Note Title 11/25/2006

# Advanced AdaBoost.

Variant of AdaBoost. (Viola & Jones)

Strong classifier: $H_M(x) = \sum_{\mu=1}^{M} \{\alpha_\mu h_\mu(x) + \beta_\mu\}$

$h_\mu(x)$ - weak classifier.

Modify the update rule:

$$D_{t+1}(i) = \frac{1}{Z_t} D_t(i)\, e^{-\omega_i\{\alpha_t h_t(x) + \beta_t\}}$$

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

$$2\{\sqrt{W_{++}W_{+-}} + \sqrt{W_{-+}W_{--}}\}$$

set $\alpha_t + \beta_t = \log \sqrt{W_{++}/W_{+-}}$

$\alpha_t - \beta_t = \log \sqrt{W_{--}/W_{-+}}$

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:

$L = \sqrt{k}$, if $\omega_i = 1$, and $H(x_i) = -1$

$L = 1/\sqrt{k}$, if $\omega_i = -1$, and $H(x_i) = 1$.

$L = 0$, otherwise.

Modify the update rule:

$$D_{t+1}(i) = \frac{1}{Z_t} D_t e^{-\omega_i(\alpha_t h_t(x_i)+\beta_t)} e^{1/N\, \omega_i \log\sqrt{k}}$$

Verify that the weighted loss is bounded.

$$(1/N)\sum_{i=1}^{N}\{\sqrt{k}\,\delta_{\omega_i,1}\,\delta_{H(x_i),-1} + (1/\sqrt{k})\,\delta_{\omega_i,-1}\,\delta_{H(x_i),1}\}$$
$$\leq \frac{1}{N}\sum_{i=1}^{N} e^{-\omega_i\sum_{\mu=1}^{n}(\alpha_\mu h_\mu(x_i)+\beta_\mu)} e^{\omega_i \log\sqrt{k}}$$

(3) This modifies the update rule by

$$\beta_t \to \beta_t + (1/\lambda) \log \sqrt{kl} \quad //$$

<u>Cascade.</u>

Motivation

(4) <u>LogitBoost.</u>

$$p(\omega_i|x_i) = \frac{e^{\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)}}{e^{\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)} + e^{-\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)}} = \frac{1}{1+e^{-2\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)}}$$

Formulate a log-likelihood

$$l(\lambda, \varphi) = \log \prod_{i=1}^{m} p(\omega_i|x_i)$$

$$= -\sum_{i=1}^{m} \log(1+e^{-2\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)})$$

This requires solving the optimization problem,

$$(\lambda,\varphi)^* = \text{ARG MAX } l(\lambda,\varphi)$$

$$= \text{ARG MIN} \sum_{i=1}^{m} \log\{1+e^{-2\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)}\}$$

This differs from AdaBoost criteria

$$(\lambda,\varphi)^* = \text{ARG MIN} \sum_{i=1}^{m} \exp\{-\omega_i \underline{\lambda}.\underline{\varphi}(\underline{x}_i)\}$$

<u>Claim</u> — these become similar as $m \to \infty$.

Replace sample mean by expectation, giving

AdaBoost $(\lambda,\varphi)^* = \text{ARG MIX } E_{p(x,y)}\{\log(1+e^{-2\omega \langle\lambda,h(x)\rangle})\}$

LogitBoost $(\lambda,\varphi)^* = \text{ARG MIN } E_{p(x,y)}\{e^{-\omega \underline{\lambda}.\underline{\varphi}(x)}\}$

(5) Claim, these problem have the same minimum at

$$F(x) = \underline{\lambda}^* \cdot \underline{\varphi}^*(x) = \frac{1}{2} \log \frac{p(\omega=+1|x)}{p(\omega=-1|x)}$$

Proof. $A(F) = \sum_{\omega \in \{\pm 1\}} \int p(x,y) \log(1+e^{-2yF(s)}) ds$

$$\frac{\delta A(F)}{\delta F} = 0$$

$$\Rightarrow p(x,y=+1) \frac{-2e^{-F(x)}}{1+e^{-2F(x)}} + p(x,y=-1) \frac{2e^{F(x)}}{1+e^{2F(x)}} = 0 \quad \forall x$$

solving gives:

$$F^*(x) = \frac{1}{2} \log \frac{p(x,\omega=+1)}{p(x,\omega=-1)} = \frac{1}{2} \log \frac{p(\omega=1|x)}{p(\omega=1|x=-1)}$$

Viola & Jones paper.