Correction des travaux dirigés de Machine learning

Cycle pluridisciplinaire d'études supérieures
Université Paris sciences et lettres

Joon Kwon

vendredi 5 juin 2020

Exercice 1. —

1) Pour $i \in [n]$, sachant que $\widehat{f}(x_i)$ et y_i sont dans $\{-1,1\}$, on a :

$$\begin{aligned}\widehat{f}(x_i) \neq y_i \quad &\iff \quad \operatorname{sign}\left(\widehat{f}^{(1:m)}(x_i)\right) \neq \operatorname{sign}(y_i) \\ &\iff \quad y_i \widehat{f}^{(1:m)}(x_i) \leqslant 0.\end{aligned}$$

On a donc :

$$\begin{aligned}\varepsilon_S(\widehat{f}) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{\widehat{f}(x_i) \neq y_i\right\}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\left\{y_i \widehat{f}^{(1:m)}(x_i) \leqslant 0\right\}} \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n} \exp(-y_i \widehat{f}^{(1:m)}(x_i)) = Z^{(m)},\end{aligned}$$

car on peut facilement se convaincre que pour tout $u \in \mathbb{R}$, $\mathbb{1}_{\{u \geqslant 0\}} \leqslant e^u$.

2) Par récurrence. Pour $k = 0$,

$$\pi^{(1)} = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \left(\frac{e^{-y_i \times 0}}{\sum_{j=1}^{n} e^{-y_i \times 0}}\right)_{i \in [n]} = \left(\frac{e^{-y_i \widehat{f}^{(1:0)}(x_i)}}{\sum_{j=1}^{n} e^{-y_j \widehat{f}^{(1:0)}(x_i)}}\right)_{i \in [n]}.$$

Pour $1 \leqslant k \leqslant m$, et $i \in [n]$, on a par hypothèse de récurrence :

$$\begin{aligned}
\pi_i^{(k)} \exp\left(-w^{(k)} y_i \widehat{f}^{(k)}(x_i)\right) &= \frac{\exp\left(-y_i \widehat{f}^{(1:(k-1))}(x_i)\right)}{\sum_{j=1}^n \exp\left(-y_j \widehat{f}^{(1:(k-1))}(x_j)\right)} \exp\left(-w^{(k)} y_i \widehat{f}^{(k)}(x_i)\right) \\
&= \frac{\exp\left(-w^{(k)} \widehat{f}^{(1:k)}(x_i)\right)}{\sum_{j=1}^n \exp\left(-y_j \widehat{f}^{(1:(k-1))}(x_j)\right)}.
\end{aligned}$$

Donc,

$$\begin{aligned}
\pi^{(k+1)} &= \left(\frac{\pi_i^{(k)} \exp\left(-w^{(k)} y_i \widehat{f}^{(k)}(x_i)\right)}{\sum_{j=1}^n \pi_j^{(k)} \exp\left(-w^{(k)} y_j \widehat{f}^{(k)}(x_j)\right)}\right)_{i\in[n]} \\
&= \left(\frac{\exp\left(-y_i \widehat{f}^{(1:k)}(x_i)\right)}{\sum_{j=1}^n \exp\left(-y_j \widehat{f}^{(1:k)}(x_j)\right)}\right)_{i\in[n]}
\end{aligned}$$

3) Pour $0 \leqslant k \leqslant m-1$, on a $\hat{f}^{(1:k+1)} = \hat{f}^{(1:k)} + w^{(k)}\hat{f}^{(k)}$, et donc :

$$\begin{aligned}
\frac{Z^{(k+1)}}{Z^{(k)}} &= \frac{\sum_{i=1}^{n} \exp\left(-y_i \hat{f}^{(1:k+1)}(x_i)\right)}{\sum_{i=1}^{n} \exp\left(-y_i \hat{f}^{(1:k)}(x_i)\right)} \\
&= \frac{\sum_{i=1}^{n} \exp\left(-y_i \hat{f}^{(1:k)}(x_i)\right) \exp\left(-y_i w^{(k+1)} \hat{f}^{(k+1)}(x_i)\right)}{\sum_{i=1}^{n} \exp\left(-y_i \hat{f}^{(1:k)}(x_i)\right)} \\
&= \sum_{i=1}^{n} \pi_i^{(k+1)} \exp\left(-y_i w^{(k+1)} \hat{f}^{(k+1)}(x_i)\right) \\
&= e^{-w^{(k+1)}} \sum_{\substack{i\in[n] \\ \hat{f}^{(k+1)}(x_i)=y_i}} \pi_i^{(k+1)} + e^{w^{(k+1)}} \sum_{\substack{i\in[n] \\ \hat{f}^{(k+1)}(x_i)\neq y_i}} \pi_i^{(k+1)} \\
&= e^{-w^{(k+1)}}(1-\varepsilon^{(k+1)}) + e^{w^{(k+1)}}\varepsilon^{(k+1)} \\
&= \frac{1}{\sqrt{1/\varepsilon^{(k+1)}-1}}(1-\varepsilon^{(k+1)}) + \sqrt{1/\varepsilon^{(k+1)}-1}\cdot\varepsilon^{(k+1)} \\
&= \sqrt{\frac{\varepsilon^{(k+1)}}{1-\varepsilon^{(k+1)}}}(1-\varepsilon^{(k+1)}) + \sqrt{\frac{1-\varepsilon^{(k+1)}}{\varepsilon^{(k+1)}}}\varepsilon^{(k+1)} \\
&= 2\sqrt{\varepsilon^{(k+1)}(1-\varepsilon^{(k+1)})}.
\end{aligned}$$

4) Un simple étude de fonction permet de voir que $u \mapsto u(1-u)$ est croissante sur $[0, 1/2]$. Comme pour $0 \leqslant k \leqslant m$, $\varepsilon^{(k+1)} \leqslant 1/2 - \gamma$ par hypthèse, on a, en remarquant que $Z^{(0)} = 1$,

$$\begin{aligned}
\varepsilon_S(\hat{f}) \leqslant Z^{(m)} &= \frac{Z^{(m)}}{Z^{(m-1)}} \times \cdots \times \frac{Z^{(1)}}{Z^{(0)}} = \prod_{k=0}^{m-1} 2\sqrt{\varepsilon^{(k+1)}(1-\varepsilon^{(k+1)})} \\
&\leqslant \prod_{k=0}^{m-1} 2\sqrt{\left(\frac{1}{2}-\gamma\right)\left(\frac{1}{2}+\gamma\right)} = (1-4\gamma^2)^{m/2} = e^{\frac{m}{2}\log(1-4\gamma^2)} \\
&\leqslant e^{\frac{m}{2}(-4\gamma^2)} = e^{-2\gamma^2 m}.
\end{aligned}$$