Imprecise Gaussian Discriminant Classification

11th International Symposium on Imprecise Probabilities: Theories and Applications

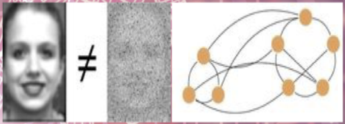
CARRANZA-ALARCON Yonatan-Carlos
Ph.D. Candidate in Computer Science

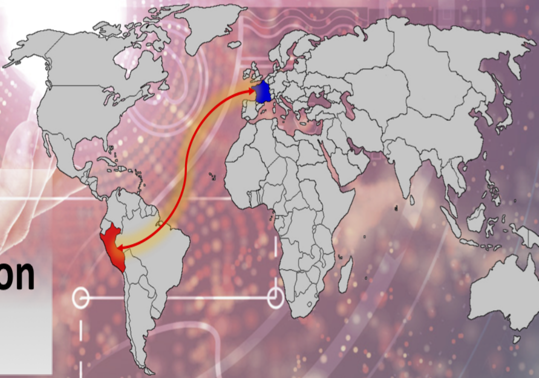
DESTERCKE Sébastien
Ph.D Director

utc
Université de Technologie Compiègne
heudiasyc
cnrs

03 Jul 2018 to 09 Jul 2019

CID Team





New challenges

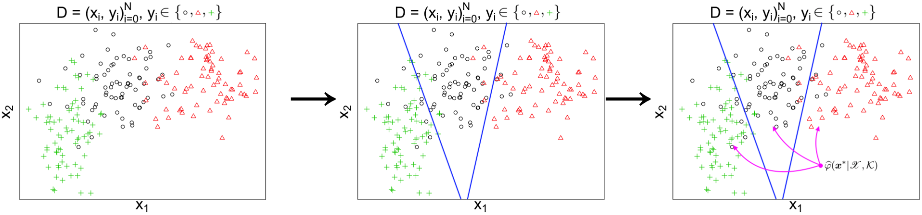
- Multi label classification
- Label Ranking

Overview

- Classification
  - Motivation
  - Precise Decision
  - Discriminant Analysis
- Imprecise Classification
  - Imprecise Gaussian discriminant analysis
  - Cautious Decision
- Evaluation
  - Cautious accuracy measure and Datasets
  - Experimental results
- Conclusions and Perspectives

Classification - Outline (Example)

☞ Data training $D = \{x_i, y_i\}_{i=0}^{N} \subseteq \mathbb{R}^p \times \mathscr{K}$



Objective

Given training data $D = \{x_i, y_i\}_{i=0}^{N}$:

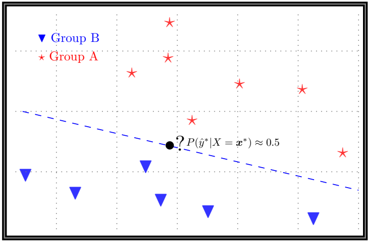
❶ learning a classification rule : $\varphi : \mathscr{X} \rightarrow \mathscr{K}$.

❷ predicting new instances $\widehat{\varphi}(\boldsymbol{x}^*)$

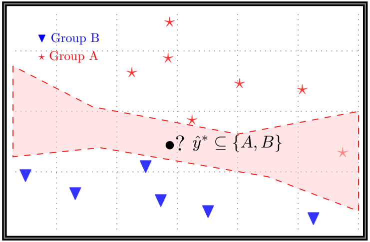
Motivation

What is the bigger problem in (precise) Classification ?

- Precise models can produce many mistakes for hard to predict unlabeled instances.



- One way to recognize such instances and avoid making such mistakes too often → **Making a cautious decision.**



Precise Classification

Step ❶ Learning the conditional probability distribution $\mathbb{P}_{Y|\boldsymbol{x}^*}$.

Step ❷ Predicting the "optimal" label amongst $\mathscr{K} = \{ m_1, ..., m_K \}$, under $\mathscr{L}_{0/1}$ loss function, for a new instance $\boldsymbol{x}^*$:

$$m_{i_K} \succ m_{i_{K-1}} \succ .... \succ m_{i_1} \iff P(y = m_{i_K} | \boldsymbol{x}^*) > .... > P(y = m_{i_1} | \boldsymbol{x}^*)$$

☞ Pick out the most preferable label m_{i_K}

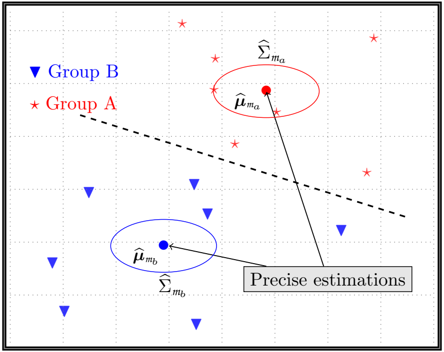
$\iff$ maximal probability plausible $P(y = m_{i_K} | \boldsymbol{x}^*)$

