Classificateur de Bayes

Introduction

- On présente une méthode de classification reposant sur une approche probabiliste basée sur la **règle de Bayes**.

- Elle permet naturellement d'intégrer **des connaissances a priori**, ce que ne permettent pas aussi facilement la plupart des méthodes que nous verrons par la suite.

- Dans cette approche, on s'intéresse souvent à estimer la **probabilité conditionnelle** $p(\mathbf{x}|y)$, qui veut dire:

*Étant donné y **vérifié** ou **observé** ($\mathbf{ex.}$ y peut être une classe), qu'elle est la probabilité d'observer la donnée $\mathbf{x}$?*

Rappels sur le théorème de Bayes

- Soit *A*, *B* et *C* trois évènements, on a alors:

Probabilité a posteriori

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

Probabilité a priori

$$p(A|B,C) = \frac{p(B|A,C)p(A|C)}{p(B|C)}$$

- $p(A|B)$ est la vraisemblance de *A* si *B* est observé.
- $p(A|B,C)$ est la vraisemblance de *A* si *B* et *C* sont observés.

Note: l'événement C représenté le contexte générale.

Classification par la règle de Bayes

- Pour la classification, supposons que:

  $y = C_k$ dénote l'événement « **j'observe la classe** C_k».
  $\mathbf{x} = x$ dénote l'événement « **j'observe la donnée** $\boldsymbol{x}$».

- $p(y = C_k|\mathbf{x} = x, \mathcal{D})$ donne la probabilité d'observer la classe C_k si on observe la donnée $\mathbf{x} = x$ sachant que l'on dispose de l'ensemble des exemple $\mathcal{D}$ (Contexte) .

- En appliquant la règle de Bayes, on aura:

$$p(y = C_k|\mathbf{x} = x, \mathcal{D}) = \frac{p(\mathbf{x} = x|y = C_k, \mathcal{D})p(y = C_k|\mathcal{D})}{p(\mathbf{x} = x|\mathcal{D})}$$

Classification par la règle de Bayes

- $p(y = C_k | \mathcal{D})$ donne la probabilité d'observer la classe C_k étant donné l'ensemble des exemples $\mathcal{D}$. Elle est estimée par la **proportion** d'exemples dans $\mathcal{D}$ qui sont de classe C_k.

- $p(\mathbf{x} = x | y = C_k, \mathcal{D})$ donne la probabilité d'observer la donnée $\mathbf{x} = x$ si elle est de classe $y = C_k$. Ce terme est plus difficile à estimer que le précédent.

- **L'hypothèse de Bayes naïve (HBN)** consiste à dire que **x** est une **conjonction de valeurs d'attributs** et que ces attributs sont **des variables aléatoires indépendantes**.

Classification par la règle de Bayes

- **L'HBN n'est pas vérifiée en générale**. Cependant, elle permet de faire des calculs simples.

- Si on applique l'HBN, en supposant que la donnée **x** est décrite par D attributs notés x_1 , x_2 ,…, x_D dont les valeurs sont notées par x_1 , x_2 ,…, x_D, on aura:

$$p(\mathbf{x} = x|y = C_k, \mathcal{D}) = \\ p(\mathrm{x}_1 = x_1|y = C_k, \mathcal{D}) \times \cdots \times p(\mathrm{x}_D = x_D|y = C_k, \mathcal{D})$$

- Chaque terme $p(\mathrm{x}_d = x_d|y = C_k, \mathcal{D})$ est estimé à l'aide d'un ensemble d'exemples.

Classification par la règle de Bayes

- $p(\mathbf{x} = x|\mathcal{D})$ est la probabilité d'observer la valeur $\mathbf{x} = x = (x_1, \dots, x_D)$ ayant l'ensemble d'exemples $\mathcal{D}$. Pour calculer ce terme, on utilise la **marginalisation** suivante:

$$p(\mathbf{x} = x|\mathcal{D}) = \sum_{j=1}^{K} p(\mathbf{x} = x, y = C_j|\mathcal{D})$$
$$= \sum_{j=1}^{K} p(\mathbf{x} = x|y = C_j, \mathcal{D}) p(y = C_j|\mathcal{D})$$

- Donc on aura:

$$p(y = C_k|\mathbf{x} = x, \mathcal{D}) = \frac{p(\mathbf{x} = x|y = C_k, \mathcal{D}) p(y = C_k|\mathcal{D})}{\sum_{j=1}^{K} p(\mathbf{x} = x|y = C_j, \mathcal{D}) p(y = C_j|\mathcal{D})}$$

