Rapport SY19 TP03 - Regle de Bayes et notion de rejet

Thomas ABASSI & Joan MOESCH

November 27, 2011

Le but de ce TP est d'appréhender la règle de bayes dans un probleme de detection (2 classes) dans le cas ou 3 choix sont possibles : choisir la classe 1, la classe 2 ou ne prendre aucune des 2. Cette decision s'effectue grace a l'extension de la regle de Bayes (rapport des probabilites a posteriori) avec la notion de rejet.

Nous verrons comment s'expriment ces fonctions de decisons dans le cas general pour l'appliquer a notre probleme de detection.

6. Regle de Bayes avec rejet

a) Expression de la fonction de decision

Dans le cas ou $c_{11} = c_{22} = 0$, $c_{01} = c_{02} = c_0$ et $c_{12} = c_{21} = 1$, la regle de bayes avec rejet s'ecrit :

$$\delta^*(x) = \begin{cases} a_0 & \text{si } \mathrm{Min}(P[Y=1|x], P[Y=2|x]) > c_0 \\ a_1 & \text{si } P[Y=1|x] > p[Y=2|x] \text{ et } P[Y=1|x] > 1-c_0 \\ a_2 & \text{si } P[Y=2|x] > p[Y=1|x] \text{ et } P[Y=2|x] > 1-c_0 \end{cases} \tag{1}$$

Ainsi, l'action de prendre la decision a_1 s'ecrit :

$$\delta^*(x) = a_1 \Leftrightarrow \begin{cases} \frac{f_1(x)}{f_2(x)} > \frac{\pi_2}{\pi_1} \\ P[Y=1|x] > 1-c_0 \end{cases}$$

De plus, nous avons le rapport de densite suivant :

$$\begin{aligned} \frac{f_1(x)}{f_2(x)} &= \frac{\frac{1}{2\pi} e^{-\frac{1}{2}(x_1+1)^2 - \frac{1}{2}(x_2)^2}}{\frac{1}{2\pi} e^{-\frac{1}{2}(x_1-1)^2 - \frac{1}{2}(x_2)^2}} \\ &= \frac{e^{-\frac{1}{2}(x_1+1)^2}}{e^{-\frac{1}{2}(x_1-1)^2}} \\ &= \frac{e^{-\frac{1}{2}(x_1^2+2x_1+1)}}{e^{-\frac{1}{2}(x_1^2-2x_1+1)}} \\ &= e^{-2x_1} \end{aligned}$$

En rappelant que la probabilite a posteriori est donnee par :

$$P[Y=1|x] = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)} \tag{2}$$

Nous obtenons donc :

$$\delta^*(x) = a_1 \Leftrightarrow \begin{cases} e^{-2x_1} > \frac{\pi_2}{\pi_1} \\ \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)} > 1-c_0 \end{cases}$$

Ce qui nous donne finalement l'expression de la regle en fonction de x_1 :

$$\delta^*(x) = a_1 \Leftrightarrow \begin{cases} x_1 < -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}) \\ x_1 < -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}\frac{1-c_0}{c_0}) \end{cases} \tag{3}$$

De la meme maniere, nous trouvons les equations de frontiere pour la classe ω_2 :

$$\delta^*(x) = a_2 \Leftrightarrow \begin{cases} x_1 > -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}) \\ x_1 > -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}\frac{c_0}{1-c_0}) \end{cases} \tag{4}$$

La droite d'equation $x = -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1})$ correspond au centre de la zone de rejet, commune aux deux equations de frontieres pour a_1 et a_2.
D'autre part, les droites d'equations $x_1 = -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}\frac{1-c_0}{c_0})$ et $x_2 = -\frac{1}{2}Ln(\frac{\pi_2}{\pi_1}\frac{c_0}{1-c_0})$ correspondent respectivement aux frontieres extremes de la zone de rejet, dans le cas ou l'on choisit la classe ω_1 ou ω_2.

b) Traces des frontieres

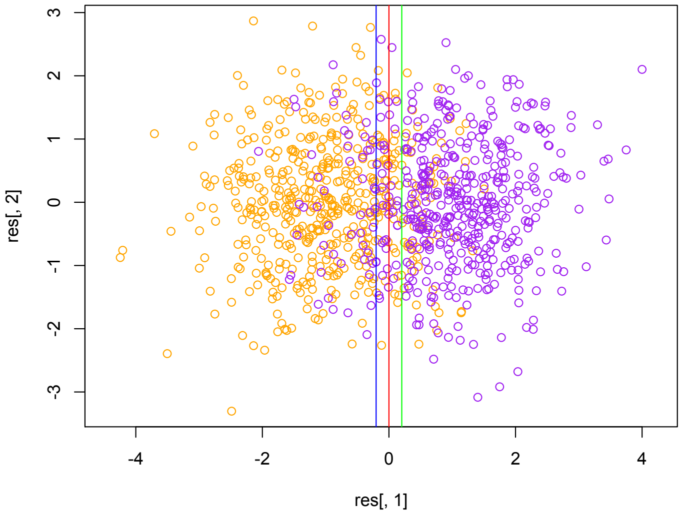
Nous ecrivons la fonction suivante :

