Uncertainty reasoning and machine learning
Some first credal classifiers

Vu-Linh Nguyen

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

AOS4 master courses

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc

Objectives

After this lecture students should be able to

- use IDM and related models in Naïve credal classifier (NCC) [3]
- use IDM and related models in decision trees [8]

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers

- Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc

How to interpret a decision tree?

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

How to interpret a decision tree?

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc
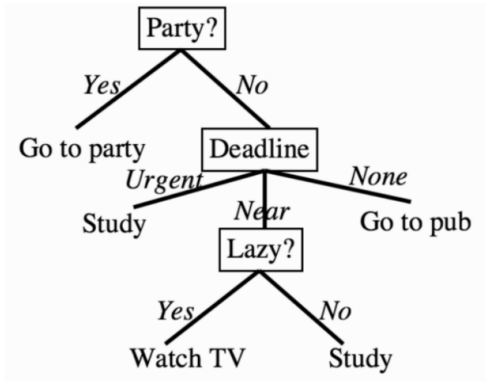
How to interpret a (feedforward) neural network?

How to interpret a (feedforward) neural network?

Without dropout

With dropout

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks
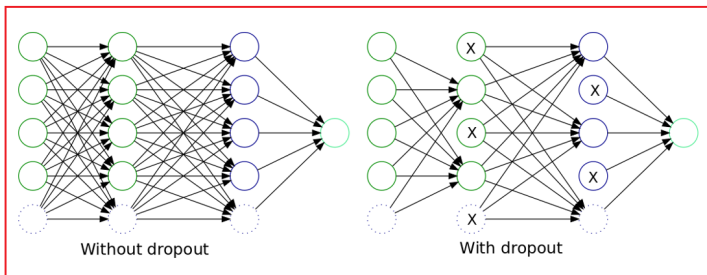
heudiasyc

Probabilistic Models: Graphical Interpretation [5, 9]
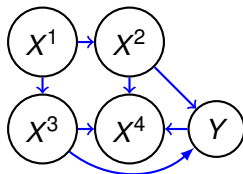
Basic setup

- A set of features $\boldsymbol{X} = \{X^1, \ldots, X^M\}$; $[M] := \{1, \ldots, M\}$
- A class variable Y whose outcome $y \in \mathcal{Y}$

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Probabilistic Models: Graphical Interpretation [5, 9]

Basic setup

- A set of features $\boldsymbol{X} = \{X^1,\ldots,X^M\}$; $[M] := \{1,\ldots,M\}$
- A class variable Y whose outcome $y \in \mathcal{Y}$
- A directed acyclic graph (DAG) connecting Y and X^m

This DAG (model structure) tells us:
- ○ $\mathrm{pa}(Y) = \{X^2, X^3\}$, $\mathrm{pa}(X^1) = \varnothing$
- ○ $\mathrm{pa}(X^2) = \{X^1\}$, $\mathrm{pa}(X^3) = \{X^1\}$
- ○ $\mathrm{pa}(X^4) = \{Y, X^2, X^3\}$

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Probabilistic Models: Graphical Interpretation [5, 9]

Basic setup

- A set of features $\boldsymbol{X} = \{X^1, \ldots, X^M\}$; $[M] := \{1, \ldots, M\}$
- A class variable Y whose outcome $y \in \mathcal{Y}$
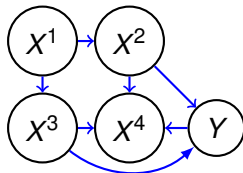- A directed acyclic graph (DAG) connecting Y and X^m

This DAG (model structure) tells us:
  - $\mathrm{pa}(Y) = \{X^2, X^3\}$, $\mathrm{pa}(X^1) = \emptyset$
  - $\mathrm{pa}(X^2) = \{X^1\}$, $\mathrm{pa}(X^3) = \{X^1\}$
  - $\mathrm{pa}(X^4) = \{Y, X^2, X^3\}$

Probabilistic Models:

- Expressing $P(Y, \boldsymbol{X})$ using the **chain rule** (probability):

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

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Probabilistic Models: Model Families [9]

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 $Y \in \text{pa}(X^p)$, $m \in [M]$.

Probabilistic Models: Model Families [9]

Probabilistic Models:

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

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

Extreme Cases:

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

Model Families:

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

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Credal (Imprecise Probability) Models

Basic setup

- A set of features $\boldsymbol{X} = \{X^1, \ldots, X^M\}$
- A class variable Y whose outcome $y \in \mathcal{Y}$

Credal Models:

- $\mathscr{P} := \{P(Y, \boldsymbol{X}) | P \text{ is compatible with knowledge/data}\}$
- Chain rule (probability):

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

Extreme Cases:

- Discriminative models: $Y \notin \mathrm{pa}(X^m)$, $m \in [M] := \{1, \ldots, M\}$
- Generative models: $\mathrm{pa}(Y) = \emptyset$ and $Y \in \mathrm{pa}(X^m)$, $m \in [M]$.

Model Families:

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

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

Assumptions and Questions

Assumption and desirable property:

A1. X^m, $m \in [M] := \{1, \ldots, M\}$, are always made available

P1. Best estimates of $P(Y|\text{pa}(Y))$ and $P(X^m|\text{pa}(X^m))$ can be found given (training) data.

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc

Assumptions and Questions

Assumption and desirable property:

A1. X^m, $m \in [M] := \{1, \ldots, M\}$, are always made available

P1. Best estimates of $P(Y|\mathrm{pa}(Y))$ and $P(X^m|\mathrm{pa}(X^m))$ can be found given (training) data.

Questions (Exercise):

- Does the P1 hold for Naïve Bayes Classifier?
- Does the P1 hold for Decision trees?

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Ba Neural Networks

heudiasyc

Assumptions and Questions

Assumption and desirable property:

A1. X^m, $m \in [M] := \{1, \dots, M\}$, are always made available

P1. Best estimates of $P(Y|\text{pa}(Y))$ and $P(X^m|\text{pa}(X^m))$ can be found given (training) data.

Questions (Exercise):

- Does the P1 hold for Naïve Bayes Classifier?
- Does the P1 hold for Decision trees?

