

**Ex. 1** On note, pour tout  $x \in \mathbb{R}$ ,  $\sigma(x) = \frac{1}{1+e^{-x}}$ .

Vérifier que :

$$\sigma' = (1 - \sigma)\sigma,$$

$$(\ln(\sigma(u)))' = (1 - \sigma)u',$$

$$(\ln(1 - \sigma(u)))' = -\sigma(u)u'.$$

*Rappel :*

$$(u(v))' = u'(v) \times v' \text{ avec comme cas particuliers : } (\ln(u))' = \frac{u'}{u}, (e^u)' = u'e^u.$$

### Maximum de vraisemblance.

Étant données des observations  $X_j \mapsto Y_j$ ,  $j \in \{1, \dots, J\}$ , avec  $X_j \in \mathbb{R}^n$  et  $Y_j \in \{0, 1\}$ , on cherche une « bonne explication » des valeurs de  $Y$  à partir de celle de  $X$ .

À cette fin, on considère que les valeurs  $Y_j$  sont en partie aléatoires, et issues des valeurs des  $X$  à travers une loi de probabilité  $P : X \mapsto P(Y = 1|X)$ , exprimant la probabilité que  $Y = 1$  étant donné des valeurs  $X$ . Le problème est alors de trouver une « bonne » loi de probabilité  $P$ .

Si les valeurs de  $Y$  sont effectivement issues de  $P$ , alors étant donné  $X_j$ , la probabilité d'observer  $Y = 1$  est  $P(Y = 1|X_j)$ , et celle d'observer  $Y = 0$  est  $1 - P(Y = 1|X_j)$ . Par conséquent

$$P(Y = 1|X_j)^{Y_j}(1 - P(Y = 1|X_j))^{1-Y_j}$$

est la probabilité d'observer  $Y = Y_j$  étant donné  $X_j$ .

Plus généralement, si chaque observation est indépendante des autres, alors la *vraisemblance* de la loi  $P$ , donnée par

$$L(P) = \prod_j P(Y = 1|X_j)^{Y_j}(1 - P(Y = 1|X_j))^{1-Y_j}$$

est la probabilité d'observer  $Y = Y_1$  (étant donné  $X_1$ ) et  $Y = Y_2$  (étant donné  $X_2$ ) et ... et  $Y = Y_J$  (étant donné  $X_J$ ).

En général, on paramétrise  $P$ , et le problème est alors de trouver la valeur des paramètres maximisant la vraisemblance de  $P$ .

**Ex. 2** On considère les observations :

$$X_1 = \begin{pmatrix} 1 & 1 \end{pmatrix} \mapsto Y_1 = 1, X_2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \mapsto Y_2 = 0, X_3 = \begin{pmatrix} 0 & 1 \end{pmatrix} \mapsto Y_3 = 0 \text{ et } X_4 = \begin{pmatrix} 0 & 0 \end{pmatrix} \mapsto Y_4 = 0.$$

On considère les lois définies, pour tout  $X = \begin{pmatrix} x_1 & x_2 \end{pmatrix}$ , par

$$P(Y = 1|X) = \sigma(b + w_1x_1 + w_2x_2)$$

avec  $b, w_1, w_2 \in \mathbb{R}$ .

1. Expliciter la log-vraisemblance, donnée par

$$\begin{aligned} \ln(L(b, w_1, w_2)) \\ = \sum_j (Y_j \ln(P(Y = 1|X_j)) + (1 - Y_j) \ln(1 - P(Y = 1|X_j))). \end{aligned}$$

pour la loi  $P : X \mapsto \sigma(b + w_1x_1 + w_2x_2)$  et pour les données  $(Y_j, X_j)$  ci-dessus.

2. Vérifier que  $(b, w_1, w_2) \mapsto \ln(L(b, w_1, w_2))$  est une fonction partout concave.  
*Indication* : La somme de fonctions concaves est concave : il suffit donc de montrer que chaque terme de la vraisemblance est une fonction concave. Par ailleurs, il est possible d'utiliser le critère de Sylvester pour montrer qu'une fonction est concave : une fonction est concave si la fonction opposée est convexe.
3. Expliciter le système d'équations à résoudre afin de déterminer l'éventuel maximum de  $\ln(L(b, w_1, w_2))$ . Est-il possible de résoudre ce système ?