Notions de Théorie de la Décision

Essai de présentation « simple »
Avril 2010
commentaires bienvenus (leroux@polytech.unice.fr)

Dans l'approche statistique de la reconnaissance de forme, on est amené à décider si un élément caractérisé par un vecteur de paramètres appartient à une classe ou à une autre, chacune de ces classes correspondant à un nuage de points qui lui sont assignés et si possible à une densité de probabilité. En pratique, cette classification se fait assez rarement sur des bases théoriques rigoureuses, mais il peut tout de même être intéressant d'avoir une idée des méthodes utilisées par les statisticiens quand ils optimisent la classification ; il y a plusieurs approches connues en théorie de la décision parmi lesquelles celle de Bayes et celle de Neyman et Pearson. Cette note tente d'en faire une présentation simpl(ist ?)e illustrée par un exemple.

Approche bayesienne

Un effet mesuré y, représenté par un nombre réel, peut être produit *uniquement* par une cause u qui ne peut prendre que les valeurs 0 ou 1 ; Les probabilités pour que u vaille 0 ou 1, q_0 et q_1, sont connues; Si la cause est $u = 0$, la loi de probabilité de y est $p(y|u = 0)$; si la cause est $u = 1$, la loi de probabilité de y est $p(y|u = 1)$ (la formule de Bayes donnant les probabilités des causes lorsqu'on connaît y est rappelée à la fin de cette note).

Le problème de la décision est le suivant: on a mesuré y ; que vaut il mieux choisir parmi les deux propositions : y a été causé par $u = 0$ ou par $u = 1$; que nous noterons $d = 0$ et $d = 1$?

Pour poser complètement le problème dans l'approche bayesienne, il faut connaître les probabilités des causes q_0 et q_1 et les probabilités conditionnelles; $p(y|u = 0)$ $p(y|u = 1)$, mais il faut aussi se donner une fonction de pénalité : les quatre coûts c associés aux situations possibles

$c(0|0)$ quand on choisit $d = 0$ et que la vraie valeur est $u = 0$
$c(0|1)$ quand on choisit $d = 0$ et que la vraie valeur est $u = 1$
$c(1|0)$ quand on choisit $d = 1$ et que la vraie valeur est $u = 0$
$c(1|1)$ quand on choisit $d = 1$ et que la vraie valeur est $u = 1$

Pour une valeur de y mesurées, on choisira $d = 0$ si le coût associé à ce choix est moins élevé que le coût associé au choix $d = 1$

calcul de la valeur moyenne du coût associé au choix $u = 0$ en tenant compte du fait que cette valeur de y a pu avoir une des deux causes

les quatre possibilité de choix

u=0 & d=0	u=0 & d=1
u=1 & d=0	u=1 & d=1

les probabilités associées

$p(y|u=0).q_0$	$p(y|u=0).q_0$
$p(y|u=1).q_1$	$p(y|u=1).q_1$

Les coûts moyens associés aux décisions sont obtenus en considérant pour chaque décision les probabilités des valeurs possibles de u :

$d = 0$	$d = 1$
$c(0\|0)\ p(y\|u=0).q_0 + c(0\|1)p(y\|u=1).q_1$	$c(1\|0)\ p(y\|u=0).q_0 + c(1\|1)p(y\|u=1).q_1$

On choisit $d = 0$ si en moyenne cela coute moins que de choisir $d = 1$, ce qui se traduit par l'inégalité

$$c(0|0)\ p(y|u=0).q_0 + c(0|1)p(y|u=1).q_1 < c(1|0)\ p(y|u=0).q_0 + c(1|1)p(y|u=1).q_1$$

qu'on réécrit

$$(c(0|1) - c(1|1))p(y|u=1).q_1 < (c(1|0) - c(0|0))\,p(y|u=0).q_0$$

et, en supposant que les couts des mauvaises décisions sont plus élevés que les couts des décisions correctes ($c(0|0) < c(1|0)$ et $c(1|1) < c(0|1)$)

