Analyzing the Training Error

- Theorem:

$$\begin{aligned}\widehat{\mathrm{err}}(H) &\le \prod_{t=1}^{T}\left[2\sqrt{\epsilon_t(1-\epsilon_t)}\right]\\ &= \prod_{t=1}^{T}\sqrt{1-4\gamma_t^2}\\ &\le \exp\left(-2\sum_{t=1}^{T}\gamma_t^2\right)\end{aligned}$$

- so: if $\forall t :\ \gamma_t \ge \gamma > 0$
  then $\widehat{\mathrm{err}}(H) \le e^{-2\gamma^2 T}$

Step 1

Claim: $D_{T+1}(i) = \dfrac{\exp(-y_i F(x_i))}{m \prod_t Z_t}$

where $F(x) = \sum_t \alpha_t h_t(x)$ [so $H(x) = \mathrm{sign}(F(x))$]

Proof:

- have $D_{t+1}(i) = D_t(i) \cdot \dfrac{e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$
- so:

$$
\begin{aligned}
D_{T+1}(i) &= \frac{1}{m} \times \frac{e^{-\alpha_1 y_i h_1(x_i)}}{Z_1} \times \cdots \times \frac{e^{-\alpha_T y_i h_T(x_i)}}{Z_T} \\
&= \frac{1}{m} \cdot \frac{\exp\left(-y_i \sum_t \alpha_t h_t(x_i)\right)}{\prod_t Z_t}
\end{aligned}
$$

Step 2

Claim: $\widehat{\text{err}}(H) \leq \prod_{t=1}^{T} Z_t$

Proof:

- $y \neq H(x) \Leftrightarrow y \neq \text{sign}(F(x)) \Leftrightarrow yF(x) \leq 0$
- so:

$$\begin{aligned}
\widehat{\text{err}}(H) &= \frac{1}{m} \sum_i \mathbf{1}\{y_i \neq H(x_i)\} \\
&= \frac{1}{m} \sum_i \mathbf{1}\{y_i F(x_i) \leq 0\} \\
&\leq \frac{1}{m} \sum_i \exp(-y_i F(x_i)) \\
&= \frac{1}{m} \underbrace{\sum_i D_{T+1}(i)}_{=1} \cdot m \prod_{t=1}^{T} Z_t
\end{aligned}$$

Step 3

Claim: $Z_t = 2\sqrt{\epsilon_t(1-\epsilon_t)}$

Proof:

$$
\begin{aligned}
Z_t &= \sum_i D_t(i)\exp(-\alpha_t\ y_i\ h_t(x_i)) \\
&= \underbrace{\sum_{i:y_i \neq h_t(x_i)} D_t(i)}_{=\epsilon_t}\ e^{\alpha_t} + \underbrace{\sum_{i:y_i = h_t(x_i)} D_t(i)}_{=1-\epsilon_t}\ e^{-\alpha_t} \\
&= \epsilon_t\ e^{\alpha_t} + (1-\epsilon_t)\ e^{-\alpha_t} \\
&= 2\sqrt{\epsilon_t(1-\epsilon_t)}
\end{aligned}
$$