

# Practice of Bayesian Networks

Data Mining Lab 4

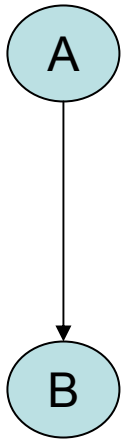
# Bayesian Networks.

## Two tasks

- Infer the structure of the network from the data (in practice, the structure of the network is identified by data experts, not by a machine)
- Fill in conditional probabilities tables

# The elementary inference types in Bayesian nets

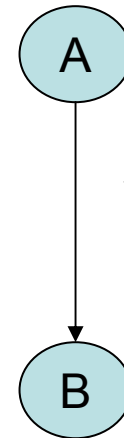
Causal



Increased probability of A makes B more likely.

A can cause B

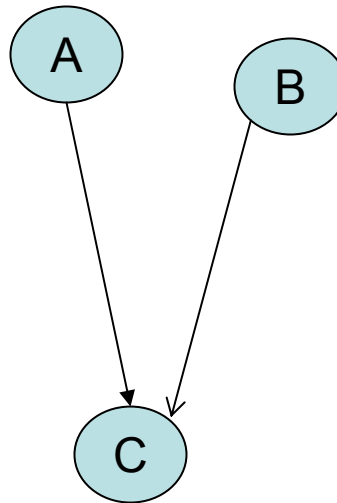
Diagnostic



Increased probability of B makes A more likely.

B is evidence for A

Intercausal

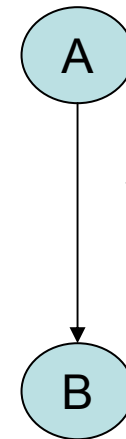


A and B can each cause C. B explains C and so is evidence against A

# Bayesian Net for Weather Data

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Diagnostic



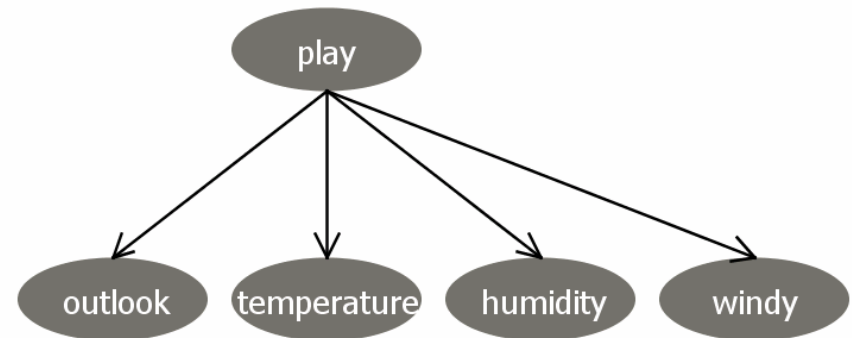
Increased probability of B makes A more likely.

B is evidence for A

# Naïve Bayes

Assuming the attributes are independent of each other, we have a Naïve Bayesian Network:

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



$$P(\text{play}=\text{yes})=9/14,$$

with Laplace correction:

$$P(\text{play}=\text{yes})=9+1/14+2=0.625$$

In general, to make Laplace correction, we add an initial count (1) to the total of all instances with a given attribute value, and we add the number of distinct values of the same attribute to the total number of instances in the group.

# Naïve Bayes

In order to fill the Conditional Probability Tables we compute conditional probabilities for each node in form:  $\Pr(\text{attribute}=\text{value} \mid \text{parents values})$  for each combinations of attributes values in parent nodes

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

$$P(\text{outlook}=\text{sunny} \mid \text{play}=\text{yes}) = (2+1)/(9+3) = 3/12$$

$$P(\text{outlook}=\text{rainy} \mid \text{play}=\text{yes}) = (3+1)/(9+3) = 4/12$$

$$P(\text{outlook}=\text{overcast} \mid \text{play}=\text{yes}) = (4+1)/(9+3) = 5/12$$

