Classification

- Complete data \( \mathcal{D} = \{(x_0^{(i)}, x^{(i)})\}_{i=1}^n \) about class variable \( X_0 \) and (discrete) features \( \mathbf{X} := (X_1, \ldots, X_n) \)

- Class label \( x_0^* \in \Omega_{X_0} \) of a new instance \( \tilde{x} \) of the features?

- Probabilistic approaches learn \( P(X_0, \mathbf{X}) \) from \( \mathcal{D} \)

- Optimal class has the highest posterior \( x_0^* := \arg \max_{x_0} P(x_0|\tilde{x}) \)

- Equivalently, dominance test: \( \forall x_0', x_0'' \in \Omega_{X_0} \) check

\[
\frac{P(x_0' | \tilde{x})}{P(x_0'' | \tilde{x})} = \frac{P(x_0', \tilde{x})}{P(x_0'', \tilde{x})} > 1
\]

- \( x_0^* \) is the only undominated class

- Posterior probabilities \( \propto \) joint probabilities
Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^n$ about class variable $X_0$ and (discrete) features $\mathbf{X} := (X_1, \ldots, X_n)$
- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{\mathbf{x}}$ of the features?

  - Probabilistic approaches learn $P(X_0, \mathbf{X})$ from $\mathcal{D}$
    - Optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{\mathbf{x}})$
  - Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check
    \[
    \frac{P(x_0' | \tilde{\mathbf{x}})}{P(x_0'' | \tilde{\mathbf{x}})} = \frac{P(x_0' , \tilde{\mathbf{x}})}{P(x_0'' , \tilde{\mathbf{x}})} > 1
    \]
    - $x_0^*$ is the only undominated class

- Posterior probabilities $\propto$ joint probabilities
Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, x^{(i)})\}_{i=1}^n$ about class variable $X_0$ and (discrete) features $X := (X_1, \ldots, X_n)$

- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{x}$ of the features?

- **Probabilistic approaches** learn $P(X_0, X)$ from $\mathcal{D}$

- Optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0|\tilde{x})$

- Equivalently, dominance test: $\forall x'_0, x''_0 \in \Omega_{X_0}$ check

\[
\frac{P(x'_0|\tilde{x})}{P(x''_0|\tilde{x})} \geq \frac{P(x'_0, \tilde{x})}{P(x''_0, \tilde{x})} > 1
\]

- $x_0^*$ is the only undominated class

- Posterior probabilities $\propto$ joint probabilities
Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, x^{(i)})\}_{i=1}^n$ about class variable $X_0$ and (discrete) features $X := (X_1, \ldots, X_n)$

- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{x}$ of the features?

- Probabilistic approaches learn $P(X_0, X)$ from $\mathcal{D}$

  - optimal class has the highest posterior $x_0^* := \arg \max_{x_0} P(x_0|\tilde{x})$

- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check

\[
\frac{P(x_0'|\tilde{x})}{P(x_0''|\tilde{x})} = \frac{P(x_0',\tilde{x})}{P(x_0'',\tilde{x})} > 1
\]

- $x_0^*$ is the only undominated class

- Posterior probabilities $\propto$ joint probabilities
Classification

- Complete data \( \mathcal{D} = \{(x_0^{(i)}, x^{(i)})\}_{i=1}^n \) about class variable \( X_0 \) and (discrete) features \( X := (X_1, \ldots, X_n) \)

- Class label \( x_0^* \in \Omega_{X_0} \) of a new instance \( \tilde{x} \) of the features?

- Probabilistic approaches learn \( P(X_0, X) \) from \( \mathcal{D} \)
  - optimal class has the highest posterior \( x_0^* := \arg \max_{x_0} P(x_0 | \tilde{x}) \)

- Equivalently, dominance test: \( \forall x_0', x_0'' \in \Omega_{X_0} \) check
  \[
  \frac{P(x_0' | \tilde{x})}{P(x_0'' | \tilde{x})} = \frac{P(x_0', \tilde{x})}{P(x_0'', \tilde{x})} > 1
  \]
  - \( x_0^* \) is the only undominated class

- Posterior probabilities \( \propto \) joint probabilities
Classification

- Complete data $\mathcal{D} = \{(x_0^{(i)}, x^{(i)})\}_{i=1}^n$ about class variable $X_0$ and (discrete) features $X := (X_1, \ldots, X_n)$

- Class label $x_0^* \in \Omega_{X_0}$ of a new instance $\tilde{x}$ of the features?