$$\frac{p(y\,|\,u=1)}{p(y\,|\,u=0)} < \frac{c(1\,|\,0)-c(0\,|\,0)}{c(0\,|\,1)-c(1\,|\,1)}\cdot\frac{q_0}{q_1}$$

ou encore

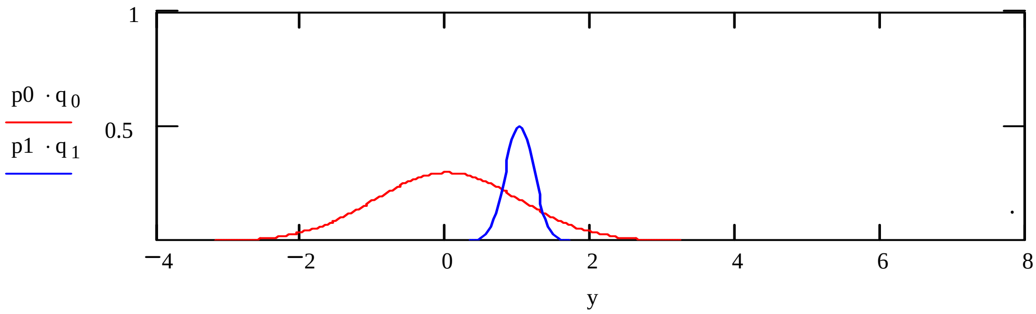
$$\boxed{\frac{p(y\,|\,u=1)}{p(y\,|\,u=0)}\cdot\frac{q_1}{q_0} < \frac{c(1\,|\,0)-c(0\,|\,0)}{c(0\,|\,1)-c(1\,|\,1)}}$$

Un exemple : deux lois de probabilités conditionnelles gaussiennes

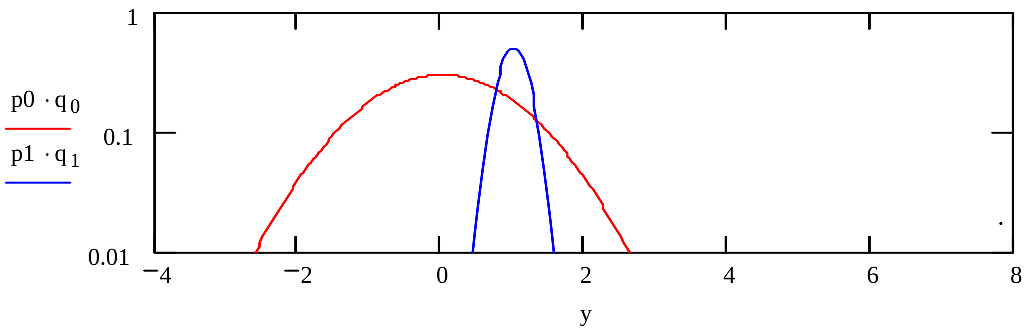
$$m_0 := 0 \quad \sigma_0 := 1 \qquad m_1 := 1 \qquad \sigma_1 := 0.2 \qquad p_0 := 0.75 \qquad p_1 := 0.25$$

$$p0_n := \frac{1}{\sigma_0\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\frac{-(y_n - m_0)^2}{2\cdot(\sigma_0)^2}\right] \qquad p1_n := \frac{1}{\sigma_1\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\frac{-(y_n - m_1)^2}{2\cdot(\sigma_1)^2}\right]$$