Sum is 1

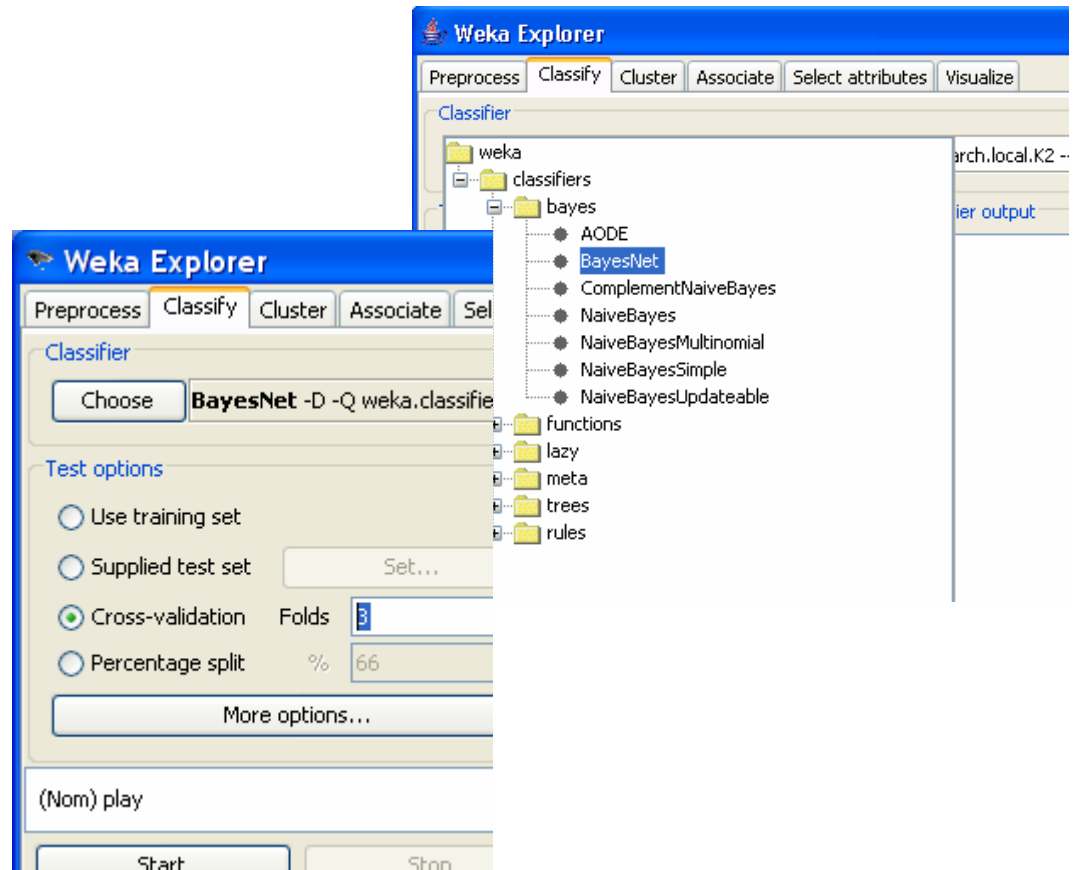
$$P(\text{outlook}=\text{sunny} \mid \text{play}=\text{yes}) = (2+1)/(9+3) = 3/12$$

$$P(\text{outlook}=\text{sunny} \mid \text{play}=\text{no}) = (3+1)/(5+3) = 4/8$$

Sum is NOT 1

# WEKA Exercise 1. Bayesian network for weather data with default parameters

- **Preprocess tab**
  - **Open file**  
*weather.nominal.arff*
  - Perform **filters** (if needed): discretize or replace missing values
- **Classify tab**
  - **Classifier->choose->classifiers->bayes->BayesNet**
  - Click on row with selected classifier, change Laplace correction (initial count) to 1 (instead of 0.5) in the **option row for the estimator**
  - **Cross-validation** change to **3 folds** (since we have only 14 instances, with 10 folds cross validation we will have test groups of size less than 2, which makes the classifier less reliable). Press **Start**



# WEKA Exercise 2. Examining the output

- In the history box, right-click and choose **visualize graph**. Check that probabilities in CPT correspond to what we calculated before (clicking on the graph node brings table of conditional probabilities)
- Naïve Bayes? Study parameters of the program. Click on **choose** line again.
- Save this model in file weather.xml for later use
- Click on **searchAlgorithm** row. Default parameters are:  
***initAsNaiveBayes=true***  
***maxNrOfParents=1***
- Change ***maxNrOfParents=2***. Run. Visualize graph
- Change to ***initAsNaiveBayes=false***. Run. Visualize graph. Change back to true.

The image shows two screenshots from the WEKA software interface. The top screenshot is a window titled "Probability Distribution Table For outlook". It displays a table of conditional probabilities for the "play" attribute based on the "outlook" attribute. The table has columns for "play" (sunny, overcast, rainy) and rows for "yes" and "no".

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

The bottom screenshot is a window titled "Weka Classifier Graph Visualizer: 18:24:24 - bayes.Bay...". It shows a directed acyclic graph (DAG) with a root node "play" and four child nodes: "outlook", "temperature", "humidity", and "windy".

Annotations in the image include:

- "Possible values of the parent node attribute" pointing to the "play" column header.
- "Possible values of the outlook attribute" pointing to the "sunny", "overcast", and "rainy" column headers.
- "Conditional probabilities" pointing to the numerical values in the table cells.
- "Click on the node to see probability tables" pointing to the "play" node in the graph.



