Probabilistic reasoning

Recall

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

$$P(X,Y) = P(X|Y) \times P(Y)$$
$$= P(Y|X) \times P(X)$$

X and Y are dependent

$$P(X|Y) = P(X)$$
$$P(X,Y) = P(X) \times P(Y)$$

X and Y are independent

Conditional Independence

- For two events, X and Y, that are conditionally independent given knowledge of a third events, here Z, the definition of the probability of a joint event and conditional probability are:

$$P(X|Y, Z) =$$
$$P(X, Y|Z) =$$

Conditional Independence

- For two events, X and Y, that are conditionally independent given knowledge of a third events, here Z, the definition of the probability of a joint event and conditional probability are:

$$P(X|Y,Z) = P(X|Z)$$
$$P(X,Y|Z) = P(X|Z) \times P(Y|Z)$$

Without conditional independence

$$P(X, Y, Z, W) = P(X|W) \times P(Y|X, W) \times P(Z|Y, X, W) \times P(W)$$

Factorization

Without conditional independence

$$P(X, Y, Z, W) = P(X|W) \times P(Y|X, W) \times P(Z|Y, X, W) \times P(W)$$

With conditional independence

$$P(X, Y, Z, W) = \underbrace{P(X|W)}_{Factor1} \times \underbrace{P(Y|W)}_{Factor2} \times \underbrace{P(Z|W)}_{Factor3} \times \underbrace{P(W)}_{Factor4}$$

Conditional Independence and chain rule

- If the event $t = l$ causes the events $\mathbf{q}[1], \ldots, \mathbf{q}[m]$ to happen then the events $\mathbf{q}[1], \ldots, \mathbf{q}[m]$ are conditionally independent of each other given knowledge of $t = l$ and the chain rule definition can be simplified as follows:

$$
\begin{aligned}
P(\mathbf{q}[1], \ldots, \mathbf{q}[m] \mid t = l) \\
&= P(\mathbf{q}[1] \mid t = l) \times P(\mathbf{q}[2] \mid t = l) \times \cdots \times P(\mathbf{q}[m] \mid t = l) \\
&= \prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l)
\end{aligned}
$$

Conditional Independence and Bayes

- Using this we can simplify the calculations in Bayes' Theorem, under the assumption of conditional independence between the descriptive features given the level l of the target feature:

$$P(t = l \mid \mathbf{q}[1], \ldots, \mathbf{q}[m]) = \frac{\left(\prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l) \right) \times P(t = l)}{P(\mathbf{q}[1], \ldots, \mathbf{q}[m])}$$

Conditional Independence and Factorization

- Assuming the descriptive features are conditionally independent of each other given MENINGITIS we only need to store four factors:

$$P(H, F, V, M) = P(M) \times P(H|M) \times P(F|M) \times P(V|M)$$

Conditional Independence and Factorization

- Assuming the descriptive features are conditionally independent of each other given MENINGITIS we only need to store four factors:

$$Factor_1 : < P(M) >$$
$$Factor_2 : < P(h|m), P(h|\neg m) >$$
$$Factor_3 : < P(f|m), P(f|\neg m) >$$
$$Factor_4 : < P(v|m), P(v|\neg m) >$$

$$P(H, F, V, M) = P(M) \times P(H|M) \times P(F|M) \times P(V|M)$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

- Calculate the factors from the data.

$$Factor_1 : < P(M) >$$

$$Factor_2 : < P(h|m), P(h|\neg m) >$$

$$Factor_3 : < P(f|m), P(f|\neg m) >$$

$$Factor_4 : < P(v|m), P(v|\neg m) >$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

$$Factor_1 : < P(m) = 0.3 >$$

$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$

$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$

$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

ID	HEADACHE	FEVER	VOMITING	MENINGITIS
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

$Factor_1 : < P(m) = 0.3 >$

$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$

$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$

$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$

HEADACHE	FEVER	VOMITING	MENINGITIS
true	true	false	?

