

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

Apply the attribute independence assumption, we have the following factorization of the conditional joint probability:

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

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

True, and Humidity = High:

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

Exercise 1-2. J48 decision tree

(a) Use J48 to find a decision tree.

Solution: The decision tree found by J48 with default 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:

$$E [I(\text{Outlook})] = \text{info}([2, 3], [4, 0], [3, 2]) = 0.693$$

$$E [I(\text{Temperature})] = \text{info}([3, 1], [2, 2], [4, 2]) = 0.911$$

$$E [I(\text{Humidity})] = \text{info}([6, 1], [3, 4]) = 0.788$$

$$E [I(\text{Windy})] = \text{info}([6, 2], [3, 3]) = 0.892$$

Information gain for each attribute:

$$G(\text{Outlook}) = \text{info}([9, 5]) - 0.693 = 0.247$$

$$G(\text{Temperature}) = \text{info}([9, 5]) - 0.911 = 0.029$$

$$G(\text{Humidity}) = \text{info}([9, 5]) - 0.788 = 0.152$$

$$G(\text{Windy}) = \text{info}([9, 5]) - 0.892 = 0.048$$

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 coverage (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 no: 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 converted 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 (ii). 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.