

IFT3395/6390

Fondements de l'apprentissage machine

Régression multiple et classification multiclasse
régions de décision
encodage one-hot

Professeur: Pascal Vincent

Les types de problèmes ou tâches classiques en apprentissage

| | supervisé | supervisé | non-supervisé |
|---|--|--|---|
| | Classification | Régression | Estimation de densité |
| Signification de la cible y | indique une classe parmi m classes. | une valeur réelle à prédire. | pas de cible y ! |
| Domaine de y | $y \in \{-1, 1\}$ ou $y \in \{1, \dots, m\}$ ou $y \in \{0, 1, \dots, m-1\}$ | $y \in \mathbb{R}$ | pas de cible y ! |
| Ce que $f(x)$ vise à prédire | la classe de x (la classe la plus probablement associée à x) | la valeur espérée de y (le y "moyen") correspondant à x . $E[Y \mid X=x]$ | la densité $p(x)$ (l'observation x est-elle fort ou peu probable?) |
| Fonction de perte (ou coût) que l'on veut habituellement minimiser. | l'erreur de classification: $L((x, y), f) = I_{\{f(x) \neq y\}}$ | l'erreur quadratique: $L((x, y), f) = (f(x) - y)^2$ | la log-vraisemblance négative: $L(x, f) = -\log f(x)$ |

Régression multiple

- **Régression multiple:** lorsque la cible n'est pas un unique scalaire réel mais un *vecteur* de *m* réels (cible multiple).

$$y \in \mathbb{R}^m$$

- On peut construire *m* régresseurs scalaires.
- On peut aussi adapter un algorithme de régression pour manipuler des vecteurs plutôt que des scalaires.

Classifieur binaire

- Pour la **classification binaire** (2 classes), on utilise souvent un classifieur qui **calcule une sortie $g(x)$ réelle scalaire**.
 g est appelée **fonction discriminante**.
- La **fonction de décision ou classification $f(x)$** consiste alors à **comparer cette sortie $g(x)$ avec un seuil s** . On décide la première classe si la *sortie* est inférieure au *seuil* et la deuxième si la sortie est supérieure.
- Un cas fréquent est d'utiliser un *seuil* de 0, lorsque la sortie est réelle, ce qui revient à ne considérer que son **signe**.
- Un autre cas fréquent, est d'utiliser un *seuil* de 0.5, lorsque la *sortie* se situe dans $[0, 1]$ et qu'on l'interprète comme la **probabilité** d'une des classes sachant l'entrée x . Notez que la probabilité de l'autre classe est alors $1 - \text{sortie}$. Un seuil de 0.5 permet ainsi de choisir la classe la plus probable.

Classifieur binaire

- Terminologie/notation

- données d'entraînement: $D_n = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n))$

- observation: $\mathbf{x}_i \in \mathbb{R}^d$

- étiquette/classe: $y_i \in \{-1, 1\}$

- fonction discriminante: $g : \mathbb{R}^d \mapsto \mathbb{R}$, souvent $g : \mathbb{R}^d \mapsto [-1, 1]$

fonction de décision:

- fonction de classification/classifieur: $f : \mathbb{R}^d \mapsto \{-1, 1\}$

- fonction discriminante \rightarrow classifieur:

$$f(\mathbf{x}) = \begin{cases} 1, & \text{si } g(\mathbf{x}) \geq 0 \\ -1, & \text{si } g(\mathbf{x}) < 0 \end{cases}$$

ou bien

$$y_i \in \{0, 1\}$$

$$g : \mathbb{R}^d \rightarrow [0, 1]$$

$$f : \mathbb{R}^d \rightarrow \{0, 1\}$$

$$f(\mathbf{x}) = \begin{cases} 1 & \text{si } g(\mathbf{x}) \geq 0.5 \\ 0 & \text{si } g(\mathbf{x}) < 0.5 \end{cases}$$

Classifieur multiclasse

- Au cas où on a m classes: le classifieur produit en sortie $g(\mathbf{x})$ un **vecteur de dimension m** , qui va contenir un “**score**” pour chaque classe.

fonctions discriminantes:

$$sortie = g(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x}))$$

- La fonction de décision ou classification $f(\mathbf{x})$ consiste alors à décider la **classe qui a obtenu le score le plus élevé**:

$$décision = \arg \max(sortie)$$

$$f(\mathbf{x}) = \arg \max(g(\mathbf{x}))$$

- Quand les sorties (scores) sont **positifs** ou nuls et **sommés à 1**, on peut interpréter la **sortie j** comme la **probabilité** de la j ème classe, sachant l'entrée.
- Peut aussi s'appliquer au cas binaire (2 classes: 2 sorties)!
- Remarque: avec deux classes, choisir la classe avec le **score** le plus élevé, revient à considérer le *signe* de la différence entre les deux scores.

Notion de régions de décision

- S'applique à toute fonction de classification (classifieur).
- La «région de décision» pour une classe est la région de l'espace d'entrée dans laquelle le classifieur prédit cette classe (comme étant la plus probable).
- Un classifieur à m classes partitionne ainsi l'espace en m régions de décision disjointes.
- Une région de décision n'est pas nécessairement contiguë. (Ex: union de régions disjointes)

Notion de frontière de décision

- C'est la bordure d'une région de décision:
- L'interface entre la région de décision d'une classe et celle(s) de(s) autre(s) classe(s).
- C'est là que le classifieur est maximalelement incertain (classe est aussi probable qu'une autre).
- Notion plus souvent appliquée au cas de 2 classes.
- Une frontière de décision n'est pas nécessairement contiguë.

