

# Variantes crédibilistes des modèles des cotes proportionnelles basées sur la vraisemblance

Sébastien Ramel, Frédéric Pichon et François Delmotte

Université d'Artois  
LGI2A

Séminaire CID (Heudiasyc)  
Compiègne, 20 avril 2021

# Classification

En fonction de la **nature de la cible  $Y$** , différents types de classifieurs sont distingués :

- **Binaire**,  $|\mathcal{Y}| = 2$ .  $Y$  est une variable de Bernoulli
- **Multiclasse**,  $|\mathcal{Y}| = k > 2$ , plus précisément
  - \* **Nominal**,  $Y$  est une variable catégorique de support **non-ordonné**, ex.  $\mathcal{Y} = \{y_1, \dots, y_k\}$ .
  - \* **Ordinal**,  $Y$  est une variable catégorique de support **ordonné**, ex.  $\mathcal{Y} = \{1, \dots, k\}$ .

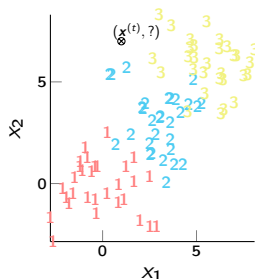


Figure – Données  $\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^n$  ordinales (monotones)

# Régression logistique ordinale

- La **régression logistique** est un classifieur binaire dont l'incertitude de classification est représentée par une distribution de **Bernoulli** t.q.  $Y \sim \mathcal{B}(\theta)$  où  $\theta := P_Y(Y = 1|\mathbf{x})$ .
  - Le **modèle des cotes proportionnelles** est une extension ordinale dont l'incertitude est donnée par une distribution **catégorique** t.q.  $Y \sim \mathcal{C}(\boldsymbol{\theta})$  où  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  et  $\theta_j := P_Y(Y = j|\mathbf{x})$ .
  - Reposent sur une combinaison linéaire de  $\mathbf{x}$  avec des paramètres  $\boldsymbol{\beta} \Rightarrow$  probabilités **monotones** et **interprétables**.
  - En revanche ils sont incapables de distinguer **différentes sources d'incertitude** (stochastique vs. épistémique = **ignorance**).
- $\Rightarrow$  Proposition d'une **variante fiable** du modèle des cotes proportionnelles formalisée dans le cadre théorique des **fonctions de croyance**.

# Plan

- 1 Théorie des fonctions de croyance
  - Représentation
  - Inférence
  - Prédiction
  - Décision
- 2 Modèle des Cotes Proportionnelles Partielles (MCP)
- 3 Extension crédibiliste
  - Prédiction d'une variable ordinale
  - Instanciation pour le MCP
  - Expériences
- 4 Conclusion

## 1 Théorie des fonctions de croyance

- Représentation
- Inférence
- Prédiction
- Décision

## 2 Modèle des Cotes Proportionnelles Partielles (MCP)

## 3 Extension crédibiliste

- Prédiction d'une variable ordinale
- Instanciation pour le MCP
- Expériences

## 4 Conclusion

## Représentation de l'information

- Soit  $Y$  une v.a. dont la réalisation  $y$  appartient au **cadre de discernement** fini  $\mathcal{Y} = \{y_i\}_{i=1}^K$ .

### Fonction de masse

Fonction  $m^{\mathcal{Y}} : 2^{\mathcal{Y}} \rightarrow [0, 1]$  t.q.  $m^{\mathcal{Y}}(\emptyset) = 0$  et  $\sum_{A \subseteq \mathcal{Y}} m^{\mathcal{Y}}(A) = 1$ .

- $m^{\mathcal{Y}}(A)$  est la part de croyance affectée exactement à  $y \in A$ .
- $A \subseteq \mathcal{Y}$ , t.q.  $m^{\mathcal{Y}}(A) > 0$  sont appelés **éléments focaux**.
- Si  $A_1 \subseteq A_2 \subseteq \dots \subseteq A_N$ , alors  $m^{\mathcal{Y}}$  est **consonante**.

