



Module 2

Théorie de décision de Bayes

Sommaire

2.1	Introduction	2
2.2	Règle de la décision de Bayes	3
2.2.1	Théorème de Bayes	3
2.2.2	Règle de décision	5
2.2.3	Généralisation de la règle de Bayes	7
2.3	Fonction coût	9
2.4	Classificateur à taux d'erreurs minimum	11
2.5	Fonctions discriminantes	12
2.6	Annexes	14

2.1 Introduction

La théorie Bayésienne de décision constitue le fondement théorique de la reconnaissance des formes. Elle porte sur le calcul de la probabilité d'un événement basé sur une connaissance *a priori*, c'est-à-dire qui prend en compte des connaissances préalables. La théorie de décision bayésienne est utilisée dans divers domaines tels que l'économie, la statistique, la psychologie, la sociologie, la politique et la philosophie.

- **Exemple 2.1** Afin d'introduire la théorie de décision de Bayes, nous commençons par cet exemple simple. Considérons un jeu ordinaire de 52 cartes constitué de quatre types de cartes : pique, coeur, carreau et trèfle. Ces cartes forment treize valeurs ordonnées dans l'ordre croissant : As, 2 à dix, valet, dame et roi.

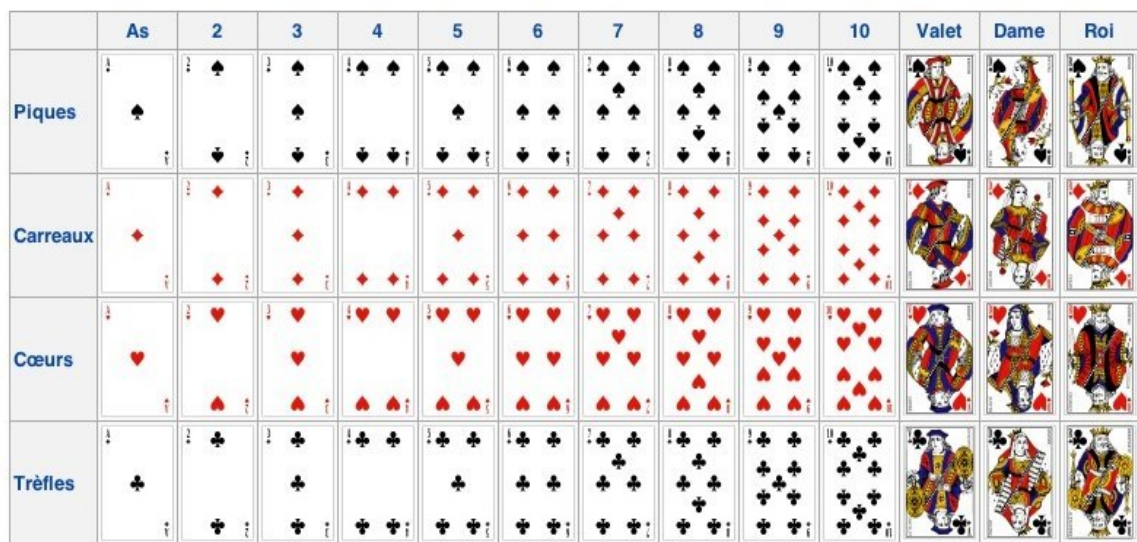


Figure 2.1: Jeu de 52 cartes [1]

Le mathématicien britannique **Thomas Bayes** (1702-1761)

est connu pour avoir formulé le théorème de Bayes. Ses découvertes en probabilités ont été résumées dans un document qui décrit la manière de résoudre un problème dans la doctrine des risques (Essay Towards Solving a Problem in the Doctrine of Chances) publié en 1763 par son ami Richard Price. (Extrait de Wikipédia [2])



On s'intéresse aux deux événements consécutifs suivants :

1. Événement A : tirer une carte contenant une figure, c'est-à-dire un roi, une dame ou un valet.
2. Événement B : tirer une carte dont la valeur est supérieure à 10 (c'est-à-dire un as ou une figure) et qui est noire (pique ou trèfle).

