Lecture 22: Probabilities and Bayes Rule

Probabilistic Inference: The computation of posterior probabilities for query propositions given observed evidence

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{catch} \vee \mathit{cavity}) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(catch \vee cavity) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(catch \vee cavity) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{catch} \vee \mathit{cavity}) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{catch} \vee \mathit{cavity}) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(catch \vee cavity) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		$\neg$*toothache*	
	catch	$\neg$*catch*	*catch*	$\neg$*catch*
cavity	0.108	0.012	0.072	0.008
$\neg$*cavity*	0.016	0.064	0.144	0.576

$$P(\mathit{catch} \vee \mathit{cavity}) = 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008$$

	toothache		¬*toothache*	
	catch	¬*catch*	*catch*	¬*catch*
cavity	0.108	0.012	0.072	0.008
¬*cavity*	0.016	0.064	0.144	0.576

$$\begin{aligned} P(\mathit{catch} \vee \mathit{cavity}) &= 0.108 + 0.016 + 0.072 + 0.144 + 0.012 + 0.008 \\ &= 0.36 \end{aligned}$$

$$
\begin{aligned}
P(\mathit{cavity}) &= P(\mathit{cavity}, \mathit{toothache}, \mathit{catch}) + P(\mathit{cavity}, \mathit{toothache}, \neg \mathit{catch}) \\
&\quad + P(\mathit{cavity}, \neg \mathit{toothache}, \mathit{catch}) + P(\mathit{cavity}, \neg \mathit{toothache}, \neg \mathit{catch}) \\
&= 0.108 + 0.012 + 0.072 + 0.008 = 0.2
\end{aligned}
$$

Marginalization

$$P(Y) = \sum_{z} P(Y, z)$$

Marginalization

$$P(Y) = \sum_{z} P(Y, z)$$

Conditioning

$$P(Y) = \sum_{z} P(Y|z)P(z)$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(Cavity|toothache) = \alpha P(Cavity, toothache)$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(Cavity|toothache) = \alpha P(Cavity, toothache)$$
$$= \alpha[P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(Cavity|toothache) = \alpha P(Cavity, toothache)$$
$$= \alpha[P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$$
$$= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\langle 0.12, 0.08\rangle$$

Normalization

$$P(cavity|toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(\neg cavity|toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(Cavity|toothache) = \alpha P(Cavity, toothache)$$
$$= \alpha[P(Cavity, toothache, catch) + P(Cavity, toothache, \neg catch)]$$
$$= \alpha[\langle 0.108, 0.016\rangle + \langle 0.012, 0.064\rangle]$$
$$= \alpha\langle 0.12, 0.08\rangle$$
$$= \langle 0.6, 0.4\rangle$$

General Inference Procedure

- X - Query Variable
- E - Evidence Variables
- Y - Remaining Unobserved Values

$$P(X|e)$$

General Inference Procedure

- X - Query Variable
- E - Evidence Variables
- Y - Remaining Unobserved Values

$$P(X|e) = \alpha P(X, e)$$

General Inference Procedure

- X - Query Variable
- E - Evidence Variables
- Y - Remaining Unobserved Values

$$\begin{aligned} P(X|e) &= \alpha P(X, e) \\ &= \alpha \sum_{y} P(X, e, y) \end{aligned}$$

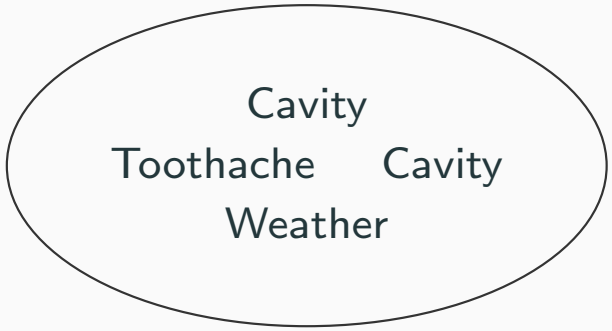
$P(toothache, cavity, catch, Weather = cloudy)$