## Autres fonctions

- Fonction de **croyance** (part de croyance **soutenant**  $y \in A$ ) et de **plausibilité** (part de croyance **ne contredisant pas**  $y \in A$ ) :

$$Bel^{\mathcal{Y}}(A) = \sum_{B \subseteq A} m^{\mathcal{Y}}(B), \quad Pl^{\mathcal{Y}}(A) = \sum_{B \cap A \neq \emptyset} m^{\mathcal{Y}}(B), \quad \forall A \subseteq \mathcal{Y}.$$

- Fonction de contour (plausibilité évaluée sur les singletons)

$$pl^{\mathcal{Y}}(y_i) = Pl^{\mathcal{Y}}(\{y_i\}), \quad \forall y_i \in \mathcal{Y}.$$

⇒ Si  $m^{\mathcal{Y}}$  est **consonante** :  $Pl^{\mathcal{Y}}(A) = \sup_{y_i \in A} pl^{\mathcal{Y}}(y_i)$ .

## Fonction de croyance basée sur la vraisemblance $Bel_z^\Theta$

- Soit  $z$  l'observation de  $Z \sim P_Z(\cdot; \theta)$ , où  $\theta \in \Theta$  est **inconnu**.
- La connaissance de  $\theta$ , donnée par  $Bel_z^\Theta$  (**consonante**) est caractérisée par

$$pl_z^\Theta(\theta) = L_z(\theta)/L_z(\hat{\theta}), \quad \forall \theta \in \Theta$$

avec

- \*  $L_z(\theta) = P_Z(z; \theta)$ , la vraisemblance,
- \*  $\hat{\theta} = \arg \max_{\theta \in \Theta} L_z(\theta)$ , t.q.  $L_z(\hat{\theta}) < \infty$
- Éléments focaux de  $Bel_z^\Theta$

$$\Gamma_z(u) = \{\theta \in \Theta \mid pl_z^\Theta(\theta) > u\}$$

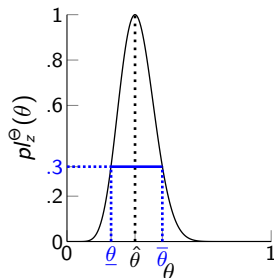


Figure – **Vraisemblance relative** d'une distribution binomiale  $\mathcal{B}(\theta, n)$ , avec  $\hat{\theta} = 1/3$  et  $n = 30$ .  
 $\Gamma_z(u) = [\underline{\theta}(u), \bar{\theta}(u)]$ .



## Fonction de croyance prédictive $Bel_z^{\mathcal{Y}}$

- Construire  $Bel_z^{\mathcal{Y}}$  à propos de  $Y \sim P_Y(\cdot; \theta)$ , sachant  $Bel_z^{\Theta}$ .

- Si  $Y$  **ordinaire**,  $F_{\theta}(Y) = V \sim \mathcal{U}([0, 1])$  d'où

$$Y = F_{\theta}^{-1}(V) = \varphi(\theta, V)$$

- Repose sur  $Y = \varphi(\Gamma_z(U), V) \subseteq \mathcal{Y}$ .

⇒ Si  $Y \sim \mathcal{B}(\theta)$  une v.a. de **Bernoulli**, alors

$$\varphi(\Gamma_z(u), v) = \begin{cases} \{1\} & \text{si } v \leq \underline{\theta}(u) \\ \{0\} & \text{si } v > \bar{\theta}(u) \\ \{0, 1\} & \text{sinon.} \end{cases}$$