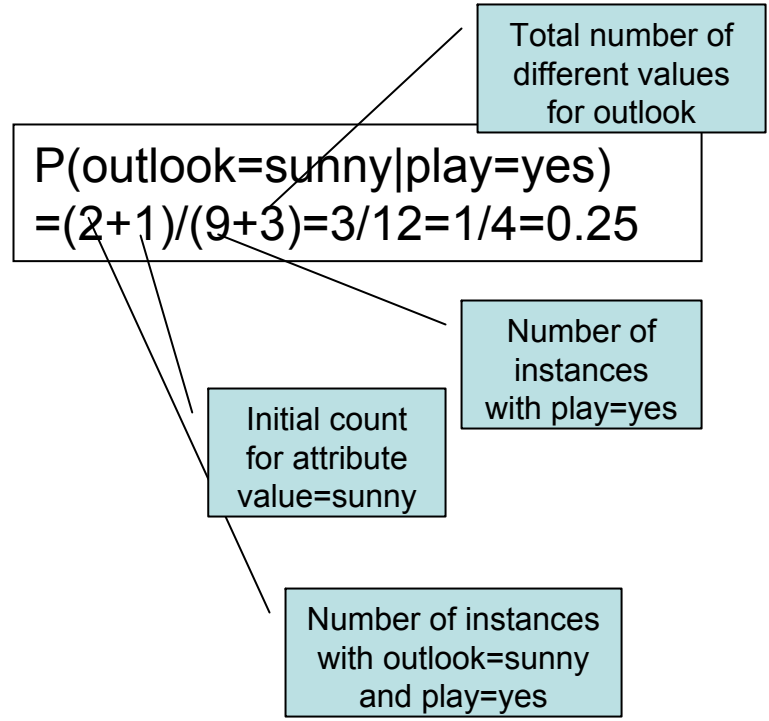
# Conditional Probabilities Tables in WEKA

- After the structure is learned, the CPT for each node are computed.
- Simple estimator computes the relative frequencies of the associated combinations of the attribute values in the training data (just like we do in our exercises).

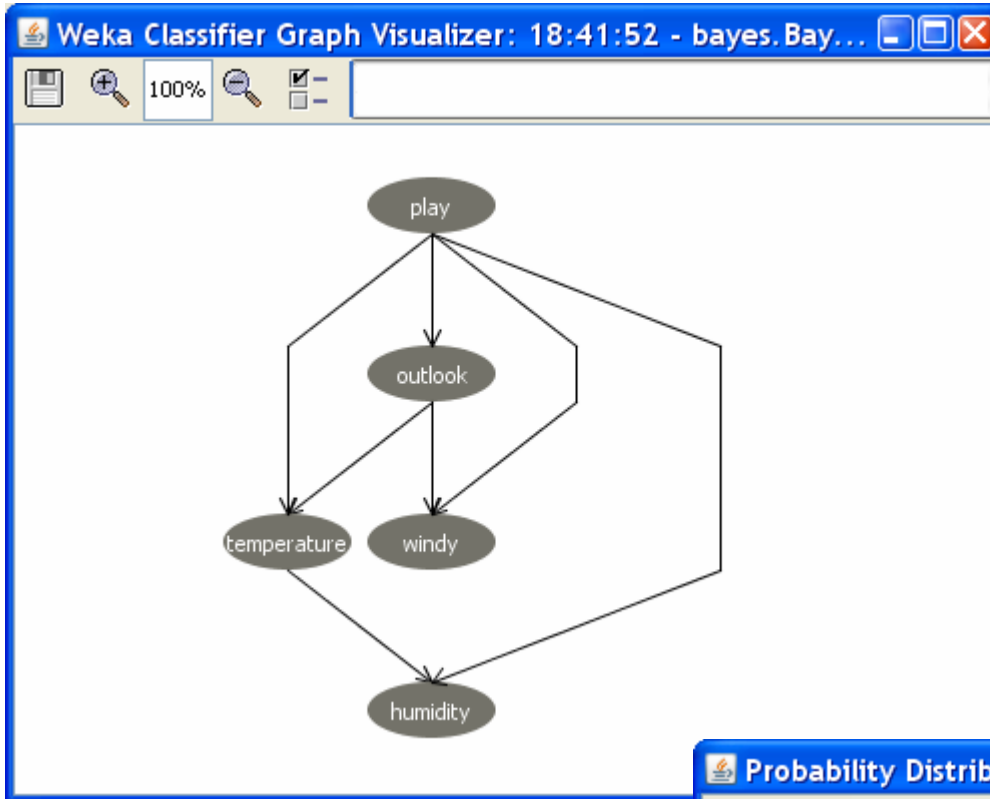
# How it was computed

play	sunny	overcast	rainy
yes	0.25	0.417	0.333
no	0.5	0.125	0.375

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



# More complex Bayesian Network for weather data (with *maxNrOfParents=2*)



Possible values of the attribute temperature

All combinations of values for 2 parent nodes: play and outlook

play	outlook	hot	mild	cool	
yes	sunny		0.2	0.4	0.4
yes	overcast		0.429	0.286	0.286
yes	rainy		0.167	0.5	0.333
no	sunny		0.5	0.333	0.167
no	overcast		0.333	0.333	0.333
no	rainy		0.2	0.4	0.4

