

TRAVAUX DIRIGÉS N° 2 : Concentration, théorie de VC

Stephan CLÉMENÇON <stephan.clemencon@telecom-paris.fr>

Ekhine IRUROZKI <irurozki@telecom-paris.fr>

EXERCICE 1. On se place dans le cadre de la classification binaire : soient un descripteur aléatoire X à valeurs dans un espace mesurable $\mathcal{X} \subset \mathbb{R}^d$ ($d \in \mathbb{N}^*$) et un label aléatoire Y valant -1 ou 1 . On considère une classe finie \mathcal{G} de classifieurs $\mathcal{X} \rightarrow \{-1, 1\}$ telle que les deux labels sont parfaitement séparables par un élément de \mathcal{G} , i.e. $\min_{g \in \mathcal{G}} L(g) = 0$ pour le risque $L : g \in \mathcal{G} \mapsto \mathbb{P}(g(X) \neq Y) \in [0, 1]$.

Soit $n \in \mathbb{N}^*$. On suppose que l'on dispose d'un échantillon i.i.d. $\{(X_i, Y_i)\}_{1 \leq i \leq n}$ suivant la même loi que (X, Y) et on note \hat{g}_n un minimiseur de l'erreur empirique de classification :

$$\hat{g}_n \in \min_{g \in \mathcal{G}} L_n(g) \quad \text{où} \quad L_n : g \in \mathcal{G} \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{g(X_i) \neq Y_i\}}.$$

- 1) Montrer que $\min_{g \in \mathcal{G}} L_n(g) = 0$ presque-sûrement.
- 2) Montrer que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|(1 - \epsilon)^n$ pour tout $\epsilon \in [0, 1]$.

En déduire que $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}|e^{-n\epsilon}$ pour tout $\epsilon > 0$.

Indication. Utiliser $\mathcal{G}_B := \{g \in \mathcal{G} : L(g) > \epsilon\}$ ainsi qu'une borne d'union.

- 3) Dédurre de la question précédente que $\mathbb{E}(L(\hat{g}_n)) \leq \frac{\log(e|\mathcal{G}|)}{n}$.

Indication. Pour toute variable aléatoire Z positive, $\mathbb{E}(Z) = \int_0^{+\infty} \mathbb{P}(Z > t) dt$.

Solution:

- 1) Avant toute chose, rappelons que l'ensemble image (aléatoire) de L_n étant inclus par construction dans l'ensemble fini (déterministe) : $\{\frac{k}{n} \in \{0, n\}\}$, la variable aléatoire $\min_{g \in \mathcal{G}} L_n(g)$ existe toujours. Il nous faut ici montrer que ce minimum est presque-sûrement nul.

Pour cela, remarquons que par hypothèse, il existe $g^* \in \mathcal{G}$ tel que $L(g^*) = \mathbb{P}(g^*(X) \neq Y) = 0$. Par conséquent,

$$\mathbb{P}(L_n(g^*) = 0) = \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{\{g^*(X_i) \neq Y_i\}} = 0\right) = \mathbb{P}(g^*(X_1) = Y_1, \dots, g^*(X_n) = Y_n)$$

$$= \mathbb{P}(g^*(X) = Y)^n = (1 - L(g^*))^n = 1.$$

Cela veut dire qu'avec probabilité 1, l'ensemble $\{L_n(g) : g \in \mathcal{G}\}$ admet bien 0 pour minimum. En d'autres termes, $\min_{g \in \mathcal{G}} L_n(g) = 0$ presque-sûrement.

- 2) (a) Remarquons tout d'abord que la variable aléatoire $L(\hat{g}_n)$ est par construction à valeurs dans $[0, 1]$, donc si $\epsilon \geq 1$ on a directement $\mathbb{P}(L(\hat{g}_n) > \epsilon) = 0$.