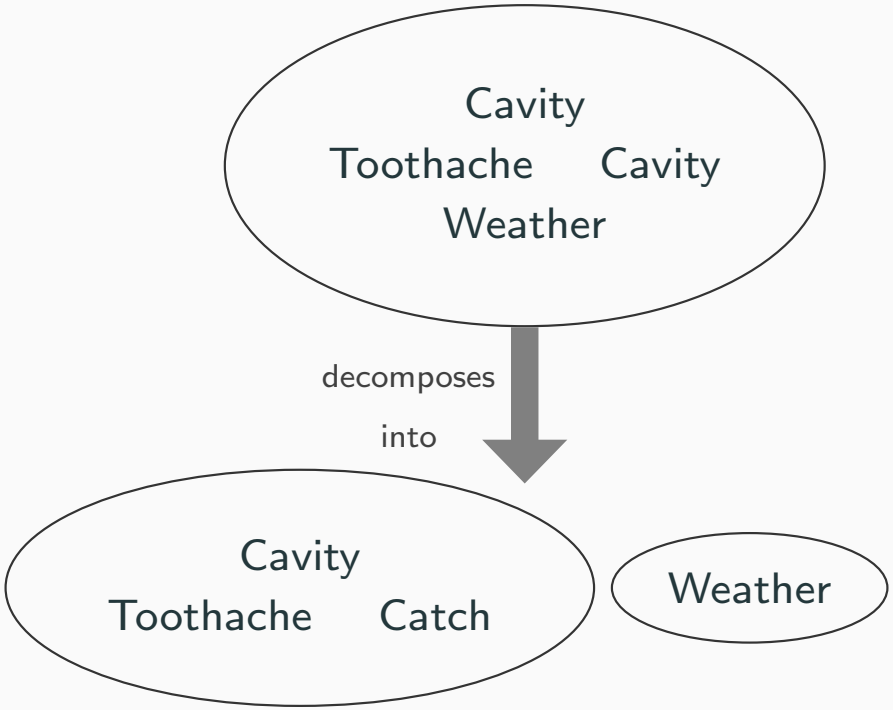
$$P(toothache, cavity, catch, Weather = cloudy)$$
$$= P(Weather = cloudy | toothache, catch, cavity) P(toothache, catch, cavity)$$

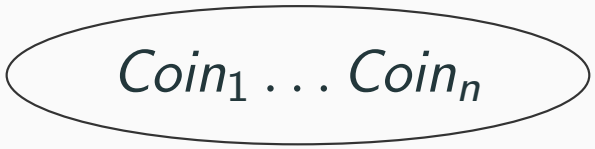
$$
\begin{aligned}
&P(\mathit{toothache}, \mathit{cavity}, \mathit{catch}, \mathit{Weather} = \mathit{cloudy}) \\
&= P(\mathit{Weather} = \mathit{cloudy} | \mathit{toothache}, \mathit{catch}, \mathit{cavity}) P(\mathit{toothache}, \mathit{catch}, \mathit{cavity}) \\
&= P(\mathit{Weather} = \mathit{cloudy}) P(\mathit{toothache}, \mathit{catch}, \mathit{cavity})
\end{aligned}
$$