Questions (which will not be discussed in this lecture):

- What may happen if X^m, $m \in [M]$, can be partially given?
- What may happen if best estimates of $P(Y|\text{pa}(Y))$ and $P(X^m|\text{pa}(X^m))$ may not be found?

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

The Next Slides

We shall elaborate on how to solve classification task using

- Naïve Bayesian classifier (NBC) (an example of generative model)
- Decision trees (DTs) (examples of discriminative model)

Graphical Interpretation of Probabilistic Models Naïve Bayesian/Credal classifiers Decision Trees Bayesian Neural Networks

heudiasyc

The Next Slides

We shall elaborate on how to solve classification task using

- Naïve Bayesian classifier (NBC) (an example of generative model)
- Decision trees (DTs) (examples of discriminative model)

How IDM (Lecture 3) can be used to generalize NBC and DTs to

- cope with the case of small and partial/missing data
- make set-valued predictions under the presence of uncertainty

The Next Slides

We shall elaborate on how to solve classification task using

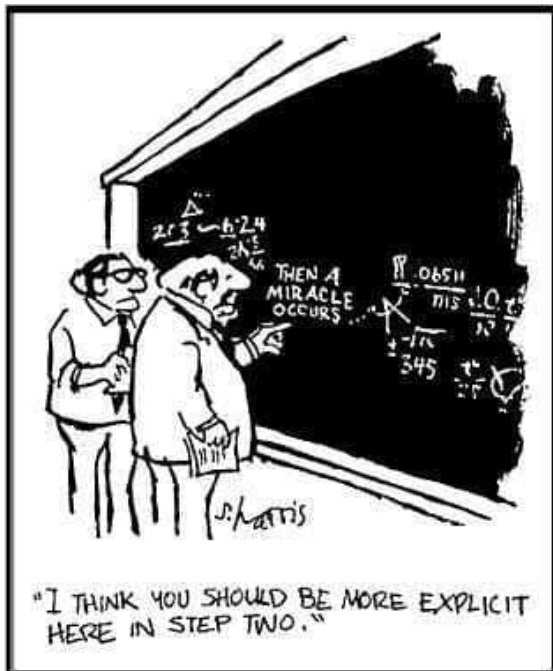- Naïve Bayesian classifier (NBC) (an example of generative model)
- Decision trees (DTs) (examples of discriminative model)

How IDM (Lecture 3) can be used to generalize NBC and DTs to

- cope with the case of small and partial/missing data
- make set-valued predictions under the presence of uncertainty

We would also discuss (if we have time) the cases of

- Ensembles (Trees, Neural Nets, etc.)
- Bayesian Neural Nets

"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - Naïve Credal classifiers

- Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - Naïve Credal classifiers

- Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Generative Models

Probabilistic Models:

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

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

Extreme Cases:

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

Model Families:

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*
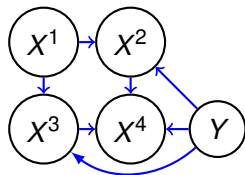
heudiasyc

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):

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
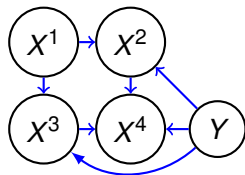*Naïve Bayesian classifier* *Naïve Credal classifiers*

heudiasyc

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

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):

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
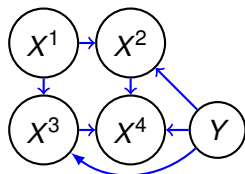*Naïve Bayesian classifier* *Naïve Credal classifiers*
heudiasyc

Generative Models: Structure (Solution 1)

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

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

Chain rule (probability):

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

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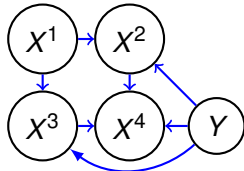- $\text{pa}(X^3) = \{Y, X^1\}$
- $\text{pa}(X^4) = \{Y, X^2, X^3\}$

Chain rule gives us

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Naïve Bayesian classifier (NBC)

Comments:

- NBC is a generative model with no arc $X' \longrightarrow X$
- Chain rule gives us

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Naïve Bayesian classifier (NBC)

Comments:

- NBC is a generative model with no arc $X' \longrightarrow X$
- Chain rule gives us

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

To solve the **classification task**,

- joint probability distribution $P(Y, \boldsymbol{X})$ is learn from training data $\boldsymbol{D}$
- conditional distribution $P(Y | \boldsymbol{X})$ is extracted using **Bayes' theorem**

$$P(y | \boldsymbol{x}) = \frac{P(y, \boldsymbol{x})}{\sum_{y' \in \mathcal{Y}} P(y', \boldsymbol{x})} = \frac{P(y) \prod_{m=1}^{M} P(x^m | y)}{\sum_{y' \in \mathcal{Y}} P(y') \prod_{m=1}^{M} P(x^m | y')}. \quad (1)$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Estimate Parameters of NBC

Basic setup:

- A class variable Y with K possible values: $\mathcal{Y} = \{y^1, \ldots y^K\}$
- M discrete features: $\boldsymbol{X} = (X^1, \ldots, X^M)$
- Feature X^m has Q_m possible values: $\mathcal{X}^m = \{x^{m,1}, \ldots x^{m,Q_m}\}$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks
Naïve Bayesian classifier   *Naïve Credal classifiers*

heudiasyc

Estimate Parameters of NBC

Basic setup:

- A class variable Y with K possible values: $\mathscr{Y} = \{y^1, \dots y^K\}$
- M discrete features: $\boldsymbol{X} = (X^1, \dots, X^M)$
- Feature X^m has Q_m possible values: $\mathscr{X}^m = \{x^{m,1}, \dots x^{m,Q_m}\}$

Task: Finding the best estimate of

- $\theta_k := P(y^k),\ k \in [K]$
- $\theta_k^{m,q_m} := P(x^{q_m,m}|y^k),\ q_m \in [Q_m],\ k \in [K],\ m \in [M]$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Estimate Parameters of NBC

Basic setup:

- A class variable Y with K possible values: $\mathcal{Y} = \{y^1, \dots y^K\}$
- M discrete features: $\boldsymbol{X} = (X^1, \dots, X^M)$
- Feature X^m has Q_m possible values: $\mathcal{X}^m = \{x^{m,1}, \dots x^{m,Q_m}\}$