Dans le cas de l'événement A, il est facile de remarquer à partir de l'image précédente que la chance de tirer une carte contenant une figure est

$$P(A) = \frac{12}{52}$$

Dans la théorie de Bayes, cette probabilité est dite *probabilité a priori* (antérieure ou préalable) parce qu'elle précède toute information sur l'événement B.

Supposons maintenant que, avant de voir le contenu de la carte tirée, nous apprenons que l'événement B s'est aussi réalisé. Cela veut dire que la carte tirée est une image et que cette image est noire (pique ou trèfle). Cette information restreint la valeur de la carte tirée à un ensemble de 8 cartes dont 6 sont des figures (et 2 sont des as). La probabilité restreinte est appelée *probabilité a posteriori*; elle est notée $P(A|B)$:

$$P(A|B) = \frac{6}{8} = \frac{6/52}{8/52}$$

2.2 Règle de la décision de Bayes

2.2.1 Théorème de Bayes

En théorie des probabilités, la probabilité conditionnelle d'un événement A sachant qu'un autre événement B de probabilité non nulle s'est réalisé (probabilité de A sachant B) est

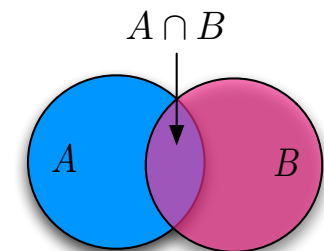


Figure 2.2: Probabilité conditionnelle.

défini par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.1)$$

$$P(A|B)P(B) = P(A \cap B) \quad (2.2)$$

Or,

$$P(A \cap B) = P(B \cap A)$$

Alors,

$$P(A|B)P(B) = P(B|A)P(A)$$

Nous obtenons ainsi, la relation (2.3) entre les probabilités conditionnelles :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.3)$$

Cette relation est appelé le **théorème de Bayes**. Elle a joué un rôle central dans les domaines de reconnaissance de formes et d'apprentissage machine comme nous allons le voir dans la suite.

Il est à noter que la théorie de la décision de Bayes est très bien expliquée dans le livre de Duda, Hart et Stork [3]. Le texte de référence de ce module s'inspire du contenu de ce livre qui est considéré comme un livre de référence en apprentissage automatique et en classification de formes. Ce livre est disponible dans le site :

<http://cns-classes.bu.edu/cn550/Readings/duda-et-al-00.pdf>.

Nous invitons les lecteurs qui désirent avoir plus de détails à consulter le chapitre 2 - *Bayesian Decision theory*, à la page 39.

- **Exemple 2.2** L'objectif de cet exemple est d'introduire la règle de la décision bayésienne. Supposons qu'une personne distribue de manière aléatoire (à l'aveuglette) des fruits à partir d'un distributeur de fruits contenant des pommes et des oranges (Figure 2.3). Dans la



Figure 2.3: Distributeur de fruits.

terminologie de la théorie de prise de décisions, nous disons que le fruit constitue la *forme* ou l'*observation* et que *Pomme* et *Orange* constituent les deux états ou classes possibles. Soit ω la classe du fruit distribué avec $\omega = \omega_1$ pour pomme et $\omega = \omega_2$ pour orange.

Étant donné le fait que la distribution se fait de manière aléatoire, la variable ω peut être décrite de façon probabiliste. Plus spécifiquement, nous supposons que la probabilité *a priori* pour que le prochain fruit soit une pomme est $P(\omega_1)$ et que la probabilité *a priori* pour que le prochain fruit soit une orange est $P(\omega_2)$.