Conditional probabilities

# How it was computed

Outlook	Temp.	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

play	outlook	hot	mild	cool
yes	sunny	0.2	0.4	0.4
yes	overcast	0.429	0.286	0.286
yes	rainy	0.167	0.5	0.333
no	sunny	0.5	0.333	0.167
no	overcast	0.333	0.333	0.333
no	rainy	0.2	0.4	0.4

$$P(\text{temperature}=\text{hot}|\text{play}=\text{yes}, \text{outlook}=\text{sunny}) \\ = (0+1)/(2+3) = 1/5 = 0.2$$

Number of instances with play=yes and outlook=sunny

Number of instances with temperature=hot, outlook=sunny and play=yes

# How WEKA infers a structure of the network

- The nodes correspond to the attributes
- Learning the structure is to find edges
  - Searching through the possible edges sets
  - For each set estimate the conditional probability tables from the data
  - Estimate quality of the network as the probability of obtaining the set of data given this network

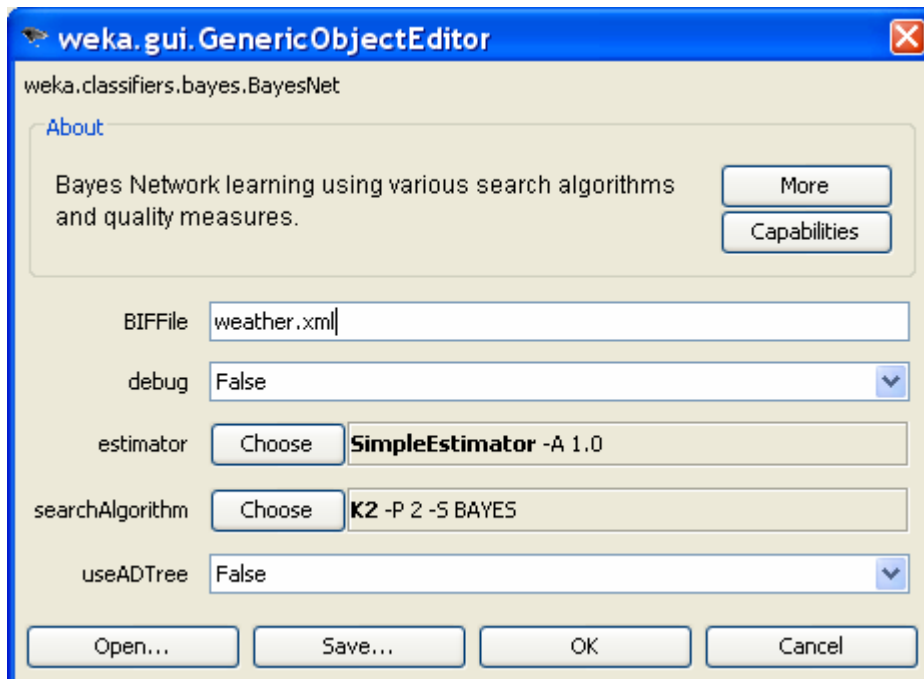
# WEKA Search Algorithms.

## Example

- By default: K2.
- Starts with a given ordering of attributes.
- Adds one node in order and considers adding edges from each previously added node to a new node.
- Then it adds the edge which maximizes the network score.
- The number of parents is restricted to a predefined maximum.
- *The Markov blanket* of a node includes all its parents, children and children parents. It is proven, that a given node is conditionally dependent only on nodes in its Markov blanket. So the edge is added from the class node to the node which is not in its Markov blanket. Otherwise the value of this attribute would be irrelevant for the class.

# WEKA Exercise 3. Improving the network supplied as a file

- Bring in the window of **classifier's options**
  - Type in the **BIFF file** box: *weather.xml*
- **Run**
- In the output window find the comparison between two networks:
  - the supplied and inferred by machine learning.



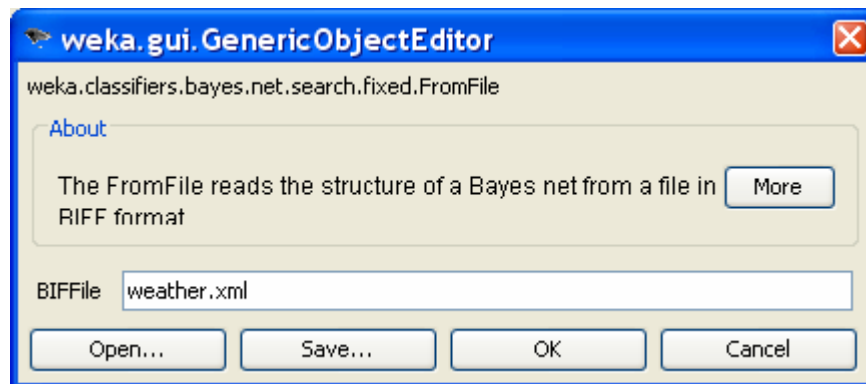
```
LogScore BDeu: -158.45430601513422
LogScore MDL: -135.38725699825952
LogScore ENTROPY: -97.12092571883828
LogScore AIC: -126.12092571883827
Missing: 0 Extra: 3 Reversed: 0
Divergence: -0.1762693031351526
```