Task: Finding the best estimate of

- $\theta_k := P(y^k)$, $k \in [K]$
- $\theta_k^{m,q_m} := P(x^{q_m,m}|y^k)$, $q_m \in [Q_m]$, $k \in [K]$, $m \in [M]$

Probability axioms:

- $\sum_{k=1}^{K} \theta_k = 1$
- $\sum_{q_m=1}^{Q_m} \theta_k^{m,q_m} = 1$ when fixing k and m

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Maximum Likelihood Estimate

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Maximum Likelihood Estimate

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}

MLE gives us the best estimates

$$\theta_k := n_k/N \tag{2}$$

$$\theta_k^{m,q_m} := n^{m,q_m}/n_k \tag{3}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

MLE (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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

MLE (Exercise 2)

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

$$n_A = 3 \qquad n_B = 2 \qquad n_C = 3$$
$$\theta_A = {}^3/8 \qquad \theta_B = {}^1/4 \qquad \theta_C = {}^3/8$$

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 (Exercise 2)

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

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

$$n_A = 3 \qquad n_B = 2 \qquad n_C = 3$$
$$\theta_A = {}^3/_8 \quad \theta_B = {}^1/_4 \quad \theta_C = {}^3/_8$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = {}^2/_3$	$\theta_A^{1,e} = {}^1/_3$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = {}^1/_2$	$\theta_B^{1,e} = {}^1/_2$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = {}^1/_3$	$\theta_C^{1,e} = {}^2/_3$
$n_A^{2,f} = ???$	$n_A^{2,g} = ???$	$\theta_A^{2,f} = ???$	$\theta_A^{2,g} = ???$
$n_B^{2,f} = ???$	$n_B^{2,g} = ???$	$\theta_B^{2,f} = ???$	$\theta_B^{2,g} = ???$
$n_C^{2,f} = ???$	$n_C^{2,g} = ???$	$\theta_C^{2,f} = ???$	$\theta_C^{2,g} = ???$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

MLE (Exercise 2)

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

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

$$n_A = 3 \qquad n_B = 2 \qquad n_C = 3$$
$$\theta_A = {}^3/_8 \qquad \theta_B = {}^1/_4 \qquad \theta_C = {}^3/_8$$

$n_A^{1,d} = 2$	$n_A^{1,e} = 1$	$\theta_A^{1,d} = {}^2/_3$	$\theta_A^{1,e} = {}^1/_3$
$n_B^{1,d} = 1$	$n_B^{1,e} = 1$	$\theta_B^{1,d} = {}^1/_2$	$\theta_B^{1,e} = {}^1/_2$
$n_C^{1,d} = 1$	$n_C^{1,e} = 2$	$\theta_C^{1,d} = {}^1/_3$	$\theta_C^{1,e} = {}^2/_3$
$n_A^{2,f} = ???$	$n_A^{2,g} = ???$	$\theta_A^{2,f} = ???$	$\theta_A^{2,g} = ???$
$n_B^{2,f} = ???$	$n_B^{2,g} = ???$	$\theta_B^{2,f} = ???$	$\theta_B^{2,g} = ???$
$n_C^{2,f} = ???$	$n_C^{2,g} = ???$	$\theta_C^{2,f} = ???$	$\theta_C^{2,g} = ???$

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

MLE (Solution 2)

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

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

$$n_A = 3 \qquad n_B = 2 \qquad n_C = 3$$
$$\theta_A = \tfrac{3}{8} \quad \theta_B = \tfrac{1}{4} \quad \theta_C = \tfrac{3}{8}$$

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Conditional Probabilities (Exercise 3)

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

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^M \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^M \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^M \theta_k^{m,q_m}. \qquad (4)$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Conditional Probabilities (Exercise 3)

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

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (4)$$

$$\theta_A = \tfrac{3}{8} \quad \theta_B = \tfrac{1}{4} \quad \theta_C = \tfrac{3}{8}$$

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

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Conditional Probabilities (Solution 3)

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

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (5)$$

$$\theta_A = {}^3/_8 \quad \theta_B = {}^1/_4 \quad \theta_C = {}^3/_8$$

$$
\begin{array}{cc|cc}
\theta_A^{1,d} = {}^2/_3 & \theta_A^{1,e} = {}^1/_3 & \theta_A^{2,f} = {}^1/_3 & \theta_A^{2,g} = {}^2/_3 \\
\theta_B^{1,d} = {}^1/_2 & \theta_B^{1,e} = {}^1/_2 & \theta_B^{2,f} = {}^1/_2 & \theta_B^{2,g} = {}^1/_2 \\
\theta_C^{1,d} = {}^1/_3 & \theta_C^{1,e} = {}^2/_3 & \theta_C^{2,f} = {}^2/_3 & \theta_C^{2,g} = {}^1/_3 \\
\end{array}
$$

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

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$
(d,f)	$1/12$	$1/18$	$1/12$
(e,h)	0	0	0

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Optimal Decision Rules (Exercise 4)

If $\ell(y^{k'}, y^k) = \mathbb{1}(y^{k'} \neq y^k)$, then (See Lecture 3+Check!)
$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \underset{y^k \in \mathscr{Y}}{\operatorname{argmax}} P'(y^k | \boldsymbol{x})$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Optimal Decision Rules (Exercise 4)

If $\ell(y^{k'}, y^k) = \mathbb{1}(y^{k'} \neq y^k)$, then (See Lecture 3+Check!)
$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \underset{y^k \in \mathscr{Y}}{\operatorname{argmax}} P'(y^k | \boldsymbol{x})$$

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$
(d, f)	$1/12$	$1/18$	$1/12$
(e, h)	0	0	0

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Optimal Decision Rules (Exercise 4)

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

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

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$
(d, f)	$1/12$	$1/18$	$1/12$
(e, h)	0	0	0

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

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = ???, \tag{6}$$

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

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = ???, \tag{7}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

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)	$1/12$	$1/18$	$1/12$
(e, h)	0	0	0

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

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \text{ either } A \text{ or } B, \tag{8}$$

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

$$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x}) = \text{ not well-defined }, \tag{9}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + MLE: Comments

- May lead to **indecision** and **not well-defined** $P(y|\boldsymbol{x})$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + MLE: Comments