Supposons maintenant $\epsilon \in [0, 1[$ et posons $\mathcal{G}_\epsilon := \{g \in \mathcal{G} : L(g) > \epsilon\}$, de telle manière que $\mathbb{P}(L(\hat{g}_n) > \epsilon) = \mathbb{P}(\hat{g}_n \in \mathcal{G}_\epsilon)$. D'après la question précédente, \hat{g}_n étant un minimiseur de l'erreur empirique de classification, on a $L_n(\hat{g}_n) = 0$ presque-sûrement. Ainsi,

$$\begin{aligned} \mathbb{P}(L(\hat{g}_n) > \epsilon) &= \mathbb{P}(\hat{g}_n \in \mathcal{G}_\epsilon, L_n(\hat{g}_n) = 0) \leq \mathbb{P}\left(\bigcup_{g \in \mathcal{G}_\epsilon} \{L_n(g) = 0\}\right) \leq \sum_{g \in \mathcal{G}_\epsilon} \mathbb{P}(L_n(g) = 0) \\ &= \sum_{g \in \mathcal{G}_\epsilon} \mathbb{P}(g(X) = Y)^n \quad (\text{comme en question 1}) \\ &= \sum_{g \in \mathcal{G}_\epsilon} (1 - L(g))^n \leq \sum_{g \in \mathcal{G}_\epsilon} (1 - \epsilon)^n \quad (\text{pour } g \in \mathcal{G}_\epsilon \text{ on a } L(g) > \epsilon) \\ &\leq |\mathcal{G}_\epsilon| (1 - \epsilon)^n \leq |\mathcal{G}| (1 - \epsilon)^n \quad (\mathcal{G}_\epsilon \subset \mathcal{G} \text{ ensemble fini}). \end{aligned}$$

Nous avons donc montré la première inégalité :

$$\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}| (1 - \epsilon)^n \mathbb{1}_{\{0 \leq \epsilon < 1\}}.$$

- (b) Par convexité de la fonction exponentielle, on a $1 - \epsilon \leq e^{-\epsilon}$ avec $e^{-\epsilon} > 0$. Ainsi, on a toujours $(1 - \epsilon)^n \mathbb{1}_{\{0 \leq \epsilon < 1\}} < e^{-n\epsilon}$. En partant de l'inégalité montrée à la question précédente, on obtient donc directement $\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq |\mathcal{G}| e^{-n\epsilon}$.

- 3) Comme la variable aléatoire $L(\hat{g}_n)$ est positive, à valeurs dans $[0, 1]$, on a

$$\mathbb{E}(L(\hat{g}_n)) = \int_0^{+\infty} \mathbb{P}(L(\hat{g}_n) > \epsilon) d\epsilon.$$

En utilisant la seconde inégalité montrée à la question précédente et le fait qu'une probabilité est toujours plus petite que 1, quel que soit $\epsilon \in \mathbb{R}^+$, on a

$$\mathbb{P}(L(\hat{g}_n) > \epsilon) \leq \min \{1, |\mathcal{G}| e^{-n\epsilon}\}.$$

Or, pour tout $\varepsilon \in \mathbb{R}^+$, on a $|\mathcal{G}|e^{-n\varepsilon} \leq 1$ si et seulement si $\varepsilon \geq \frac{1}{n} \ln(|\mathcal{G}|)$. Ainsi,

$$\begin{aligned} \mathbb{E}(L(\hat{g}_n)) &\leq \int_0^{\frac{\ln(|\mathcal{G}|)}{n}} 1 d\varepsilon + \int_{\frac{\ln(|\mathcal{G}|)}{n}}^{+\infty} |\mathcal{G}|e^{-n\varepsilon} d\varepsilon = \frac{\ln(|\mathcal{G}|)}{n} + |\mathcal{G}| \left[\frac{-e^{-n\varepsilon}}{n} \right]_{\frac{\ln(|\mathcal{G}|)}{n}}^{+\infty} \\ &= \frac{\ln(|\mathcal{G}|)}{n} + \frac{|\mathcal{G}|}{n} \frac{1}{|\mathcal{G}|} = \frac{1}{n} (\ln(|\mathcal{G}|) + 1) = \frac{1}{n} \ln(e|\mathcal{G}|). \end{aligned}$$

On retrouve donc bien l'inégalité recherchée.

Remarque. Par construction, la variable aléatoire $L(\hat{g}_n)$ est bornée, à valeurs dans $[0, 1]$. Il en est donc de même pour son espérance. La borne obtenue n'a donc d'intérêt que si $\frac{1}{n} \ln(e|\mathcal{G}|) \leq 1$, c'est-à-dire si $|\mathcal{G}| \leq e^{n-1}$.

