
THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Uncertainty reasoning and machine learning

Some first probabilistic and credal classifiers

Vu-Linh Nguyen

Chaire de Professeur Junior, Laboratoire Heudiasyc
Université de technologie de Compiègne

AOS4 master courses

Outline

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - Naïve Credal classifiers
- Decision Trees

Outline

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - Naïve Credal classifiers
- Decision Trees

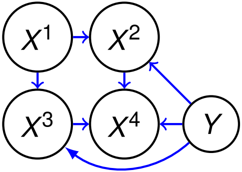
Generative Models: Structure (Exercise 1)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\mathrm{pa}(Y)) \prod_{m=1}^{M} P(X^m|\mathrm{pa}(X^m)).$$



- $\mathrm{pa}(Y) = \emptyset$, $\mathrm{pa}(X^1) = \emptyset$
- $\mathrm{pa}(X^2) = ???$
- $\mathrm{pa}(X^3) = ???$
- $\mathrm{pa}(X^4) = ???$

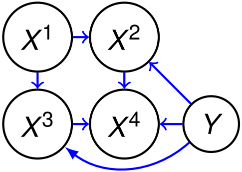
Generative Models: Structure (Exercise 1)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$



- $\text{pa}(Y) = \emptyset$, $\text{pa}(X^1) = \emptyset$
- $\text{pa}(X^2) =???$
- $\text{pa}(X^3) =???$
- $\text{pa}(X^4) =???$

Chain rule gives us

$$P(Y, \boldsymbol{X}) =???.$$

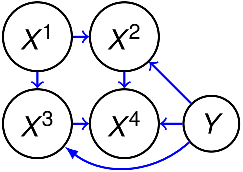
Generative Models: Structure (Solution 1)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$



- $\text{pa}(Y) = \emptyset$, $\text{pa}(X^1) = \emptyset$
- $\text{pa}(X^2) = \{Y, X^1\}$
- $\text{pa}(X^3) = \{Y, X^1\}$
- $\text{pa}(X^4) = \{Y, X^2, X^3\}$

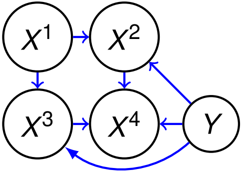
Generative Models: Structure (Solution 1)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$



- $\text{pa}(Y) = \emptyset$, $\text{pa}(X^1) = \emptyset$
- $\text{pa}(X^2) = \{Y, X^1\}$
- $\text{pa}(X^3) = \{Y, X^1\}$
- $\text{pa}(X^4) = \{Y, X^2, X^3\}$

Chain rule gives us

$$P(Y, \boldsymbol{X}) = P(Y)P(X^1)P(X^2|Y, X^1)P(X^3|Y, X^1)P(X^4|Y, X^2, X^3).$$

MLE with Examples (Exercise 2)

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

MLE with Examples (Exercise 2)

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

MLE with Examples (Exercise 2)

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = 2/3$	$\theta_A^{1,e} = 1/3$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = 1/2$	$\theta_B^{1,e} = 1/2$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = 1/3$	$\theta_C^{1,e} = 2/3$
$n_A^{2,f} = ???$	$n_A^{2,g} = ???$	$\theta_A^{2,f} = ???$	$\theta_A^{2,g} = ???$
$n_B^{2,f} = ???$	$n_B^{2,g} = ???$	$\theta_B^{2,f} = ???$	$\theta_B^{2,g} = ???$
$n_C^{2,f} = ???$	$n_C^{2,g} = ???$	$\theta_C^{2,f} = ???$	$\theta_C^{2,g} = ???$

MLE with Examples (Exercise 2)