- Probabilistic approaches learn $P(X_0, X)$ from $\mathcal{D}$
  
  optimal class has the highest posterior $x_0^* := \arg\max_{x_0} P(x_0 | \tilde{x})$

- Equivalently, dominance test: $\forall x_0', x_0'' \in \Omega_{X_0}$ check
  
  $$\frac{P(x_0' | \tilde{x})}{P(x_0'' | \tilde{x})} = \frac{P(x_0', \tilde{x})}{P(x_0'', \tilde{x})} > 1$$

  $x_0^*$ is the only undominated class

- Posterior probabilities $\propto$ joint probabilities
Classification with Bayesian networks

- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, \mathbf{X})$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, \mathbf{x}) = \prod_{i=0}^{n} P(x_i|\pi_i)$$

- Dominance test rewrites as

$$\frac{P(x'_0, \tilde{x})}{P(x'_0, \tilde{x})} = \prod_{i=0}^{n} \frac{P(x'_i|\pi'_i)}{P(x''_i|\pi''_i)} > 1$$

- Factors not including $X_0$ are equal to one focusing on the Markov blanket of $X_0$
Classification with Bayesian networks

- Bayesian networks: a graph $G$ to depict conditional independencies in $(X_0, X)$
- $G$ induces a **factorization** in the joint

\[
P(x_0, x) = \prod_{i=0}^{n} P(x_i | \pi_i)
\]

- Dominance test rewrites as

\[
\frac{P(x'_0, \tilde{x})}{P(x''_0, \tilde{x})} = \prod_{i=0}^{n} \frac{P(x'_i | \pi'_i)}{P(x''_i | \pi''_i)} > 1
\]

- Factors not including $X_0$ are equal to one focusing on the **Markov blanket** of $X_0$
Classification with Bayesian networks

- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, X)$
- $\mathcal{G}$ induces a factorization in the joint

$$
P(x_0, x) = \prod_{i=0}^{n} P(x_i | \pi_i)
$$

- Dominance test rewrites as

$$
\frac{P(x'_0, \tilde{x})}{P(x'_0, \tilde{x})} = \prod_{i=0}^{n} \frac{P(x'_i | \pi'_i)}{P(x''_i | \pi''_i)} > 1
$$

- Factors not including $X_0$ are equal to one focusing on the Markov blanket of $X_0$
Classification with Bayesian networks

- Bayesian networks: a graph $\mathcal{G}$ to depict conditional independencies in $(X_0, X)$
- $\mathcal{G}$ induces a factorization in the joint

$$P(x_0, x) = \prod_{i=0}^{n} P(x_i | \pi_i)$$

- Dominance test rewrites as

$$\frac{P(x'_0, \tilde{x})}{P(x'_0, \tilde{x})} = \prod_{i=0}^{n} \frac{P(x'_i | \pi'_i)}{P(x''_i | \pi''_i)} > 1$$

- Factors not including $X_0$ are equal to one focusing on the Markov blanket of $X_0$
Classification with Bayesian networks

- Bayesian networks: a graph \( \mathcal{G} \) to depict conditional independencies in \((X_0, X)\)
- \( \mathcal{G} \) induces a factorization in the joint

\[
P(X_0, X) = \prod_{i=0}^{n} P(X_i | \pi_i)
\]

- Dominance test rewrites as

\[
\prod_{X_i \in \text{Blank}(X_0)} \frac{P(X'_i | \pi'_i)}{P(X''_i | \pi''_i)} > 1
\]