Time taken to build model: 0.02 seconds

```
=== Stratified cross-validation ===
=== Summary ===
```

# WEKA Exercise 4. Structure supplied by the user

- Bring in the window of *classifier's options*
- In *searchAlgorithm* row press **Choose** button
- Choose search->fixed->FromFile. OK
- Press *searchAlgorithm* row to define parameters
- Type in the **BIFF file** box: *weather.xml* (Do NOT use the button Open...)
- **Run**
- Check that WEKA has produced the Naïve Bayes, as it was supplied in your file

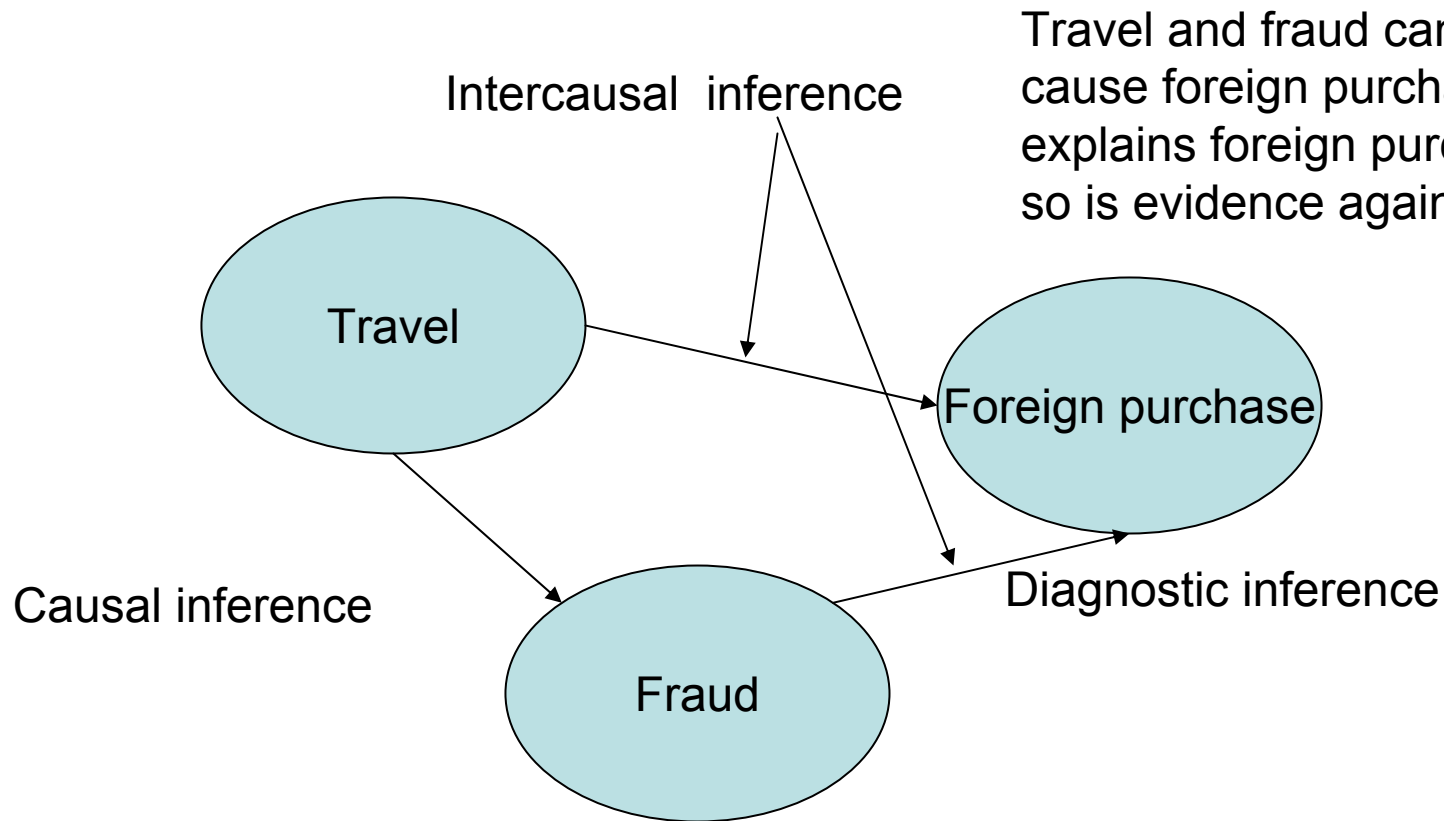




# Tutorial Exercise 1. Build Bayesian Network

- Suppose you are working for a financial institution and you are asked to build a fraud detection system. You plan to use the following information:
  - When the card holder is traveling abroad, fraudulent transactions are more likely since tourists are prime targets for thieves. More precisely, 1% of transactions are fraudulent when the card holder is traveling, whereas only 0.2% of the transactions are fraudulent when he is not traveling. On average, 5% of all transactions happen while card holder is traveling. If a transaction is fraudulent, then the likelihood of a foreign purchase increases, unless the card holder happens to be traveling. More precisely, when the card holder is not traveling, 10% of the fraudulent transactions are foreign purchases, whereas only 1% of the legitimate transactions are foreign purchases. On the other hand, when the card holder is traveling, 90% of the transactions are foreign purchases regardless of the legitimacy of the transactions.

