1 Question 1

Nous avons L classifieurs $(c_l)_{l=1}^L$ indépendants qui ont tous la même proba de bien prédire $p = 0.7$. Ainsi les variables $Z_l = \mathbb{1}_{\{c_l \text{ ne se trompe pas}\}}$ suivent des lois de Bernoulli indépendantes de paramètre p. La proba de bien prédire est plutôt faible. Cependant, en combinant tous les classifieurs, on peut obtenir un nouveau classifieur C_* qui a une proba (bien) plus élevée de bien prédire. Nous construisons ce nouveau classifieur de la manière suivante: $C_* = 1$ si la majorité des $(c_l)_{l=1}^L$ renvoie 1 et $C_* = -1$ si la majorité des $(c_l)_{l=1}^L$ renvoie -1. Quelle est la probabilité que C_* ne se trompe pas? Cette probabilité est en fait la probabilité que la majorité des $(c_l)_{l=1}^L$ ne se trompent pas. Soit $S = \sum_{l=1}^L Z_l$ et supposons pour simplifier que L est impair (pour éviter d'avoir à traiter les cas où il y a autant de -1 que de 1 prédits). Nous avons au final

$$\begin{aligned}
\mathbb{P}(C_* \text{ ne se trompe pas}) &= \mathbb{P}(S \geq L/2) \\
&= \sum_{l=L/2+1}^{L} \mathbb{P}(S = l) \\
&= \sum_{l=L/2+1}^{L} \binom{L}{l} p^l (1-p)^{L-l}
\end{aligned}$$

Quand $p > 0.5$, cette proba se rapproche très vite de 1 lorsque L augmente et donc C_* est bien meilleur que chacun des classifieurs $(c_l)_{l=1}^L$!

Idée à retenir: combiner des classifieurs simples (*i.e* avec une proba de ne pas se tromper pas très élevée) peut permettre de construire un classifieur bien meilleur in fine. Ceci est une illustration théorique simplifiée de ce qui se produit avec les méthodes d'ensemble learning (bagging, forêts aléatoires, boosting...).

2 Question 11

Supposons la question 10 admise. Nous avons donc que $\mathbb{E}[\exp(-Yf(X))]$ est minimisée en $f^*_{\exp} : x \mapsto \frac{1}{2}\ln\left(\frac{\eta(x)}{1-\eta(x)}\right)$. Nous remarquons que $f^*_{\exp}(x)$ est positive si et seulement si $\eta(x) > 1/2$. Ainsi nous avons bien que le classifieur de Bayes associé à $f^*_{\exp}(x)$, qui vaut $\operatorname{sign}\left(f^*_{\exp}(x)\right)$, est identique à celui associé à la perte $\mathbb{1}_{\{-yf(x)\geq 0\}}$. Ainsi, en ce qui concerne la qualité de la prédiction, nous ne perdons rien à remplacer la perte indicatrice par la perte exponentielle. Nous sommes même gagnants d'un point de vue de pratique car la perte exponentielle est plus régulière (possibilité de faire une descente de gradient).

3 Question 10

Soit $\mathcal{F}$ l'ensemble de toutes les fonctions mesurables. Nous souhaitons résoudre le problème suivant

$$\inf_{f\in\mathcal{F}} \mathbb{E}[\exp(-Yf(X))].$$

Nous avons (par la loi des espérances itérées)

$$\inf_{f\in\mathcal{F}} \mathbb{E}[\exp(-Yf(X))] = \inf_{f\in\mathcal{F}} \mathbb{E}\left[\mathbb{E}[\exp(-Yf(X)) \mid X]\right].$$

Nous allons dans un premier temps résoudre pour tout $x \in \mathcal{X}$

$$\inf_{f \in \mathcal{F}} \mathbb{E}[\exp(-Yf(x)) \mid X = x],$$

ce qui suffira à résoudre le problème initial.

Comme Y prend uniquement les valeurs -1 et 1, nous pouvons écrire pour tout $f \in \mathcal{F}$

$$\begin{aligned}\mathbb{E}[\exp(-Yf(x)) \mid X = x] &= \exp(-1 \times f(x))\underbrace{\mathbb{P}(Y = 1 \mid X = x)}_{=:\eta(x)} + \exp(1 \times f(x))\underbrace{\mathbb{P}(Y = -1 \mid X = x)}_{=:1-\eta(x)} \\ &= \exp(-f(x))\eta(x) + \exp(f(x))(1 - \eta(x)).\end{aligned}$$

Astuce de champion: comme $f(x) \in \mathbb{R}$ pour tout $f \in \mathcal{F}$ et tout $x \in \mathcal{X}$, nous pouvons écrire

$$\begin{aligned}\inf_{z \in \mathbb{R}} \exp(-z)\eta(x) + \exp(z)(1 - \eta(x)) &\leq \inf_{f \in \mathcal{F}} \{\exp(-f(x))\eta(x) + \exp(f(x))(1 - \eta(x))\} \\ &= \inf_{f \in \mathcal{F}} \mathbb{E}[\exp(-Yf(x)) \mid X = x].\end{aligned}$$

Les fonctions $z \mapsto \exp(-z)$ et $z \mapsto \exp(z)$ sont convexes et continûment différentiables. De plus, pour tout $x \in \mathcal{X}$, $\eta(x) \in [0, 1]$. Ainsi pour tout $x \in \mathcal{X}$, la fonction

$$g : z \mapsto \exp(-z)\eta(x) + \exp(z)(1 - \eta(x))$$

est convexe et continûment différentiable. Dans notre cadre, s'il existe z_* tel que $\partial_z g(z_*) = 0$, alors z_* est l'unique minimiseur de g. Nous trouvons

$$z_* = z_*(x) = \frac{1}{2} \ln\left(\frac{\eta(x)}{1 - \eta(x)}\right).$$

La fonction $x \mapsto z_*(x)$ fait partie de l'ensemble $\mathcal{F}$ et donc cette fonction est solution de

$$\inf_{f \in \mathcal{F}} \mathbb{E}[\exp(-Yf(x)) \mid X = x].$$

A fortiori, elle est également solution du problème initial, à savoir $\inf_{f \in \mathcal{F}} \mathbb{E}[\exp(-Yf(X))]$.