Conditional Independence and Factorization

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

$$P(m|h, f, \neg v) = \frac{P(h|m) \times P(f|m) \times P(\neg v|m) \times P(m)}{\sum_i}$$

Conditional Independence and Factorization

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

$$P(m|h, f, \neg v) = \frac{P(h|m) \times P(f|m) \times P(\neg v|m) \times P(m)}{\sum_i P(h|M_i) \times P(f|M_i) \times P(\neg v|M_i) \times P(M_i)}$$

Conditional Independence and Factorization

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

$$P(m|h, f, \neg v) = \frac{P(h|m) \times P(f|m) \times P(\neg v|m) \times P(m)}{\sum_i P(h|M_i) \times P(f|M_i) \times P(\neg v|M_i) \times P(M_i)}$$

$$\frac{0.6666 \times 0.3333 \times 0.3333 \times 0.3}{(0.6666 \times 0.3333 \times 0.3333 \times 0.3) + (0.7143 \times 0.4286 \times 0.4286 \times 0.7)} = 0.1948$$

Conditional Independence and Factorization

$$P(m|h, f, \neg v) = 0.1948$$

$$P(\neg m|h, f, \neg v) =$$

Conditional Independence and Factorization

$$P(m|h, f, \neg v) = 0.1948$$

$$P(\neg m|h, f, \neg v) = 0.8052$$

The Naive Bayes' Classifier

Table: A dataset from a loan application fraud detection domain.

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

Naive Bayes Classifier

- Naive Bayes classifiers are based on applying Bayes' theorem with strong (naive) independence assumptions between the features.
- The independence assumptions are often wrong. However, the naive Bayes classifier has several properties that make it useful in practice.
- Naive Bayes models are known under a variety of names, including simple Bayes and independence Bayes.

Naive Bayes Classifier

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

$$P(t \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid t) \times P(t)}{P(\mathbf{x})}$$

Naive Bayes Classifier

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

$$P(t \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid t) \times P(t)}{P(\mathbf{x})}$$

$$P(t = l \mid x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n \mid t = l) \times P(t = l)}{P(x)}$$

Naive Bayes Classifier

$$\text{posterior} \;=\; \frac{\text{likelihood} \times \;\text{prior}}{\text{evidence}}$$

$$P(t \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid t) \times P(t)}{P(\mathbf{x})}$$

$$P(t = l \mid x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n \mid t = l) \times P(t = l)}{P(x)}$$

$$P(t = l \mid x_1, x_2, \ldots, x_n) \;\propto\; \underline{P(x_1, x_2, \ldots, x_n, l)}$$

Here, $\propto$ denotes proportionality since we omited denominator $P(\mathbf{x})$.

Consider the **chain rule**

$$
\begin{aligned}
& P\left(x_1, \ldots, x_n, l\right) \\
& =P\left(x_1 \mid x_2, \ldots, x_n, l\right) P\left(x_2 \mid x_3, \ldots, x_n, l\right) \cdots P\left(x_{n-1} \mid x_n, l\right) P\left(x_n \mid l\right) P(l)
\end{aligned}
$$

Consider the **chain rule**

$$\begin{aligned}
&P\left(x_1, \ldots, x_n, l\right) \\
&= P\left(x_1 \mid x_2, \ldots, x_n, l\right) P\left(x_2 \mid x_3, \ldots, x_n, l\right) \cdots P\left(x_{n-1} \mid x_n, l\right) P\left(x_n \mid l\right) P(l)
\end{aligned}$$

Naive conditional independence assumptions: all features in $\mathbf{x}$ are mutually independent, conditional on the category l. Under this assumption,

$$P\left(x_i \mid x_{i+1}, \ldots, x_n, l\right) = P\left(x_i \mid l\right).$$

Consider the **chain rule**

$$
\begin{aligned}
&P(x_1, \dots, x_n, l) \\
&= P(x_1 \mid x_2, \dots, x_n, l)\, P(x_2 \mid x_3, \dots, x_n, l) \cdots P(x_{n-1} \mid x_n, l)\, P(x_n \mid l)\, P(l)
\end{aligned}
$$

Naive conditional independence assumptions: all features in $\mathbf{x}$ are mutually independent, conditional on the category l. Under this assumption,

$$P(x_i \mid x_{i+1}, \dots, x_n, l) = P(x_i \mid l) \,.$$

Thus, the joint model can be expressed as

$$P(l \mid x_1, \dots, x_n) \propto \; P(l, x_1, \dots, x_n)$$

Consider the **chain rule**

$$
\begin{aligned}
& P(x_1, \ldots, x_n, l) \\
& = P(x_1 \mid x_2, \ldots, x_n, l)\, P(x_2 \mid x_3, \ldots, x_n, l) \cdots P(x_{n-1} \mid x_n, l)\, P(x_n \mid l)\, P(l)
\end{aligned}
$$