# Network Structure



Travel and fraud can each cause foreign purchase. Travel explains foreign purchase and so is evidence against fraud

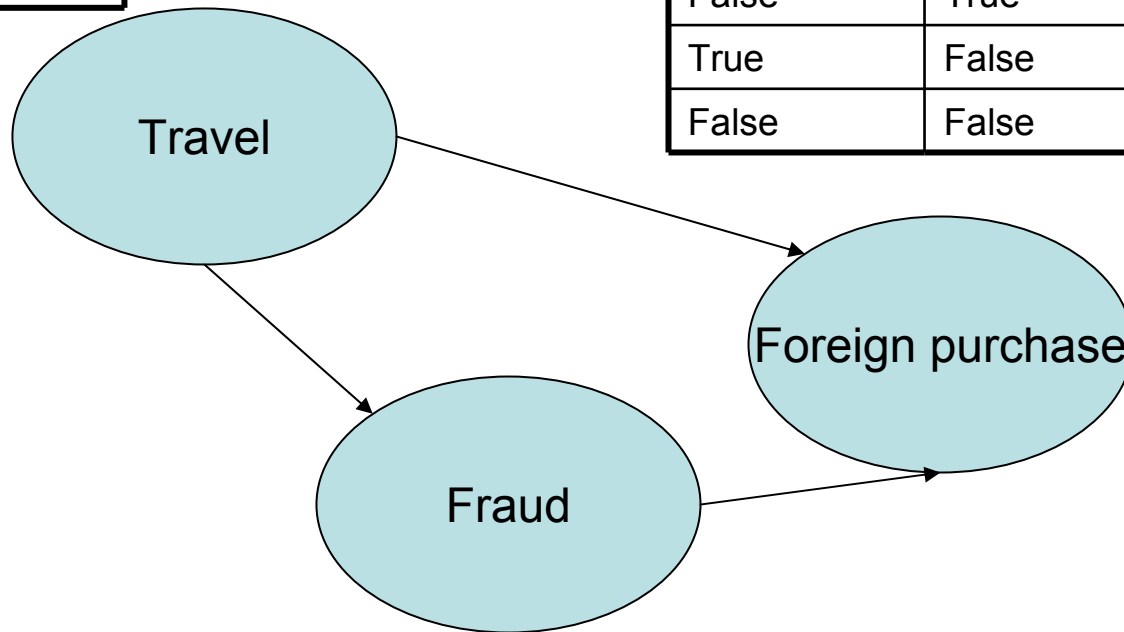
Increased probability of travel makes fraud more likely. Travel can cause fraud.

Increased probability of foreign purchase makes fraud more likely. Foreign purchase is evidence for fraud.

# Conditional probabilities

True	False
0.05	0.95

Travel	Fraud	True	False
True	True	0.90	0.10
False	True	0.10	0.90
True	False	0.90	0.10
False	False	0.01	0.99



Travel	True	False
True	0.01	0.99
False	0.002	0.998

# Tutorial Exercise 2.

## Classify with hidden variables

- System has detected the foreign purchase. What is the probability of a fraud if we don't know whether the card holder is traveling or not?
- It is equivalent to: classify with hidden variables  
**[travel=?,foreign\_purchase=true, fraud=?]**

$P(\text{fraud}=\text{true}|\text{foreign-purchase}=\text{true})=$

$\alpha * \sum_{\text{travel}} [P(\text{fraud}=\text{true}|\text{travel}) * P(\text{foreign-purchase} = \text{true} | \text{travel}, \text{fraud}=\text{true})$   
 $* P(\text{travel})]$

$= \alpha * [P(\text{fraud}=\text{true} | \text{travel}=\text{true})$

$* P(\text{foreign-purchase} = \text{true} | \text{travel}=\text{true}, \text{fraud}=\text{true}) * P(\text{travel}=\text{true}) +$

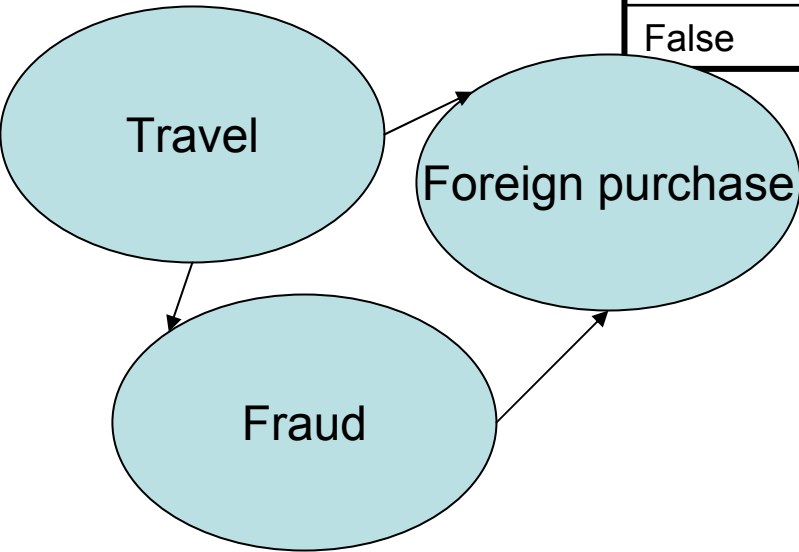
$P(\text{fraud}=\text{true} | \text{travel}=\text{false})$