(Precise) Gaussian Discriminant Analysis

Applying Baye's rules to $P(Y = m_a | X = \boldsymbol{x}^*)$:

$$P(y = m_k | X = \boldsymbol{x}^*) = \frac{P(X = \boldsymbol{x}^* | y = m_k) P(y = m_k)}{\sum_{m_l \in \mathcal{K}} P(X = \boldsymbol{x}^* | y = m_l) P(y = m_l)}$$

Normality $P_{X|Y=m_k} \sim \mathcal{N}(\mu_{m_k}, \Sigma_{m_k})$ and precise marginal $\pi_{m_k} := P_{Y=m_k}$.



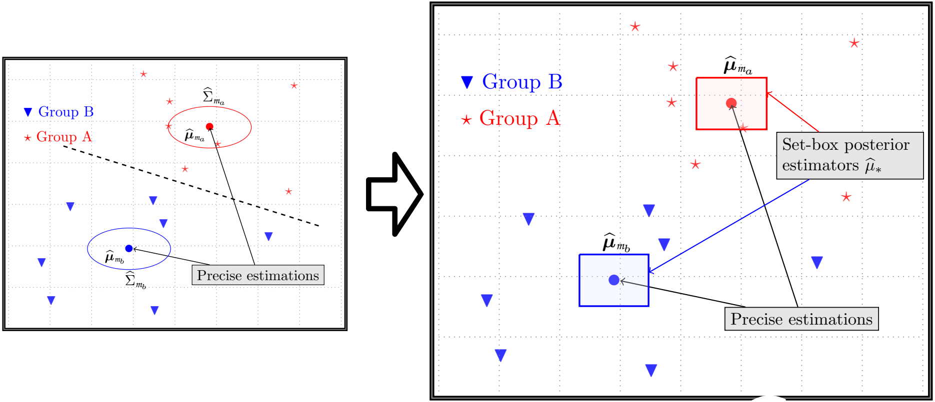
Overview

- Classification
  - Motivation
  - Precise Decision
  - Discriminant Analysis
- Imprecise Classification
  - Imprecise Gaussian discriminant analysis
  - Cautious Decision
- Evaluation
  - Cautious accuracy measure and Datasets
  - Experimental results
- Conclusions and Perspectives

Imprecise Gaussian Discriminant Analysis (IGDA)

Objective : *Making imprecise the parameter mean* μ_k *of each Gaussian distribution family* $\mathscr{G}_k := P_{X|Y=m_k} \sim \mathcal{N}(\mu_k, \widehat{\Sigma}_{m_k})$

Proposition : Using a set of posterior distribution $\mathscr{P}$ ([4, eq 17]).



Decision Making in Imprecise Probabilities

Definition (Partial Ordering by Maximality [1])

Under $\mathscr{L}_{0/1}$ loss function and let $\mathscr{P}_{Y|\boldsymbol{x}^*}$ a set of probabilities then m_a is preferred to m_b if and only if

$$\inf_{\mathbb{P}_{Y|\boldsymbol{x}^*} \in \mathscr{P}_{Y|\boldsymbol{x}^*}} P(Y = m_a|\boldsymbol{x}^*) - P(Y = m_b|\boldsymbol{x}^*) > 0 \quad (1)$$

- ☞ This definition give us a partial order $\succ_M$
- ☞ The maximal element of partial order is the cautious decision :

$$Y_M = \{m_a \in \mathscr{K} \mid \nexists m_b : m_a \succ_M m_b\}$$

Decision Making in IGDA

- Using the Bayes' rule on the criterion of maximality :

$$\inf_{\mathbb{P}_{Y|\boldsymbol{x}^*} \in \mathscr{P}_{Y|\boldsymbol{x}^*}} P(Y = m_a|\boldsymbol{x}^*) - P(Y = m_b|\boldsymbol{x}^*) > 0 \quad (2)$$

- We can reduce it to solving two different optimization problems :

$$\sup_{P \in \mathscr{P}_{X|m_b}} P(\boldsymbol{x}^*|y = m_b) \iff \overline{\mu}_{m_b} = \arg\max_{\mu_{m_b} \in \mathscr{P}_{\mu_{m_b}}} -\frac{1}{2}(\boldsymbol{x}^* - \mu_{m_b})^T \widehat{\Sigma}_{m_b}^{-1} (\boldsymbol{x}^* - \mu_{m_b}) \quad \text{(BQP)}$$

$$\inf_{P \in \mathscr{P}_{X|m_a}} P(\boldsymbol{x}^*|y = m_a) \iff \underline{\mu}_{m_a} = \arg\min_{\mu_{m_a} \in \mathscr{P}_{\mu_{m_a}}} -\frac{1}{2}(\boldsymbol{x}^* - \mu_{m_a})^T \widehat{\Sigma}_{m_a}^{-1} (\boldsymbol{x}^* - \mu_{m_a}) \quad \text{(NBQP)}$$

☞ First problem box-constrained quadratic problem (BQP).
☞ Second problem non-convex BQP
→ solved through Branch and Bound method.