En utilisant le caractère borné de $L(\hat{g}_n)$, on peut même raffiner cette borne. Si $|\mathcal{G}| \leq e^n$, alors

$$\begin{aligned} \mathbb{E}(L(\hat{g}_n)) &\leq \int_0^{\frac{\ln(|\mathcal{G}|)}{n}} 1 d\varepsilon + \int_{\frac{\ln(|\mathcal{G}|)}{n}}^1 |\mathcal{G}|e^{-n\varepsilon} d\varepsilon \\ &= \frac{1}{n} (\ln(|\mathcal{G}|) + 1 - |\mathcal{G}|e^{-n}). \end{aligned}$$

On retrouve la première borne en remarquant que

$$\begin{aligned} \frac{1}{n} (\ln(|\mathcal{G}|) + 1 - |\mathcal{G}|e^{-n}) &\leq \frac{1}{n} (\ln(|\mathcal{G}|) + 1 - \ln(|\mathcal{G}|)e^{-n}) \quad (|\mathcal{G}| > \ln(|\mathcal{G}|)) \\ &= \frac{1}{n} (\ln(|\mathcal{G}|)(1 - e^{-n}) + 1) \\ &\leq \frac{1}{n} \ln(e|\mathcal{G}|). \quad \text{for } 1 - e^{-n} \leq 1 \text{ pour } n \geq 0. \end{aligned}$$

Interprétation. Tous ces résultats indiquent qu'à nombre d'observations n fixé, on peut d'autant mieux contrôler l'erreur empirique de classification que la classe \mathcal{G} considérée est restreinte, i.e. que son cardinal est faible.

Alternativement, plus on enrichit \mathcal{G} , plus il faut de données (i.e. d'information) pour garantir une faible erreur empirique de classification. En particulier, on déduit de la question 2 que pour tous $\delta \in]0, 1[$ et $\epsilon > 0$, dès que $n \geq \frac{1}{\epsilon} \ln \frac{|\mathcal{G}|}{\delta}$, on a $\mathbb{P}(L(\hat{g}_n) \leq \epsilon) \geq 1 - \delta$.

EXERCICE 2. On se place dans le cadre de la classification binaire. On utilisera les mêmes notations que dans l'exercice précédent. On pose $L^* := L(g^*)$ avec $g^* : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta(x) \geq 1/2\}} - 1$ et on note $\eta : x \in \mathcal{X} \mapsto \mathbb{P}(Y = 1 \mid X = x) \in [0, 1]$ la fonction de régression. Soit $(\eta_n)_{n \in \mathbb{N}^*}$ une suite de fonctions définies sur \mathcal{X} à valeurs dans $]0, 1[$. Pour tout $n \in \mathbb{N}^*$ on considère le classifieur $g_n : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta_n(x) \geq 1/2\}} - 1$.

1) On suppose qu'il existe $\delta > 0$ tel que $|\eta(x) - 1/2| \geq \delta$ pour tout $x \in \mathcal{X}$. Montrer que

$$L(g_n) - L^* \leq \frac{2 \mathbb{E} ((\eta_n(X) - \eta(X))^2)}{\delta}.$$

2) Montrer que si $L^* = 0$, alors quel que soit $q \in [1, +\infty[$

$$L(g_n) \leq 2^q \mathbb{E} (|\eta_n(X) - \eta(X)|^q).$$

Soient maintenant $\eta' : \mathcal{X} \rightarrow]0, 1[$ et $g : x \in \mathcal{X} \mapsto 2\mathbb{1}_{\{\eta'(x) \geq 1/2\}} - 1$.

3) On suppose que $\mathbb{P}\{\eta'(X) = 1/2\} = 0$ et que $\mathbb{E} (|\eta_n(X) - \eta'(X)|) \rightarrow 0$ lorsque $n \rightarrow +\infty$. Montrer que $L(g_n) \rightarrow L(g)$ lorsque $n \rightarrow +\infty$.

4) On suppose que le label Y n'est plus observable, mais qu'une variable Z à valeurs dans $\{-1, +1\}$ l'est, telle que :

$$\mathbb{P}(Z = 1 \mid Y = -1, X) = \mathbb{P}(Z = 1 \mid Y = -1) = a < 1/2,$$

$$\mathbb{P}(Z = -1 \mid Y = 1, X) = \mathbb{P}(Z = -1 \mid Y = 1) = b < 1/2.$$

On pose à présent $\eta' : x \in \mathcal{X} \mapsto \mathbb{P}(Z = +1 \mid X = x)$. Montrer que :

$$L(g) \leq L^* \left(1 + \frac{2|a-b|}{1-2\max(a,b)} \right).$$

Que peut-on en déduire lorsque $a = b$?