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = 2/3$	$\theta_A^{1,e} = 1/3$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = 1/2$	$\theta_B^{1,e} = 1/2$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = 1/3$	$\theta_C^{1,e} = 2/3$
$n_A^{2,f} = ???$	$n_A^{2,g} = ???$	$\theta_A^{2,f} = ???$	$\theta_A^{2,g} = ???$
$n_B^{2,f} = ???$	$n_B^{2,g} = ???$	$\theta_B^{2,f} = ???$	$\theta_B^{2,g} = ???$
$n_C^{2,f} = ???$	$n_C^{2,g} = ???$	$\theta_C^{2,f} = ???$	$\theta_C^{2,g} = ???$

$$n_A^{2,h} = ??? \quad n_B^{2,h} = ??? \quad n_C^{2,h} = ???$$
$$\theta_A^{2,h} = ??? \quad \theta_B^{2,h} = ??? \quad \theta_C^{2,h} = ???$$

MLE with Examples (Solution 2)

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = 2/3$	$\theta_A^{1,e} = 1/3$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = 1/2$	$\theta_B^{1,e} = 1/2$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = 1/3$	$\theta_C^{1,e} = 2/3$
$n_A^{2,f} = 1$	$n_A^{2,g} = 2$	$\theta_A^{2,f} = 1/3$	$\theta_A^{2,g} = 2/3$
$n_B^{2,f} = 1$	$n_B^{2,g} = 1$	$\theta_B^{2,f} = 1/2$	$\theta_B^{2,g} = 1/2$
$n_C^{2,f} = 2$	$n_C^{2,g} = 1$	$\theta_C^{2,f} = 2/3$	$\theta_C^{2,g} = 1/3$

$$n_A^{2,h} = 0 \quad n_B^{2,h} = 0 \quad n_C^{2,h} = 0$$
$$\theta_A^{2,h} = 0 \quad \theta_B^{2,h} = 0 \quad \theta_C^{2,h} = 0$$

Conditional Probabilities (Exercise 3)

Given $\boldsymbol{x} = (x^{1,q_1}, \ldots, x^{M,q_M})$, for any $y^k \in \mathscr{Y}$:

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (1)$$

Conditional Probabilities (Exercise 3)

Given $\boldsymbol{x} = (x^{1,q_1}, \dots, x^{M,q_M})$, for any $y^k \in \mathscr{Y}$:

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (1)$$

$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

$$\begin{array}{cc|cc}
\theta_A^{1,d} = 2/3 & \theta_A^{1,e} = 1/3 & \theta_A^{2,f} = 1/3 & \theta_A^{2,g} = 2/3 \\
\theta_B^{1,d} = 1/2 & \theta_B^{1,e} = 1/2 & \theta_B^{2,f} = 1/2 & \theta_B^{2,g} = 1/2 \\
\theta_C^{1,d} = 1/3 & \theta_C^{1,e} = 2/3 & \theta_C^{2,f} = 2/3 & \theta_C^{2,g} = 1/3 \\
\hline
\end{array}$$

$$\theta_A^{2,h} = 0 \quad \theta_B^{2,h} = 0 \quad \theta_C^{2,h} = 0$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$
(d,f)	???	???	???
(e,h)	???	???	???

Conditional Probabilities (Solution 3)

Given $\boldsymbol{x} = (x^{1,q_1}, \dots, x^{M,q_M})$, for any $y^k \in \mathscr{Y}$:

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (2)$$

$$\theta_A = 3/8 \quad \theta_B = 1/4 \quad \theta_C = 3/8$$

$$\begin{array}{cc|cc}
\theta_A^{1,d} = 2/3 & \theta_A^{1,e} = 1/3 & \theta_A^{2,f} = 1/3 & \theta_A^{2,g} = 2/3 \\
\theta_B^{1,d} = 1/2 & \theta_B^{1,e} = 1/2 & \theta_B^{2,f} = 1/2 & \theta_B^{2,g} = 1/2 \\
\theta_C^{1,d} = 1/3 & \theta_C^{1,e} = 2/3 & \theta_C^{2,f} = 2/3 & \theta_C^{2,g} = 1/3 \\
\hline
\end{array}$$

$$\theta_A^{2,h} = 0 \quad \theta_B^{2,h} = 0 \quad \theta_C^{2,h} = 0$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$
(d,f)	$1/12$	$1/16$	$1/12$
(e,h)	0	0	0

Optimal Decision Rules (Exercise 4)

If $\ell(y^{k'}, y^k) = \mathbb{1}(y^{k'} \neq y^k)$, then (See Lecture 3+Check!)