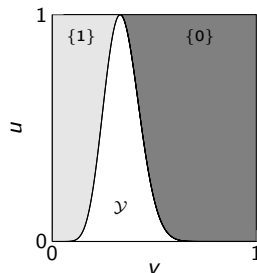


Figure –  $m_z^{\mathcal{Y}}$  de  $Y \sim \mathcal{B}(\theta)$ .  
 ■  $m_z^{\mathcal{Y}}(\{1\})$ , ■  $m_z^{\mathcal{Y}}(\{0\})$ ,  
 □  $m_z^{\mathcal{Y}}(\mathcal{Y})$

## Règles de décision

- Soit  $c_j(i)$  le coût de la prédiction  $y = j$  pour la vérité  $i$ . Risque inférieur et supérieur :

$$* \underline{R}(j) = \sum_{A \subseteq \mathcal{Y}} m_z^{\mathcal{Y}}(A) \min_{i \in A} c_j(i),$$

$$* \overline{R}(j) = \sum_{A \subseteq \mathcal{Y}} m_z^{\mathcal{Y}}(A) \max_{i \in A} c_j(i).$$

⇒ Règle **optimiste** :  $a = \arg \min_{j \in \mathcal{Y}} \underline{R}(j) \in \mathcal{Y}$ .

⇒ Règle de la **dominance d'intervalles (ID)** :

$$A = \{j \in \mathcal{Y} \mid \nexists i : j \succ_{ID} i\} \subseteq \mathcal{Y},$$

d'après les relations d'ordre

$$j \succ_{ID} i \iff \overline{R}(j) < \underline{R}(i), \quad \forall (j, i) \in \mathcal{Y}^2.$$

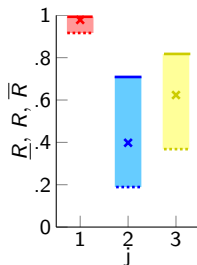


Figure – Risques menant à  $A = \{2, 3\}$  via ID.

# Plan

- 1 Théorie des fonctions de croyance
  - Représentation
  - Inférence
  - Prédiction
  - Décision
- 2 Modèle des Cotes Proportionnelles Partielles (MCP)
- 3 Extension crédibiliste
  - Prédiction d'une variable ordinale
  - Instanciation pour le MCP
  - Expériences
- 4 Conclusion

## Modèles des Cotes Proportionnelles

Soit  $Y \sim C(\theta)$  de support  $\mathcal{Y} = \{1, \dots, k\}$  **ordonné**.

### Modèle des Cotes Proportionnelles Partielles (MCP)

$$\log \left( \frac{P_Y(Y \leq j | \mathbf{x})}{1 - P_Y(Y \leq j | \mathbf{x})} \right) = \alpha_j + \boldsymbol{\sigma}^\top \mathbf{x} + \boldsymbol{\gamma}_j^\top \boldsymbol{\tau}_x, \quad j = 1, \dots, k-1,$$

avec

- \*  $\boldsymbol{\sigma} \in \mathbb{R}^p$ , des coefficients de  $\mathbf{x} \in \mathbb{R}^p$ , indépendants de  $j$  (=Hypothèse de Proportionnalité **HP**),
- \*  $\alpha_j$  t.q.  $\alpha_1 \leq \dots \leq \alpha_{k-1}$ , des ordonnées à l'origine,
- \*  $\boldsymbol{\tau}_x \in \mathbb{R}^q$  les attributs de  $\mathbf{x}$  rejetant **HP**,
- \*  $\boldsymbol{\gamma}_j \in \mathbb{R}^q$  des coefficients de  $\boldsymbol{\tau}_x$  dépendants de  $j$ .

si  $q = 0 \Rightarrow$  Modèle des Cotes Proportionnelles (MCP),

si  $q = p \Rightarrow$  Modèle des Cotes Non-Proportionnelles (MCNP).