Solution:

1) Soit $n \in \mathbb{N}^*$. Tout d'abord, l'inégalité à démontrer est trivialement vraie si $L(g_n) = L^*$. Supposons donc que $L(g_n) \neq L^*$, i.e. que $L(g_n) - L^* > 0$ (par définition de L^*), alors

$$\begin{aligned} L(g_n) - L^* &= \mathbb{P}(g_n(X) \neq Y) - \mathbb{P}(g^*(X) \neq Y) = \mathbb{E} (\mathbb{1}_{\{g_n(X) \neq Y\}} - \mathbb{1}_{\{g^*(X) \neq Y\}}) \\ &= \mathbb{E} (\mathbb{1}_{\{Y=-1\}} (\mathbb{1}_{\{g_n(X)=1\}} - \mathbb{1}_{\{g^*(X)=1\}}) + \mathbb{1}_{\{Y=1\}} (\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g^*(X)=-1\}})) \\ &= \mathbb{E} ((1 - \eta(X)) (\mathbb{1}_{\{g_n(X)=1\}} - \mathbb{1}_{\{g^*(X)=1\}}) + \eta(X) (\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g^*(X)=-1\}})) \\ &= \mathbb{E} ((2\eta(X) - 1) (\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g^*(X)=-1\}})) \end{aligned}$$

Or pour tout $x \in \mathcal{X}$, si $g^*(x) = -1$ alors $2\eta(x) - 1 \leq 0$ et $\mathbb{1}_{\{g_n(x)=-1\}} - \mathbb{1}_{\{g^*(x)=-1\}} \leq 0$, puis

si $g^*(x) = 1$ alors $2\eta(x) - 1 > 0$ et $\mathbb{1}_{\{g_n(x)=-1\}} - \mathbb{1}_{\{g^*(x)=-1\}} \geq 0$, d'où $(2\eta(x) - 1)(\mathbb{1}_{\{g_n(x)=-1\}} - \mathbb{1}_{\{g^*(x)=-1\}}) \geq 0$. Ainsi,

$$\begin{aligned} L(g_n) - L^* &= \mathbb{E} \left[|2\eta(X) - 1| \left| \mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g^*(X)=-1\}} \right| \right] \\ &= 2\mathbb{E} \left[\left| \eta(X) - \frac{1}{2} \right| \mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right]. \end{aligned} \quad (1)$$

On remarque que pour tout $x \in X$, si $g_n(x) \neq g^*(x)$ alors $\eta_n(x)$ et $\eta(x)$ sont d'un côté et de l'autre de $\frac{1}{2}$, d'où $|\eta_n(x) - \eta(x)| \geq \left| \eta(x) - \frac{1}{2} \right|$. On en déduit que

$$L(g_n) - L^* \leq 2\mathbb{E} \left[|\eta_n(X) - \eta(X)| \mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right] \quad (2)$$

$$\leq 2\mathbb{E} \left[(\eta_n(X) - \eta(X))^2 \right]^{\frac{1}{2}} \mathbb{E} \left[(\mathbb{1}_{\{g_n(X) \neq g^*(X)\}})^2 \right]^{\frac{1}{2}} \quad (\text{Cauchy-Schwartz}).$$

$$= 2\mathbb{E} \left[(\eta_n(X) - \eta(X))^2 \right]^{\frac{1}{2}} \mathbb{E} \left[\mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right]^{\frac{1}{2}} \quad (3)$$

En outre, par hypothèse, il existe $\delta > 0$ tel que $|\eta(x) - \frac{1}{2}| \geq \delta$ pour tout $x \in X$. Alors en repartant de l'Eq. (1), on obtient

$$L(g_n) - L^* = 2\mathbb{E} \left[\left| \eta(X) - \frac{1}{2} \right| \mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right] \geq 2\delta \mathbb{E} \left[\mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right] \geq 0$$

d'ou

$$\mathbb{E} \left[\mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right]^{\frac{1}{2}} \leq \left(\frac{1}{2\delta} (L(g_n) - L^*) \right)^{\frac{1}{2}}$$