- May lead to **indecision** and **not well-defined** $P(y|\boldsymbol{x})$

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$
(d,f)	$1/12$	$1/18$	$1/12$
(e,h)	0	0	0

- May suffer from small numbers of observations
  - ○ n_k: Number of training instances with label y^k
  - ○ n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + MLE: Comments

- May lead to **indecision** and **not well-defined** $P(y|\boldsymbol{x})$

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$
(d,f)	$1/12$	$1/18$	$1/12$
(e,h)	0	0	0

- May suffer from small numbers of observations
  - n_k: Number of training instances with label y^k
  - n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}
- Does not (naturally) take into account missing/partial data

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks

Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + Dirichlet Model (DM)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + Dirichlet Model (DM)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

DM gives Bayesian estimates

$$\theta_k := (n_k + \alpha_k)/(N+s) = (n_k + sf_k)/(N+s) \tag{10}$$

$$\theta_k^{m,q_m} := (n_k^{m,q_m} + \alpha_k^{m,q_m})/(n_k+s) = (n_k^{m,q_m} + sf_k^{m,q_m})/(n_k+s) \tag{11}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + Dirichlet Model (DM)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

DM gives Bayesian estimates

$$\theta_k := (n_k + \alpha_k)/(N+s) = (n_k + sf_k)/(N+s) \tag{10}$$

$$\theta_k^{m,q_m} := (n_k^{m,q_m} + \alpha_k^{m,q_m})/(n_k+s) = (n_k^{m,q_m} + sf_k^{m,q_m})/(n_k+s) \tag{11}$$

Advocators	α_v (= y^k or x^{m,q_m})	s
Haldane (1948)	0	0
Perks (1947)	$1/|\mathcal{V}|$	1
Jeffreys (1946, 1961)	$1/2$	$|\mathcal{V}|/2$
Bayes-Laplace	1	$|\mathcal{V}|$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + DM (Exercise 5)

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

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

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

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

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

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

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

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

$$n_A^{2,h} = 0 \qquad n_B^{2,h} = 0 \qquad n_C^{2,h} = 0$$
$$\theta_A^{2,h} = 1/12 \qquad \theta_B^{2,h} = 1/9 \qquad \theta_C^{2,h} = 1/12$$

Graphical Interpretation of Probabilistic Models   **Naïve Bayesian/Credal classifiers**   Decision Trees   Bayesian Neural Networks
Naïve Bayesian classifier   *Naïve Credal classifiers*

heudiasyc

Conditional Probabilities (Exercise 6)

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

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y^{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (12)$$

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

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

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Conditional Probabilities (Solution 6)

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

$$P(y^k|\boldsymbol{x}) = \frac{\theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}}{\sum_{y_{k'} \in \mathscr{Y}} \theta_{k'} \prod_{m=1}^{M} \theta_{k'}^{m,q_m}} \propto P'(y^k|\boldsymbol{x}) = \theta_k \prod_{m=1}^{M} \theta_k^{m,q_m}. \qquad (13)$$

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

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

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

$\boldsymbol{x}$	$P'(A|\boldsymbol{x})$	$P'(B|\boldsymbol{x})$	$P'(C|\boldsymbol{x})$	$y_\ell^{\boldsymbol{\theta}}(\boldsymbol{x})$
(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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + DM: Comments

- May lead to **indecision**, but can avoid **not well-defined** $P(y|x)$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Ba... Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + DM: Comments

- May lead to **indecision**, but can avoid **not well-defined** $P(y|\boldsymbol{x})$

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

- May suffer from small numbers of observations
  - n_k: Number of training instances with label y^k
  - n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NBC + DM: Comments

- May lead to **indecision**, but can avoid **not well-defined** $P(y|\boldsymbol{x})$

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

- May suffer from small numbers of observations
  - n_k: Number of training instances with label y^k
  - n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}
- Does not (naturally) take into account missing/partial data

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers
  - Naïve Bayesian classifier
  - **Naïve Credal classifiers**

- Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Naïve Credal classifiers (NCC)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Naïve Credal classifiers (NCC)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

Imprecise Dirichlet model (IDM) gives

$$\underline{\theta}_k := n_k/(N+s) \qquad (14) \qquad \overline{\theta}_k := (n_k+s)/(N+s) \qquad (16)$$

$$\underline{\theta}_k^{m,q_m} := n_k^{m,q_m}/(n_k+s) \qquad (15) \qquad \overline{\theta}_k^{m,q_m} := (n_k^{m,q_m}+s)/(n_k+s) \qquad (17)$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier Naïve Credal classifiers

heudiasyc

Naïve Credal classifiers (NCC)

Basic setup: Given training data $\boldsymbol{D} = \{(y_1, \boldsymbol{x}_1), \ldots, (y_N, \boldsymbol{x}_N)\}$, count

- n_k: Number of training instances with label y^k
- n_k^{m,q_m}: Number of instances with $Y = y^k$ and feature $X^m = x^{m,q_m}$

Imprecise Dirichlet model (IDM) gives

$$\underline{\theta}_k := n_k/(N+s) \qquad (14) \qquad\qquad \overline{\theta}_k := (n_k+s)/(N+s) \qquad (16)$$

$$\underline{\theta}_k^{m,q_m} := n_k^{m,q_m}/(n_k+s) \qquad (15) \qquad \overline{\theta}_k^{m,q_m} := (n_k^{m,q_m}+s)/(n_k+s) \qquad (17)$$

IDM + ϵ regularization [2]

$$\underline{\theta}_k := (n_k+s\underline{\epsilon}_k)/(N+s) \qquad (18) \qquad \overline{\theta}_k := (n_k+s\overline{\epsilon}_k)/(N+s) \qquad (20)$$

$$\underline{\theta}_k^{m,q_m} := (n_k^{m,q_m}+s\underline{\epsilon}_k^{m,q_m})/(n_k+s) \qquad (19) \qquad \overline{\theta}_k^{m,q_m} := (n_k^{m,q_m}+s\overline{\epsilon}_k^{m,q_m})/(n_k+s) \qquad (21)$$

