Q2 : a. Une distribution de vraisemblance correspond à la probabilité $P(\vec{x} \mid t)$ avec $\vec{x}$ une donnée et t la cible.
On pourrait la calculer à l'aide de la formule de Bayes :

$$P(\vec{x} \mid t) = \frac{P(t \mid \vec{x}) \times p(\vec{x})}{P(t)}$$

$$= \frac{P(t, \vec{x})}{p(t)}$$

b. La distribution a priori correspond à $P(t)$.

On a : $P(t=i) = \frac{\text{nombre de données ayant pour cible } i}{\text{nombre de données}}$

c. Il est raisonnable d'émettre l'hypothèse que P(consommation d'essence | voiture de sport) est gaussienne. En effet, les voitures de sport sont une sous catégorie très spécifique de voitures.
On peut supposer, puisque toutes ces voitures sont en compétition sur un marché restreint, qu'elles vont avoir des caractéristiques très similaires, notamment en terme de consommation d'essence.

Salomé Gomez (19 156 930)
Julien Bracher (19 156 832)
Arnold Godart (19 156 869)

Démonstration 1 :

$$E_D(W) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln(y_{nk})$$

et $y_k(o) = \dfrac{\exp(o_k)}{\sum_j \exp(o_j)}$

Donc $\nabla_{o_j} E_D(W) = -\sum_{n=1}^{N} \nabla_{o_j} \cdot \left(\sum_{\substack{k=1 \\ k \neq j}}^{K} t_{kn} \ln(y_{nk}) + t_{jn} \ln(y_{nj}) \right)$ (1)

Or $\dfrac{\partial y_k}{\partial o_j} = \dfrac{\partial}{\partial o_j} \cdot \dfrac{\exp(o_k)}{\sum_c \exp(o_c)}$ (2)

avec $\dfrac{\partial \exp(o_k)}{\partial o_j} = \begin{cases} \exp(o_j) & \text{si } k = j \\ 0 & \text{sinon} \end{cases}$

donc $\dfrac{\partial \exp(o_k)}{\partial o_j} = I_{kj} \cdot \exp(o_k)$

Ainsi (2) $\Leftrightarrow \dfrac{\partial y_k}{\partial o_j} = \dfrac{I_{kj} \cdot \exp(o_k) \cdot \sum_c \exp(o_c) - \exp(o_k) \cdot \exp(o_j)}{\left(\sum_c \exp(o_c)\right)^2}$

$$= I_{kj}\, y_k - y_k \cdot y_j$$

$$= y_k \left(I_{kj} - y_j \right)$$

$$= \begin{cases} -y_k \cdot y_j & \text{si } k \neq j \\ y_k(1 - y_j) & \text{si } k = j \end{cases}$$

Salomé Gomez (19 156 930)
Julien Brodier (19 156 832)
Arnould Godart (19 156 869)

Donc 1 $\Leftrightarrow$

$$\nabla_j E_j(W) = -\sum_{n=1}^{N} \left(\sum_{\substack{k=1 \\ k \neq j}}^{K} t_{kn} \cdot \frac{-y_{kn} y_{jn}}{y_{kn}} + t_{jn} \cdot \frac{y_{jn}(1-y_{jn})}{y_{jn}} \right)$$

$$= -\sum_{n=1}^{N} \left(\sum_{\substack{k=1 \\ k \neq j}}^{K} t_{kn}(-y_{jn}) + t_{jn} - t_{jn} y_{jn} \right)$$

$$= -\sum_{n=1}^{N} \left(-y_{jn} \sum_{k=1}^{K} t_{kn} + t_{jn} \right)$$

$$= \sum_{n=1}^{N} y_{jn} - t_{jn}$$

Salomé Gomez (19 156 930)
Julien Brochier (19 156 832)
Arnaud Godart (19 156 869)

Démo 3

* Formules de la position, de la vitesse et de l'accélération

$$a = \frac{v_1 - v_0}{t_1 - t_0} \Leftrightarrow v_1 = a(t_1 - t_0) + v_0$$

Pour $a(t_1 - t_0) \to 0$ $a(t_1 - t_0)$ devient ∇a
De la même façon $v(t_1 - t_0)$ devient ∇v

$$p_1 = v(t_1 - t_0) + p_0$$

* Formules du momentum :

$$v_{t+1} = \mu v_t - \nabla E_{\vec{x}_n}(w_t)$$

$$w_{t+1} = w_t + \eta v_{t+1}$$

Le vecteur **w** peut être vue comme une position sur le graphique de la descente de gradient.

À chaque itération, nous calculons son accélération ∇e, sa vitesse v_t et nous mettons à jour sa position.

Ici, le gradient est négatif car nous descendons la courbe alors qu'il est positif dans le cas de l'accélération d'un objet qui chute. Cela explique pourquoi nous soustrayons le gradient (l'accélération) au lieu de l'ajouter.

ρ peut symboliser la friction parce qu'il freine tout déplacement (dans notre cas toute valeur non nulle de v).