En injectant ce dernier résultat dans l'Eq.(3), on obtient finalement

$$\begin{aligned} L(g_n) - L^* &\leq 2\mathbb{E} \left[(\eta_n(X) - \eta(X))^2 \right]^{\frac{1}{2}} \left(\frac{1}{2\delta} (L(g_n) - L^*) \right)^{1/2} \\ \Leftrightarrow (2\delta(L(g_n) - L^*))^{1/2} &\leq 2\mathbb{E} \left[(\eta_n(X) - \eta(X))^2 \right]^{\frac{1}{2}} \\ \Leftrightarrow L(g_n) - L^* &\leq \frac{2}{\delta} \mathbb{E} \left[(\eta_n(X) - \eta(X))^2 \right]. \quad (\text{tout est positif}) \end{aligned}$$

- 2) Soit $n \in \mathbb{N}^*$ et supposons $L^* = 0$. Cela signifie que $g^*(X) = Y$ presque-sûrement. Comme à la question précédente, l'inégalité à démontrer est trivialement vraie si $L(g_n) = 0$. Supposons donc que $L(g_n) > 0$, alors en repartant de l'Eq. (2) et en appliquant l'inégalité de Hölder, pour tout $q \in [1, +\infty[$ on a

$$\begin{aligned}
L(g_n) &\leq 2\mathbb{E} [|\eta_n(X) - \eta(X)|^q]^{\frac{1}{q}} \mathbb{E} \left[(\mathbb{1}_{\{g_n(X) \neq g^*(X)\}})^{\frac{q}{q-1}} \right]^{\frac{q-1}{q}} \\
&= 2\mathbb{E} [|\eta_n(X) - \eta(X)|^q]^{\frac{1}{q}} \mathbb{E} \left[\mathbb{1}_{\{g_n(X) \neq g^*(X)\}} \right]^{\frac{q-1}{q}} \\
&= 2\mathbb{E} [|\eta_n(X) - \eta(X)|^q]^{\frac{1}{q}} \mathbb{E} \left[\mathbb{1}_{\{g_n(X) \neq Y\}} \right]^{\frac{q-1}{q}} \\
&= 2\mathbb{E} [|\eta_n(X) - \eta(X)|^q]^{\frac{1}{q}} L(g_n)^{\frac{q-1}{q}}
\end{aligned}$$

Ainsi,

$$\begin{aligned}
L(g_n) &\leq 2\mathbb{E} [|\eta_n(X) - \eta(X)|^q]^{\frac{1}{q}} L(g_n)^{\frac{q-1}{q}} \\
&\iff L(g_n)^{1-\frac{q-1}{q}} = L(g_n)^{\frac{1}{q}} \leq 2\mathbb{E} (|\eta_n(X) - \eta(X)|^q)^{1/q} \quad (L(g_n) > 0 \text{ par hypothèse}) \\
&\iff L(g_n) \leq 2^q \mathbb{E} (|\eta_n(X) - \eta(X)|^q). \quad (\text{Tout est positif})
\end{aligned}$$

3) Soit $n \in \mathbb{N}^*$. En reprenant les mêmes calculs qu'à la question 1, on obtient

$$\begin{aligned}
|L(g_n) - L(g)| &= 2 \left| \mathbb{E} \left((\eta(X) - 1/2) (\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g(X)=-1\}}) \right) \right| \\
&\leq 2\mathbb{E} (|\eta(X) - 1/2| |\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g(X)=-1\}}|) \quad (\text{inégalité triangulaire}) \\
&\leq \mathbb{E} (|\mathbb{1}_{\{g_n(X)=-1\}} - \mathbb{1}_{\{g(X)=-1\}}|) \quad (|\eta(X) - 1/2| \leq 2/2 \text{ par définition}) \\
&= \mathbb{E} (\mathbb{1}_{\{g_n(X) \neq g(X)\}}) = P(g_n(X) \neq g(X)) \\
&= P\left(\eta_n(X) > \frac{1}{2}, \eta(X) \leq \frac{1}{2}\right) + P\left(\eta_n(X) \leq \frac{1}{2}, \eta(X) > \frac{1}{2}\right)
\end{aligned}$$