Représentations de ces lois pondérées par les probabilités a priori des causes



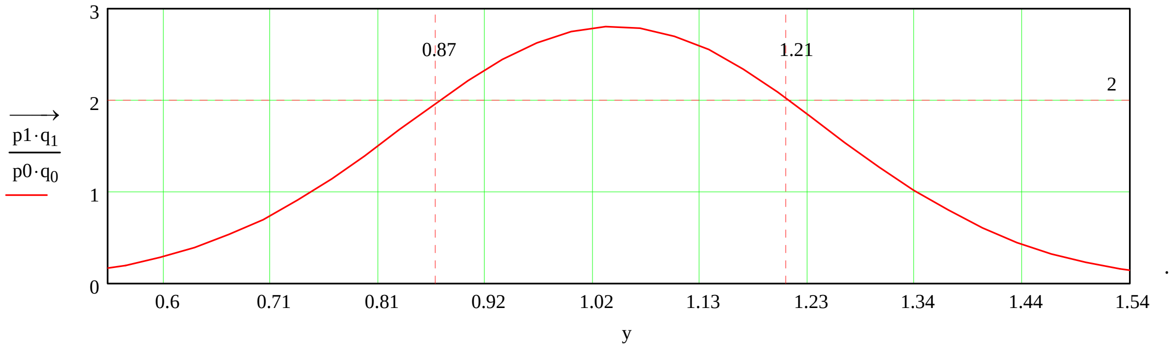
les mêmes présentées sur une échelle logarithmique des ordonnées



On choisit des pénalités telles que

$$\frac{c(1\,|\,0) - c(0\,|\,0)}{c(0\,|\,1) - c(1\,|\,1)} = 2$$

Pour minimiser le critère, on choisira l'hypothèse u = 1 si y est dans l'intervalle (ymin=0.87 , ymax=1.21) ; si y est en dehors de cet intervalle, on choisira u = 0.



La probabilité de détection correcte et la probabilité d'erreur (probabilité de fausse alarme) sont

$$p(d=1|u=1)= \int_{0.87}^{1.21} \frac{1}{\sigma_1\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\frac{-(y-m_1)^2}{2\cdot(\sigma_1)^2}\right] dy = 0.595$$

$$p(d=1|u=0)= \int_{0.87}^{1.21} \frac{1}{\sigma_0\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\frac{-(y-m_0)^2}{2\cdot(\sigma_0)^2}\right] dy = 0.079$$

Remarque : comme la variance liées à l'hypothèse u = 1 est plus petite que celle liée à la loi u = 0, il apparaît que le domaine des valeurs de y où on est amené à choisir l'hypothèse u = 0 est séparé en deux régions disjointes.

Règle de Bayes : la définition des probabilités conditionnelles est

$$p(u=0|y)=\frac{p(u=0 \text{ et } y)}{p(y)}.$$

On peut écrire de deux manières différentes

$$p(u=0 \text{ et } y)=p(u=0|y).p(y)=p(y|u=0).p(u=0),$$

si bien que p(u = 0| y) peut être écrit en fonction de p(y|u = 0)

$$p(u=0|y)=\frac{p(u=0 \text{ et } y)}{p(y)}=\frac{p(y|u=0)\,p(u=0)}{p(y)}.$$

la probabilité $p(y)$ s'écrit en fonction des probabilités conditionnelles

$$p(y)=p(y|u=0)\,p(u=0)+p(y|u=1)\,p(u=1),$$

et on en déduit

$$\boxed{p(u=0|y)=\frac{p(y|u=0)\,p(u=0)}{p(y|u=0)\,p(u=0)+p(y|u=1)\,p(u=1)}}.$$

si on connaît les probabilités *a priori* des causes et les probabilités conditionnelles de l'effet relativement aux cause, on peut en déduire les probabilités conditionnelles des causes relativement à l'effet.

Critère de Neyman Pearson

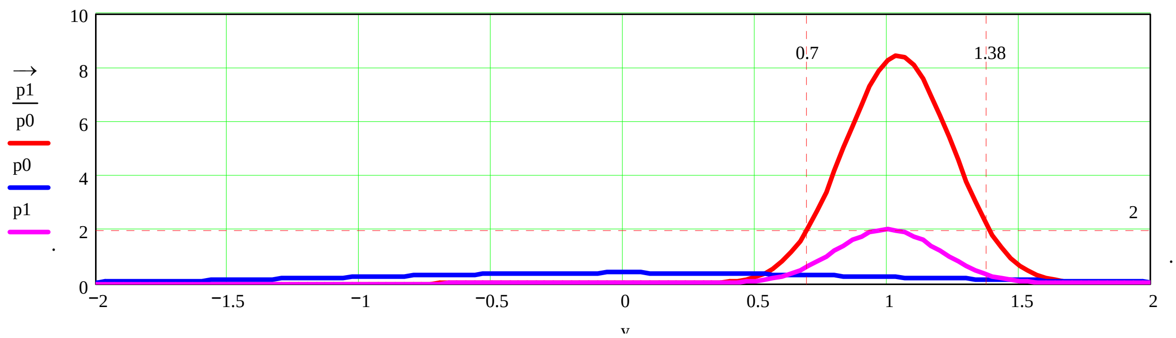
Dans cette approche, probabilités a priori des causes $q_0 = p(u{=}0)$ et $q_1 = p(u{=}1)$ ne sont pas connues.