Naive conditional independence assumptions: all features in $\mathbf{x}$ are mutually independent, conditional on the category l. Under this assumption,

$$
P(x_i \mid x_{i+1}, \ldots, x_n, l) = P(x_i \mid l).
$$

Thus, the joint model can be expressed as

$$
\begin{aligned}
P(l \mid x_1, \ldots, x_n) \propto\ & P(l, x_1, \ldots, x_n) \\
& = P(l)\, P(x_1 \mid l)\, P(x_2 \mid l)\, P(x_3 \mid l) \ldots \\
& = P(l) \prod_{i=1}^{n} P(x_i \mid l),
\end{aligned}
$$

Consider the **chain rule**

$$
\begin{aligned}
&P(x_1, \ldots, x_n, l) \\
&= P(x_1 \mid x_2, \ldots, x_n, l)\, P(x_2 \mid x_3, \ldots, x_n, l) \cdots P(x_{n-1} \mid x_n, l)\, P(x_n \mid l)\, P(l)
\end{aligned}
$$

Naive conditional independence assumptions: all features in $\mathbf{x}$ are mutually independent, conditional on the category l. Under this assumption,

$$
P(x_i \mid x_{i+1}, \ldots, x_n, l) = P(x_i \mid l).
$$

Thus, the joint model can be expressed as

$$
\begin{aligned}
P(l \mid x_1, \ldots, x_n) \propto\ & P(l, x_1, \ldots, x_n) \\
&= P(l)\, P(x_1 \mid l)\, P(x_2 \mid l)\, P(x_3 \mid l) \ldots \\
&= P(l) \prod_{i=1}^{n} P(x_i \mid l),
\end{aligned}
$$

How to define a classifier using above?

$$\text{Naive Bayes classifier}(\mathbf{x}) = \underset{l \in \text{ labels } (t)}{\operatorname{argmax}} \left(\prod_{i=1}^{n} P(\mathbf{x}_i \mid t = l) \right) \times P(t = l)$$

assigns a class label for the max value of probability

$$\text{Naive Bayes classifier}(\mathbf{x}) = \underset{l \in \text{ labels } (t)}{\operatorname{argmax}} \left(\prod_{i=1}^{n} P(\mathbf{x}_i \mid t = l) \right) \times P(t = l)$$

assigns a class label for the max value of probability

Naive Bayes' is simple to train!

1. calculate the priors for each of the target levels
2. calculate the conditional probabilities for each feature given each target level.

Table: A dataset from a loan application fraud detection domain.

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

$P(fr)$	$=$	$P(\neg fr)$	$=$
$P(\mathrm{CH} = \textit{'none'} \mid fr)$	$=$	$P(\mathrm{CH} = \textit{'none'} \mid \neg fr)$	$=$
$P(\mathrm{CH} = \textit{'paid'} \mid fr)$	$=$	$P(\mathrm{CH} = \textit{'paid'} \mid \neg fr)$	$=$
$P(\mathrm{CH} = \textit{'current'} \mid fr)$	$=$	$P(\mathrm{CH} = \textit{'current'} \mid \neg fr)$	$=$
$P(\mathrm{CH} = \textit{'arrears'} \mid fr)$	$=$	$P(\mathrm{CH} = \textit{'arrears'} \mid \neg fr)$	$=$
$P(\mathrm{GC} = \textit{'none'} \mid fr)$	$=$	$P(\mathrm{GC} = \textit{'none'} \mid \neg fr)$	$=$
$P(\mathrm{GC} = \textit{'guarantor'} \mid fr)$	$=$	$P(\mathrm{GC} = \textit{'guarantor'} \mid \neg fr)$	$=$
$P(\mathrm{GC} = \textit{'coapplicant'} \mid fr)$	$=$	$P(\mathrm{GC} = \textit{'coapplicant'} \mid \neg fr)$	$=$
$P(\mathrm{ACC} = \textit{'own'} \mid fr)$	$=$	$P(\mathrm{ACC} = \textit{'own'} \mid \neg fr)$	$=$
$P(\mathrm{ACC} = \textit{'rent'} \mid fr)$	$=$	$P(\mathrm{ACC} = \textit{'rent'} \mid \neg fr)$	$=$
$P(\mathrm{ACC} = \textit{'free'} \mid fr)$	$=$	$P(\mathrm{ACC} = \textit{'free'} \mid \neg fr)$	$=$

Table: The probabilities needed by a Naive Bayes prediction model calculated from the dataset. Notation key: FR=FRAUDULENT, CH=CREDIT HISTORY, GC = GUARANTOR/COAPPLICANT, ACC = ACCOMODATION, T=*'true'*, F=*'false'*.