Supposons que la personne qui distribue les fruits soit obligée, à un instant donné, de prendre une décision (prédire) sur le type du prochain fruit distribué. Dans ce cas, la prise de décision se basera uniquement sur la valeur des probabilités a priori selon la règle suivante :

$$\begin{cases} \text{Décider} & \omega = \omega_1 \text{ si } P(\omega_1) > P(\omega_2) \\ \text{Sinon} & \omega = \omega_2 \end{cases} \quad (2.4)$$

Cette décision est très simpliste parce qu'elle ne considère aucune information sur la forme considérée. Ainsi, nous serons amenés à prendre toujours la même décision ($\omega = \omega_1$ si $P(\omega_1) > P(\omega_2)$ sinon $\omega = \omega_2$ si $P(\omega_2) > P(\omega_1)$). D'où le besoin d'introduire une caractéristique, c'est-à-dire une mesure qui décrit la forme, par exemple, le poids, le périmètre ou le volume du fruit.

2.2.2 Règle de décision

Soit x une mesure sur laquelle nous nous baserons pour la prise de décision. Notons $p(\omega_i|x)$ la probabilité de x si x appartient à ω_i . En considérant la mesure x , la règle de décision de l'équation (2.4) peut être reformulée comme suit :

$$\begin{cases} \text{Décider} & \omega = \omega_1 \text{ si } P(\omega_1|x) > P(\omega_2|x) \\ \text{Sinon} & \omega = \omega_2 \end{cases} \quad (2.5)$$

Or d'après l'équation (2.3), nous pouvons écrire :

- pour la classe ω_1 :

$$P(\omega_1|x) = \frac{p(x|\omega_1).P(\omega_1)}{p(x)}$$

- pour la classe ω_2 :

$$P(\omega_2|x) = \frac{p(x|\omega_2).P(\omega_2)}{p(x)}$$

avec

$$p(x) = p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)$$

$P(\omega_1|x)$ est la probabilité de x si $x \in \omega_1$. (respectivement pour $P(\omega_2|x)$).

$p(x|\omega_1)$ est la vraisemblance, c'est-à-dire la probabilité que x provienne de la classe ω_1 . (respectivement pour $p(x|\omega_2)$).

$P(\omega_1)$ est l'a *priori*, c'est-à-dire la probabilité a priori de la classe ω_1 . (respectivement pour $P(\omega_2)$).

$p(x)$ est l'évidence ; il s'agit tout simplement d'un facteur de normalisation.

La règle de décision de l'équation (2.5) devient alors :

$$\begin{cases} \text{Décider} & \omega = \omega_1 \text{ si } p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2) \\ \text{Sinon} & \omega = \omega_2 \end{cases}$$

Cas de classes équiprobables

Lorsque nous sommes en présence de deux classes équiprobables $P(\omega_1) = P(\omega_2)$, la règle de décision devient :

$$\begin{cases} \text{Décider} & \omega = \omega_1 \text{ si } p(x|\omega_1) > p(x|\omega_2) \\ \text{Sinon} & \omega = \omega_2 \end{cases} \quad (2.6)$$

Cas de vraisemblance égales

Si $p(x|\omega_1) = p(x|\omega_2)$ cela veut dire qu'une observation particulière de la mesure x n'ajoute aucune information sur la forme, la règle de décision devient :

$$\begin{cases} \text{Décider} & \omega = \omega_1 \text{ si } P(\omega_1) > P(\omega_2) \\ \text{Sinon} & \omega = \omega_2 \end{cases}$$

ce qui permet de revenir à la règle de décision de départ décrite par l'équation 2.4

2.2.3 Généralisation de la règle de Bayes

La règle de Bayes peut être généralisée à plusieurs classes :

$$P(\omega_i|x) = \frac{p(x|\omega_i) \cdot P(\omega_i)}{p(x)} \quad i = 1, 2, \dots, c. \quad (2.7)$$

c étant le nombre de classes.

$P(\omega_i|x)$ est la probabilité de x si $x \in \omega_i$.

$p(x|\omega_i)$ est la vraisemblance, c'est-à-dire la probabilité que x provienne de la classe ω_i .