On cherche à décider si une mesure x correspond à l'émission d'une donnée $u = 0$, et dans ce cas la densité de probabilité de x est $p_0(x)$; ou si elle correspond à l'émission $u = 1$, et dans ce cas la densité de probabilité de x est $p_1(x)$; on décidera que $d = 1$ si le rapport $r(x) = \dfrac{p_1(x)}{p_0(x)}$ dépasse un seuil s qui est donné de la manière suivante :

On considère la probabilité p_{fa} d'une **fausse alarme** ($d = 1$ alors que $u = 0$) pour chaque mesure x ; on cherche à maximiser la probabilité p_{dc} de **détection correcte** ($d = 1$ quand $u = 1$). la probabilité p_{em} d'un **événement manqué** ($d = 0$ alors que $u = 1$) vaut $1 - p_{dc}$;

La probabilité de fausse alarme est la probabilité que $u = 0$ alors que le rapport $r(x)$ dépasse le seuil s, c'est ainsi l'intégrale de la densité de probabilité $p_0(x)$ calculée pour l'ensemble des valeurs (domaine D) de x pour lequel ce seuil est dépassé

$$p_{fa} = \int_D p_0(x)dx$$



Dans cet exemple, où les probabilités a priori des causes ne sont pas prises en compte (ou si elles sont identiques), le seuil s étant égal à 2, on choisit $d = 1$ lorsque x est dans l'intervalle (0.7,1.4) et la probabilité de fausse alarme est

$$p_{fa} = \frac{1}{\sigma_0 . \sqrt{2\pi}} \int_{0.7}^{1.38} \exp - \frac{x^2}{2\sigma_0^2} dx = 0.158,$$

la probabilité de détection correcte est de

$$\int_{0.7}^{1.38} \frac{1}{\sigma_1 \cdot (2 \cdot \pi)^{0.5}} \cdot \exp\left[\frac{-(y - m_1)^2}{2 \cdot (\sigma_1)^2}\right] dy = 0.904 .$$

Dans l'approche de Neyman Pearson, **on se donne un seuil** α que cette probabilité de fausse alarme p_{fa} ne doit pas dépasser et on en déduit le seuil *s* utilisé dans la décision. On ne s'appuie pas sur la connaissance des probabilités *a priori* des causes q_0 et q_1.

Considérons le cas simple où $r(x)$ est une fonction croissante puis décroissante comme dans l'exemple précédent ; dans ce cas le domaine *D* se réduira à un segment (que nous supposerons fini) dont la borne inférieure sera notée *xmin* et la borne supérieure notée *xmax*.

$$p_{dc} = \int_{x\min}^{x\max} p_1(x)dx$$

xmin et *xmax* sont tels que dans l'intervalle [*xmin*, *xmax*] : $r(x) < s$.

Le problème se pose alors de la manière suivante : comment ajuster *s* et par conséquent les bornes *xmin* et *xmax* pour ***maximiser* la probabilité de décision correcte**, tout en assurant que la probabilité de fausse alarme $p_{fa} = \int_{x\min}^{x\max} p_0(x)dx$ ne dépasse pas le seuil α.