Table: A dataset from a loan application fraud detection domain.

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	
$P(\text{CH} = \textit{'none'} \mid fr)$	$=$		$P(\text{CH} = \textit{'none'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'paid'} \mid fr)$	$=$		$P(\text{CH} = \textit{'paid'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'current'} \mid fr)$	$=$		$P(\text{CH} = \textit{'current'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'arrears'} \mid fr)$	$=$		$P(\text{CH} = \textit{'arrears'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'none'} \mid fr)$	$=$		$P(\text{GC} = \textit{'none'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'guarantor'} \mid fr)$	$=$		$P(\text{GC} = \textit{'guarantor'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'coapplicant'} \mid fr)$	$=$		$P(\text{GC} = \textit{'coapplicant'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'own'} \mid fr)$	$=$		$P(\text{ACC} = \textit{'own'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'rent'} \mid fr)$	$=$		$P(\text{ACC} = \textit{'rent'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'free'} \mid fr)$	$=$		$P(\text{ACC} = \textit{'free'} \mid \neg fr)$	$=$	

Table: The probabilities needed by a Naive Bayes prediction model calculated from the dataset. Notation key: FR=FRAUDULENT, CH=CREDIT HISTORY, GC = GUARANTOR/COAPPLICANT, ACC = ACCOMODATION, T=*'true'*, F=*'false'*.

Table: A dataset from a loan application fraud detection domain.

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \textit{'none'} \mid fr)$	$=$		$P(\text{CH} = \textit{'none'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'paid'} \mid fr)$	$=$		$P(\text{CH} = \textit{'paid'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'current'} \mid fr)$	$=$		$P(\text{CH} = \textit{'current'} \mid \neg fr)$	$=$	
$P(\text{CH} = \textit{'arrears'} \mid fr)$	$=$		$P(\text{CH} = \textit{'arrears'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'none'} \mid fr)$	$=$		$P(\text{GC} = \textit{'none'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'guarantor'} \mid fr)$	$=$		$P(\text{GC} = \textit{'guarantor'} \mid \neg fr)$	$=$	
$P(\text{GC} = \textit{'coapplicant'} \mid fr)$	$=$		$P(\text{GC} = \textit{'coapplicant'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'own'} \mid fr)$	$=$	0.6666	$P(\text{ACC} = \textit{'own'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'rent'} \mid fr)$	$=$		$P(\text{ACC} = \textit{'rent'} \mid \neg fr)$	$=$	
$P(\text{ACC} = \textit{'free'} \mid fr)$	$=$		$P(\text{ACC} = \textit{'free'} \mid \neg fr)$	$=$	

Table: The probabilities needed by a Naive Bayes prediction model calculated from the dataset. Notation key: FR=FRAUDULENT, CH=CREDIT HISTORY, GC = GUARANTOR/COAPPLICANT, ACC = ACCOMODATION, T=*'true'*, F=*'false'*.

