Weighted Majority Voting (WMV)

$$p_t = 1 - \varepsilon_t$$

Recall from Adaboost

$$H(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \qquad h_t(x) \in \{\pm 1\}$$

error performance

$$\alpha_t = \frac{1}{2}\log\frac{1-\varepsilon_t}{\varepsilon_t} = \frac{1}{2}\log\frac{p_t}{1-p_t}$$

$H(x)$ is a weighted majority vote with the T classifiers. Why do we choose the weights in this way?

- → Boosting derived the weights
- → Is there another way to look at the WMV?

Thm: Consider an ensemble of T independent classifiers with performances $p_1, p_2 \ldots p_T$. The outputs are a weighted vote. The accuracy of the ensemble is maximized when

$$w_i \propto \boxed{\log\frac{p_i}{1-p_i}} \qquad j^* = \arg\max_j \sum_{i=1}^{T} w_i c_{ij}$$

Proof Denote $s = [s_1, \dots, s_T]$ as a vector of output labels of each classifier. The Bayes optimal decision function is given by

$p(s_1, s_2, \dots, s_T \mid \omega_j)$

\# of classes

j^{th}-class

$$g_j(x) = \log \{ P(\omega_j)\, P(s \mid \omega_j) \} \qquad j = 1, \dots, c$$

$$= \log \{ P(\omega_j) \prod_{i=1}^{T} p(s_i \mid \omega_j) \}$$

$$= \log \{ P(\omega_j) \} + \boxed{\log \{ \prod_{i=1}^{T} p(s_i \mid \omega_j) \}}$$

$$= \log \{ P(\omega_j) \} + \log \{ \prod_{i: s_i = \omega_j} p(s_i \mid \omega_j) \prod_{i: s_i \neq \omega_j} p(s_i \mid \omega_j) \}$$

$$= \log \{ P(\omega_j) \} + \log \{ \underbrace{\prod_{i: s_i = \omega_j} p_i}_{6} \cdot \frac{1-p_i}{1-p_i} \underbrace{\prod_{i: s_i \neq \omega_j} (1-p_i)}_{\cancel{4}\ 10 = T} \}$$

$$= \log \{ P(\omega_j) \} + \log \{ \prod_{i: s_i = \omega_j} \frac{p_i}{1-p_i} \prod_{i=1}^{T} (1-p_i) \}$$

$$= \log \{ P(\omega_j) \} + \sum_{i: s_i = \omega_j} \log \frac{p_i}{1-p_i} + \sum_{i=1}^{T} \log \frac{p_i}{1-p_i}$$