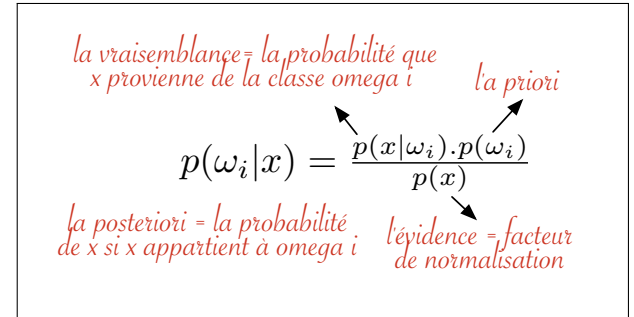
$P(\omega_i)$ est la probabilité a priori de la classe ω_i .

$p(x)$ est l'évidence ; elle est donnée par

$$p(x) = \sum_i p(x|\omega_i) \cdot P(\omega_i).$$

Le théorème de Bayes peut être exprimé informellement selon :

$$\text{Prob. a posteriori} = \frac{\text{vraisemblance} \times \text{a priori}}{\text{Evidence}}$$


$$p(\omega_i|x) = \frac{p(x|\omega_i) \cdot P(\omega_i)}{p(x)}$$

Densité normale

Si une variable aléatoire $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\text{Alors } f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

► **Exemple 2.3** L'objectif de cet exemple est d'appliquer la règle de la décision bayésienne. Supposons que la distribution des tailles des filles ($\omega_1 = F$) dans la population étudiante canadienne suive une loi normale de moyenne 1.65 m et d'écart-type 0.16 m et que celle des garçons ($\omega_2 = G$) suit une loi normale de moyenne 1.75 m et d'écart-type 0.15 m (figure 2.4). Nous avons alors :

$$p(t|F) = \frac{1}{0.16\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t - 1.65}{0.16}\right)^2\right) \quad (2.8)$$

$$p(t|G) = \frac{1}{0.15\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{t - 1.75}{0.15}\right)^2\right) \quad (2.9)$$

Supposons également qu'il y ait autant de filles que de garçons dans la population considérée. Ceci veut dire que les deux classes sont équiprobables :

$$P(F) = P(G) = 0.5$$

La question qui se pose est alors : étant donné un individu de cette population qui mesure 1.60 m, quelle est la probabilité pour que cet individu soit une femme ?

La formule de Bayes fournit directement la réponse en attribuant à la caractéristique x la valeur 1.60 :

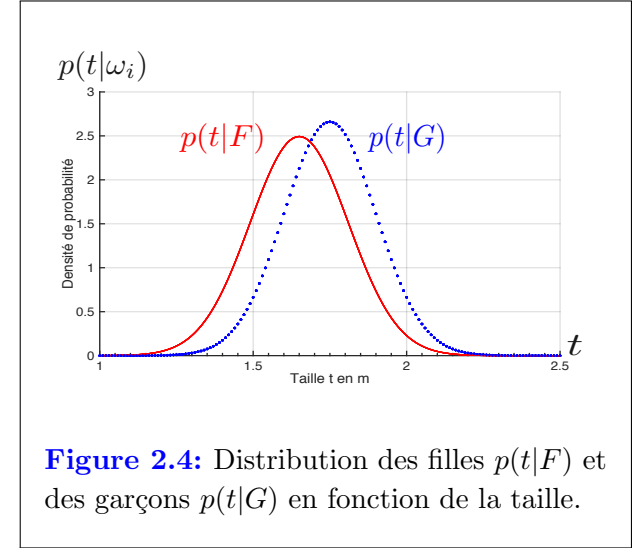
$$P(F|1.60) = \frac{0.5p(1.60|F)}{0.5p(1.60|F) + 0.5p(1.60|G)} = \frac{0.5 \times 2.38}{0.5 \times 2.38 + 0.5 \times 1.61} \approx 60\%$$