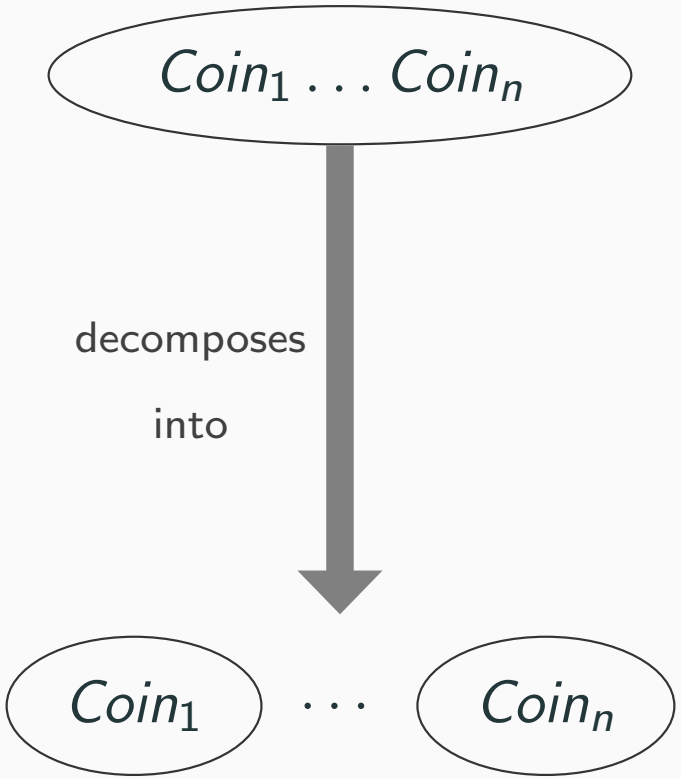
Independence

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


Cavity
Toothache
Cavity
Weather


Cavity
Toothache
Cavity
Weather
decomposes
into
Cavity
Toothache
Catch
Weather


$Coin_1 \ldots Coin_n$


$Coin_1 \ldots Coin_n$
decomposes
into
$Coin_1$
$\cdots$
$Coin_n$

Bayes Rule

$$\left.\begin{aligned} P(a \wedge b) &= P(a|b)P(b) \\ P(a \wedge b) &= P(b|a)P(a) \end{aligned}\right\} \Rightarrow P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

Bayes Rule

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

Applying Bayes Rule Example

Meningitis causes stiff neck with 50%

Prior probability of Meningitis (m) 1/50000

Prior probability of stiff neck (s) 1/20

Applying Bayes Rule Example

Meningitis causes stiff neck with 50%

Prior probability of Meningitis (m) 1/50000

Prior probability of stiff neck (s) 1/20

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{\frac{1}{2} \times \frac{1}{50000}}{\frac{1}{20}} = \frac{1}{5000}$$

Applying Bayes Rule Example

Meningitis causes stiff neck with 50%

Prior probability of Meningitis (m) 1/50000

Prior probability of stiff neck (s) 1/20

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{\frac{1}{2} \times \frac{1}{50000}}{\frac{1}{20}} = \frac{1}{5000}$$

$$P(M|s) = \alpha\langle P(s|m)P(m), P(s|\neg m)P(\neg m)\rangle$$

$$P(\mathit{Cavity}|\mathit{toothache} \wedge \mathit{catch}) = \alpha\langle 0.108, 0/016\rangle \approx \langle 0.871, 0.129\rangle$$

$$P(Cavity|toothache \wedge catch) = \alpha\langle 0.108, 0/016\rangle \approx \langle 0.871, 0.129\rangle$$

$$P(Cavity|toothache \wedge catch) = \alpha P(toothache \wedge catch|Cavity)P(Cavity)$$

$$P(Cavity|toothache \wedge catch) = \alpha\langle 0.108, 0/016\rangle \approx \langle 0.871, 0.129\rangle$$

$$P(Cavity|toothache \wedge catch) = \alpha P(toothache \wedge catch|Cavity)P(Cavity)$$

$$P(toothache \wedge catch|Cavity) = P(toothache|Cavity)P(catch|Cavity)$$

Conditional Independence

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

Conditional Independence

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

Naive Bayes Model

$$P(Cause, Effect_1, Effect_2, \ldots, Effect_n) = P(Cause) \prod_i P(Effect_i | Cause)$$

Table 1. Classification accuracies and sample standard deviations, averaged over 20 random training/test splits. "Bayes" is the Bayesian classifier with discretization and "Gauss" is the Bayesian classifier with Gaussian distributions. Superscripts denote confidence levels for the difference in accuracy between the Bayesian classifier and the corresponding algorithm, using a one-tailed paired t test: 1 is 99.5%, 2 is 99%, 3 is 97.5%, 4 is 95%, 5 is 90%, and 6 is below 90%.