Or, quel que soit $\lambda \in [0, \frac{1}{2}]$, on a les relations d'événements suivantes :

$$\begin{aligned}
\left\{ \eta'(X) > \frac{1}{2} \right\} &= \left\{ \eta'(X) \geq \frac{1}{2} + \lambda \right\} \cup \left\{ \frac{1}{2} < \eta'(X) < \frac{1}{2} + \lambda \right\}, \text{ puis} \\
\left\{ \eta_n(X) \leq \frac{1}{2} \right\} \cap \left\{ \eta'(X) \geq \frac{1}{2} + \lambda \right\} &\subset \{ |\eta_n(X) - \eta'(X)| \geq \lambda \}, \\
\left\{ \eta_n(X) \leq \frac{1}{2} \right\} \cap \left\{ \frac{1}{2} < \eta'(X) < \frac{1}{2} + \lambda \right\} &\subset \left\{ \left| \eta'(X) - \frac{1}{2} \right| < \lambda \right\}.
\end{aligned}$$

d'où

$$\mathbb{P} \left(\eta_n(X) \leq \frac{1}{2}, \eta'(X) > \frac{1}{2} \right) \leq \mathbb{P} \left(\left| \eta'(X) - \frac{1}{2} \right| < \lambda \right) + \mathbb{P} (|\eta_n(X) - \eta'(X)| \geq \lambda),$$

$$\begin{aligned}\left\{\eta'(X) \leq \frac{1}{2}\right\} &= \left\{\eta'(X) \leq \frac{1}{2} + \lambda\right\} \cup \left\{\frac{1}{2} < \eta'(X) \leq \frac{1}{2} + \lambda\right\}, \text{ puis} \\ \left\{\eta_n(X) > \frac{1}{2}\right\} \cap \left\{\eta'(X) \leq \frac{1}{2} + \lambda\right\} &\subset \{|\eta_n(X) - \eta'(X)| \geq \lambda\}, \\ \left\{\eta_n(X) > \frac{1}{2}\right\} \cap \left\{1 < \eta'(X) \leq \frac{1}{2} + \lambda\right\} &\subset \left\{\left|\eta'(X) - \frac{1}{2}\right| < \lambda\right\}.\end{aligned}$$

d'où

$$\mathbb{P}\left(\eta_n(X) > \frac{1}{2}, \eta'(X) \leq \frac{1}{2}\right) \leq \mathbb{P}\left(\left|\eta'(X) - \frac{1}{2}\right| < \lambda\right) + \mathbb{P}(|\eta_n(X) - \eta'(X)| \geq \lambda),$$

Par conséquent, pour tout $\lambda \in]0, \frac{1}{2}[$ on a

$$|L(g_n) - L(g)| \leq 2\mathbb{P}\left(\left|\eta'(X) - \frac{1}{2}\right| < \lambda\right) + 2\mathbb{P}(|\eta_n(X) - \eta'(X)| \geq \lambda),$$

Soit $\epsilon > 0$, il s'agit maintenant de trouver $N \in \mathbb{N}^*$ tel que pour tout $n \geq N$, on a $|L(g_n) - L(g)| < \epsilon$.

Tout d'abord, puisque $\mathbb{P}(\eta'(X) = \frac{1}{2}) = \mathbb{P}(|\eta'(X) - \frac{1}{2}| = 0) = 0$ par hypothèse, on peut choisir $\lambda \in (0, \frac{1}{2})$ tel que $\mathbb{P}(|\eta'(X) - \frac{1}{2}| < \lambda) \leq \frac{\epsilon}{4}$ (si on avait une masse $m \in]0, 1]$ en 0, on aurait toujours $\mathbb{P}(|\eta'(X) - \frac{1}{2}| < \lambda) \geq m$ et on ne pourrait pas descendre en dessous).

Ensuite, d'après l'inégalité de Markov, $\mathbb{P}(|\eta_n(X) - \eta'(X)| \geq \lambda) \leq \frac{1}{\lambda} \mathbb{E}(|\eta_n(X) - \eta'(X)|)$ et par hypothèse, il existe $N = N(\epsilon, \lambda) \in \mathbb{N}^*$ tel que pour tout $n \geq N$, $\mathbb{E}(|\eta_n(X) - \eta'(X)|) \leq \frac{\lambda\epsilon}{4}$, et finalement $|L(g_n) - L(g)| < \epsilon$.

Nous avons donc bien montré que $L(g_n) \rightarrow L(g)$ quand $n \rightarrow +\infty$.

