In this lecture we consider the classification problem, i.e. $\mathcal{Y} = \{-1, +1\}$.

Consider a family of weak classifiers

$$\mathcal{H} = \{h \colon \mathcal{X} \to \{-1, +1\}\}.$$

Let the empirical minimizer be

$$h_0 = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^{n} I(h(X_i) \neq Y_i)$$

and assume its expected error,

$$\frac{1}{2} > \varepsilon = Error(h_0),\ \varepsilon > 0$$

Examples:

- $\mathcal{X} = \mathbb{R}^d$, $\mathcal{H} = \{\operatorname{sign}(wx + b) \colon w \in \mathbb{R}^d, b \in \mathbb{R}\}$
- Decision trees: restrict depth.
- Combination of simple classifiers:

$$f = \sum_{t=1}^{T} \alpha_t h_t(x),$$

where $h_t \in \mathcal{H}$, $\sum_{t=1}^{T} \alpha_t = 1$. For example,

$h_1 =$

1	-1
1	-1

, $h_2 =$

1	1
-1	-1

, $h_3 =$

1	1
1	1

$f = \frac{1}{7}(h_1 + 3h_2 + 3h_3) =$

7	5
1	-1

, $\operatorname{sign}(f) =$

1	1
1	-1

AdaBoost

Assign weight to training examples $w_1(i) = 1/n$.

for $t = 1..T$

1) find "good" classifier $h_t \in \mathcal{H}$; Error $\varepsilon_t = \sum_{i=1}^{n} w_t(i) I(h(X_i) \neq Y_i)$

2) update weight for each i:

$$w_{t+1}(i) = \frac{w_t(i) e^{-\alpha_t Y_i h_t(X_i)}}{Z_t}$$

$$Z_t = \sum_{i=1}^{n} w_t(i) e^{-\alpha_t Y_i h_t(X_i)}$$

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t} > 0$$

3) t = t+1

end

Output the final classifier: $f = \text{sign}(\sum \alpha_t h_t(x))$.

Theorem 2.1. *Let $\gamma_t = 1/2 - \varepsilon_t$ (how much better h_t is than tossing a coin). Then*

$$\frac{1}{n}\sum_{i=1}^{n} I(f(X_i) \neq Y_i) \leq \prod_{t=1}^{T} \sqrt{1 - 4\gamma_t^2}$$

Proof.

$$I(f(X_i) \neq Y_i) = I(Y_i f(X_i) = -1) = I(Y_i \sum_{t=1}^{T} \alpha_t h_t(X_i) \leq 0) \leq e^{-Y_i \sum_{t=1}^{T} \alpha_t h_t(X_i)}$$

Consider how weight of example i changes:

$$\begin{aligned}
w_{T+1}(i) &= \frac{w_T(i) e^{-Y_i \alpha_T h_T(X_i)}}{Z_t} \\
&= \frac{e^{-Y_i \alpha_T h_T(X_i)}}{Z_t} \frac{w_{T-1}(i) e^{-Y_i \alpha_{T-1} h_{T-1}(X_i)}}{Z_{T-1}} \\
&\dots \\
&= \frac{e^{-Y_i \sum_{t=1}^{T} \alpha_t h_t(X_i)}}{\prod_{t=1}^{t} Z_t} \frac{1}{n}
\end{aligned}$$

Hence,

$$w_{T+1}(i) \prod Z_t = \frac{1}{n} e^{-Y_i \sum_{t=1}^{T} \alpha_t h_t(X_i)}$$

and therefore

$$\frac{1}{n}\sum_{i=1}^{n} I(f(X_i) \neq Y_i) \leq \frac{1}{n}\sum_{i=1}^{n} e^{-Y_i \sum_{t=1}^{T} \alpha_t h_t(X_i)} = \prod_{t=1}^{T} Z_t \sum_{i=1}^{n} w_{T+1}(i) = \prod_{t=1}^{T} Z_t$$

$$\begin{aligned}
Z_t &= \sum w_t(i) e^{-\alpha_t Y_i h_t(X_i)} \\
&= \sum_{i=1}^{n} w_t(i) e^{-\alpha_t} I(h_t(X_i) = Y_i) + \sum_{i=1}^{n} w_t(i) e^{+\alpha_t} I(h_t(X_i) \neq Y_i) \\
&= e^{+\alpha_t} \sum_{i=1}^{n} w_t(i) I(h_t(X_i) \neq Y_i) + e^{-\alpha_t} \sum_{i=1}^{n} w_t(i)(1 - I(h_t(X_i) \neq Y_i)) \\
&= e^{\alpha_t} \varepsilon_t + e^{-\alpha_t}(1 - \varepsilon_t)
\end{aligned}$$

Minimize over α_t to get

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \varepsilon_t}{\varepsilon_t}$$

and

$$e^{\alpha_t} = \left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)^{1/2}.$$

Finally,

$$
\begin{aligned}
Z_t &= \left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)^{1/2}\varepsilon_t + \left(\frac{\varepsilon_t}{1-\varepsilon_t}\right)^{1/2}(1-\varepsilon_t) \\
&= 2(\varepsilon_t(1-\varepsilon_t))^{1/2} = 2\sqrt{(1/2-\gamma_t)(1/2+\gamma_t)} \\
&= \sqrt{1-4\gamma_t^2}
\end{aligned}
$$

□