Naive Credal Classifier (with IDM, [4])

Credal Classifiers (CCs)

- **Classification**: Class variable \( C \) with generic value \( c \in \mathcal{C} \)
- **Features**: \( \{ \mathcal{F}_1, \ldots, \mathcal{F}_n \} \), values \( j_i \in \mathcal{F}_i, i = 1, \ldots, m \)
- **Bayesian Classifiers**: Learn joint distribution \( P(C, \mathcal{F}) \)
  - Assign to \( \mathcal{F} \) the most probable class label \( \arg \max_{c \in \mathcal{C}} P(c|\mathcal{F}) \)
- **Naive Credal Classifiers**: Learn joint credal set \( P(C, \mathcal{F}) \)
  - Set of optimal classes (e.g., according to maximality)
    \[ \{ c' \in \mathcal{C} | \mathcal{F} \in \mathcal{P} : P(c'|\mathcal{F}) > P(c|\mathcal{F}) \} \]
    - This defines a credal classifier, i.e.,
      \[ (\mathcal{F}_1, \ldots, \mathcal{F}_n) \rightarrow \mathcal{C}^n \]
    - May return more than a single class label!

Bayesian learning extended to imprecision

- Set of Dirichlet modelling prior near-ignorance
- Bounds of the posterior is
  \[ P(c|\mathcal{F}) \in \left[ \frac{n(c)}{N + 2}, \frac{n(c) + 1}{N + 3} \right] \]
- \( n(c) \) counting function
- Real parameter \( \alpha \) as equivalent sample size
- Typical choices \( \alpha \in [1, 2] \)
- If \( n(c) \rightarrow +\infty \), vacuous intervals
- If \( n(c) \rightarrow 0 \), ML estimator

Concepts for imprecision

- **IDM (optimality)**: Accuracy over instances classified
  \[ \alpha \in [0, 1] \]
  - \( \alpha = 0 \): maximum accuracy
  - \( \alpha = 1 \): maximum optimality

Optimization problem

- **IMD vs. Likelihood**

Learning Credal Sets (Likelihood)

- Frequentist (ML) learning extended to imprecision
- With complete (multinomial) data, likelihood is unimodal
- Replace the single ML estimator, with the set of models whose likelihood is behind a threshold (\( \alpha \) times the ML)
  \[ P^{(\alpha)}(C) = \{ P(C) \in \mathcal{P}(C) : \forall \mathcal{F} \in \mathcal{P} \} \]

Naive Credal Classifier

- **Naive Classifiers**
  - **given class** \( C \), features \( \{ \mathcal{F}_1, \ldots, \mathcal{F}_n \} \) are conditionally independent
  - Often unrealistic, but good for classification!
  - E.g., NBC (naive Bayes classifier) [3]
  - Can be extended to the imprecise case with both concepts of strong independence and epistemic irrelevance. Same inferences are obtained!

Learning Credal Sets (IDM)

- **Bayesian learning extended to imprecision**
  - **Set** of Dirichlet modelling prior near-ignorance
  - **Bounds** of the posterior is
    \[ P(c|\mathcal{F}) \in \left[ \frac{n(c)}{N + 2}, \frac{n(c) + 1}{N + 3} \right] \]
  - **\( n(c) \)** counting function
  - **Real parameter** \( \alpha \) as equivalent sample size
  - **Typical choices** \( \alpha \in [1, 2] \)
  - If \( n(c) \rightarrow +\infty \), vacuous intervals
  - If \( n(c) \rightarrow 0 \), ML estimator

Coping with zero-couants

- Just add to the likelihood record \( (c \rightarrow \mathcal{F} = f) \) with \( C \) missing-at-random!
- This is called LNCCaf.

Experiments

- **Discounted accuracy**: rewards a set-valued classification with \( 1/\alpha \) or 0, depending on whether the set contains or not the correct class; Single Accuracy is the accuracy of the classifier when it return a single class.
- **The performance of the two classifiers is very close, when the determinacy is comparable!**

References