4) Commençons par remarquer que pour tout $x \in \mathcal{X}$, on a

$$\begin{aligned}\eta'(x) &= \mathbb{P}(Z = 1 \mid X = x) \\ &= \mathbb{P}(Z = 1 \mid Y = 1, X = x)\mathbb{P}(Y = 1 \mid X = x) + \mathbb{P}(Z = 1 \mid Y = -1, X = x)\mathbb{P}(Y = -1 \mid X = x) \\ &= (1 - b)\eta(x) + a(1 - \eta(x)) = a + (1 - b - a)\eta(x).\end{aligned}$$

Maintenant, d'après l'Eq.(2), on a

$$\begin{aligned}L(g) - L^* &\leq 2\mathbb{E}(|\eta'(X) - \eta(X)| \mathbb{1}_{\{g(X) \neq g^*(X)\}}) \\ &= 2\mathbb{E}(|\eta'(X) - \eta(X)| \mid g(X) \neq g^*(X)) \mathbb{P}(g(X) \neq g^*(X)).\end{aligned}$$

Il s'agit de majorer les deux termes à droite de cette dernière inégalité (l'espérance conditionnelle et la probabilité).

Soit $x \in \mathcal{X}$, alors $g(x) \neq g^*(x)$ signifie (i) $\eta(x) \geq \frac{1}{2} > \eta'(x)$ ou (ii) $\eta(x) < \frac{1}{2} \leq \eta'(x)$. Puisque $\eta'(x) = a + (1 - a - b)\eta(x)$, cela donne

(a) $\frac{1 - 2a}{2(1 - a - b)} > \eta(x) \geq \frac{1}{2}$, ce qui n'est possible que si $a < b$, et dans ce cas

$$|\eta'(x) - \eta(x)| = \eta(x) - \eta'(x) = \eta(x)(a + b) - a$$

$$\frac{1 - 2a}{2(1 - a - b)}(a + b) - a = \frac{b - a}{2(1 - a - b)} = \frac{|a - b|}{2(1 - a - b)}.$$

(b) $\frac{1 - 2a}{2(1 - a - b)} \leq \eta(x) < \frac{1}{2}$, ce qui n'est possible que si $a > b$, et dans ce cas

$$|\eta'(x) - \eta(x)| = \eta'(x) - \eta(x) = a - \eta(x)(a + b)$$

$$a - \frac{1 - 2a}{2(1 - a - b)}(a + b) = \frac{a - b}{2(1 - a - b)} = \frac{|a - b|}{2(1 - a - b)}.$$

On peut donc d'ores et déjà majorer l'espérance conditionnelle :

$$\mathbb{E}(|\eta'(X) - \eta(X)| | g(X) \neq g^*(X)) \leq \frac{|a - b|}{2(1 - a - b)}.$$

Pour majorer la probabilité, rappelons-nous que d'après le TD1 on a

$$\begin{aligned} L^* &= \mathbb{E}(\eta(X) \wedge (1 - \eta(X))) \\ &\geq \mathbb{E}(\eta(X) \wedge (1 - \eta(X)) \mathbb{1}_{g(X) \neq g^*(X)}) \\ &= \mathbb{E}(\eta(X) \wedge (1 - \eta(X)) | g(X) \neq g^*(X)) \mathbb{P}(g(X) \neq g^*(X)) \\ &\geq \begin{cases} \frac{1 - 2b}{2(1 - a - b)} \mathbb{P}(g(X) \neq g^*(X)) & \text{si } a < b \\ \frac{1 - 2a}{2(1 - b - a)} \mathbb{P}(g(X) \neq g^*(X)) & \text{si } a > b \end{cases} \\ &= \frac{1 - 2(a \vee b)}{2(1 - a - b)} \mathbb{P}(g(X) \neq g^*(X)). \end{aligned}$$

d'où

$$\mathbb{P}(g(X) \neq g^*(X)) \leq 2L^* \frac{1 - a - b}{1 - 2(a \vee b)}.$$

Finalement, on obtient

$$L(g) - L^* \leq \frac{|a - b|}{1 - a - b} 2L^* \frac{1 - a - b}{1 - 2(a \vee b)} = L^* \frac{2|a - b|}{1 - 2(a \vee b)}.$$

d'où le résultat recherché

$$L(g) \leq L^* \left(1 + \frac{2|a-b|}{1-2(a \vee b)} \right)$$

Lorsque $a = b$, on a $\eta'(x) = a + (1-2a)\eta(x)$. Ainsi, $\eta'(x) > \frac{1}{2} \iff \eta(x) > \frac{1}{2} \quad \frac{1-2a}{1-2a} = 1$ et donc $g = g^*$; on retombe sur le classifieur de Bayes.
