Exercice 1 (Modèle de régression logistique) On considère le problème de statistique suivant appelé problème de régression logistique. Dans le cas du problème de régression classique, on a des observations $Y_i = X_i^t\beta + \epsilon_i$ pour $i = 1, \ldots, n$ où on suppose que les ϵ_i sont i.i.d. et suivent une loi $\mathcal{N}(0, \sigma^2)$. Il arrive souvent que les observations Y_i ne soient pas de ce type et que pourtant on veuille expliquer les Y_i par les variables X_i. Un cas très fréquent est les cas où les observations sont 0 ou 1. Dans ce cas, on suppose que les Y_i suivent une loi de Bernoulli $\mathcal{B}(1, p_i)$ où chaque p_i est une fonction de X_i. Le modèle logistique est spécifié par la relation particulière

$$p_i = \frac{\exp(X_i^t\beta)}{1 + \exp(X_i^t\beta)}. \tag{1}$$

Il s'agit alors de trouver le vecteur de regression β.

a. Ecrire la vraisemblance pour ce modèle puis son logarithme.
b. Donner le gradient de la log-vraisemblance puis sa matrice Hessienne.
c. La log-vraisemblance est-elle une fonction convexe ?
d. Ecrire la condition d'optimalité pour l'estimateur $\hat{\beta}$ au sens du maximum de vraisemblance.
e. Peut-on trouver $\hat{\beta}$ à la main ?

Exercice 2 (Canal d'information et systèmes de communications). On considère un message $X(t)$, $t = 1, 2, \ldots$, le temps étant compté en fraction de seconde, les valeurs de $X(t)$ appartenant à l'ensemble $\{1, \ldots, n\}$ et, ce message traversant ce qu'on appelle un canal, c'est-à-dire par exemple étant transformé par codage, puis traversant l'air, une fibre optique, etc ..., on observe à la réception le signal $Y(t)$, $t = 1, 2, \ldots$ prenant ses valeurs dans $\{1, \ldots, m\}$.

La relation entre l'entrée et la sortie n'est pas certaine mais on suppose connues les probabilités

$$P(Y(t) = j \mid X(t) = i), \quad i = 1, \ldots, n, \; j = 1, \ldots, m. \tag{2}$$

On a déjà un peu entendu parler de Claude Shannon en théorie de Fourier. Ce grand chercheur est aussi très connu pour un fameux théorème sur ce type de modèle dont il est l'inventeur ou presque car on ne sait jamais trop avec le foisonnement des idées à une époque donnée.

On note $\mathcal{P}$ l'ensemble des vecteurs à composantes positives ou nulles et dont la somme des composantes est égale à 1. On note π la loi de $X(t)$, c'est à dire le vecteur dont les composantes sont $\pi_i = P(X(t) = i)$, $i = 1, \ldots, n$. On appelle alors information mutuelle entre $X(t)$ et $Y(t)$ la quantité

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} \pi_i P_{ij} \log_2\left(\frac{P_{ij}}{\sum_{k=1}^{n} \pi_k P_{kj}}\right) \tag{3}$$

La capacité du canal de transmission est alors définie comme étant la quantité :

$$C = \max_{\pi\in\mathcal{P}} I(X;Y). \tag{4}$$

Montrer que la capacité C est la valeur optimale d'un problème d'optimisation convexe.