$* P(\text{foreign-purchase} = \text{true} | \text{travel} = \text{false}, \text{fraud}=\text{true}) * P(\text{travel}=\text{false}) ]$

# Tutorial Exercise 2.

True	False
0.05	0.95

Travel	Fraud	True	False
True	True	0.90	0.10
False	True	0.10	0.90
True	False	0.90	0.10
False	False	0.01	0.99



Travel	True	False
True	0.01	0.99
False	0.002	0.998

$$P(\text{fraud=true}|\text{foreign-purchase=true})=$$

$$= \alpha * [P(\text{fraud=true} | \text{travel=true}) * P(\text{foreign-purchase} = \text{true} | \text{travel=true, fraud=true}) * P(\text{travel=true}) + P(\text{fraud=true} | \text{travel=false}) * P(\text{foreign-purchase} = \text{true} | \text{travel} = \text{false, fraud=true}) * P(\text{travel=false}) ] =$$

$$\alpha * [0.01 * 0.90 * 0.05 + 0.002 * 0.10 * 0.95] = \alpha * [0.00045 + 0.00019] = 0.00064 \alpha$$

$$P(\text{fraud=false} | \text{foreign-purchase=true})=$$

$$= \alpha * [P(\text{fraud=false} | \text{travel=true}) * P(\text{foreign-purchase} = \text{true} | \text{travel=true, fraud=false}) * P(\text{travel=true}) + P(\text{fraud=false} | \text{travel=false}) * P(\text{foreign-purchase} = \text{true} | \text{travel} = \text{false, fraud=false}) * P(\text{travel=false}) ] = \alpha * [0.99 * 0.90 * 0.05 + 0.998 * 0.01 * 0.95] = \alpha * [0.04455 + 0.009481] = 0.054031 \alpha$$

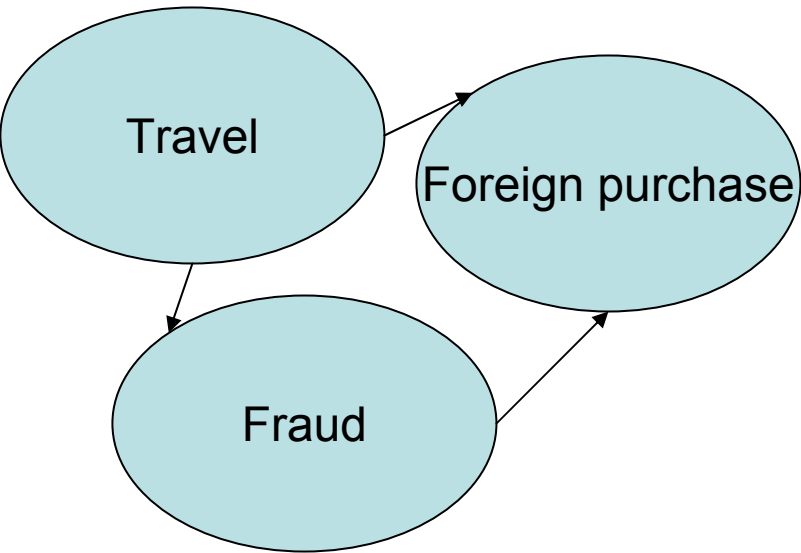
$$\alpha = 1 / (0.00064 + 0.054031)$$

$$P(\text{fraud=true}|\text{foreign-purchase=true})=1.1\%$$

# Tutorial Exercise 3.

## Classify without the hidden variables

True	False
0.05	0.95



Travel	True	False
True	0.01	0.99
False	0.002	0.998

Travel	Fraud	True	False
True	True	0.90	0.10
False	True	0.10	0.90
True	False	0.90	0.10
False	False	0.01	0.99

Suppose that probability more than 1% causes an agent to call the client to confirm the transaction. An agent calls but the card holder is not at home. Her spouse confirms that she is out of town on a business trip. How does the probability of the fraud changes based on this new piece of information?