$bayesRejet <- function(n, pi1, pi2, L0)$
$\quad x0 = -log(pi2/pi1)/2$
$\quad x1 = -1/2 * log(pi2 * (1 - L0)/(pi1 * L0))$
$\quad x2 = -1/2 * log(pi2 * L0/(pi2 * (1 - L0)))$
$\quad echantillon(n)$
$\quad abline(v = x0, col = "red")$
$\quad abline(v = x1, col = "blue")$
$\quad abline(v = x2, col = "green")$

Ou la fonction *echantillon* trace le nuage de points en coloriant chaque individus selon sa classe.
Pour chaque cas, voici les resultats obtenus :

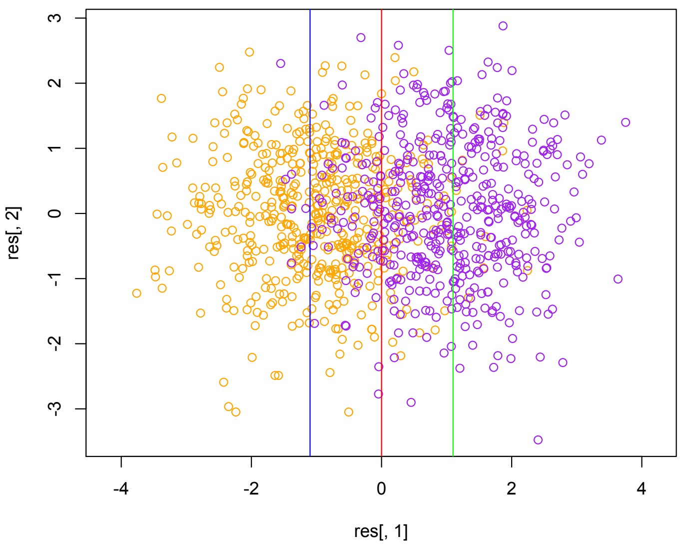
Cas 1

Nous prenons les valeurs $\pi_1 = \pi_2$, $c_0 = 0.4$



Cas 2

Nous prenons les valeurs $\pi_1 = \pi_2$, $c_0 = 0.1$



Dans chaque cas, les donnnes à gauche de la droite bleu sont classees dans w_1, ceux à droite de la verte dans w_2 et aucune classe n'est attribuee dans le cas des donnees entre la bleu et la rouge. La droite rouge represente l'axe central de la zone de rejet.

On remarque que plus le cout de rejet c_0 est faible, plus la zone de rejet a tendance à s'agrandir. Ceci entrainera evidemment une baisse de l'erreur de decision mais provoquera egalement une baisse du nombre de decisions (plus de rejet).

c) Estimation des erreurs de decision

Par definition, les probabilites de mauvaises decisions α et β sont donnes par :
$\alpha = P(\delta^*(X) = a_2|\omega_1)$
$\beta = P(\delta^*(X) = a_1|\omega_1)$

On peut donc estimer ces deux types d'erreurs en comptant le nombre d'individus de la classe ω_1 qui ne sont a droite de "l'axe vert" et le nombre d'individus de la classe ω_2 qui ne sont à gauche de "l'axe bleu".

En posant x_1 et x_2 les abcisses des droites bleue et verte (extremites de la zone de rejet), les estimateurs s'ecrivent:
$\hat{\alpha} = \frac{1}{n_1} \sum_{\omega_1} \mathbb{1}_{x_i > x_2}(i)$
$\hat{\beta} = \frac{1}{n_2} \sum_{\omega_2} \mathbb{1}_{x_i < x_1}(i)$

Dans le cas $c_0 = 0.4$, nous obtenons les valeurs d'erreurs de classification suivantes :
$\hat{\alpha} = 0.126$
$\hat{\beta} = 0.106$

Dans le cas $c_0 = 0.1$, nous obtenons les valeurs d'erreurs de classification suivantes :
$\hat{\alpha} = 0.02$
$\hat{\beta} = 0.014$

La valeur du seuil de rejet permet donc de diminuer les erreurs de classification.

d) Probabilites de rejet et d'erreur

La probabilite de rejet de la decision correspond a la quantite des individus qui se trouvent dans la region de rejet, c'est a dire entre x_1 et x_2, un bon estimateur de cette probabilite serait donc :

$\hat{r} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_1 < x_i < x_2}(i)$

Ceci nous donne, dans le cas $c_0 = 0.4$, une estimation :
$\hat{r} = 0.098$

Dans le cas $c_0 = 0.1$, nous avons :
$\hat{r} = 0.538$

Ceci confirme bien les theories avancees precedemment : plus le cout c_0 est faible, plus les erreurs α et β diminuent mais plus la probabilite de rejet augmente. Ceci est facilement comprehensible dans le sens ou la region de rejet augmente lorsque la valeur de c_0 diminue.

D'autre part, l'estimateur de la probabilite d'erreur est donne par : $\epsilon(\delta^*) = \frac{\hat{\alpha}+\hat{\beta}}{n_1+n_2}$

Nous obtenons donc l'estimation de la probabilite d'erreur :
Dans le cas $c_0 = 0.4$, $\hat{\epsilon}(\delta^*) = 2.32.10^{-4}$
Dans le cas $c_0 = 0.1$, $\hat{\epsilon}(\delta*) = 0.34.10^{-4}$