## Modèle des Cotes Proportionnelles Partielles (MCP)

Soit  $\beta_j = (\alpha_j, \sigma^\top, \gamma_j^\top)$ ,  $\phi_x = (1, x^\top, \tau_x^\top) \Rightarrow \beta_j^\top \phi_x = \alpha_j + \sigma^\top x + \gamma_j^\top \tau_x$

$\Rightarrow P_j := P_Y(Y \leq j | x) = h(\beta_j^\top \phi_x)$ , où  
 $h(z) = (1 + e^{-z})^{-1}$  (=logistique)

$\Rightarrow \theta_j := P_Y(Y = j | x)$

$$= \begin{cases} h(\beta_j^\top \phi_x) & \text{si } j = 1, \\ 1 - h(\beta_{j-1}^\top \phi_x) & \text{si } j = k, \\ h(\beta_j^\top \phi_x) - h(\beta_{j-1}^\top \phi_x) & \text{sinon.} \end{cases}$$

$\Rightarrow L_D(\beta) = \prod_i \prod_j (\theta_j^{(i)})^{t_j^{(i)}}$ , avec  
 $t_j^{(i)} = \mathbb{1}(y^{(i)} = j)$ .

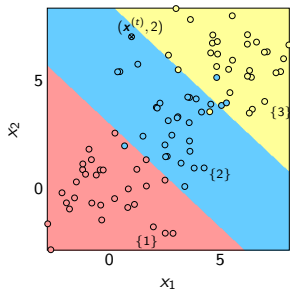


Figure – Décisions de MCP via un coût (0,1).

# Plan

- 1 Théorie des fonctions de croyance
  - Représentation
  - Inférence
  - Prédiction
  - Décision
- 2 Modèle des Cotes Proportionnelles Partielles (MCP)
- 3 Extension crédibiliste
  - Prédiction d'une variable ordinale
  - Instanciation pour le MCP
  - Expériences
- 4 Conclusion

## Génération d'une v.a. ordinale $Y \sim C(\theta)$

- Considérons

- \*  $\mathbf{P} = (P_0, \dots, P_k)$  avec  $P_0 = 0$ ,  $P_j = \sum_{i=1}^j \theta_i$ ,  $\forall j \in \mathcal{Y}$ .
- \*  $V \sim \mathcal{U}([0, 1])$  une v.a. **pivot** de distribution uniforme.

- $Y$  peut être **générée** à partir de  $\mathbf{P}$  et  $V$  par

$$\varphi(\mathbf{P}, V) = j, \quad \text{si} \quad P_{j-1} \leq V < P_j, \quad \forall j \in \mathcal{Y}.$$

$$\Rightarrow P_Y(\varphi(\mathbf{P}, V) = j) = \theta_j.$$

Figure – Génération d'une variable **ordinaire**  $Y = \varphi(\mathbf{P}, V)$ .



## Extension imprécise basée sur la vraisemblance

### Inférence de $P$

- Ayant observé  $\mathbf{z}$  succès,  $P$  est **incertain**.

$$* pl_z^\Theta(\theta) = \prod_{j=1}^k \left( \frac{\theta}{\hat{\theta}_j} \right)^{z_j}, \text{ où } \hat{\theta}_j = \frac{z_j}{\sum_i z_i}.$$

$$* \hat{P}_l = \sum_{i=1}^l \hat{\theta}_i, l \in [1, k] \text{ et } \hat{P}_0 = 0.$$

⇒ Incertitude  $Bel_z^{P_l}$  caractérisée par

$$pl_z^{P_l}(P) = \begin{cases} \begin{cases} 1 & \text{si } P = \hat{P}_l \\ 0 & \text{sinon} \end{cases} & \text{si } l \in \{0, k\} \\ \sup_{\theta \in A_l} pl_z^\Theta(\theta) & \text{sinon} \end{cases}$$

$$\text{où } A_l = \{\theta \in \Theta \mid \sum_{i=1}^l \theta_i = P\}.$$

$$\Rightarrow \Gamma_z^l(u) = [P_l(u), \bar{P}_l(u)].$$

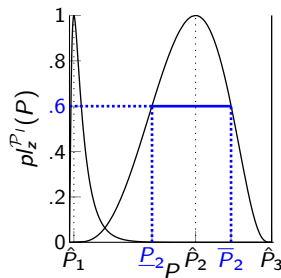


Figure – Contours  $pl_z^{P_l}$   
pour  $l = 0, \dots, 3$ .  
 $\Gamma_z^2(u) = [P_2(u), \bar{P}_2(u)]$