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \operatorname*{argmax}_{y^k \in \mathscr{Y}} P'(y^k|\boldsymbol{x})$$

Optimal Decision Rules (Exercise 4)

If $\ell(y^{k'}, y^k) = \mathbb{1}(y^{k'} \neq y^k)$, then (See Lecture 3+Check!)

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \operatorname*{argmax}_{y^k \in \mathcal{Y}} P'(y^k | \boldsymbol{x})$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$
(d,f)	$1/12$	$1/16$	$1/12$
(e,h)	0	0	0

Optimal Decision Rules (Exercise 4)

If $\ell(y^{k'}, y^{k}) = \mathbb{1}(y^{k'} \neq y^{k})$, then (See Lecture 3+Check!)

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \operatorname*{argmax}_{y^k \in \mathcal{Y}} P'(y^k|\boldsymbol{x})$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$
(d,f)	$1/12$	$1/16$	$1/12$
(e,h)	0	0	0

- If $\boldsymbol{x} = (d,f)$, then

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = ???, \tag{3}$$

- If $\boldsymbol{x} = (e,h)$, then

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = ???, \tag{4}$$

Optimal Decision Rules (Solution 4)

If $\ell(y^{k'}, y^k) = \mathbb{1}(y^{k'} \neq y^k)$, then (See Lecture 3+Check!)

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \underset{y^k \in \mathcal{Y}}{\operatorname{argmax}} P'(y^k|\boldsymbol{x})$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$
(d,f)	$1/12$	$1/16$	$1/12$
(e,h)	0	0	0

- If $\boldsymbol{x} = (d,f)$, then

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \text{ either } A \text{ or } C, \tag{5}$$

- If $\boldsymbol{x} = (e,h)$, then

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \text{ not well-defined },\tag{6}$$

NBC + DM (Exercise 5)

$$\theta_k := {}^{(n_k+1/3)}\big/_{(N+1)}, \qquad \theta_k^{m,q_m} := {}^{(n_k^{m,q_m}+1/|\mathscr{X}^m|)}\big/_{(n_k+1)}.$$

NBC + DM (Exercise 5)

$$\theta_k := (n_k + 1/3)/(N+1), \qquad \theta_k^{m,q_m} := (n_k^{m,q_m} + 1/|\mathscr{X}^m|)/(n_k+1).$$

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = 10/27 \quad \theta_B = 7/27 \quad \theta_C = 10/27$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = 5/8$	$\theta_A^{1,e} = 3/8$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = 3/6$	$\theta_B^{1,e} = 3/6$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = 3/8$	$\theta_C^{1,e} = 5/8$
$n_A^{2,f} = 1$	$n_A^{2,g} = 2$	$\theta_A^{2,f} = ???$	$\theta_A^{2,g} = ???$
$n_B^{2,f} = 1$	$n_B^{2,g} = 1$	$\theta_B^{2,f} = ???$	$\theta_B^{2,g} = ???$
$n_C^{2,f} = 2$	$n_C^{2,g} = 1$	$\theta_C^{2,f} = ???$	$\theta_C^{2,g} = ???$

$$n_A^{2,h} = 0 \quad n_B^{2,h} = 0 \quad n_C^{2,h} = 0$$
$$\theta_A^{2,h} = ??? \quad \theta_B^{2,h} = ??? \quad \theta_C^{2,h} = ???$$

NBC + DM (Solution 5)

$$\theta_k := {}^{(n_k+1/3)}\!/_{(N+1)}, \qquad \theta_k^{m,q_m} := {}^{(n_k^{m,q_m}+1/|\mathscr{X}^m|)}\!/_{(n_k+1)}.$$