$$
\begin{array}{rclrcl}
P(fr) & = & 0.3 & P(\neg fr) & = & 0.7 \\
P(\text{CH} = \textit{'none'} \mid fr) & = & 0.1666 & P(\text{CH} = \textit{'none'} \mid \neg fr) & = & 0 \\
P(\text{CH} = \textit{'paid'} \mid fr) & = & 0.1666 & P(\text{CH} = \textit{'paid'} \mid \neg fr) & = & 0.2857 \\
P(\text{CH} = \textit{'current'} \mid fr) & = & 0.5 & P(\text{CH} = \textit{'current'} \mid \neg fr) & = & 0.2857 \\
P(\text{CH} = \textit{'arrears'} \mid fr) & = & 0.1666 & P(\text{CH} = \textit{'arrears'} \mid \neg fr) & = & 0.4286 \\
P(\text{GC} = \textit{'none'} \mid fr) & = & 0.8334 & P(\text{GC} = \textit{'none'} \mid \neg fr) & = & 0.8571 \\
P(\text{GC} = \textit{'guarantor'} \mid fr) & = & 0.1666 & P(\text{GC} = \textit{'guarantor'} \mid \neg fr) & = & 0 \\
P(\text{GC} = \textit{'coapplicant'} \mid fr) & = & 0 & P(\text{GC} = \textit{'coapplicant'} \mid \neg fr) & = & 0.1429 \\
P(\text{ACC} = \textit{'own'} \mid fr) & = & 0.6666 & P(\text{ACC} = \textit{'own'} \mid \neg fr) & = & 0.7857 \\
P(\text{ACC} = \textit{'rent'} \mid fr) & = & 0.3333 & P(\text{ACC} = \textit{'rent'} \mid \neg fr) & = & 0.1429 \\
P(\text{ACC} = \textit{'free'} \mid fr) & = & 0 & P(\text{ACC} = \textit{'free'} \mid \neg fr) & = & 0.0714
\end{array}
$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$\left(\prod^{m} P(\mathbf{q}[k] \mid fr) \right) \times P(fr) =$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$
\begin{aligned}
P(fr) &= 0.3 \\
P(\text{CH} = \textit{'paid'} \mid fr) &= 0.1666 \\
P(\text{GC} = \textit{'none'} \mid fr) &= 0.8334 \\
P(\text{ACC} = \textit{'rent'} \mid fr) &= 0.3333
\end{aligned}
$$

$$
\left(\prod^{m} P(\mathbf{q}[k] \mid fr)\right) \times P(fr) =
$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$P(fr)$	$=$	0.3
$P(\text{CH} = \textit{'paid'} \mid fr)$	$=$	0.1666
$P(\text{GC} = \textit{'none'} \mid fr)$	$=$	0.8334
$P(\text{ACC} = \textit{'rent'} \mid fr)$	$=$	0.3333

$$\left(\prod^{m} P\left(\mathbf{q}\left[k\right] \mid fr\right)\right) \times P\left(fr\right) = 0.0139$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \textit{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{GC} = \textit{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{ACC} = \textit{'rent'} \mid \neg fr)$	$=$	0.1429

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid fr\right)\right) \times P\left(fr\right) = 0.0139$$

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid \neg fr\right)\right) \times P(\neg fr) =$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$
\begin{aligned}
P(\neg fr) &= 0.7 \\
P(\text{CH} = \textit{'paid'} \mid \neg fr) &= 0.2857 \\
P(\text{GC} = \textit{'none'} \mid \neg fr) &= 0.8571 \\
P(\text{ACC} = \textit{'rent'} \mid \neg fr) &= 0.1429
\end{aligned}
$$

$$
\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid fr\right)\right) \times P\left(fr\right) = 0.0139
$$

$$
\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid \neg fr\right)\right) \times P(\neg fr) = 0.0245
$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid fr\right)\right) \times P\left(fr\right) = 0.0139$$

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid \neg fr\right)\right) \times P(\neg fr) = 0.0245$$

CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	?

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid fr\right)\right) \times P\left(fr\right) = 0.0139$$

$$\left(\prod_{k=1}^{m} P\left(\mathbf{q}\left[k\right] \mid \neg fr\right)\right) \times P(\neg fr) = 0.0245$$

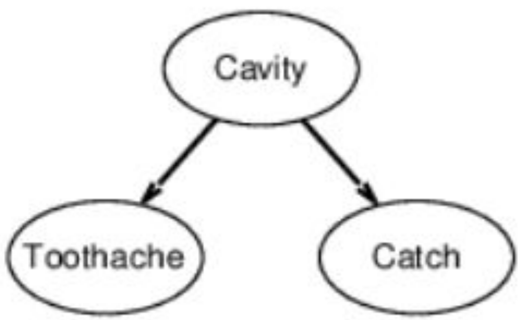
CREDIT HISTORY	GUARANTOR/COAPPLICANT	ACCOMODATION	FRAUDULENT
paid	none	rent	*'false'*

