Solution 1: AdaBoost - Empirical Risk

Write $\tilde{w}^{[m](i)}$ for the unnormalized weight and $w^{[m](i)}$ for the normalized weight of instance $i = 1, \dots, n$ in iteration step $m = 1, \dots, M$. Thus,

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

and

$$w^{[m+1](i)} = \frac{\tilde{w}^{[m](i)}}{\sum_{i=1}^{n} \tilde{w}^{[m](i)}} = \frac{w^{[m](i)} \cdot \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\sum_{i=1}^{n} w^{[m](i)} \cdot \exp\left(-\beta^{[m]} y^{(i)} \hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}. \tag{2}$$

(a) Recall that

$$\mathrm{err}^{[m]} = \sum_{i=1}^{n} w^{[m](i)} \cdot \mathbb{1}_{\{y^{(i)} \neq \hat{b}^{[m]}(\mathbf{x}^{(i)})\}}$$

is the weighted error of $\hat{b}^{[m]}$. Random guessing has an error of approximately[1] $\frac{1}{2}$, so that $\gamma^{[m]} = \frac{1}{2} - \mathrm{err}^{[m]}$ tells us how much better $\hat{b}^{[m]}$ (in terms of the error) is compared to random guessing.

(b) By means of (1) it holds that $W^{[m]} = \sum_{i=1}^{m} \tilde{w}^{[m](i)}$ is the total weigh in iteration m **before** normalizing the weights for any $m = 1, \dots, M$. With this,

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

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

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

[1] If the data set is balanced.

Using this for our representation of $W^{[m]}$ we obtain

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

As a side note: $\beta^{[m]}$ is chosen such that $\exp\left(\beta^{[m]}\right)\mathrm{err}^{[m]} + \exp\left(-\beta^{[m]}\right)(1-\mathrm{err}^{[m]})$ is minimal. This is due to (we will see this below):

$$
\frac{\mathcal{R}_{\mathrm{emp}}(\hat{h})}{n} \leq \prod_{m=1}^{M} W^{[m]} = \prod_{m=1}^{M} \exp\left(\beta^{[m]}\right)\mathrm{err}^{[m]} + \exp\left(-\beta^{[m]}\right)(1-\mathrm{err}^{[m]}).
$$

(c) Using (2) repeatedly, we obtain

$$
\begin{aligned}
w^{[M+1](i)} &= w^{[M](i)} \cdot \frac{\exp\left(-\beta^{[M]}y^{(i)}\hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{\sum_{i=1}^{n} w^{[M](i)} \cdot \exp\left(-\beta^{[M]}y^{(i)}\hat{b}^{[M]}(\mathbf{x}^{(i)})\right)} && \text{(Using (2))} \\
&= w^{[M](i)} \cdot \frac{\exp\left(-\beta^{[M]}y^{(i)}\hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{W^{[M]}} && \text{(Definition of } W^{[M]}\text{)} \\
&= w^{[M-1](i)} \cdot \frac{\exp\left(-\beta^{[M-1]}y^{(i)}\hat{b}^{[M-1]}(\mathbf{x}^{(i)})\right)}{W^{[M-1]}} \cdot \frac{\exp\left(-\beta^{[M]}y^{(i)}\hat{b}^{[M]}(\mathbf{x}^{(i)})\right)}{W^{[M]}} && \text{(Using (2) again)} \\
&= w^{[1](i)} \cdot \frac{\prod_{m=1}^{M}\exp\left(-\beta^{[m]}y^{(i)}\hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}} && \text{(Using (2) again and again)} \\
&= w^{[1](i)} \cdot \frac{\exp\left(-y^{(i)}\sum_{m=1}^{M}\beta^{[m]}\hat{b}^{[m]}(\mathbf{x}^{(i)})\right)}{\prod_{m=1}^{M} W^{[m]}} \\
&= \frac{w^{[1](i)}\exp(-y^{(i)}\hat{f}\left(\mathbf{x}^{(i)}\right))}{\prod_{m=1}^{M} W^{[m]}}. && \text{(Since } \textstyle\sum_{m=1}^{M}\beta^{[m]}\hat{b}^{[m]}(\mathbf{x}^{(i)}) = \hat{f}\left(\mathbf{x}^{(i)}\right)\text{)}
\end{aligned}
$$

(d) For any $(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y}$ it holds that

$$
\begin{aligned}
\hat{h}(\mathbf{x}) \neq y \quad &\Leftrightarrow \quad \mathrm{sign}(\hat{f}(\mathbf{x})) \neq y \\
&\Leftrightarrow \quad \hat{f}(\mathbf{x})y < 0 \\
&\Leftrightarrow \quad -\hat{f}(\mathbf{x})y > 0 \\
&\Leftrightarrow \quad \exp(-\hat{f}(\mathbf{x})y) > \exp(0) = 1 = \mathbb{1}_{[\hat{h}(\mathbf{x})\neq y]}.
\end{aligned}
$$

(e) We show the desired result by using (b), (c) and (d):

$$
\begin{aligned}
\frac{\mathcal{R}_{\text{emp}}(\hat{h})}{n} &= \frac{\sum_{i=1}^{n} \mathbb{1}_{[\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}]}}{n} \\
&= \sum_{i=1}^{n} \frac{1}{n} \mathbb{1}_{[\hat{h}(\mathbf{x}^{(i)}) \neq y^{(i)}]} \\
&\leq \sum_{i=1}^{n} \frac{1}{n} \exp\left(-y^{(i)} \hat{f}\left(\mathbf{x}^{(i)}\right)\right) && \text{(Using (d))} \\
&= \sum_{i=1}^{n} w^{[1](i)} \exp\left(-y^{(i)} \hat{f}\left(\mathbf{x}^{(i)}\right)\right) && \text{(Definition of } w^{[1]}\text{)} \\
&= \sum_{i=1}^{n} w^{[M+1](i)} \prod_{m=1}^{M} W^{[m]} && \text{(Using (c))} \\
&= \prod_{m=1}^{M} W^{[m]} \underbrace{\sum_{i=1}^{n} w^{[M+1](i)}}_{=1} \\
&\leq \prod_{m=1}^{M} \sqrt{1 - 4(\gamma^{[m]})^2}. && \text{(Using (b))}
\end{aligned}
$$