- $\mathscr{Y} = \{A, B, C\}$
- $\mathscr{X}^1 = \{d, e\}$
- $\mathscr{X}^2 = \{f, g, h\}$

n	Y	X^1	X^2
1	A	d	f
2	A	d	g
3	A	e	g
4	B	d	f
5	B	e	g
6	C	d	f
7	C	e	f
8	C	e	g

$$n_A = 3 \quad n_B = 2 \quad n_C = 3$$
$$\theta_A = {}^{10}\!/_{27} \quad \theta_B = {}^{7}\!/_{27} \quad \theta_C = {}^{10}\!/_{27}$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = {}^5\!/_8$	$\theta_A^{1,e} = {}^3\!/_8$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = {}^3\!/_6$	$\theta_B^{1,e} = {}^3\!/_6$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = {}^3\!/_8$	$\theta_C^{1,e} = {}^5\!/_8$
$n_A^{2,f} = 1$	$n_A^{2,g} = 2$	$\theta_A^{2,f} = {}^4\!/_{12}$	$\theta_A^{2,g} = {}^7\!/_{12}$
$n_B^{2,f} = 1$	$n_B^{2,g} = 1$	$\theta_B^{2,f} = {}^4\!/_9$	$\theta_B^{2,g} = {}^4\!/_9$
$n_C^{2,f} = 2$	$n_C^{2,g} = 1$	$\theta_C^{2,f} = {}^7\!/_{12}$	$\theta_C^{2,g} = {}^4\!/_{12}$

$$n_A^{2,h} = 0 \quad n_B^{2,h} = 0 \quad n_C^{2,h} = 0$$
$$\theta_A^{2,h} = {}^1\!/_{12} \quad \theta_B^{2,h} = {}^1\!/_9 \quad \theta_C^{2,h} = {}^1\!/_{12}$$

Conditional Probabilities (Exercise 6)

Given $\boldsymbol{x} = (x^{1,q_1}, \ldots, x^{M,q_M})$, for any $y^k \in \mathscr{Y}$:

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \tag{7}$$

$$\theta_A = {}^{10}\!/_{27} \quad \theta_B = {}^{7}\!/_{27} \quad \theta_C = {}^{10}\!/_{27}$$

$\theta_A^{1,d} = 5/8$	$\theta_A^{1,e} = 3/8$	$\theta_A^{2,f} = 4/12$	$\theta_A^{2,g} = 7/12$
$\theta_B^{1,d} = 3/6$	$\theta_B^{1,e} = 3/6$	$\theta_B^{2,f} = 4/9$	$\theta_B^{2,g} = 4/9$
$\theta_C^{1,d} = 3/8$	$\theta_C^{1,e} = 5/8$	$\theta_C^{2,f} = 7/12$	$\theta_C^{2,g} = 4/12$

$$\theta_A^{2,h} = 1/12 \quad \theta_B^{2,h} = 1/9 \quad \theta_C^{2,h} = 1/12$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$	$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x})$
(d,f)	???	???	???	???
(e,h)	???	???	???	???

Conditional Probabilities (Solution 6)

Given $\boldsymbol{x} = (x^{1,q_1}, \dots, x^{M,q_M})$, for any $y^k \in \mathscr{Y}$:

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^M \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^M \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^M \theta_k^{m,q_m}. \tag{8}$$

$$\theta_A = {}^{10}/_{27} \quad \theta_B = {}^{7}/_{27} \quad \theta_C = {}^{10}/_{27}$$

$$\begin{array}{cc|cc}
\theta_A^{1,d} = {}^5/_8 & \theta_A^{1,e} = {}^3/_8 & \theta_A^{2,f} = {}^4/_{12} & \theta_A^{2,g} = {}^7/_{12} \\
\theta_B^{1,d} = {}^3/_6 & \theta_B^{1,e} = {}^3/_6 & \theta_B^{2,f} = {}^4/_9 & \theta_B^{2,g} = {}^4/_9 \\
\theta_C^{1,d} = {}^3/_8 & \theta_C^{1,e} = {}^5/_8 & \theta_C^{2,f} = {}^7/_{12} & \theta_C^{2,g} = {}^4/_{12} \\
\hline
\end{array}$$

$$\theta_A^{2,h} = {}^1/_{12} \quad \theta_B^{2,h} = {}^1/_9 \quad \theta_C^{2,h} = {}^1/_{12}$$