# Extension imprécise basée sur la vraisemblance

## Prédiction crédibiliste d'une variables ordinale $Y$

### Prédiction d'une variable ordinale $Y$

Ayant observé  $\mathbf{z} \in \mathbb{N}^k$  succès, la fonction de masse prédictive de  $Y$  est définie par  $m_z^Y([i, j]) = A_{ij} - B_{ij}$  avec

$$A_{ij} = \begin{cases} \hat{P}_j - \hat{P}_{j-1} & \text{if } i = j \\ \int_{P_{i-1,j-1}^*}^{P_{i,j-1}^*} pl_z^{\mathcal{P}_{j-1}}(P) dP + \int_{P_{i,j-1}^*}^{P_{i,j}^*} pl_z^{\mathcal{P}_i}(P) dP & \text{else} \end{cases}$$

$$B_{ij} = \int_{P_{i-1,j-1}^*}^{P_{i-1,j}^*} pl_z^{\mathcal{P}_{i-1}}(P) dP + \int_{P_{i-1,j}^*}^{P_{i,j}^*} pl_z^{\mathcal{P}_j}(P) dP,$$

et  $P_{i,j}^*$ , l'unique valeur de l'intervalle unité telle que

$pl_{z_1}^{\mathcal{P}_i}(P_{i,j}^*) = pl_{z_1}^{\mathcal{P}_j}(P_{i,j}^*)$  pour  $0 \leq i < j \leq k$ ,  $P_{i,j}^* = \hat{P}_i$  pour  $j = i$ , et  $P_{0,k}^*$  est quelconque.

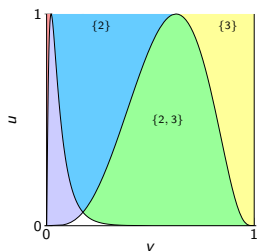


Figure –  $m_z^Y$  de  $Y \sim \mathcal{C}(\theta)$ .  $\square m_z^Y(\{1\})$ ,  $\blacksquare m_z^Y(\{2\})$ ,  $\blacksquare m_z^Y(\{3\})$ ,  $\blacksquare m_z^Y(\{1, 2\})$ ,  $\blacksquare m_z^Y(\{2, 3\})$ ,  $\square m_z^Y(Y)$ .

## MCP)P Crédibiliste (MCP)PC)

- Soit  $\mathcal{D}$  et  $(\mathbf{x}, y)$  où  $y \in \mathcal{Y}$  **inconnue**, est représentée par la v.a. (ordinaire)  $Y \sim \mathcal{C}(\theta)$ , définie par  $\mathbf{P} = (0, h(\beta_1^\top \phi_{\mathbf{x}}), \dots, h(\beta_{k-1}^\top \phi_{\mathbf{x}}), 1)$

$$* pl_{\mathcal{D}}^B(\beta) = L_{\mathcal{D}}(\beta) / L_{\mathcal{D}}(\hat{\beta}),$$

$$* \hat{P}_l = h(\hat{\beta}_l^\top \phi_{\mathbf{x}}), l = 1, \dots, k-1.$$

⇒ Incertitude  $Bel_{\mathcal{D}, \mathbf{x}}^{P_l}$ , caractérisée par

$$pl_z^{P_l}(\mathbf{P}) = \sup_{\beta^* \in B_l^x} pl_{\mathcal{D}, \mathbf{x}}^B(\beta^*), l \in [1, k-1]$$

$$\text{où } B_l^x = \{\beta \in B \mid h(\beta_l^\top \phi_{\mathbf{x}}) = P\}.$$

⇒ Prediction  $m_{\mathcal{D}, \mathbf{x}}^{\mathcal{Y}}$ , donnée par les surfaces induites par  $pl_{\mathcal{D}}^{P_l}$ .