- Factors not including \( X_0 \) are equal to one focusing on the Markov blanket of \( X_0 \)
Classification with credal networks

- Few data ⇒ unreliable learning of $P(X_i | \pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $G$)
- How to classify instances?
- $x'_0$ dominates $x''_0$ iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0, X) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i | \pi'_i)}{P(x''_i | \pi''_i)} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data \(\Rightarrow\) unreliable learning of \(P(X_i|\pi_i)\)
- More reliable with sets of prob functions
- A credal set of joint \(P(X_0, X)\), instead of a single \(P(X_0, X)\) (all Bayesian nets over \(G\))
- How to classify instances?
- \(x'_0\) dominates \(x''_0\) iff this happens for each Bayesian net (maximality), i.e.,

\[
\min_{P(X_0, X) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i|\pi'_i)}{P(x''_i|\pi''_i)} > 1
\]

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data ⇒ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
  - A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $G$)
  - How to classify instances?
  - $x_0'$ dominates $x_0''$ iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0, X) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x_i'|\pi_i')}{P(x_i''|\pi_i'')} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $\mathcal{G}$)

How to classify instances?
- $x'_0$ dominates $x''_0$ iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0, X) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i|\pi'_i)}{P(x''_i|\pi''_i)} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data ⇒ unreliable learning of $P(X_i | \pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $G$)
- How to classify instances?
  - $x'_0$ dominates $x''_0$ iff this happens for each Bayesian net (maximality), i.e.,
    \[
    \min_{P(X_0, x) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i | \pi'_i)}{P(x''_i | \pi''_i)} > 1
    \]
- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
- $x'_0$ dominates $x''_0$ iff this happens for each Bayesian net (maximality), i.e.,

$$\min_{P(X_0, X) \in P(X_0)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i|\pi'_i)}{P(x''_i|\pi''_i)} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
Classification with credal networks

- Few data $\Rightarrow$ unreliable learning of $P(X_i|\pi_i)$
- More reliable with sets of prob functions
- A credal set of joint $P(X_0, X)$, instead of a single $P(X_0, X)$ (all Bayesian nets over $\mathcal{G}$)
- How to classify instances?
  - $x'_0$ dominates $x''_0$ iff this happens for each Bayesian net (maximality), i.e.,
  
  $$\min_{P(X_0, X) \in P(X_0, X)} \prod_{X_i \in \text{Blank}(X_0)} \frac{P(x'_i|\pi'_i)}{P(x''_i|\pi''_i)} > 1$$

- Not always a single optimal class, can be also a set of undominated classes
- This is a credal classifier possibly assigning multiple classes to test instance
TREE-AUGMENTED NAIVE

Class

Feat1 -> Feat2
Feat2 -> Feat3
Feat3 -> Feat4
ONE-DEPENDENCE (ODE)
Using IDM

Simplest case: naive with (local) IDM

\[
\frac{P(c', f_1, f_2)}{P(c'', f_1, f_2)} = \frac{P(c')P(f_1|c')P(f_2|c')}{P(c'')P(f_1|c')P(f_2|c'')} > 1
\]

- IDM: \( P(c) = \frac{n(c)+st(c)}{N+s} \) with \( 0 \leq t(c) \leq 1, \sum_c t(c) = 1 \)

\[
\min_t \frac{[n + st(c')]}{[n(c'') + st(c'')]} \cdot \ldots > 1
\]

- Efficient optimization
  NAIVE CREDAL CLASSIFIER (Zaffalon & Corani)

http://www.idsia.ch/ giorgio/jncc2.html
Accuracy (% of correct classification) is not the only descriptor

- Single-Accuracy (accuracy if a single class is returned)
- Set-Accuracy (accuracy if multiple classes are returned)
- Determinacy (% number of instances with a single class)
- Average output size (average number of classes)
- Bayes-I (accuracy of the Bayesian if credal indeterminate)
- Bayes-D = Single-Accuracy
- Zaffalon et al. proposed a new utility-based measure (ISIPTA ’09)