Il est à noter que $P(1.60|F) = 2.38$ a été calculée à partir de l'Eq. 2.10 en remplaçant t par 1.60.

$$p(1.60|F) = \frac{1}{0.16\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{1.60 - 1.65}{0.16}\right)^2\right) \quad (2.10)$$

Nous pouvons, par la suite, déduire que $p(H|1.60) = 40\%$ (c-à-d 1-60%). Nous pouvons, également, calculer cette probabilité :

$$P(G|1.60) = \frac{0.5p(1.60|G)}{0.5p(1.60|F) + 0.5p(1.60|G)} = \frac{1.61}{2.38 + 1.61} \approx 40\%$$



Nous avons $P(F|1.60) > P(G|1.60)$, alors nous pouvons décider que l'individu décrit par $t = 1.60$ sera affecté à la classe F .

La figure 2.5, permet intuitivement de prévoir ce résultat puisque à $t = 1.6$ la valeur de la courbe rouge est supérieure à celle de la courbe bleue. Nous attirons votre attention à ce niveau, que nous ne pouvons tirer une telle conclusion qu'en raison du fait deux classes sont équiprobables.

Supposons maintenant que l'individu observé soit choisi au hasard dans les tribunes d'un stade au cours d'un match de football américain. Dans ce cas, il y a généralement plus de garçons que de filles. Admettons que les proportions soient de 30% de filles et 70% de garçons. Cela veut dire que $p(F) = 0.3$ et $P(G) = 0.7$.

Dans ce cas l'application de la formule de Bayes donne :

$$p(F|1.60) = \frac{0.3p(1.60|F)}{0.3p(1.60|F) + 0.3p(1.60|H)} \approx 39\%$$

et

$$p(G|1.60) = 61\%$$

Compte tenu de ces résultats, l'individu décrit par $t = 1.60$ sera affecté à la classe G .

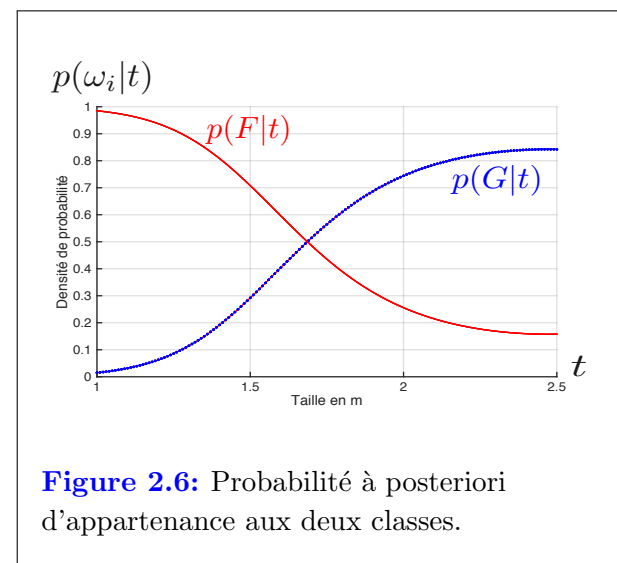
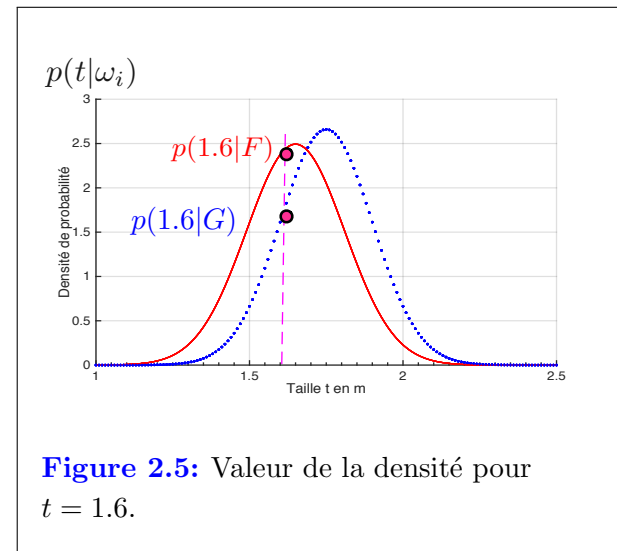
2.3 Fonction coût