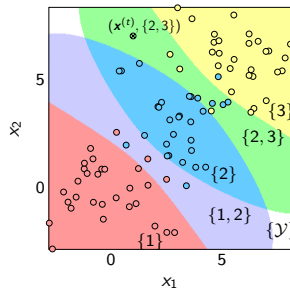
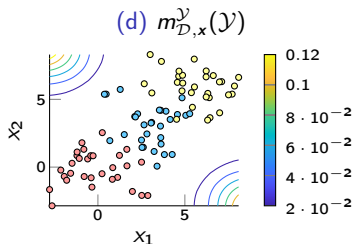
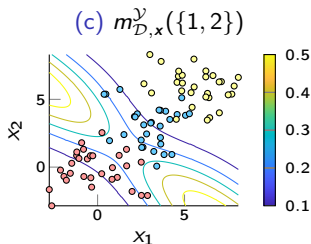
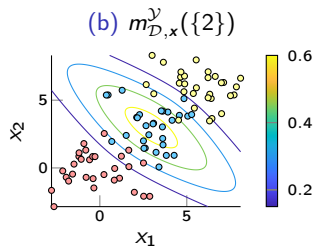
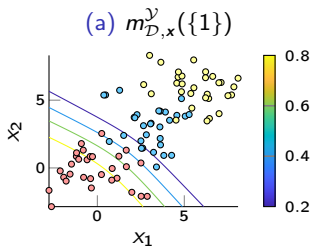


Figure – Décisions fiables de MCP)C via ID et un coût (0,1).

## Fonction de masse prédictive



## Configuration

- 9 jeux de données **monotones** partitionnés en
  - \* un ensemble d'apprentissage  $\mathcal{D}$ ,
  - \* un ensemble de test  $\mathcal{T}$ .
- Modèles : MCP, MCNP, MCP et leurs **variantes crédibilistes MCPC, MCNPC, MCPPC** entraînés sur  $\mathcal{D}$ .
- Performances évaluées sur  $\mathcal{T}$  et moyennées sur 100 permutations
  - \* taux de décision  $a \in \mathcal{Y}$  correcte
$$\text{Acc} = |\mathcal{T}|^{-1} \sum_i \mathbb{1}_{a^{(i)}}(y^{(i)}) ,$$
  - \* taux de décision  $A \subseteq \mathcal{Y}$  correcte réduit par utilité
$$\text{Acc}_{U_{65}} = |\mathcal{T}|^{-1} \sum_i \mathbb{1}_{A^{(i)}}(y^{(i)}) \cdot u_{65}(|A^{(i)}|^{-1}) .$$
- HP **vérifiée** pour MCP et MCPPC via le **test de Brant-Wald**.

## Résultats

Taux de décision (précise) correcte (Acc)

Jeux	MCP	MCP	MCNP	MCPC	MCP	MCNPC
BAL	.8821(5)	.8822(4)	.8819(6)	.9019(2)	.9022(1)	.8983(3)
ERA	.5671(4)	.5665(5)	.5642(6)	.5687(1.5)	.5685(3)	.5687(1.5)
ESL	.8732(1)	.8729(2)	.8719(6)	.8724(5)	.8726(3)	.8725(4)
HAM	.3303(5)	.3281(6)	.3431(2)	.3372(4)	.3394(3)	.3464(1)
HOU	.7416(3.5)	.7417(2)	.7425(1)	.7414(5)	.7412(6)	.7416(3.5)
LEV	.6120(1)	.6109(4)	.6094(6)	.6117(2)	.6111(3)	.6104(5)
MPG	.8268(1)	.8265(2)	.8245(5)	.8262(3)	.8260(4)	.8234(6)
SWD	.5715(3)	.5712(4)	.5752(1)	.5709(5)	.5703(6)	.5745(2)
WIN	.6962(3.5)	.6962(3.5)	.6914(6)	.6968(1)	.6963(2)	.6932(5)
Rang	3.00	3.61	4.33	3.17	3.44	3.44

