Advanced AdaBoost.

Variant of AdaBoost (Viola & Jones)

Strong classifier:
\[ H_n(x) = \sum_{\mu=1}^{m} \left[ \alpha_\mu h_\mu(x) + \beta_\mu \right] \]

\[ h_\mu(x) - \text{weak classifier} \]

Modify the update rule:

\[ D_{t+1}(i) = \frac{1}{Z_t} D_t(i) e^{-\omega_i \left( \alpha_t h_\mu(x) + \beta_t \right)} \]

Let \( W_pq \) be the sum of the weights if the weak class is \( p \) and the true class is \( q \).

Pick weak classifier to minimize

\[ 2 \left( \sqrt{w_+ w_-} + \sqrt{w_- w_+} \right) \]

Set
\[ \alpha_t - \beta_t = \log \frac{\sqrt{w_+} + \sqrt{w_-}}{\sqrt{w_-} + \sqrt{w_+}} \]

Proof - natural extension of basic proof.
(2) We may want to penalize false positives and false negatives in a different way. For example, detecting faces in images.

In a typical image, there will far more non-faces than faces. So we want to be very certain before labelling a window as a face.

**Loss function:**

\[
L = \begin{cases} 
\sqrt{|k|}, & \text{if } \omega_i = 1, \text{ and } H(x_i) = -1 \\
\sqrt{|k|}, & \text{if } \omega_i = -1, \text{ and } H(x_i) = 1 \\
0, & \text{otherwise} 
\end{cases}
\]

**Modify the update rule:**

\[
D_{t+1}(c) = \frac{1}{Z_t} e^{-\omega_i (d_{tk} + c_k + \beta_t)} e^{\omega_i \log |k|}
\]

**Verify that the weighted loss is bounded.**

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \sqrt{|k|} S_{w,1} S_{H(x_i),1} + (\sqrt{|k|}) S_{w,1} S_{H(x_i),1} \right)
\leq \frac{1}{N} \sum_{i=1}^{N} e^{-\omega_i \sum_{k=1}^{K} (d_{tk} + c_k + \beta_t)} e^{\omega_i \log |k|}
\]
(3) This modifies the update rule by

\[ \beta_t \rightarrow \beta_t + (K_t \log \sqrt{n} ) \]

Cascade.

Motivation
\[ P(x_i|\alpha) = \frac{e^{w_i^* \alpha}}{e^{w_i^* \alpha} + e^{-w_i^* \alpha}} \]

Formulate a log-likelihood:

\[ l(\lambda, \phi) = \log \prod_{i=1}^{m} P(x_i|\alpha) \]

\[ = -\frac{1}{2} \sum_{i=1}^{m} \log (1 + e^{-2w_i \lambda \phi(x_i)}) \]

This requires solving the optimization problem:

\[ (\lambda, \phi)^* = \arg \max l(\lambda, \phi) \]

\[ = \arg \min \sum_{i=1}^{m} \log (1 + e^{-2w_i \lambda \phi(x_i)}) \]

This differs from AdaBoost criteria:

\[ (\lambda, \phi)^* = \arg \min \sum_{i=1}^{m} \exp \left(-w_i \lambda \phi(x_i)\right) \]

Claim — they become similar as \( m \rightarrow \infty \).

Replace sample mean by expectation:

\[ \text{AdaBoost} \quad (\lambda, \phi)^* = \lambda \log E_{p(x)} \log (1 + e^{-2\lambda \phi(x)}) \]

\[ \text{LogitBoost} \quad (\lambda, \phi)^* = \lambda \log E_{p(x)} \left( e^{-\lambda \phi(x)} \right) \]
Claim: these problem have the same
mean at

\[ F(x) = \sum_{(\omega, y)} \theta(\omega) q(\omega) = \frac{1}{Z} \sum_{(\omega, y)} \log \frac{p(\omega=1|x)}{p(\omega=-1|x)} \]

Proof: \[ A(F) = \frac{1}{Z} \int P(x, y) \log \left( \frac{1}{2} e^{-2y F(s)} \right) ds \]

Since \[ A(F) = 0 \]

\[ p(x, y=+1) \frac{2}{1 + e^{-2F(x)}} + p(x, y=-1) \frac{2}{1 + e^{2F(x)}} = 0 \]

Solving gives:

\[ F^*(x) = \frac{1}{Z} \log \frac{p(\omega=+1|x)}{p(\omega=-1|x)} \]

Viola & Jones paper.