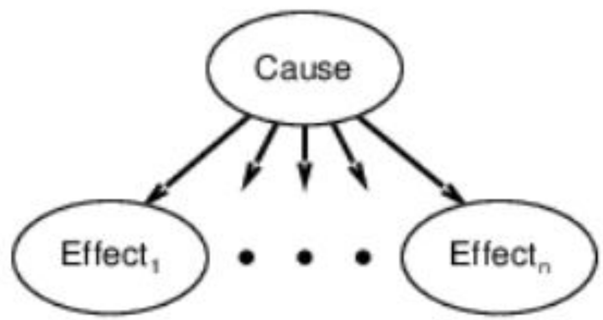
Bayesian Belief Networks

Bayesian networks

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

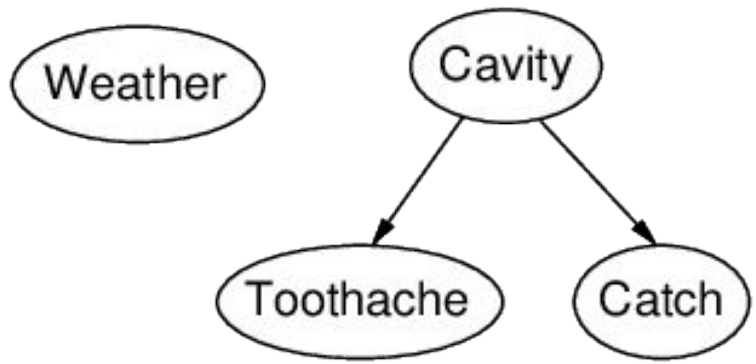
$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$





Example

Topology of network encodes conditional independence assertions:



$Weather$ is independent of the other variables

$Toothache$ and $Catch$ are conditionally independent given $Cavity$

Belief Networks

Let $X_1, \ldots, X_n$ be discrete random variables.

A *belief network (or Bayesian network)* for $X_1, \ldots, X_n$ is a graph with m nodes such that

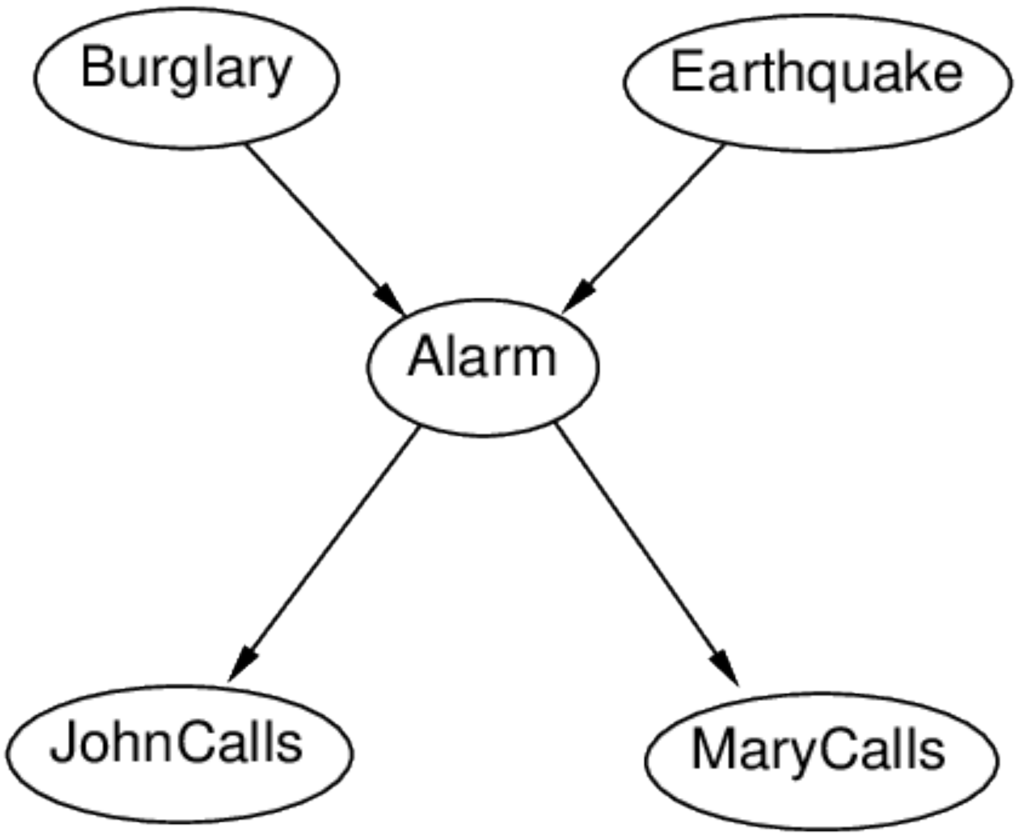
- there is a node for each X_i
- all the edges between two nodes are directed
- there are no cycles
- each node has a conditional probability table (CPT), given in terms of its parents

The intuitive meaning of an edge from a node X_i to a node X_j is that X_i has a direct influence on X_j

Example

- You have a new burglar alarm installed at home.
- It is fairly reliable at detecting a burglary, but is occasionally set off by minor earthquakes.
- You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm.
- John nearly always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too.
- Mary, on the other hand, likes rather loud music and often misses the alarm altogether.
- Given the evidence of who has or has not called, we would like to estimate the probability of a burglary.