Soit $\mathbf{x}$ un vecteur de caractéristiques de dimension d . Comme nous l'avons vu précédemment, la probabilité a posteriori est donnée par la formule de Bayes selon :

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i).P(\omega_i)}{p(\mathbf{x})} \quad i = 1, 2, \dots, c.$$

avec

$$p(\mathbf{x}) = \sum_i p(\mathbf{x}|\omega_i).P(\omega_i)$$



Supposons que nous observons un vecteur $\mathbf{x}$ particulier et que nous envisageons prendre des décisions aussi appelées actions α_i . Si la classe réelle est ω_j , nous encourons une perte $\lambda(\alpha_i|\omega_j)$. Or $P(\omega_j|\mathbf{x})$ est la probabilité de $\mathbf{x}$ si $\mathbf{x}$ provient de ω_j . Donc la perte attendue associée à une action α_i est simplement

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x}) \quad (2.11)$$

$R(\alpha_i|\mathbf{x})$ est le *risque conditionnel* associé à α_i sur la base de l'observation $\mathbf{x}$.

Cas de deux classes

Soit α_1 l'action à prendre si la décision correspond à ω_1 et α_2 l'action à prendre si la décision correspond à ω_2 .

Soit $\lambda_{ij} = \lambda(\alpha_i|\omega_j)$, le coût associé à la décision α_i quand $\mathbf{x}$ appartient à ω_j . Nous pouvons alors écrire pour chacune des classes :

$$R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$$

$$R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$$

La règle de décision s'exprime alors :

$$\begin{cases} \text{Décider } \omega = \omega_1 & \text{si } R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x}) \\ \text{Sinon} & \omega = \omega_2 \end{cases} \quad (2.12)$$

Ou bien encore :

$$\text{Décider } \omega = \omega_1 \text{ si } \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x}) > \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$$

C'est-à-dire :

$$\text{Décider } \omega = \omega_1 \text{ si } (\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x})$$

En utilisant la formule de Bayes, on obtient :

$$\text{Décider } \omega = \omega_1 \text{ si } (\lambda_{21} - \lambda_{11})p(\mathbf{x}|\omega_1)P(\omega_1) > (\lambda_{12} - \lambda_{22})(\mathbf{x}|\omega_2)P(\omega_2)$$

Supposons que $\lambda_{21} > \lambda_{11}$. Cette supposition est très raisonnable parce qu'elle veut dire que nous désirons que la perte en faisant une erreur soit supérieure à la perte lorsque nous ne faisons pas d'erreur. Dans ce cas :

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)} \quad (2.13)$$

Le rapport $p(\mathbf{x}|\omega_1)/p(\mathbf{x}|\omega_2)$ est appelé le *rapport de vraisemblance*.

La règle de décision de Bayes peut être interprétée par : décider ω_1 si le rapport de vraisemblance est supérieur à un certain seuil indépendant de l'observation $\mathbf{x}$.

2.4 Classificateur à taux d'erreurs minimum

Dans cette section, nous considérons un problème de classification de formes dans lequel nous disposons de c classes. L'action α_i est généralement : décider que la forme appartient à la classe ω_i . Nous pouvons donc considérer la fonction de coût suivante :

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad (2.14)$$

Cette fonction est appelée fonction perte 0-1 (*0-1 loss function*). Elle affecte la valeur nulle (aucune perte) à une bonne décision ($i = j$) et une valeur de 1 à une mauvaise décision ($i \neq j$).

