Exercise of Supervised Learning: Boosting Part 1

Yawei Li

yawei.li@stat.uni-muenchen.de

December 22, 2023

Exercise 1: AdaBoost - Empirical Risk

Let $\hat{f}(\mathbf{x}) = \sum_{m=1}^{M} \hat{\beta}^{[m]} \hat{b}^{[m]}(\mathbf{x})$ be the scoring function after running AdaBoost for $M \in \mathbb{N}$ iterations. Show that the average empirical risk (on $\mathcal{D}_{\text{train}}$) of the corresponding classifier $h(\mathbf{x}) = \text{sign}(\hat{f}(\mathbf{x}))$ is bounded as follows

$$\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} = \frac{\sum_{i=1}^{n} \mathbb{I}_{[\hat{h}\left(\mathbf{x}^{(i)}\right) \neq y^{(i)}]}}{n} \leq \prod_{m=1}^{M} \sqrt{1 - 4\left(\hat{\gamma}^{[m]}\right)^2}, \tag{1}$$

where $\hat{\gamma}^{[m]} = \frac{1}{2} - \mathrm{err}^{[m]}$. For this purpose, proceed as follows:

(a) Given an interpretation of $\hat{\gamma}^{[m]}$.

Solution to Exercise 1 (a)

- Recall that $\mathrm{err}^{[m]} = \sum_{i=1}^{n} w^{[m](i)} \cdot \boldsymbol{I}_{y^{(i)} \neq \hat{b}^{[m]}(\mathbf{x}^{(i)})}$ is the weighted error of $\hat{b}^{[m]}$.
- Random guessing has an error of approx. $\frac{1}{2}$.
- So, $\hat{\gamma}^{[m]} = \frac{1}{2} - \mathrm{err}^{[m]}$ tells us how better $\hat{b}^{[m]}$ is compared to random guessing.

Exercise 1 (b)

(b) For any $m = 1, \ldots, M$ let $W^{[m]} = \sum_{i=1}^{n} \tilde{w}^{[m](i)}$ be the total weight in iteration m before normalizing the weights. Show that $W^{[m]} = \sqrt{1 - 4\left(\hat{\gamma}^{[m]}\right)^2}$.

Hint:

- $\tilde{w}^{[m](i)} = w^{[m](i)} \cdot \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)$.
- $\mathrm{err}^{[m]} = \sum_{i=1}^{n} w^{[m](i)} \cdot \mathbb{I}_{y^{(i)} \neq \hat{b}^{[m]}(\mathbf{x}^{(i)})}$

Solution to Exercise 1 (b)

$$
\begin{aligned}
W^{[m]} &= \sum_{i=1}^{n} \tilde{w}^{[m](i)} \\
&= \sum_{i=1}^{n} w^{[m](i)} \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right) \\
&= \underbrace{\sum_{i: y^{(i)} \neq \hat{b}^{[m]}(\mathbf{x}^{(i)})}}_{\text{incorrect pred.}} w^{[m](i)} \cdot \exp\left(\beta^{[m]}\right) + \underbrace{\sum_{i: y^{(i)} = \hat{b}^{[m]}(\mathbf{x}^{(i)})}}_{\text{correct pred.}} w^{[m](i)} \cdot \exp\left(-\beta^{[m]}\right) \\
&= \exp\left(\beta^{[m]}\right) \underbrace{\sum_{i: y^{(i)} \neq \hat{b}^{[m]}(\mathbf{x}^{(i)})} w^{[m](i)}}_{\mathrm{err}^{[m]}} + \exp\left(-\beta^{[m]}\right) \underbrace{\sum_{i: y^{(i)} = \hat{b}^{[m]}(\mathbf{x}^{(i)})} w^{[m](i)}}_{1 - \mathrm{err}^{[m]}} \\
&= \exp\left(\beta^{[m]}\right) \mathrm{err}^{[m]} + \exp\left(-\beta^{[m]}\right) \left(1 - \mathrm{err}^{[m]}\right)
\end{aligned}
$$

Solution to Exercise 1 (b): Continued

$$W^{[m]} = \sum_{i=1}^{n} \tilde{w}^{[m](i)} = \exp\left(\beta^{[m]}\right) \mathrm{err}^{[m]} + \exp\left(-\beta^{[m]}\right) \left(1 - \mathrm{err}^{[m]}\right) \tag{2}$$

Recall that $\beta^{[m]} = \frac{1}{2} \log\left(\frac{1-\mathrm{err}^{[m]}}{\mathrm{err}^{[m]}}\right)$, so that

$$\exp\left(\beta^{[m]}\right) = \sqrt{\frac{1 - \mathrm{err}^{[m]}}{\mathrm{err}^{[m]}}}, \qquad \text{and} \qquad \exp\left(-\beta^{[m]}\right) = \sqrt{\frac{\mathrm{err}^{[m]}}{1 - \mathrm{err}^{[m]}}}. \tag{3}$$

We can then plug (2) into (3) and elimnate the terms related to $\beta^{[m]}$.