Nous illustrons cette approche en prenant l'exemple

$$m_0 := 0 \quad \sigma_0 := 1 \qquad m_1 := 1 \qquad \sigma_1 := 0.2$$

Quand *s* est fixé, on peut sur cet exemple calculer les valeurs *xmin* et *xmax* entre lesquelles on décidera d=1. Le dépassement du seuil est donné par l'inégalité

$$\frac{\sigma_0}{\sigma_1}\exp\left(\frac{x^2}{2\sigma_0^2} - \frac{(x-m_1)^2}{2\sigma_1^2}\right) > s\,,$$

soit, en termes de logarithmes :

$$\log(\frac{\sigma_0}{\sigma_1.s}) + \left(\frac{x^2}{2\sigma_0^2} - \frac{(x-m_1)^2}{2\sigma_1^2}\right) > 0$$

$$\frac{\sigma_0^2-\sigma_1^2}{\sigma_0^2\sigma_1^2}x^2 - 2\frac{m_1}{\sigma_1^2}x + \frac{m_1^2}{\sigma_1^2} + 2\log(\frac{\sigma_1.s}{\sigma_0}) > 0\,.$$

Les deux valeurs du dépassement du seuil sont ainsi les racines d'une équation du deuxième degré

$$\Delta = \frac{m_1^2}{\sigma_1^4} - \frac{\sigma_0^2-\sigma_1^2}{\sigma_0^2\sigma_1^2}.\left(\frac{m_1^2}{\sigma_1^2} + 2\log(\frac{\sigma_1.s}{\sigma_0})\right),$$

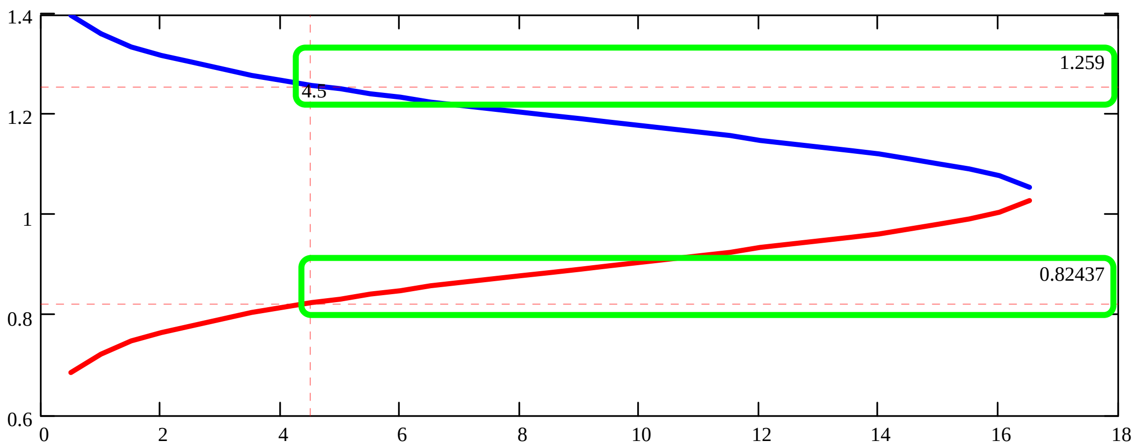
$$x\min = \frac{\sigma_0^2\sigma_1^2}{\sigma_0^2-\sigma_1^2}\left(\frac{m_1}{\sigma_1^2} - \Delta^{0.5}\right),$$

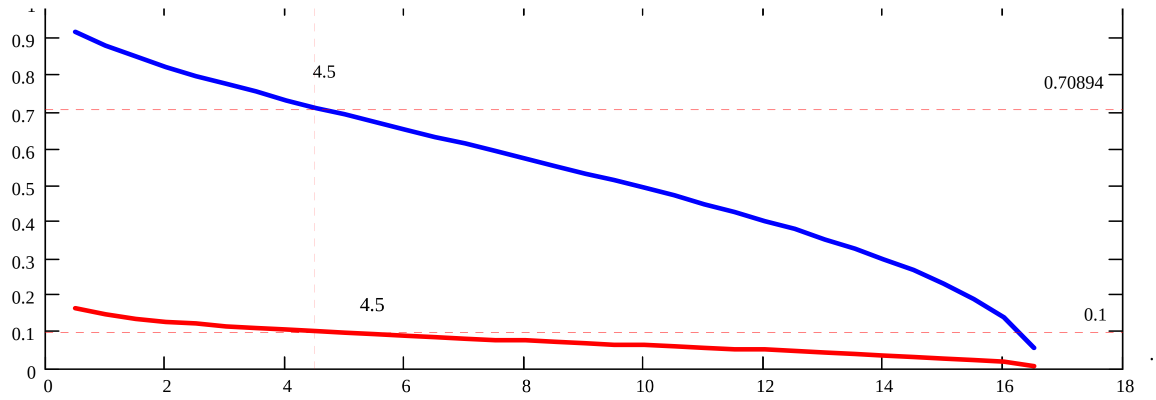
$$x\max = \frac{\sigma_0^2\sigma_1^2}{\sigma_0^2-\sigma_1^2}\left(\frac{m_1}{\sigma_1^2} + \Delta^{0.5}\right).$$