Classification par la règle de Bayes

- Une fois que l'on a calculé la probabilité d'appartenance $p(y = C_k|\mathbf{x} = x, \mathcal{D})$ à chacune des classes $k \in \{1, \dots, K\}$ pour la donnée $\mathbf{x} = x$, on peut prédire la classe de x en prenant le **maximum de probabilité a posteriori**:

Probabilité a posteriori

$$y_{MAP} = \operatorname{argmax}_{y=C_k} \{p(y = C_k|\mathbf{x} = x, \mathcal{D})\}$$

$$y_{MAP} = \operatorname{argmax}_{y=C_k} \{p(\mathbf{x} = x|y = C_k, \mathcal{D})p(y = C_k|\mathcal{D})\}$$

- Si l'on ne tient pas compte de $p(y = C_k|\mathcal{D})$, on aura:

$$y_{ML} = \operatorname{argmax}_{y=C_k} \{p(\mathbf{x} = x|y = C_k, \mathcal{D})\}$$

- Qu'on appellera **le maximum de vraisemblance**.

Exemple

D = 4 K = 2

Jour	x_1 Ciel	x_2 Température	x_3 Humidité	x_4 Vent	$C_1 = Oui, C_2 = Non$ Jouer au tennis ?
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

N = 14

$x = (Ciel{=}Couvert, Température{=}Chaud, Humidité{=}Normale, Vent{=}Faible)$
$y =$ Oui

Classification par la règle de Bayes

Exemple: Pour notre exemple **jouer au tennis?**,

x = (Ciel, Température, Humidité, Vent), y Є {oui, non} .

On voudrait prédire la classe de la donnée x = (Ciel = *ensoleillé*, Température = *fraiche*, Humidité = *élevée*, Vent = *fort*). Si on utilise la règle de Bayes, on aura:

$$p(jouer = oui|x, \mathcal{D}) = \frac{p(x|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{p(x|\mathcal{D})}$$

$$= \frac{p(x|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{\sum_{a \in \{oui,non\}} p(x|jouer = a, \mathcal{D})p(jouer = a|\mathcal{D})}$$

Classification par la règle de Bayes

Jour	Ciel	Température	Humidité	Vent	Jouer
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

attribut	valeur	J = oui	J = non
Ciel	Ensoleillé		
	Couvert		
	Pluie		
Température	Chaude		
	Tiède		
	Fraiche		
Humidité	Élevée		
	Normale		
Vent	Fort		
	Faible		
Jouer			

Classification par la règle de Bayes

Jour	Ciel	Température	Humidité	Vent	Jouer
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

attribut	valeur	J = oui	J = non
Ciel	Ensoleillé	2	
	Couvert		
	Pluie		
Température	Chaude		
	Tiède		
	Fraiche		
Humidité	Élevée		
	Normale		
Vent	Fort		
	Faible		
Jouer			

Classification par la règle de Bayes

Jour	Ciel	Température	Humidité	Vent	Jouer
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

attribut	valeur	J = oui	J = non
Ciel	Ensoleillé	2	3
	Couvert		
	Pluie		
Température	Chaude		
	Tiède		
	Fraiche		
Humidité	Élevée		
	Normale		
Vent	Fort		
	Faible		
Jouer			

Classification par la règle de Bayes

Jour	Ciel	Température	Humidité	Vent	Jouer
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

attribut	valeur	J = oui	J = non
Ciel	Ensoleillé	2	3
	Couvert	4	0
	Pluie	3	2
Température	Chaude		
	Tiède		
	Fraiche		
Humidité	Élevée		
	Normale		
Vent	Fort		
	Faible		
Jouer		9	5

Classification par la règle de Bayes

attribut	valeur	Jouer = oui	Jouer = non
Ciel	Ensoleillé	2	3
	Couvert	4	0
	Pluie	3	2
Température	Chaude	2	2
	Tiède	4	2
	Fraiche	3	1
Humidité	Élevée	3	4
	Normale	6	1
Vent	Fort	3	3
	Faible	6	2
Jouer		**9**	**5**

Classification par la règle de Bayes

Suivant le tableau, on a:

- $p(jouer = oui|\mathcal{D}) = \frac{9}{14}$.
- $p(jouer = non|\mathcal{D}) = \frac{5}{14}$.
- $p(x|jouer = oui, \mathcal{D}) \overset{\textbf{HBN}}{=} p(\text{Ciel}= \textit{ensoleillé}|jouer = oui, \mathcal{D})$

$$\times\, p(\text{Température} = \textit{fraiche}|jouer = oui, \mathcal{D})$$

$$\times\, p(\text{Humidité} = \textit{élevée}|jouer = oui, \mathcal{D})$$

$$\times\, p(\text{Vent} = \textit{fort}|jouer = oui, \mathcal{D})$$

$$= \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} = \frac{2}{243}$$

Classification par la règle de Bayes

$$p(x|jouer = non, \mathcal{D}) = p(\text{Ciel} = \textit{ensoleillé}|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Température} = \textit{fraiche}|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Humidité} = \textit{élevée}|jouer = non, \mathcal{D})$$

$$\times\, p(\text{Vent} = \textit{fort}|jouer = non, \mathcal{D})$$

$$= \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} = \frac{36}{625}$$

$$p(jouer = oui|x, \mathcal{D}) = \frac{\frac{2}{243} \times \frac{9}{14}}{\frac{2}{243} \times \frac{9}{14} + \frac{36}{625} \times \frac{5}{14}} \approx 0.205$$

$$p(jouer = non|x, \mathcal{D}) = 0.795$$

Attributs numériques

- Si **l'attribut est numérique**, la détermination de la probabilité d'observation d'une valeur particulière ne peut pas se faire comme on vient de le présenter.

- On suppose que la **distribution** de la valeur de l'attribut x_d est **normale**: $\mathrm{x}_d \sim \mathcal{N}(\mu_d, \sigma_d)$

$$p(\mathrm{x}_d = x_d | \mu_d, \sigma_d) = \frac{1}{\sigma_d\sqrt{2\pi}} \exp\left(-\frac{(x_d - \mu_d)^2}{2\sigma_d{}^2}\right)$$

- On calcule la moyenne et l'écart-type de chaque attribut numérique et pour chaque valeur de l'attribut cible.

Exemple

Jour	Ciel	Température	Humidité	Vent	Jouer au tennis ?
1	Ensoleillé	$27,5$	85	Faible	Non
2	Ensoleillé	25	90	Fort	Non
3	Couvert	$26,5$	86	Faible	Oui
4	Pluie	20	96	Faible	Oui
5	Pluie	19	80	Faible	Oui
6	Pluie	$17,5$	70	Fort	Non
7	Couvert	17	65	Fort	Oui
8	Ensoleillé	21	95	Faible	Non
9	Ensoleillé	$19,5$	70	Faible	Oui
10	Pluie	$22,5$	80	Faible	Oui
11	Ensoleillé	$22,5$	70	Fort	Oui
12	Couvert	21	90	Fort	Oui
13	Couvert	$25,5$	75	Faible	Oui
14	Pluie	$20,5$	91	Fort	Non

Attributs numériques

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}$$

	Température		Humidité	
	oui	non	oui	non
	26,5	27,5	86	85
	20	25	96	90
	19	17,5	80	70
	17	21	65	95
	19,5	20,5	70	91
	22,5		80	
	22,5		70	
	21		90	
	25,5		75	
moyenne	21,5	22,3	79,1	86,2
écart-type	2,91	3,53	10,2	9,7

Attributs numériques

- On peut alors calculer par exemple:

$$p(\text{Température = 18}|jouer = oui, \mathcal{D}) = \frac{1}{\sqrt{2\pi(2.91)^2}} e^{-\frac{(18-21.5)^2}{2(2.91)^2}}$$

$$\approx 0.0665$$

$$p(\text{Température = 18}|jouer = non, \mathcal{D}) = \frac{1}{\sqrt{2\pi(3.53)^2}} e^{-\frac{(18-22.3)^2}{2(3.53)^2}}$$

$$\approx 0.0538$$

Attributs numériques

$$p(\text{Humidité} = 90 | jouer = oui, \mathcal{D}) = \frac{1}{\sqrt{2\pi(10.2)^2}} e^{-\frac{(90-79.1)^2}{2(10.2)^2}}$$

$$\approx 0.0221$$

$$p(\text{Humidité} = 90 | jouer = non, \mathcal{D}) = \frac{1}{\sqrt{2\pi(9.70)^2}} e^{-\frac{(90-86.2)^2}{2(9.70)^2}}$$

$$\approx 0.0381$$

Donc pour une journée: x = (Ciel= *ensoleillé*, Température = *18*, Humidité = *90*, Vent = *fort*), on aura:

Attributs numériques

$$p(jouer = oui|x, \mathcal{D}) = \frac{2/9 \times 0.0665 \times 0.0221 \times 3/9 \times 9/14}{p(x|\mathcal{D})}$$

$$\approx \frac{2.1 \times 10^{-4}}{p(x|\mathcal{D})}$$

$$p(jouer = non|x, \mathcal{D}) = \frac{3/5 \times 0.0538 \times 0.0381 \times 3/5 \times 5/14}{p(x|\mathcal{D})}$$

$$\approx \frac{2.635 \times 10^{-4}}{p(x|\mathcal{D})}$$

Attributs numériques

En ayant:

$$p(x|\mathcal{D}) = 2.1 \times 10^{-4} + 2.635 \times 10^{-4}$$

On obtient:

$$p(jouer = oui|x, \mathcal{D}) = \frac{2.1 \times 10^{-4}}{2.1 \times 10^{-4} + 2.635 \times 10^{-4}} \approx 0.44$$

$$p(jouer = non|x, \mathcal{D}) = \frac{2.635 \times 10^{-4}}{2.1 \times 10^{-4} + 2.635 \times 10^{-4}} \approx 0.56$$

Valeurs d’attributs manquantes

- La **méthode de Bayes** n'est pas applicable telle quelle si une certaine valeur d'attribut n'apparaît pas dans l’ensemble d'entraînement $\mathcal{D}$.

- En effet, dans ce cas, on voit apparaitre des probabilités estimées à 0. Comme **les probabilités sont multipliées les unes par les autres**, on obtient 0 à la fin du compte.

Exemple: Il n’y a pas d’exemples négatifs pour la valeur de **Ciel** = *Couvert*. L’estimation est $p($**Ciel** = *Couvert*$|\text{jouer}=\text{non}, \mathcal{D}) = 0$. D'un point de vue conceptuel, cela n'a pas de sens. D'un point de vue pratique, le 0 pose problème.

Valeurs d’attributs manquantes

- Il est du au manque d'exemples correspondants. Il est naturel de se dire que 0 **cache en réalité une valeur très petite**.

- On peut remplacer ce 0 **par une petite valeur**; tout en gardant les règles sur les probabilités vérifiées.

- La solution consiste à s'arranger pour que cette estimation ne soit pas nulle. Pour cela, on utilise la technique dite de **l'estimateur de Laplace** .

- On ajoute une valeur (par exemple 1) à chaque décompte.

Valeurs d'attributs manquantes

- Par exemple, pour l'attribut « **Ciel** », au lieu d'avoir les probabilités estimées $\frac{3}{5}, \frac{0}{5}$ et $\frac{2}{5}$, on aurait $\frac{4}{8}, \frac{1}{8}$ et $\frac{3}{8}$.

- Plus généralement, on peut ajouter une valeur μ à chaque dénominateur pour un attribut donné et ajouter μ-arité de l'attribut considéré au numérateur.

- μ-arité de l'attribut considéré peut être vue comme une ***probabilité a priori*** de l'observation de chacune des valeurs de l'attribut .

Valeurs d'attributs manquantes

$$\mu = 3$$

$$p(jouer = oui|\text{Ciel} = \text{couvert}, \mathcal{D}) = \frac{p(\text{Ciel} = \text{couvert}|jouer = oui, \mathcal{D})p(jouer = oui|\mathcal{D})}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

La probabilité 0 est remplacé par une petite valeur

$$= \frac{4/9 \times 9/14}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

$$p(jouer = non|\text{Ciel} = \text{couvert}, \mathcal{D}) = \frac{1/8 \times 5/14}{p(\text{Ciel} = \text{couvert}|\mathcal{D})}$$

$$p(jouer = oui|\text{Ciel} = \text{couvert}, \mathcal{D}) = \frac{\frac{4}{9} \times \frac{9}{14}}{\frac{4}{9} \times \frac{9}{14} + \frac{1}{8} \times \frac{5}{14}} \approx 0.8649$$

Validation d'un classificateur

Validation d'un Classificateur de Bayes

- Une fois un CB construit, il est essentiel de le valider en estimant **la probabilité que la classe prédite** pour une donnée quelconque soit correcte. Cette quantité est donc une variable aléatoire dont il faut estimer la valeur.

- **L'erreur de classification** *E* d'un classificateur est la probabilité que ce dernier ne prédise pas correctement la classe d'une donnée de $\mathcal{D}$. **Le taux de succès** est égal a 1 - *E*.

- **L'erreur apparente** E_{app} est mesurée avec les exemples $\mathcal{D}_{app}$: c'est la proportion d'exemples dont la classe est mal prédite par le classificateur.

Validation d'un Classificateur de Bayes

Remarques:

- $\boldsymbol{E}_{\text{app}}$ n'est **pas un bon estimateur de l'erreur qui serait commise face à de nouvelles données**.

- L'apprentissage doit pouvoir être généralisé à de nouvelles données: **l'enjeu est bien là !**

- Un algorithme ne peut être qualifié **de bon algorithme d'apprentissage** que s'il est capable de **généraliser** ce qu'il a appris **à de nouvelles données**.