## Résultats

Taux de décision (imprécise) correcte réduit par utilité ( $\text{Acc}_{U_{65}}$ )

Jeux	MCP	MCP	MCNP	MCPC	MCP	MCNP
BAL	.8821(5)	.8822(4)	.8819(6)	.9065(2)	.9065(2)	.9065(2)
ERA	.5671(4)	.5665(5)	.5642(6)	.6102(1)	.6098(2)	.6096(3)
ESL	.8732(4)	.8729(5)	.8719(6)	.8802(1)	.8784(2)	.8771(3)
HAM	.3303(5)	.3281(6)	.3431(4)	.4617(1)	.4616(2)	.4615(3)
HOU	.7416(6)	.7417(5)	.7425(4)	.7587(3)	.7588(2)	.7612(1)
LEV	.6120(4)	.6109(5)	.6094(6)	.6382(1)	.6379(2)	.6368(3)
MPG	.8268(4)	.8265(5)	.8245(6)	.8412(1.5)	.8412(1.5)	.8405(3)
SWD	.5715(5)	.5712(6)	.5752(4)	.6043(3)	.6049(2)	.6069(1)
WIN	.6962(4.5)	.6962(4.5)	.6914(6)	.7139(1)	.7136(2)	.7085(3)
Rang	4.61	5.06	5.33	1.61	1.94	2.44

## Conclusion et perspectives

- Conclusion : proposition d'une **variante crédibiliste** (fiable) de MCP basée sur la vraisemblance.
- Perspectives : **généralisation** de l'approche crédibiliste
  - \* au cas où  $Y \sim \mathcal{C}(\theta)$  est une variable **nominale** (sans ordre),  
⇒ modèle d'échantillonnage de Dempster étendu à des variables nominales (**échantillonnage dans un simplex**)
  - \* aux modèles des réseaux de neurones artificiels,  
⇒ **approximation uni-modale** de la vraisemblance régularisée, ex. approximation de Laplace.

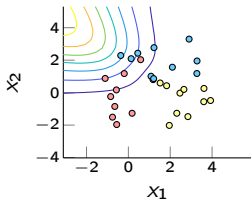
Merci pour votre attention.



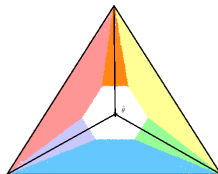
# Prédiction d'une variable nominale

## Application à la régression logistique multinomiale

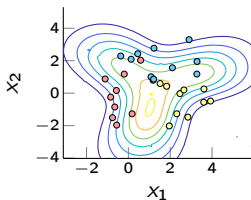
(a)  $m_{\mathcal{D},x}^{\mathcal{Y}}(\{1,2\})$



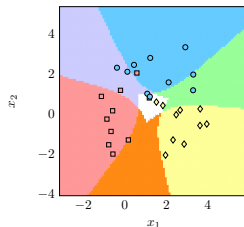
(b) Échantillonnage dans le simplexe



(c)  $m_{\mathcal{D},x}^{\mathcal{Y}}(\mathcal{Y})$

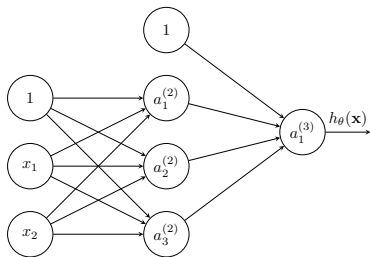


(d) Décisions fiables



# Extension au réseau de neurones

(a) Réseau de neurones utilisé



(b)  $m_{D,x}^{\mathcal{Y}}(\mathcal{Y})$