A Belief Network



Conditional Probability Tables

Each node X_i in a belief network has an associated CPT expressing the probability of X_i, given its parents as evidence

Example:

CPT for $Alarm$:

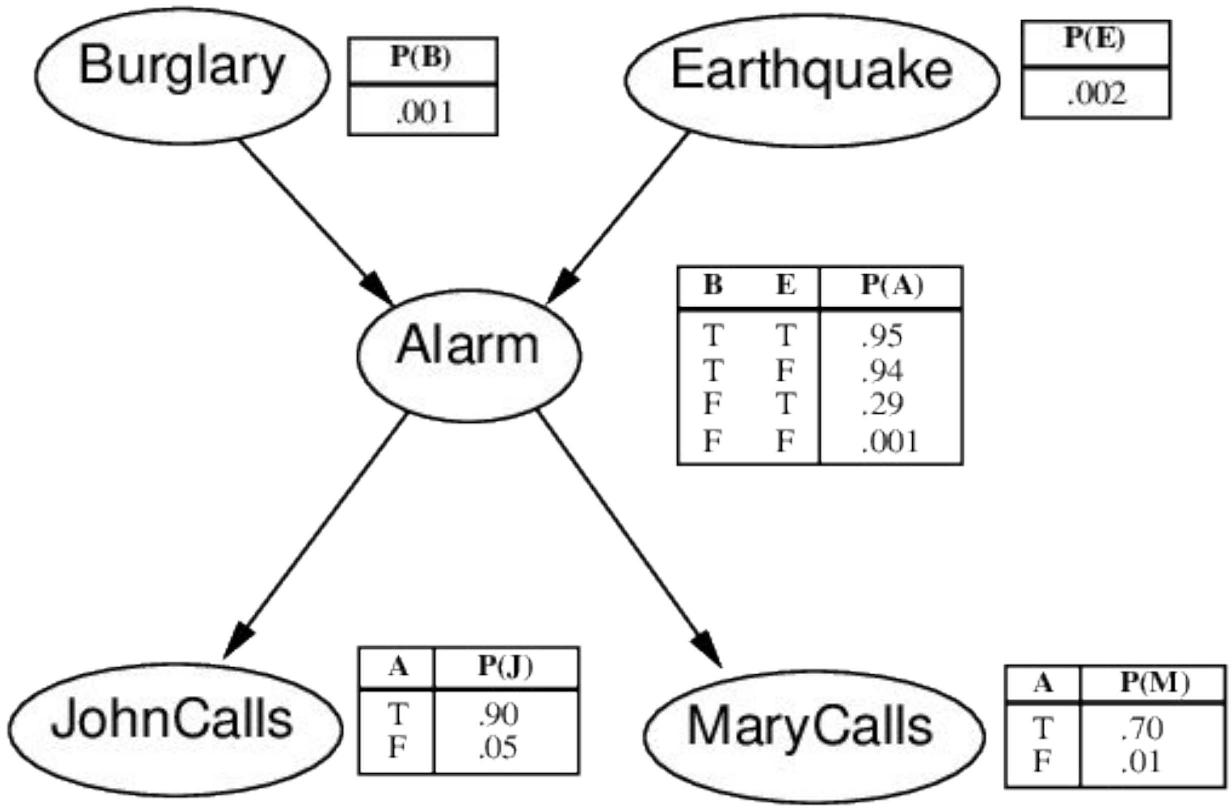
$Burglary$	$Earthquake$	$Alarm$ T	$Alarm$ F
T	T	0.950	0.050
T	F	0.940	0.060
F	T	0.290	0.710
F	F	0.001	0.999

$P(alarm \mid burglary \wedge earthquake) = 0.950$

$P(\neg alarm \mid \neg burglary \wedge earthquake) = 0.710$

...

A Belief Network with CPTs



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Note: The tables only show $P(X = true)$ here because $P(X = false) = 1 - P(X = true)$