$\boldsymbol{x}$	$P'(A\|\boldsymbol{x})$	$P'(B\|\boldsymbol{x})$	$P'(C\|\boldsymbol{x})$	$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x})$
(d,f)	$\frac{10}{27}\frac{5}{8}\frac{4}{12}$	$\frac{7}{27}\frac{3}{6}\frac{4}{9}$	$\frac{10}{27}\frac{3}{8}\frac{7}{12}$	C
(e,h)	$\frac{10}{27}\frac{3}{8}\frac{1}{12}$	$\frac{7}{27}\frac{3}{6}\frac{1}{9}$	$\frac{10}{27}\frac{5}{8}\frac{1}{12}$	C

Outline

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - Naïve Credal classifiers
- Decision Trees

Interval Conditional Probabilities (Exercise 7)

$$1/\overline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\underline{\epsilon}_k}{n_k + s\overline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}} \right),$$

$$1/\underline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\overline{\epsilon}_k}{n_k + s\underline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}} \right).$$

Interval Conditional Probabilities (Exercise 7)

$$1/\overline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\underline{\epsilon}_k}{n_k + s\overline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}} \right),$$

$$1/\underline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\overline{\epsilon}_k}{n_k + s\underline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}} \right).$$

$s = 1 \quad \underline{\epsilon}_k = 0.01 \quad \overline{\epsilon}_k = 0.99 \quad n_A = 3 \quad n_B = 2 \quad n_C = 3$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$n_A^{2,f} = 1$	$n_A^{2,g} = 2$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$n_B^{2,f} = 1$	$n_B^{2,g} = 1$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$n_C^{2,f} = 2$	$n_C^{2,g} = 1$

$n_A^{2,h} = 0 \quad n_B^{2,h} = 0 \quad n_C^{2,h} = 0$

$\boldsymbol{x}$	$\underline{P}(A\|\boldsymbol{x})$	$\underline{P}(B\|\boldsymbol{x})$	$\underline{P}(C\|\boldsymbol{x})$	$\overline{P}(A\|\boldsymbol{x})$	$\overline{P}(B\|\boldsymbol{x})$	$\overline{P}(C\|\boldsymbol{x})$
(d,f)	???	???	???	???	???	???
(e,h)	???	???	???	???	???	???

Interval Conditional Probabilities (Solution 7)

$$
1/\overline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\underline{\epsilon}_k}{n_k + s\overline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}} \right),
$$

$$
1/\underline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\overline{\epsilon}_k}{n_k + s\underline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}} \right).
$$

Interval Conditional Probabilities (Solution 7)

$$1/\overline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\underline{\epsilon}_k}{n_k + s\overline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}} \right),$$

$$1/\underline{P}(y^k|\boldsymbol{x}) - 1 = \sum_{k' \neq k} \left(\frac{n_{k'} + s\overline{\epsilon}_k}{n_k + s\underline{\epsilon}_k} \left(\frac{n_k + s}{n_{k'} + s} \right)^M \prod_{m=1}^{M} \frac{n_{k'}^{q_m,m} + s\overline{\epsilon}_k^{m,q_m}}{n_k^{q_m,m} + s\underline{\epsilon}_k^{m,q_m}} \right).$$

$s = 1 \quad \underline{\epsilon}_k = 0.01 \quad \overline{\epsilon}_k = 0.99 \quad n_A = 3 \quad n_B = 2 \quad n_C = 3$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$n_A^{2,f} = 1$	$n_A^{2,g} = 2$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$n_B^{2,f} = 1$	$n_B^{2,g} = 1$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$n_C^{2,f} = 2$	$n_C^{2,g} = 1$

$n_A^{2,h} = 0 \quad n_B^{2,h} = 0 \quad n_C^{2,h} = 0$

$\boldsymbol{x}$	$\underline{P}(A\|\boldsymbol{x})$	$\underline{P}(B\|\boldsymbol{x})$	$\underline{P}(C\|\boldsymbol{x})$	$\overline{P}(A\|\boldsymbol{x})$	$\overline{P}(B\|\boldsymbol{x})$	$\overline{P}(C\|\boldsymbol{x})$
(d,f)	???	???	???	???	???	???
(e,h)	???	???	???	???	???	???