Data Set	**Bayes**	**Gauss**	**C4.5**	**PEBLS**	**CN2**	**Def.**
Audiology	73.0±6.1	73.0±6.1^{6}	72.5±5.8^{6}	75.8±5.4^{3}	71.0±5.1^{5}	21.3
Annealing	95.3±1.2	84.3±3.8^{1}	90.5±2.2^{1}	98.8±0.8^{1}	81.2±5.4^{1}	76.4
Breast cancer	71.6±4.7	71.3±4.3^{6}	70.1±6.8^{5}	65.6±4.7^{1}	67.9±7.1^{1}	67.6
Credit	84.5±1.8	78.9±2.5^{1}	85.9±2.1^{3}	82.2±1.9^{1}	82.0±2.2^{1}	57.4
Chess endgames	88.0±1.4	88.0±1.4^{6}	99.2±0.1^{1}	96.9±0.7^{1}	98.1±1.0^{1}	52.0
Diabetes	74.5±2.4	75.2±2.1^{6}	73.5±3.4^{5}	71.1±2.4^{1}	73.8±2.7^{6}	66.0
Echocardiogram	69.1±5.4	73.4±4.9^{1}	64.7±6.3^{1}	61.7±6.4^{1}	68.2±7.2^{6}	67.8
Glass	61.9±6.2	50.6±8.2^{1}	63.9±8.7^{6}	62.0±7.4^{6}	63.8±5.5^{6}	31.7
Heart disease	81.9±3.4	84.1±2.8^{1}	77.5±4.3^{1}	78.9±4.0^{1}	79.7±2.9^{3}	55.0
Hepatitis	85.3±3.7	85.2±4.0^{6}	79.2±4.3^{1}	79.0±5.1^{1}	80.3±4.2^{1}	78.1
Horse colic	80.7±3.7	79.3±3.7^{1}	85.1±3.8^{1}	75.7±5.0^{1}	82.5±4.2^{2}	63.6
Hypothyroid	97.5±0.3	97.9±0.4^{1}	99.1±0.2^{1}	95.9±0.7^{1}	98.8±0.4^{1}	95.3
Iris	93.2±3.5	93.9±1.9^{6}	92.6±2.7^{6}	93.5±3.0^{6}	93.3±3.6^{6}	26.5
Labor	91.3±4.9	88.7±10.6^{6}	78.1±7.9^{1}	89.7±5.0^{6}	82.1±6.9^{1}	65.0
Lung cancer	46.8±13.3	46.8±13.3^{6}	40.9±16.3^{5}	42.3±17.3^{6}	38.6±13.5^{3}	26.8
Liver disease	63.0±3.3	54.8±5.5^{1}	65.9±4.4^{1}	61.3±4.3^{6}	65.0±3.8^{3}	58.1
LED	62.9±6.5	62.9±6.5^{6}	61.2±8.4^{6}	55.3±6.1^{1}	58.6±8.1^{2}	8.0
Lymphography	81.6±5.9	81.1±4.8^{6}	75.0±4.2^{1}	82.9±5.6^{6}	78.8±4.9^{3}	57.3
Post-operative	64.7±6.8	67.2±5.0^{3}	70.0±5.2^{1}	59.2±8.0^{2}	60.8±8.2^{4}	71.2
Promoters	87.9±7.0	87.9±7.0^{6}	74.3±7.8^{1}	91.7±5.9^{3}	75.9±8.8^{1}	43.1
Primary tumor	44.2±5.5	44.2±5.5^{6}	35.9±5.8^{1}	30.9±4.7^{1}	39.8±5.2^{1}	24.6
Solar flare	68.5±3.0	68.2±3.7^{6}	70.6±2.9^{1}	67.6±3.5^{6}	70.4±3.0^{2}	25.2
Sonar	69.4±7.6	63.0±8.3^{1}	69.1±7.4^{6}	73.8±7.4^{1}	66.2±7.5^{5}	50.8
Soybean	100.0±0.0	100.0±0.0^{6}	95.0±9.0^{3}	100.0±0.0^{6}	96.9±5.9^{3}	30.0
Splice junctions	95.4±0.6	95.4±0.6^{6}	93.4±0.8^{1}	94.3±0.5^{1}	81.5±5.5^{1}	52.4
Voting records	91.2±1.7	91.2±1.7^{6}	96.3±1.3^{1}	94.9±1.2^{1}	95.8±1.6^{1}	60.5
Wine	96.4±2.2	97.8±1.2^{3}	92.4±5.6^{1}	97.2±1.8^{6}	90.8±4.7^{1}	36.4
Zoology	94.4±4.1	94.1±3.8^{6}	89.6±4.7^{1}	94.6±4.3^{6}	90.6±5.0^{1}	39.4

On the Optimality of the Simple Bayesian Classifier under Zero-One Loss

Summary

- Probabilistic inference with full JPDs
- Independence / Conditional Independence
- Bayes' Rule. Applications with simple techniques