Decision Making in IGDA

- Using the Bayes' rule on the criterion of maximality :

$$\inf_{\mathbb{P}_{Y|\boldsymbol{x}^*} \in \mathscr{P}_{Y|\boldsymbol{x}^*}} P(Y = m_a|\boldsymbol{x}^*) - P(Y = m_b|\boldsymbol{x}^*) > 0 \quad (2)$$

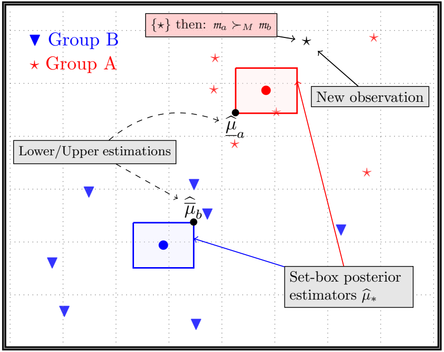
- We can reduce it to solving two different optimization problems :

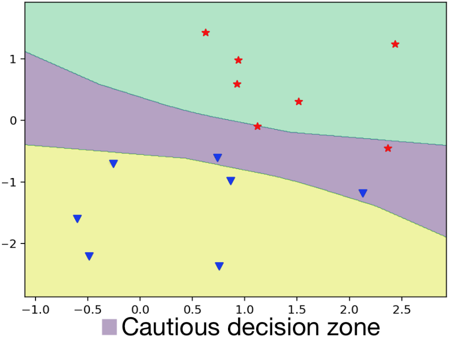
$$\sup_{P \in \mathscr{P}_{X|m_b}} P(\boldsymbol{x}^*|y = m_b) \iff \overline{\mu}_{m_b} = \arg\max_{\mu_{m_b} \in \mathscr{P}_{\mu_{m_b}}} -\frac{1}{2}(\boldsymbol{x}^* - \mu_{m_b})^T \widehat{\Sigma}_{m_b}^{-1} (\boldsymbol{x}^* - \mu_{m_b}) \quad \text{(BQP)}$$

$$\inf_{P \in \mathscr{P}_{X|m_a}} P(\boldsymbol{x}^*|y = m_a) \iff \underline{\mu}_{m_a} = \arg\min_{\mu_{m_a} \in \mathscr{P}_{\mu_{m_a}}} -\frac{1}{2}(\boldsymbol{x}^* - \mu_{m_a})^T \widehat{\Sigma}_{m_a}^{-1} (\boldsymbol{x}^* - \mu_{m_a}) \quad \text{(NBQP)}$$

☞ First problem box-constrained quadratic problem (BQP).
☞ Second problem non-convex BQP
→ solved through Branch and Bound method.

Cautious decision zone (example with 2 class)





☞ **Note the non-linearity boundary decision ! !**

Overview

- Classification
  - Motivation
  - Precise Decision
  - Discriminant Analysis
- Imprecise Classification
  - Imprecise Gaussian discriminant analysis
  - Cautious Decision
- Evaluation
  - Cautious accuracy measure and Datasets
  - Experimental results
- Conclusions and Perspectives

Datasets and experimental setting

- 9 data sets issued from UCI repository [2].
- 10×10-fold cross-validation procedure.
- Utility-discounted accuracy measure proposed to*Zaffalon et al* on [3].

$$u(y,\widehat{Y}_M) = \begin{cases} 0 & \text{if } y \notin \widehat{Y}_M \\ \frac{\alpha}{|\widehat{Y}_M|} - \frac{1-\alpha}{|\widehat{Y}_M|^2} & \text{else} \end{cases}$$

Goal : reward cautiousness to some degree α :

- $\alpha = 1$: cautiousness = randonness
- $\alpha \to \infty$: best classifier vacuous

#	name	# instances	# features	# labels
a	iris	150	4	3
b	wine	178	13	3
c	forest	198	27	4
d	seeds	210	7	3
e	dermatology	385	34	6
f	vehicle	846	18	4
g	vowel	990	10	11
h	wine-quality	1599	11	6
i	wall-following	5456	24	4