Le risque conditionnel correspondant à cette fonction coût est déterminé à partir de

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

le rapport de vraisemblance *Rapport indépendant de la valeur de la mesure : dépend des fonctions de perte et de la priori*

l'équation (2.11) :

$$\begin{aligned} R(\alpha_i|\mathbf{x}) &= \sum_{j=1}^c \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j|\mathbf{x}) \end{aligned} \quad (2.15)$$

$$= 1 - P(\omega_i|\mathbf{x}) \quad (2.16)$$

Afin de minimiser la moyenne de la probabilité d'erreur (le risque conditionnel), il faut choisir i qui maximise la probabilité a posteriori $P(\omega_i|\mathbf{x})$. Autrement dit, le classificateur à taux d'erreurs minimum implique :

$$\text{Décider } \omega_i \text{ Si } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \quad \forall i \neq j \quad (2.17)$$

Cette équation est très importante parce qu'elle permet de retrouver dans la règle de décision la règle de l'équation 2.5.

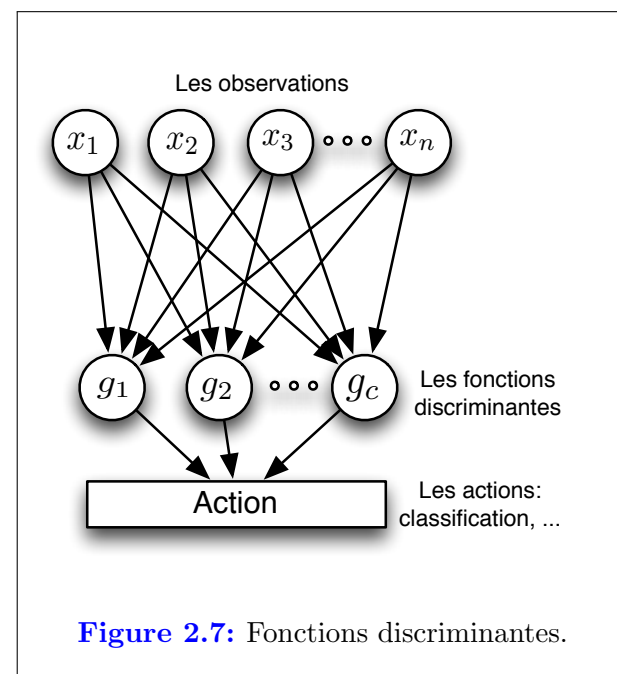
2.5 Fonctions discriminantes

Il existe différentes manières de représenter un classificateur. Une des plus utilisées est une représentation sous la forme d'un ensemble de *fonctions discriminantes* $g_i(\mathbf{x})$, $i = 1, \dots, c$. c étant le nombre de classes. Le classificateur affecte le vecteur de caractéristiques $\mathbf{x}$ à la classe ω_i si :

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall j \neq i \quad (2.18)$$

Le classificateur bayésien peut être représenté facilement et naturellement sous la forme d'une fonction discriminante. Ainsi, en exprimant la minimisation du risque conditionnel en termes de fonctions discriminantes, nous pouvons écrire :

$$g_i(\mathbf{x}) = -R(\alpha_i|\mathbf{x}) \quad (2.19)$$



D'après l'équation 2.15, on obtient l'équation 2.20 qui indique que la fonction discriminante maximale correspond à la probabilité à posteriori maximale :

$$g_i(\mathbf{x}) = p(\omega_i|\mathbf{x}) - 1 \quad (2.20)$$

Il est clair que le choix des fonctions discriminantes n'est pas unique. Nous pouvons donc multiplier toutes les fonctions discriminantes par la même constante positive ou les transférer par la même constante additive sans influencer la décision. Les fonctions discriminantes suivantes sont, alors, équivalentes :

1. $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$
2. $g_i(\mathbf{x}) = 2.P(\omega_i|\mathbf{x})$
3. $g_i(\mathbf{x}) = \ln P(\omega_i|\mathbf{x})$

Les fonctions discriminantes engendrent des régions de décision $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_c$ dans l'espace de caractéristiques qui sont décrits par :

Si $g_i(\mathbf{x}) > g_j(\mathbf{x})$ pour tout $j \neq i$ alors $\mathbf{x} \in R_i$ et la règle de décision implique que $\omega = \omega_i$.

