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

Rappel de cours : **Maximum de vraisemblance.**

Étant données des observations $X_j \mapsto Y_j$, $j \in \{1,\ldots,J\}$, avec $X_j \in \mathbb{R}^n$ et $Y_j \in \{0{,}1\}$, on cherche à expliquer les valeurs de Y à partir de celle de X.

À cette fin, on suppose que les valeurs Y_j sont issues d'une loi de probabilité conditionnelle $P : X \mapsto \Pr(Y=1|X)$, associant à chaque X la probabilité que $Y=1$.

L'analyse de données vise à déterminer 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 $1-P(Y=1|X_j)$. Plus généralement, si chaque observation est indépendante des autres, alors la *vraisemblance* de la loi P, donnée par

$$L(P) = \prod_{j:Y_j=1} P(Y=1|X_j)^{Y_j} \prod_{j:Y_j=0} (1-P(Y=1|X_j))$$

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. C'est la probabilité d'observer les données effectivement observées.

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 :
$X_1 = \begin{pmatrix} 1 & 1 \end{pmatrix} \mapsto Y_1 = 1$, $X_2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \mapsto Y_2 = 0$, $X_3 = \begin{pmatrix} 0 & 1 \end{pmatrix} \mapsto Y_3 = 0$ et $X_4 = \begin{pmatrix} 0 & 0 \end{pmatrix} \mapsto Y_4 = 0$.
On considère la famille de lois de probabilité paramétrée par b, w_1, w_2, 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

$$\begin{aligned}&\ln(L(b,w_1,w_2))\\ &= \sum_{j:Y_j=1} \big(\ln(P(Y = 1|X_j)\big) + \sum_{j:Y_j=0} \big(\ln(1 - P(Y = 1|X_j)\big).\end{aligned}$$

   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. Puisque $\ln(L(b,w_1,w_2))$ est partout concave, si elle atteint un maximum alors en ce maximum, $[\mathrm{grad} \ln(L)](b,w_1,w_2) = 0$.
   Expliciter le système d'équations $[\mathrm{grad} \ln(L)](b,w_1,w_2) = 0$. Ce sytème possède-t-il une solution ?