TABLE – Data sets used in the experiments

Experimental results

	LDA	ILDA		QDA	IQDA		Avg. time (sec.)
#	acc.	u_{80}	u_{65}	acc	u_{80}	u_{65}	
a	97.96	**98.38**	97.16	97.29	**98.08**	97.13	0.56
b	98.85	**98.99**	98.95	99.03	**99.39**	99.09	1.49
c	94.61	**94.56**	94.05	89.43	**91.77**	88.90	12.14
d	96.35	**96.59**	96.51	94.64	**95.20**	94.72	1.50
e	96.58	**97.06**	96.94	82.47	**84.24**	84.05	19.24
f	77.96	**81.98**	79.59	85.07	**87.96**	86.13	3.10
g	60.10	**67.45**	62.41	87.83	**89.96**	88.40	4.95
h	59.25	**65.83**	60.31	55.62	**65.85**	60.36	34.85
i	67.96	**71.34**	66.65	65.87	**71.79**	69.75	10.77
avg.	83.68	**86.05**	84.03	80.34	**87.16**	85.33	10.1

TABLE – AVERAGE UTILITY-DISCOUNTED ACCURACIES (%)

Overview

- Classification
  - Motivation
  - Precise Decision
  - Discriminant Analysis
- Imprecise Classification
  - Imprecise Gaussian discriminant analysis
  - Cautious Decision
- Evaluation
  - Cautious accuracy measure and Datasets
  - Experimental results
- Conclusions and Perspectives

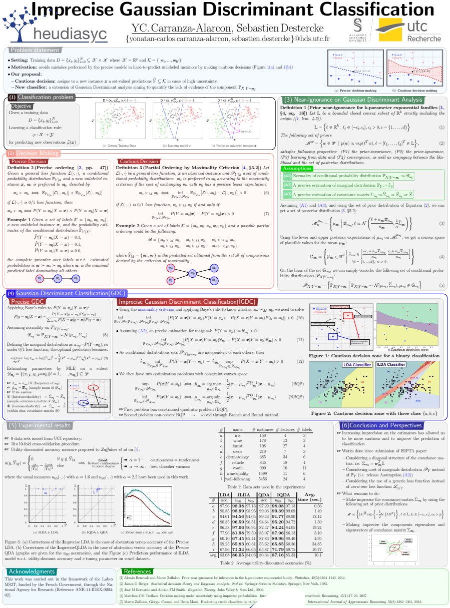
Conclusions and Perspectives
Imprecise Gaussian Discriminant Classification

1. Works done since submission of ISIPTA paper :
   - ✓ Considering the diagonal structure of the covariance matrix.
   - ✓ Releasing precise estimation of marginal distribution $\mathbb{P}_Y$ to convex set of distributions $\mathscr{P}_Y$.
   - ✓ Considering a generic loss function $\mathscr{L}$ instead of $\mathscr{L}_{0/1}$.
2. What remains to do
   - ✗ Make imprecise the covariance matrix Σ_{m_k} by using a set of prior distributions (cf.Poster).
   - ✗ Making imprecise the components eigenvalues and eigenvectors of covariance matrix Σ_{m_k}.

Poster


ANY QUESTIONS?
Please
come to my poster :)


Thank You!


Imprecise Gaussian Discriminant Classification
heudiasyc
YC. Carranza-Alarcon, Sebastien Destercke
{yonatan-carlos.carranza-alarcon, sebastien.destercke}@hds.utc.fr
utc Recherche
Problem statement
(1) Classification problem
(3) Near-Ignorance on Gaussian Discriminant Analysis
(2) Decision Making
Precise Decision
Cautious Decision
Assumptions
(4) Gaussian Discriminant Classification(GDC)
Precise GDC
Imprecise Gaussian Discriminant Classification(IGDC)
Figure 1: Cautious decision zone for a binary classification
LDA Classifier
ILDA Classifier
Figure 2: Cautious decision zone with three class {a, b, c}
(5) Experimental results
Table 1: Data sets used in the experiments
Table 2: Average utility-discounted accuracies (%)
(6)Conclusion and Perspectives
Acknowledgements
References

References

Matthias CM Troffaes. "Decision making under uncertainty using imprecise probabilities". In : *International Journal of Approximate Reasoning* 45.1 (2007), p. 17-29.

A. Frank et A. Asuncion. *UCI Machine Learning Repository*. 2010. URL : http://archive.ics.uci.edu/ml.

Marco Zaffalon, Giorgio Corani et Denis Mauá. "Evaluating credal classifiers by utility-discounted predictive accuracy". In : *International Journal of Approximate Reasoning* 53.8 (2012), p. 1282-1301.

Alessio Benavoli et Marco Zaffalon. "Prior near ignorance for inferences in the k-parameter exponential family". In : *Statistics* 49.5 (2014), p. 1104-1140.