Outline

- Naïve Bayesian/Credal classifiers
- Decision Trees
  - Decision Trees

Outline

- Naïve Bayesian/Credal classifiers
- Decision Trees
  - Decision Trees

Discriminative Models

Probabilistic Models:

- Estimate $P(Y, \boldsymbol{X})$
- Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$

Extreme Cases:

- Discriminative models: $Y \notin \text{pa}(X^m)$, $m \in [M]$
- Generative models: $\text{pa}(Y) = \emptyset$ and $Y \in \text{pa}(X^p)$, $m \in [M]$.

Model Families:

- How to encode/parametrize $P(Y|\text{pa}(Y))$ and $P(X^m|\text{pa}(X^m))$.
- How to estimate $P(Y, \boldsymbol{X}) = P(Y|\boldsymbol{X})P(\boldsymbol{X})$ from training data.

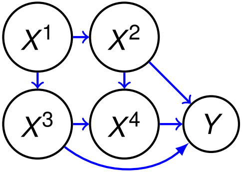
Discriminative Models: Structure (Exercise 8)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\mathrm{pa}(Y)) \prod_{m=1}^{M} P(X^m|\mathrm{pa}(X^m)).$$



- $\mathrm{pa}(Y) = ???$,
- $\mathrm{pa}(X^1) = \emptyset$, $\mathrm{pa}(X^2) = ???$
- $\mathrm{pa}(X^3) = ???$
- $\mathrm{pa}(X^4) = ???$

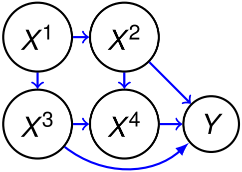
Discriminative Models: Structure (Exercise 8)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$



- $\text{pa}(Y) = ???$,
- $\text{pa}(X^1) = \emptyset$, $\text{pa}(X^2) = ???$
- $\text{pa}(X^3) = ???$
- $\text{pa}(X^4) = ???$

Chain rule gives us

$$P(Y, \boldsymbol{X}) = ???.$$

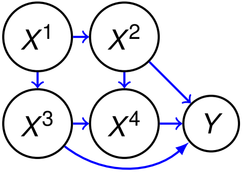
Discriminative Models: Structure (Solution 8)

Let's start with an example where one wishes to model

$$P(Y,\boldsymbol{X}) = P(Y,X^1,X^2,X^3,X^4).$$

Chain rule (probability):

$$P(Y,\boldsymbol{X}) = P(Y|\text{pa}(Y))\prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$



- $\text{pa}(Y) = \{X^2, X^3, X^4\}$
- $\text{pa}(X^1) = \emptyset$, $\text{pa}(X^2) = \{X^1\}$
- $\text{pa}(X^3) = \{X^1\}$
- $\text{pa}(X^4) = \{X^2, X^3\}$

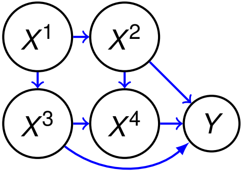
Discriminative Models: Structure (Solution 8)

Let's start with an example where one wishes to model

$$P(Y, \boldsymbol{X}) = P(Y, X^1, X^2, X^3, X^4).$$

Chain rule (probability):

$$P(Y, \boldsymbol{X}) = P(Y|\mathrm{pa}(Y)) \prod_{m=1}^{M} P(X^m|\mathrm{pa}(X^m)).$$



- $\mathrm{pa}(Y) = \{X^2, X^3, X^4\}$
- $\mathrm{pa}(X^1) = \emptyset$, $\mathrm{pa}(X^2) = \{X^1\}$
- $\mathrm{pa}(X^3) = \{X^1\}$
- $\mathrm{pa}(X^4) = \{X^2, X^3\}$