Solution to Exercise 1 (b): Continued

$$
\begin{aligned}
W^{[m]} &= \exp\left(\beta^{[m]}\right)\mathrm{err}^{[m]} + \exp\left(-\beta^{[m]}\right)(1-\mathrm{err}^{[m]}) \\
&= 2\sqrt{(1-\mathrm{err}^{[m]})\mathrm{err}^{[m]}} \\
&= 2\sqrt{\left(\frac{1}{2}+\hat{\gamma}^{[m]}\right)\left(\frac{1}{2}-\hat{\gamma}^{[m]}\right)} \qquad\qquad \left(\hat{\gamma}^{[m]} = \frac{1}{2} - \mathrm{err}^{[m]}\right) \\
&= 2\sqrt{\frac{1}{4} - \left(\hat{\gamma}^{[m]}\right)^2} \\
&= \sqrt{1-4(\hat{\gamma}^{[m]})^2}.
\end{aligned}
$$

Exercise 1 (c)

(c) Show that

$$w^{[M+1](i)} = \frac{w^{[1](i)} \exp\left(-y^{(i)}\hat{f}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}},$$

where $w^{[M+1](i)}$ is the **normalized** weight if we would run AdaBoost for $M+1$ iterations.
Hint:

$$w^{[m+1](i)} = \frac{\tilde{w}^{[m](i)}}{\sum\limits_{i=1}^{n} \tilde{w}^{[m](i)}} = \frac{w^{[m](i)} \cdot \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\sum\limits_{i=1}^{n} w^{[m](i)} \cdot \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}$$

Solution to Exercise 1 (c)

$$
\begin{aligned}
w^{[M+1](i)} &= w^{[M](i)} \cdot \frac{\exp\left(-\beta^{[M]} y^{(i)} \hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{\sum\limits_{i=1}^{n} w^{[M](i)} \cdot \exp\left(-\beta^{[M]} y^{(i)} \hat{b}^{[M]}(\mathbf{x}^{(i)})\right)} \\
&= w^{[M](i)} \cdot \frac{\exp\left(-\beta^{[M]} y^{(i)} \hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{W^{[M]}} \qquad \text{(Definition of } W^{[M]}\text{)} \\
&= w^{[M-1](i)} \cdot \frac{\exp\left(-\beta^{[M-1]} y^{(i)} \hat{b}^{[M-1]}(\mathbf{x}^{(i)})\right)}{W^{[M-1]}} \cdot \frac{\exp\left(-\beta^{[M]} y^{(i)} \hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{W^{[M]}} \qquad \text{(Use hint)} \\
&= \ldots \qquad \text{(Repeatedly use the hint)} \\
&= w^{[1](i)} \cdot \frac{\prod_{m=1}^{M} \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}} = w^{[1](i)} \frac{\exp\left(-y^{(i)} \sum_{m=1}^{M} \beta^{[m]} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}} \\
&= \frac{w^{[1](i)} \exp\left(-y^{(i)} \hat{f}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}} \qquad \left(\text{Since } \sum_{m=1}^{M} \beta^{[m]} \hat{b}^{[m]}(\mathbf{x}^{(i)}) = \hat{f}(\mathbf{x}^{(i)})\right)
\end{aligned}
$$

Exercise 1 (d)

(d) Argue that $\boldsymbol{I}_{[\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}]} \leq \exp\left(-y\hat{f}(\mathbf{x})\right)$ for any $(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$.
Hint: What happens to $\exp(-y\hat{f}(\mathbf{x}))$ if $y^{(i)} \neq \hat{h}(\mathbf{x}^{(i)})$?

Solution to Exercise 1 (d)

$$
\begin{aligned}
\hat{h}(\mathbf{x}) \neq y &\Leftrightarrow \operatorname{sign}(\hat{f}(\mathbf{x})) \neq y \\
&\Leftrightarrow -y\hat{f}(\mathbf{x}) > 0 \\
&\Leftrightarrow \exp(-y\hat{f}(\mathbf{x})) > \exp(0) = 1 = \boldsymbol{I}_{[\hat{h}(\mathbf{x}) \neq y]}
\end{aligned}
$$

Exercise 1 (e)

(e) Combine everything to conclude (1).
Hint: Since for any x it holds that $1 + x \leq \exp(x)$, we can infer from (1) that

$$\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} \leq \exp\left(-2\sum_{m=1}^{M}\left(\hat{\gamma}^{[m]}\right)^2\right) = \exp\left(-2\sum_{m=1}^{M}\left(\frac{1}{2} - \mathrm{err}^{[m]}\right)^2\right),$$

i.e., the average empirical risk is decreasing exponentially in the number of used iterations (provided $\mathrm{err}^{[m]} < 1/2$).

Solution to 1 (e)

$$\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} = \frac{\sum_{i=1}^{n} \boldsymbol{I}_{\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}}}{n} = \sum_{i=1}^{n} \frac{1}{n} \cdot \boldsymbol{I}_{\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}} \leq \sum_{i=1}^{n} \frac{1}{n} \cdot \exp\left(-y^{(i)} \hat{f}(\mathbf{x}^{(i)})\right) \qquad \text{(Use (d))}$$

$$= \sum_{i=1}^{n} w^{[1](i)} \exp\left(-y^{(i)} \hat{f}(\mathbf{x}^{(i)})\right) \qquad \text{(Definition of } w^{[1](i)}\text{)}$$

$$= \sum_{i=1}^{n} w^{[M+1](i)} \prod_{m=1}^{M} W^{[m]} \qquad \left(\text{Use (c): } w^{[M+1](i)} = \frac{w^{[1](i)} \exp\left(-y^{(i)} \hat{f}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}}\right)$$

$$= \prod_{m=1}^{M} W^{[m]} \underbrace{\sum_{i=1}^{n} w^{[M+1](i)}}_{=1} \leq \prod_{m=1}^{M} \sqrt{1 - 4\left(\hat{\gamma}^{[m]}\right)^2} \qquad \text{(Use (b))}$$