

**Question 1 :** Les images sont en noir et blanc, de dimension  $28 \times 28$ . On peut donc choisir  $\mathcal{X} = [0; 1]^{28 \times 28}$ . De plus, le problème impose  $\mathcal{Y} = \{-1, 1\}$ .  $\mathcal{X}$  (resp.  $\mathcal{Y}$ ) est inclus dans un espace de dimension 784 (resp. 1).

**Question 2 :**

(a) - Le risque empirique associé à la 0-1 loss est :

$$\hat{R}_n(f) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{f(X_i) \neq Y_i}$$

Le risque réel associé à la 0-1 loss est :

$$R(f) = \int_{[0;1]^{28 \times 28} \times \{-1,1\}} \mathbb{1}_{f(x) \neq y} d\rho(x, y)$$

où  $\rho$  est la distribution de probabilité théorique reliant les images aux lettres qu'elles représentent.

Dans ce cas, la minimisation du risque empirique est compliquée car la fonction mise en jeu n'est pas convexe, ni même régulière. Les techniques usuelles d'optimisation ne s'appliquent donc pas.

(b) - Utiliser un second jeu de données afin de tester les performances permet d'éviter l'overfitting. En effet, les données sur lesquelles on s'entraîne sont bruitées, donc à partir d'un certain seuil de précision, on ne s'approche plus de la distribution réelle mais plutôt de la distribution d'entraînement bruitée.

(c) - Le problème d'optimisation associé à la linear least square regression consiste à calculer l'estimateur suivant :

$$\hat{\theta}_n := \arg \min_{\theta \in \mathbb{R}^{28 \times 28}} \frac{1}{n} \sum_{i=1}^n (\theta^\top X_i - Y_i)^2$$

De même, le problème d'optimisation associé à la linear logistic regression consiste à calculer l'estimateur suivant :

$$\hat{\beta}_n := \arg \min_{\beta \in \mathbb{R}^{28 \times 28}} \frac{1}{n} \sum_{i=1}^n \log(1 + e^{-\beta^\top X_i Y_i})$$