$$P(\text{fraud}=\text{true}|\text{foreign-purchase}=\text{true}, \text{travel}=\text{true})= \alpha * 0.00045$$

$$P(\text{fraud}=\text{false}|\text{foreign-purchase}=\text{true}, \text{travel}=\text{true}) = \alpha * 0.04455$$

$$\alpha=1/(0.00045+ 0.04455)$$

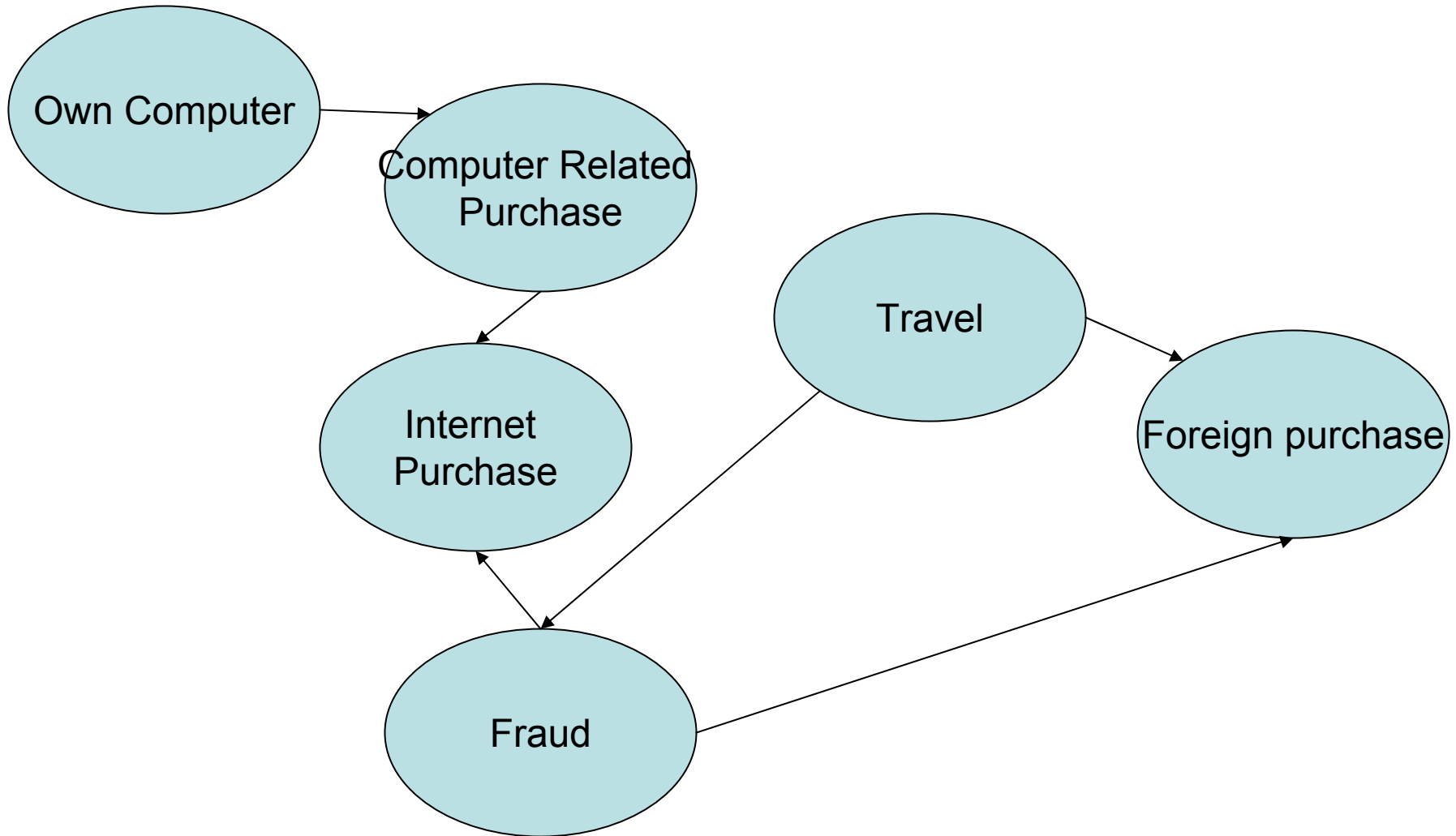
$$P(\text{fraud}=\text{true}|\text{foreign-purchase}=\text{true}, \text{travel}=\text{true})=1.0\%$$

# Tutorial Exercise 4

## (Try it)

- We need to add more information to the network of exercise 1.
- Purchases made over the internet are more likely to be fraudulent. This is especially true for card holders which do not own a computer. Currently, 60% of the population owns a computer and for those cardholders 1% of their legitimate transactions are done over the internet, but this number increases to 1.1% for fraudulent transactions. Unfortunately, the credit company doesn't know whether a cardholder owns a computer, however it can usually guess by verifying whether any of the recent transactions involve the purchase of computer related accessories. In any given week, 10% of those who own a computer purchase with their credit card at least one computer related item as opposed to just 0.01% of those who don't own any computer. Incorporate this information into your system.

# Expanded Bayes net for fraud detection





# Tutorial Exercise 5

## (Try it)

- Suppose the thief just stole a credit card. He knows how the fraud detection system was set up, but he still wants to make an important purchase over the internet. What can he do prior to his internet purchase to reduce the risk that the transaction will be rejected as a possible fraud?