Formellement

Région de décision de la classe c

induite par un classifieur défini sur l'espace d'entrée \mathbb{R}^d

$$R_c = \{x \in \mathbb{R}^d \mid f(x) = c\}$$

Classifieur multiclasse (produit un vecteur de m scores):

$$R_c = \{x \in \mathbb{R}^d \mid \operatorname{argmax}[g(x)] = c\}$$

Classifieur binaire (une sortie scalaire) avec seuil s (ex: 0 ou 0.5):

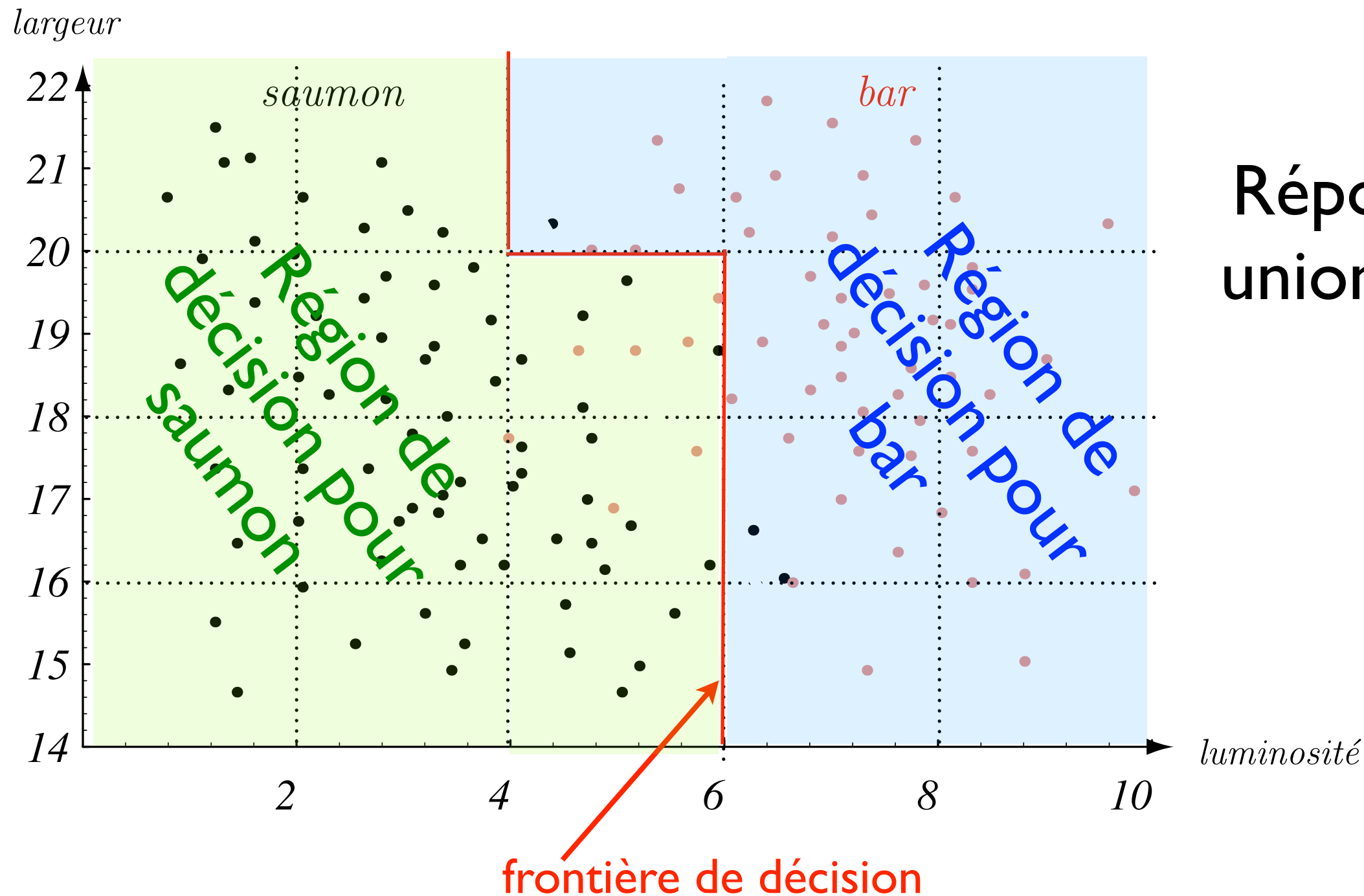
$$R_{\text{première classe}} = \{x \in \mathbb{R}^d \mid g(x) < s\}$$

$$R_{\text{deuxième classe}} = \{x \in \mathbb{R}^d \mid g(x) \geq s\}$$

$$\text{Frontière de décision: } F = \{x \in \mathbb{R}^d \mid g(x) = s\}$$

Ex: Régions de décision d'un classifieur de type histogramme

À quoi ressemblent-t-elles?



Réponse: à des unions de blocs

Encodage “one-hot” un parmi plusieurs

- Pour une **variable catégorique**, indiquant une parmi **m** catégories (classes)
- Donne un **vecteur de bits** de dimension **m** dont tous les éléments sont à 0 sauf l'élément correspondant à la classe. Celui là vaut 1

- Ex:**
- $\text{onehot}_4(1)=(1,0,0,0)$
 - $\text{onehot}_4(2)=(0,1,0,0)$
 - $\text{onehot}_6(3)=(0,0,1,0,0,0)$
 - $\text{onehot}_{10}(9)= \dots$

Souvent utilisé pour:

- **Prétraitement:** pour l'encodage de **caractéristiques catégoriques**
- **Représentation d'une classe cible** (y) sous forme d'un vecteur approprié pour certains algorithmes.