Graphical Interpretation of Probabilistic Models   **Naïve Bayesian/Credal classifiers**   Decision Trees   Bayesian Neural Networks
Naïve Bayesian classifier   **Naïve Credal classifiers**

heudiasyc

Interval Conditional Probabilities

Given a query instance $\boldsymbol{x} = (x^{q_1,1}, x^{q_2,2}, \ldots, x^{q_M,M})$, we have

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier ***Naïve Credal classifiers***

heudiasyc

Interval Conditional Probabilities

Given a query instance $\boldsymbol{x} = (x^{q_1,1}, x^{q_2,2}, \ldots, x^{q_M,M})$, we have

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

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

$$
\mathscr{P}(\mathscr{Y}|\boldsymbol{x}) := \left\{ P(\mathscr{Y}|\boldsymbol{x}) | P(y^k|\boldsymbol{x}) \in [\underline{P}(y^k|\boldsymbol{x}), \overline{P}(y^k|\boldsymbol{x})], \sum_{k=1}^{K} P(y^k|\boldsymbol{x}) = 1 \right\}.
$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Interval Conditional Probabilities (Exercise 7)

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier **Naïve Credal classifiers**

heudiasyc

Interval Conditional Probabilities (Exercise 7)

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

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

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

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

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Interval Conditional Probabilities (Solution 7)

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Interval Conditional Probabilities (Solution 7)

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

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

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

$$\begin{array}{cc|cc}
n_A^{1,d} = 2 & n_A^{1,e} = 1 & n_A^{2,f} = 1 & n_A^{2,g} = 2 \\
n_B^{1,d} = 1 & n_B^{1,e} = 1 & n_B^{2,f} = 1 & n_B^{2,g} = 1 \\
n_C^{1,d} = 1 & n_C^{1,e} = 2 & n_C^{2,f} = 2 & n_C^{2,g} = 1 \\
\end{array}$$

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Making Set-Valued Predictions (Recap)

For **each instance** x, let

- $\theta \longleftarrow P(\mathcal{Y}|x)$ and $\Theta \longleftarrow \mathcal{P}(\mathcal{Y}|x)$

E-admissibility Rule:

- An optimal prediction is

$$\mathbf{Y}_{\ell,\Theta}^{E} = \{y \in \mathcal{Y} \,|\, \exists\, \theta \in \Theta \text{ s.t. } y = y_{\ell}^{\theta}\}.$$

- Computation: Solving linear programs [10], etc.

Maximality Rule:

- An optimal prediction is

$$\mathbf{Y}_{\ell,\Theta}^{M} = \{y \in \mathcal{Y} \,|\, \nexists\, y' \text{ s.t. } y' \succ_{\ell,\Theta} y\}.$$

- Computation: Solving linear programs [10], Iterating over the extreme points of Θ [10], exploiting the properties of NCC [3].

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

Illustrative Examples → Lines of Code Would be Useful!

f /techindustan   🐦 /techindustan   📷 /techindustan

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NCC: Comments

NCC inherits properties of IDM [3]:

- May lead to **set-valued predictions**
- ϵ-regularization can avoid **not well-defined** $P(y|\boldsymbol{x})$

- May provide reliable interval probabilities when seeing small numbers of observations
    - n_k: Number of training instances with label y^k
    - n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** Decision Trees Bayesian Neural Networks
Naïve Bayesian classifier *Naïve Credal classifiers*

heudiasyc

NCC: Comments

NCC inherits properties of IDM [3]:

- May lead to **set-valued predictions**
- ϵ-regularization can avoid **not well-defined** $P(y|\boldsymbol{x})$

- May provide reliable interval probabilities when seeing small numbers of observations
    - n_k: Number of training instances with label y^k
    - n_k^{m,q_m}: Number of training instances with label y^k and feature X^m takes value x^{m,q_m}
- Provide tools to (naturally) take into account missing/partial data
    - Naive solutions are computationally expensive (in exponential time)
    - More efficient (polynomial-time) procedure exists

Graphical Interpretation of Probabilistic Models   **Naïve Bayesian/Credal classifiers**   Decision Trees   Bayesian Neural Networks
Naïve Bayesian classifier   *Naïve Credal classifiers*

heudiasyc

NCC: Technical Details + Performance

Learning Reliable Classifiers From Small or Incomplete Data Sets: The Naive Credal Classifier 2

Giorgio Corani   GIORGIO@IDSIA.CH
Marco Zaffalon   ZAFFALON@IDSIA.CH
IDSIA
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
CH-6928 Manno (Lugano), Switzerland

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers

- Decision Trees
  - ○ Decision Trees
  - ○ Credal Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   Credal Decision Trees

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers

- Decision Trees
  - Decision Trees
  - Credal Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

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.

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

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):

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

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):

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

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

Chain rule gives us

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

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):

$$P(Y, \boldsymbol{X}) = P(Y|\text{pa}(Y)) \prod_{m=1}^{M} P(X^m|\text{pa}(X^m)).$$
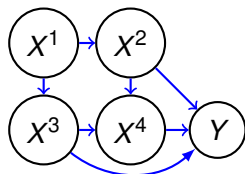
- $\text{pa}(Y) = \{X^2, X^3, X^4\}$
- $\text{pa}(X^1) = \emptyset$, $\text{pa}(X^2) = \{X^1\}$
- $\text{pa}(X^3) = \{X^1\}$
- $\text{pa}(X^4) = \{X^2, X^3\}$

Chain rule gives us

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

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Classification Task (Exercise + Solution 9)

Prove that

$$P(y|\boldsymbol{x}) = P(y|\mathrm{pa}(y)). \tag{22}$$

Graphical Interpretation of Probabilistic Models   **Naïve Bayesian/Credal classifiers**   **Decision Trees**   **Bayesian Neural Networks**
Decision Trees   *Credal Decision Trees*

heudiasyc

Classification Task (Exercise + Solution 9)