Chain rule gives us

$$P(Y, \boldsymbol{X}) = {\color{blue}P(Y|X^2, X^3, X^4)}{\color{red}P(X^1)}P(X^2|X^1)P(X^3|X^1)P(X^4|X^2, X^3).$$

Classification Task (Exercise + Solution 9)

Prove that

$$P(y|\boldsymbol{x}) = P(y|\mathrm{pa}(y)). \tag{9}$$

Classification Task (Exercise + Solution 9)

Prove that

$$P(y|\boldsymbol{x}) = P(y|\text{pa}(y)). \tag{9}$$

We have

$$P(y|\boldsymbol{x}) = \frac{P(y,\boldsymbol{x})}{\sum_{y' \in \mathcal{Y}} P(y',\boldsymbol{x})} \tag{10}$$

$$= \frac{P(y|\text{pa}(y)) \prod_{m=1}^{M} P(x^m|\text{pa}(x^m))}{\sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y')) \prod_{m=1}^{M} P(x^m|\text{pa}(x^m))} \tag{11}$$

$$= \frac{\prod_{m=1}^{M} P(x^m|\text{pa}(x^m)) P(y|\text{pa}(y))}{\prod_{m=1}^{M} P(x^m|\text{pa}(x^m)) \sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y'))} \tag{12}$$

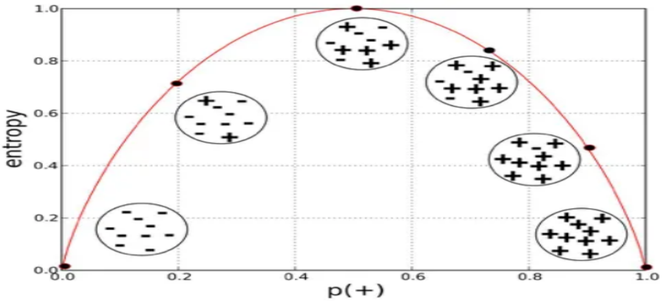
$$= \frac{P(y|\text{pa}(y))}{\sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y'))} \tag{13}$$

$$= P(y|\text{pa}(y)). \tag{14}$$

Compute Entropy (Exercise 10)

- Entropy of a node $\boldsymbol{D}_h \subset \boldsymbol{D}$ with $P(\mathcal{Y}|\boldsymbol{D}_h)$

$$U_E(P(y|\boldsymbol{D}_h)) = -\sum_{y\in\mathcal{Y}} P(y|\boldsymbol{D}_h)\log_2(P(y|\boldsymbol{D}_h)) .$$

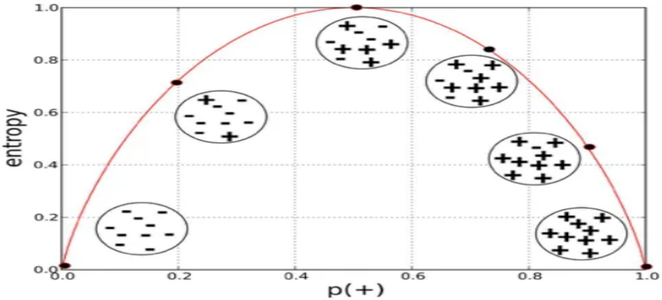


- Entropy of the bottom left node is ???
- Entropy of the top node is ???

Compute Entropy (Solution 10)

- Entropy of a node $\boldsymbol{D}_h \subset \boldsymbol{D}$ with $P(\mathscr{Y}|\boldsymbol{D}_h)$

$$U_E(P(y|\boldsymbol{D}_h)) = -\sum_{y\in\mathscr{Y}} P(y|\boldsymbol{D}_h)\log_2(P(y|\boldsymbol{D}_h)).$$



- Entropy of the bottom left node is 0
- Entropy of the top node is $-0.5\log_2(0.5)+0.5\log_2(0.5)=1$

References I