Validation d'un Classificateur de Bayes

Pour la validation, on distingue :

- Le **jeu d'exemples d'apprentissage** $\mathcal{D}_{\text{app}}$ ou **d'entrainement** avec lequel le CB est construit.

- Le **jeu d'exemples de test** $\mathcal{D}_{\text{test}}$ pour lesquels, on connait les valeurs des classes. On les classe avec le CB construit avec $\mathcal{D}_{\text{app}}$, puis on regarde s'ils sont classés correctement.

- Si on ne dispose pas d'un jeu de test, on découpe $\mathcal{D}_{\text{app}}$ en deux parties; l'une constituera **le jeu d'apprentissage effectif** pour construire le CB; l'autre servira à évaluer (technique de ***holdout***).

Mesure de qualité d'un classificateur

- On suppose que l'on a construit un classificateur (un CB ou un autre) à partir d'un jeu d'exemples d'apprentissage $\mathcal{D}_{\text{app}}$.

- Dans le cas de **classification binaire** (sachant que les définitions sont aisément extensibles aux autres cas), on définit :

➢ ***vp*** : le nombre **de vrais positifs**, c.-à-d. les exemples de classe positive dont la classe est prédite comme positive;

➢ ***vn*** : le nombre **de vrais négatifs**, c.-à-d. les exemples de classe négative dont la classe est prédites comme négative ;

Mesure de qualité d'un classificateur

- ***fp*** : le nombre **de faux positifs**, c.-à-d. les exemples de classe négative dont la classe est prédite comme positive ;

- ***fn*** : le nombre **de faux négatifs**, c.-à-d. les exemples de classe positive dont la classe négative.

	+	−	← classe prédite
+	VP	FN	
−	FP	VN	
↑ classe			

Tableau de contingence (pour K > 2, on l'appelle **matrice de confusion**)

Mesure de qualité d’un classificateur

- S'il n'y a des nombres non nuls que sur la diagonale principale, c'est qu'aucun exemple n'est mal classe (**bonne classification**).

- On peut définir aussi deux statistiques, la **précision** et le **rappel** :

  - ***Précision*** *pour les positifs* = $\frac{vp}{vp + fp}$

  - ***précision*** *pour les négatifs* = $\frac{vn}{vn + fn}$

  - ***Rappel*** *pour les positifs* = $\frac{vp}{vp + fn}$

  - ***Rappel*** *pour les négatifs* = $\frac{vn}{vn + fp}$

Mesure de qualité d'un classificateur

- La **précision** mesure la proportion d'exemples vraiment positifs (resp. négatifs) parmi ceux qui sont classés comme positifs (resp. négatifs).
- Le **rappel** mesure la proportion d'exemples vraiment positifs (resp. négatifs) parmi tous les exemples de classes positive (resp. négative).

- Il est toujours plus pratique de manipuler un seul nombre qui synthétise les autres. Ainsi la **mesure -*F*** est définie par :

$$F = \frac{2 \times rappel \times précision}{rappel + précision} = \frac{2 \times vp}{2 \times vp + fp + fn}$$

Validation croisée

- Une méthode plus sophistiquée est la **validation croisée**. Pour cela, **on découpe l'ensemble** des exemples en *n* sous-ensembles mutuellement disjoints.

- Il faut prendre garde à ce que **chaque classe apparaisse avec la même fréquence** dans les *n* sous-ensembles (**stratification des échantillons**).

- Si *n* = 3, on aura 3 ensembles A, B et C. On construit le CB $CB_{A \cup B}$ avec $A \cup B$ et on mesure le taux d'erreur sur C: E_C, c.-à-d. le nombre d'exemples de C dont la classe est mal prédite par $CB_{A \cup B}$.

Validation croisée

- Ensuite, on construit le CB $CB_{C \cup B}$ avec $C \cup B$ et on mesure le taux d'erreur sur A: E_A.

- Enfin, on construit le CB $CB_{A \cup C}$ avec $A \cup C$ et on mesure le taux d'erreur sur B: E_B.

- Le taux d'erreur E est alors estimé par la moyenne de ces trois erreurs, comme suit:

$$E = \frac{E_A + E_B + E_C}{3}$$

Références

1. M. S. Allili. Techniques d’apprentissage automatique (Cours de 2e cycle). Université du Québec en Outaouais (UQO), Québec, Canada. Hivers 2015.
2. S. Rogers et M Girolami. A first Course in machine learning, CRC press, 2012.
3. C. Bishop. Pattern Recognition and Machine learning. Springer 2006.
4. R. Duda, P. Storck et D. Hart. Pattern Classification. Prentice Hall, 2002.