Prove that

$$P(y|\boldsymbol{x}) = P(y|\text{pa}(y)). \tag{22}$$

We have

$$P(y|\boldsymbol{x}) = \frac{P(y, \boldsymbol{x})}{\sum_{y' \in \mathcal{Y}} P(y', \boldsymbol{x})} \tag{23}$$

$$= \frac{P(y|\text{pa}(y)) \prod_{m=1}^{M} P(x^m|\text{pa}(x^m))}{\sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y')) \prod_{m=1}^{M} P(x^m|\text{pa}(x^m))} \tag{24}$$

$$= \frac{\prod_{m=1}^{M} P(x^m|\text{pa}(x^m)) P(y|\text{pa}(y))}{\prod_{m=1}^{M} P(x^m|\text{pa}(x^m)) \sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y'))} \tag{25}$$

$$= \frac{P(y|\text{pa}(y))}{\sum_{y' \in \mathcal{Y}} P(y'|\text{pa}(y'))} \tag{26}$$

$$= P(y|\text{pa}(y)). \tag{27}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Classification Task: Comments

$P(Y|\mathbf{X})$ is extracted using **Bayes' theorem**

$$P(y|\mathbf{x}) = P(y|\text{pa}(y)). \tag{28}$$

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Classification Task: Comments

$P(Y|\boldsymbol{X})$ is extracted using **Bayes' theorem**

$$P(y|\boldsymbol{x}) = P(y|\text{pa}(y)). \tag{28}$$

- Features outside pa(Y) are redundant
- To solve the classification task, we only need $P(Y|\text{pa}(Y))$
- Commonly used assumption: pa(Y) = ($X^1, \ldots, X^M$)

Classification Task: Comments

$P(Y|\boldsymbol{X})$ is extracted using **Bayes' theorem**

$$P(y|\boldsymbol{x}) = P(y|\text{pa}(y)). \qquad (28)$$

- Features outside pa(Y) are redundant
- To solve the classification task, we only need $P(Y|\text{pa}(Y))$
- Commonly used assumption: pa(Y) = $(X^1, \ldots, X^M)$
- $P(Y|\text{pa}(Y))$ can be defined either globally or locally:
  - Logistic regression, neural nets, etc., define $P(Y|\text{pa}(Y))$ globally
  - Decision tree, model trees, etc., define $P(Y|\text{pa}(Y))$ locally
  - Decision tree does not require pa(Y) = $(X^1, \ldots, X^M)$

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks

Decision Trees   *Credal Decision Trees*

heudiasyc

Decision Trees: Example [11]
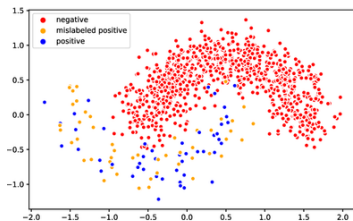
(a) Half-moons data set (ground truth)
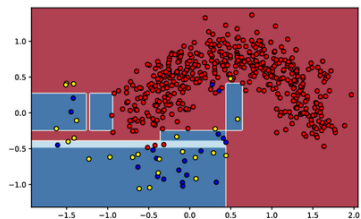
(b) PU-Hellinger Decision Tree on the test set
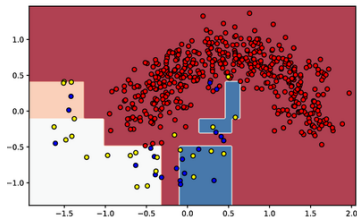
(c) Hellinger Decision Tree on the test set
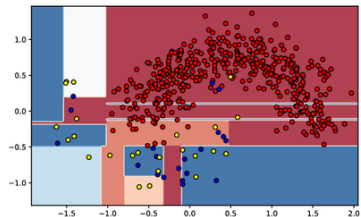
(d) CART on the test set

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Decision Trees: (Informal+Probabilistic) Definition

A decision trees is

- a collection of non-overlapping leaves $L_1, \ldots, L_H$
- where $L_1 \cap \ldots \cap L_H = \mathcal{X}$
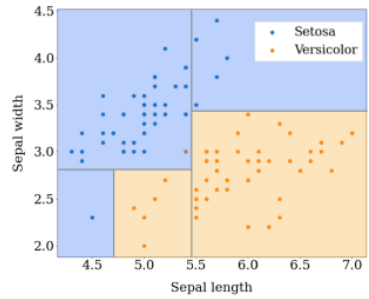- and each leaf L_h has its own $P_h(Y|\text{pa}(Y))$

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Decision Trees: (Informal+Probabilistic) Definition

A decision trees is

- a collection of non-overlapping leaves $L_1, \ldots, L_H$
- where $L_1 \cap \ldots \cap L_H = \mathcal{X}$
- and each leaf L_h has its own $P_h(Y|\text{pa}(Y))$

Learning an optimal decision tree from training data

- can be extremely hard (due to huge numbers of possible trees)
- and is often done approximately (top-down induction, bottom-up induction, etc.)

Top-Down Induction (Example) [4]
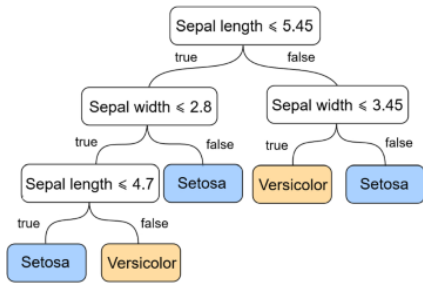
(a) Tree visualization

(b) Partitioning visualization

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Top-down induction: Steps

Basic Setup:

- Training data $\boldsymbol{D} = \{(y^n, \boldsymbol{x}^n)|n \in [N]\}$
- Local hypothesis space $P(Y|\text{pa}(Y)) \in \mathscr{P}(Y|\text{pa}(Y))$
- An uncertainty measure U or a loss function $\ell \longleftarrow$ assess how good/bad each **local classifier** is

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Top-down induction: Steps

Basic Setup:

- Training data $\boldsymbol{D} = \{(y^n, \boldsymbol{x}^n) | n \in [N]\}$
- Local hypothesis space $P(Y|\text{pa}(Y)) \in \mathscr{P}(Y|\text{pa}(Y))$
- An uncertainty measure U or a loss function $\ell \longleftarrow$ assess how good/bad each **local classifier** is

Induction protocol:

- Recursively partition the feature space $\mathscr{X}$
- From the current node, choose the best split which improves the evaluation criterion
- Evaluation criteria: Information gain, entropy, Gini score, etc.,
- Stopping criteria: No more gain on evaluation criterion U or ℓ

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Splitting Criteria: Entropy (Frequentist)

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

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

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Splitting Criteria: Entropy (Frequentist)