Plus particulièrement, dans le cas du classificateur à taux d'erreurs minimum :

$$\begin{aligned} g_i(\mathbf{x}) &= P(\omega_i|\mathbf{x}) \\ &= p(\mathbf{x}|\omega_i)P(\omega_i) \end{aligned} \quad (2.21)$$

La figure 2.8 illustre les frontières de décision dans le cas de deux classes lorsque nous considérons la fonction discriminante de l'équation (2.21).

Cas de deux classes

Lorsque nous disposons de deux classes de formes, il est plus utile d'utiliser une seule fonction discriminante telle que :

$$g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x})$$

Dans ce cas, la règle de décision utilisée est :

$$\begin{cases} \text{Décider } \omega = \omega_1 & \text{si } g(\mathbf{x}) > 0 \\ \text{Sinon } & \omega = \omega_2 \end{cases}$$

En utilisant la forme $g_i(\mathbf{x})$ suivante :

$$g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}),$$

on obtient :

$$\begin{aligned} g(\mathbf{x}) &= P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x}) \\ &= \ln \frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)} \end{aligned}$$

où $\ln$ désigne le logarithme népérien.

2.6 Annexes

Dans cette section, nous citons des fonctions R utiles pour la théorie de la décision de Bayes.

- **e1071** : Le package (*Misc Functions of the Department of Statistics, Probability Theory Group*) comprend plusieurs fonctions d'analyse de données comme la transformée de Fourier, la classification floue, les machines à support de vecteurs et la classification Bayésienne.

<https://cran.r-project.org/web/packages/e1071/index.html>

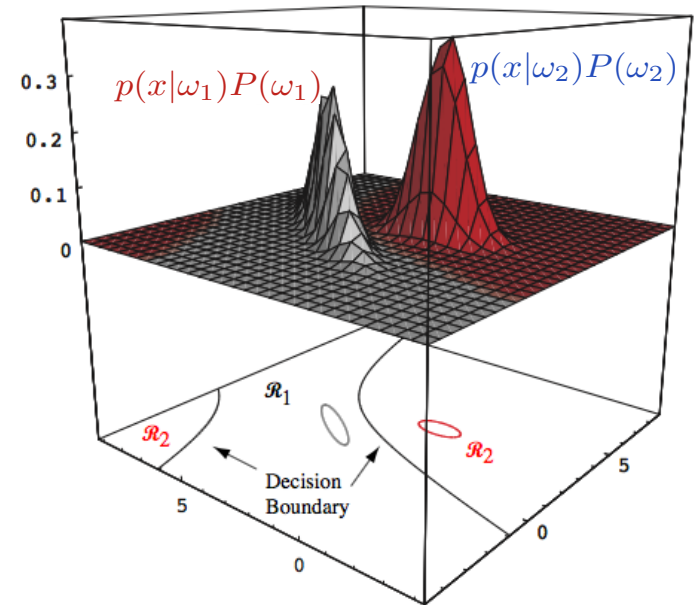


Figure 2.8: Frontières de décision (extrait de [3] et adapté).

- `naiveBayes` : Permet de calculer les probabilités conditionnelles a posteriori d'une variable en utilisant la règle de Bayes .
<http://ugrad.stat.ubc.ca/R/library/e1071/html/naiveBayes.html>
- `dnorm` : Permet de générer une densité normale
<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Normal.html>
- `set.seed` : Permet de reproduire la génération des mêmes nombres aléatoires de R.

Références

- [1] Jeu de 52 cartes. https://fr.wikipedia.org/wiki/Jeu_de_52_cartes. Consulté : 2 janvier 2019.
- [2] Thomas bayes. https://fr.wikipedia.org/wiki/Thomas_Bayes. Consulté : 2 janvier 2019.
- [3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.