Une fois ces valeurs trouvées, on calcule les bornes *xmin* et *xmax* et les pfa et pdc associées qui sont ainsi des fonctions de *s :*

$$
\text{probas}(s) := \left|
\begin{array}{l}
\Delta \leftarrow \dfrac{(m_1)^4}{(\sigma_1)^4} - \left[\dfrac{(m_1)^2}{(\sigma_1)^2} + \left(2\cdot\log\left(s\cdot\dfrac{\sigma_1}{\sigma_0}\right)\right)\right]\cdot\left[\dfrac{(\sigma_0)^2 - (\sigma_1)^2}{(\sigma_0)^2\cdot(\sigma_1)^2}\right] \\
\text{xmin} \leftarrow \left[\dfrac{m_1}{(\sigma_1)^2} - \Delta^{0.5}\right]\cdot\dfrac{(\sigma_0)^2\cdot(\sigma_1)^2}{(\sigma_0)^2 - (\sigma_1)^2} \\
\text{xmax} \leftarrow \left[\dfrac{m_1}{(\sigma_1)^2} + \Delta^{0.5}\right]\cdot\dfrac{(\sigma_0)^2\cdot(\sigma_1)^2}{(\sigma_0)^2 - (\sigma_1)^2} \\
\text{pfa} \leftarrow \displaystyle\int_{\text{xmin}}^{\text{xmax}} \dfrac{1}{\sigma_0\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\dfrac{-(x)^2}{2\cdot(\sigma_0)^2}\right] dx \\
\text{pdc} \leftarrow \displaystyle\int_{\text{xmin}}^{\text{xmax}} \dfrac{1}{\sigma_1\cdot(2\cdot\pi)^{0.5}}\cdot\exp\left[\dfrac{-(x - m_1)^2}{2\cdot(\sigma_1)^2}\right] dx \\
\text{return } \ \text{pile}(\text{xmin}, \text{xmax}, \text{pfa}, \text{pdc})
\end{array}
\right.
$$

Voici les valeur des limites xmin et xmax du domaine de décision *d* =1, en fonction du seuil





valeur de la probabilité de fausse alarme en rouge et le la probabilité de détection correcte en bleu en fonction du seuil qu'on fait varier de 0 à 17 . Si la probabilité de fausse alarme est de 0.1, on choisira un seuil de décision à 4.5, ce qui correspondra aux bornes xmin = 0.824 et xmax = 1.259 et une probabilité de décision correcte de 0.709

dans une approche élémentaire, on peut effectuer les opérations suivantes coûteuse en calcul: si on envisage une pfa de 0.1, on cherchera le domaine d'intégration (xmin – xmax) en fonction de s, ce qui permettra d'en déduire les valeurs du pfa toujours en fonction de s ; on trouve ainsi la valeur de s qui donne ce pfa et donc les bornes du domaine correspondantes.

La suite de cette approche (plus difficile à exposer, toute indication est bienvenue ... leroux@polytech.unice.fr) est de proposer une technique (notion de statistique suffisante) afin de réduire sous certaines hypothèses, la quantité de calcul à effectuer.

un certain nombre de résultats sont présentés sur le site

http://www.aiaccess.net/French