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

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

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc
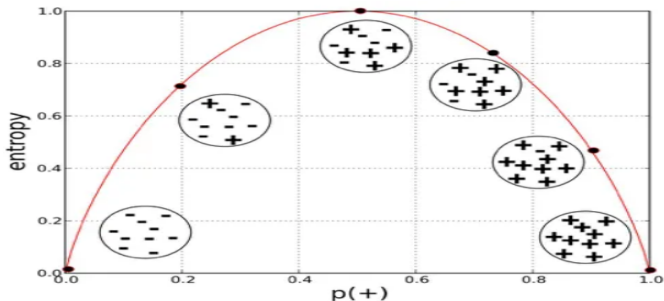
Splitting Criteria: Entropy (Frequentist)

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

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

- For each possible split $\boldsymbol{D}_h = \boldsymbol{D}_h^1 \cup \boldsymbol{D}_h^2$, its entropy is

$$U_E(\boldsymbol{D}_h^1 \cup \boldsymbol{D}_h^2) = U_E(P(\mathscr{Y}|\boldsymbol{D}_h)) + U_E(P(\mathscr{Y}|\boldsymbol{D}_h)).$$

Compute Entropy (Exercise 10)

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

$$U_E(P(y|\boldsymbol{D}_h)) = - \sum_{y \in \mathcal{Y}} P(y|\boldsymbol{D}_h) \log_2\left(P(y|\boldsymbol{D}_h)\right).$$
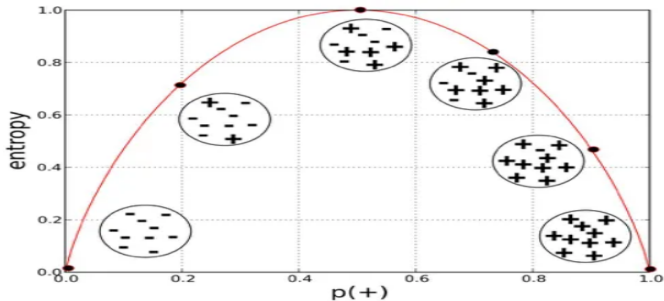
- Entropy of the bottom left node is ???
- Entropy of the top node is ???

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Compute Entropy (Solution 10)

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

$$U_E(P(y|\boldsymbol{D}_h)) = -\sum_{y\in\mathscr{Y}} P(y|\boldsymbol{D}_h)\log_2\left(P(y|\boldsymbol{D}_h)\right).$$
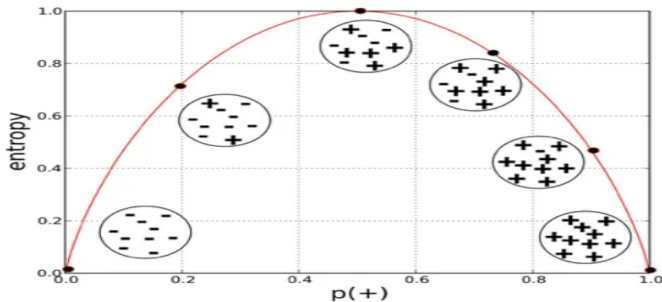
- Entropy of the bottom left node is 0
- Entropy of the top node is $-0.5\log_2(0.5) + 0.5\log_2(0.5) = 1$

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   *Credal Decision Trees*

heudiasyc

Splitting Criteria: Bayesian

In principle, we can employ Dirichlet models (DM) to

- derive Bayesian estimates of $P(\mathscr{Y}|\boldsymbol{D}_h)$ and/or $U(P(y|\boldsymbol{D}_h))$
- and modify the top-down induction steps.

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Splitting Criteria: Bayesian

In principle, we can employ Dirichlet models (DM) to

- derive Bayesian estimates of $P(\mathscr{Y}|\boldsymbol{D}_h)$ and/or $U(P(y|\boldsymbol{D}_h))$
- and modify the top-down induction steps.

- I haven't seen such decision decision trees
- I hope I can find some reference soon ...

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   **Decision Trees**   Bayesian Neural Networks
Decision Trees   Credal Decision Trees

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers

- Decision Trees
  - Decision Trees
  - Credal Decision Trees

- Bayesian Neural Networks

Graphical Interpretation of Probabilistic Models   **Naïve Bayesian/Credal classifiers**   **Decision Trees**   **Bayesian Neural Networks**
Decision Trees   *Credal Decision Trees*

heudiasyc

Where and How to be Imprecise?

In principle, we can employ Imprecise Dirichlet models (IDM) to

- derive interval estimates of $P(\mathcal{Y}|\boldsymbol{D}_h)$ and/or $U(P(y|\boldsymbol{D}_h))$
- and modify the top-down induction steps.

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**
Decision Trees *Credal Decision Trees*

heudiasyc

Where and How to be Imprecise?

In principle, we can employ Imprecise Dirichlet models (IDM) to

- derive interval estimates of $P(\mathcal{Y}|\boldsymbol{D}_h)$ and/or $U(P(y|\boldsymbol{D}_h))$
- and modify the top-down induction steps.

Credal Decision Trees [1, 7]

- Use IDM to derive interval estimates $\mathscr{P}(\mathcal{Y}|\boldsymbol{D}_h)$ of $P(\mathcal{Y}|\boldsymbol{D}_h)$
- Seek the highest entropy

$$U(\mathscr{P}(\mathcal{Y}|\boldsymbol{D}_h)) = \max_{P \in \mathscr{P}} U(P(\mathcal{Y}|\boldsymbol{D}_h)). \tag{29}$$

- Each leaf is equipped a $\mathscr{P}(\mathcal{Y}|\boldsymbol{L}_h) \rightarrow$ Precise predictions.

Credal Decision Trees: Performance [7]

Table 7

Average accuracy results of IGT, IGRT, CDT and CCDT when are applied on data sets with percentage of random noise equal to 0%, 5%, 10%, 20% and 30%.

Tree	noise 0%	noise 5%	noise 10%	noise 20%	noise 30%
IGT	78.96	78.00	77.49	76.02	74.76
IGRT	78.97	78.09	77.66	76.38	75.14
CDT	79.56	78.84	78.65	77.51	76.72
CCDT	79.63	78.99	78.66	77.57	76.74

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**

heudiasyc

Outline

- Graphical Interpretation of Probabilistic Models

- Naïve Bayesian/Credal classifiers

- Decision Trees

- Bayesian Neural Networks

Artificial neural networks vs Bayesian Neural Networks
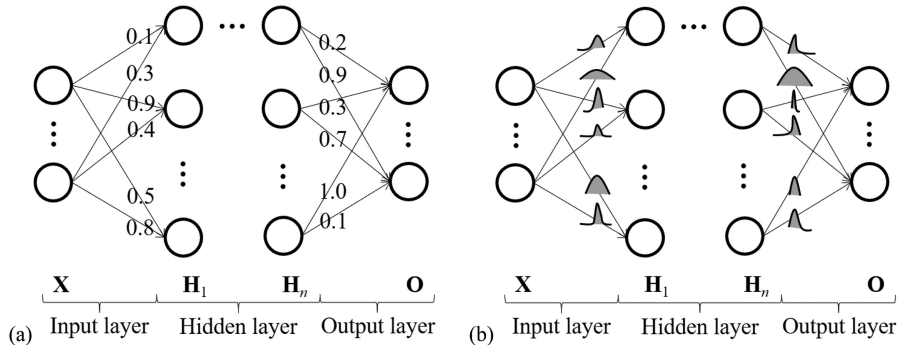
Graphical interpretation of (a) ANN and (b) BNN

Inference Problems [6]

Algorithm 1 Inference procedure for a BNN.

Define $p(\theta|D) = \dfrac{p(D_y|D_x, \theta)\, p(\theta)}{\int_\theta p(D_y|D_x, \theta')\, p(\theta')\, d\theta'}$;

for $i = 0$ **to** N **do**

  Draw $\theta_i \sim p(\theta|D)$;

  $y_i = \Phi_{\theta_i}(x)$;

end for

return $Y = \{y_i | i \in [0, N)\}$, $\Theta = \{\theta_i | i \in [0, N)\}$;

Graphical Interpretation of Probabilistic Models **Naïve Bayesian/Credal classifiers** **Decision Trees** **Bayesian Neural Networks**

heudiasyc

Inference Problems [6]

Algorithm 1 Inference procedure for a BNN.

Define $p(\theta|D) = \dfrac{p(D_y|D_x, \theta)\, p(\theta)}{\int_\theta p(D_y|D_x, \theta')\, p(\theta')\, d\theta'}$;

for $i = 0$ **to** N **do**

  Draw $\theta_i \sim p(\theta|D)$;

  $y_i = \Phi_{\theta_i}(x)$;

end for

return $Y = \{y_i \,|\, i \in [0, N)\}$, $\Theta = \{\theta_i \,|\, i \in [0, N)\}$;

- We need some way to aggregate the set of outputs
- You might want to try with MLE (Frequentist), DM (Bayesian), IDM (IP), etc.

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   **Bayesian Neural Networks**

heudiasyc

Original software publication

BNNpriors: A library for Bayesian neural network inference with different prior distributions

Vincent Fortuin [a,1,*], Adrià Garriga-Alonso [b,1], Mark van der Wilk [c,2], Laurence Aitchison [d,2]

[a] *ETH Zürich, Zürich, Switzerland*
[b] *University of Cambridge, Cambridge, UK*
[c] *Imperial College London, London, UK*
[d] *University of Bristol, Bristol, UK*

A R T I C L E   I N F O

A B S T R A C T

Bayesian neural networks have shown great promise in many applications where calibrated uncertainty estimates are crucial and can often also lead to a higher predictive performance. However, it remains challenging to choose a good prior distribution over their weights. While isotropic Gaussian priors are often chosen in practice due to their simplicity, they do not reflect our true prior beliefs well and can lead to suboptimal performance. Our new library, *BNNpriors*, enables state-of-the-art Markov Chain Monte Carlo inference on Bayesian neural networks with a wide range of predefined priors, including heavy-tailed ones, hierarchical ones, and mixture priors. Moreover, it follows a modular approach that eases the design and implementation of new custom priors. It has facilitated foundational discoveries on the nature of the cold posterior effect in Bayesian neural networks and will hopefully catalyze future research as well as practical applications in this area.

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc

References I

[1] J. Abellán and S. Moral.
Building classification trees using the total uncertainty criterion.
International Journal of Intelligent Systems, 18(12):1215–1225, 2003.

[2] A. Antonucci, G. Corani, and S. Bernaschina.
Active learning by the naive credal classifier.
In *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models (PGM)*, pages 3–10, 2012.

[3] G. Corani and M. Zaffalon.
Learning reliable classifiers from small or incomplete data sets: The naive credal classifier 2.
Journal of Machine Learning Research, 9(4), 2008.

[4] V. G. Costa and C. E. Pedreira.
Recent advances in decision trees: An updated survey.
Artificial Intelligence Review, 56(5):4765–4800, 2023.

[5] N. Friedman, D. Geiger, and M. Goldszmidt.
Bayesian network classifiers.
Machine learning, 29:131–163, 1997.

[6] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun.
Hands-on bayesian neural networksa tutorial for deep learning users.
IEEE Computational Intelligence Magazine, 17(2):29–48, 2022.

[7] C. J. Mantas and J. Abellán.
Analysis and extension of decision trees based on imprecise probabilities: Application on noisy data.
Expert Systems with Applications, 41(5):2514–2525, 2014.

Graphical Interpretation of Probabilistic Models   Naïve Bayesian/Credal classifiers   Decision Trees   Bayesian Neural Networks

heudiasyc

References II

[8]   C. J. Mantas and J. Abellan.
      Credal-c4. 5: Decision tree based on imprecise probabilities to classify noisy data.
      Expert Systems with Applications, 41(10):4625–4637, 2014.

[9]   V.-L. Nguyen, Y. Yang, and C. P. de Campos.
      Probabilistic multi-dimensional classification.
      In *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 1522–1533, 2023.

[10] V.-L. Nguyen, H. Zhang, and S. Destercke.
      Learning sets of probabilities through ensemble methods.
      In *Proceedings of the 17th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, 2023.

[11] C. Ortega Vázquez, S. vanden Broucke, and J. De Weerdt.
      Hellinger distance decision trees for pu learning in imbalanced data sets.
      Machine Learning, pages 1–32, 2023.