

Adaboost.M1
exp-loss forward
stage-wise modelling

[derivation "stolen/borrowed" from STRIKIN Y300
by Riccardo De Bin]

$$Y = \{-1, 1\}$$

$$X \in \mathbb{R}^P$$

exponential loss



0) Loss-function $L(y, f(x)) = \exp(-y f(x))$

$$\sum_{i=1}^N L(y_i, f(x_i))$$

Forward stage-wise modelling:

$$\arg \min_{\beta, b} \sum_{i=1}^N L(y_i, \underbrace{\hat{f}^{(m-1)}(x_i)}_{\text{current classifier - not to change}} + \underbrace{\beta b(x_i, \delta)}_{\text{to be changed}})$$

1) What does this look like at step m ?

$$(\beta_m, b_m) = \arg \min_{\beta, b} \sum_{i=1}^N \exp \left(-y_i \sum_{k=1}^m \beta^{(k)} b^{(k)}(x_i) \right)$$

$$= \arg \min_{\beta, b} \sum_{i=1}^N \exp \left\{ -y_i \left(\sum_{k=1}^{m-1} \beta^{(k)} \hat{f}^{(k-1)}(x_i) + \beta b(x_i) \right) \right\}$$

$$= \arg \min_{\beta, b} \sum_{i=1}^N w_i^{(m-1)} \cdot \exp(-y_i \beta b(x_i))$$

$$w_i^{(m-1)} = \exp \left\{ -y_i \hat{f}^{(m-1)}(x_i) \right\}$$

known - from last iteration -

Observe: if $y_i \neq b(x_i)$ then $y_i b(x_i) = +1$
 \neq $\div 1$

Next: two step procedure

where a) minimize wrt b
 b) minimize wrt β

$$2) \quad b_m = \underset{f}{\operatorname{argmin}} \sum_{i=1}^N w_i^{(m-1)} \cdot \exp(-y_i \beta b(x_i))$$

$$= \underset{f}{\operatorname{argmin}} \left\{ \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} \cdot e^{-\beta} + \sum_{i: y_i = b(x_i)} w_i^{(m-1)} e^{\beta} \right\}$$

add and subtract

$$\sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} e^{-\beta}$$

$$= \underset{b}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \underbrace{w_i^{(m-1)} e^{-\beta}}_{\text{no } b \text{ here}} + (e^{\beta} - e^{-\beta}) \sum_{i: y_i = b(x_i)} \underbrace{w_i^{(m-1)}}_{\text{here}} \right\}$$

$$\hat{b}_m = \underset{f}{\operatorname{argmin}} \left\{ \sum_{i=1}^N w_i^{(m-1)} I(y_i \neq b(x_i)) \right\} \quad I(y_i \neq b(x_i))$$

$$3) \quad \beta^{(m)} = \underset{\beta}{\operatorname{argmin}} \quad \sum_{i=1}^N w_i^{(m-1)} \exp \{-y_i \beta b(x_i)\}$$

$$y_i = b(x_i) \rightarrow \sum_{i: y_i = b(x_i)} w_i^{(m-1)} e^{-\beta}$$

$$y_i \neq b(x_i) \rightarrow \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} e^{\beta}$$

$$L = \sum_{i: y_i = b(x_i)} w_i^{(m-1)} e^{-\beta} + \sum_{i: y_i \neq b(x_i)} w_i e^{\beta}$$

$$\frac{\partial L}{\partial \beta} = - \sum_{i: y_i = b(x_i)} w_i^{(m-1)} e^{-\beta} + \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} e^{\beta} = 0$$

multiply e^{β}

$$- \sum_{i: y_i = b(x_i)} w_i^{(m-1)} + e^{2\beta} \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} = 0$$

$$e^{2\beta} \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)} = \sum_{i=1}^N w_i^{(m-1)} - \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)}$$

$$e^{2\beta} = \frac{\sum_{i=1}^N w_i^{(m-1)} - \sum_{i: y_i \neq b(x_i)} w_i^{(m-1)}}{\sum_{i: y_i \neq b(x_i)} w_i^{(m-1)}}$$

$$\sum_{i: y_i \neq b(x_i)} w_i^{(m-1)}$$

scale each term with $\sum_{i=1}^N w_i^{(m-1)}$ to get

$$e^{2\beta} = \frac{1 - \text{err}_m}{\text{err}_m}$$

$$2\hat{\beta}_m = \ln \left(\frac{1 - \text{err}_m}{\text{err}_m} \right)$$

$$\text{where } \text{err}_m = \frac{\sum_{i: y_i \neq b(x_i)} w_i^{(m-1)}}{\sum_{i=1}^N w_i^{(m-1)}}$$

In the Adaboost.11: $\alpha_m = 2\beta_m$

$$\hat{\alpha}_m = \ln \left(\frac{1 - \text{err}_m}{\text{err}_m} \right)$$

4) In 2) we found that \hat{b}_m is

minimizing the "weighed misclassification error" and in 3) we found

$$\hat{\beta}_m = \frac{1}{2} \ln \left(\frac{1 - \text{err}_m}{\text{err}_m} \right), \text{ or } \hat{\alpha}_m = \ln \left(\frac{1 - \text{err}_m}{\text{err}_m} \right)$$

This gives

Our classifier is updated as

$$\hat{f}^{(m)}(x_i) = \hat{f}^{(m-1)}(x_i) + \hat{\beta}_m \cdot \hat{b}_m(x_i)$$

and the weights (from 1)

$$w_i^{(m)} = w_i^{(m-1)} \exp(-y_i \hat{\beta}_m \hat{b}_m(x_i))$$

But in the algorithm the weight update is written differently " $w_i \leftarrow w_i \cdot \exp(\alpha_m \cdot I(y_i \neq \hat{b}_m(x_i)))$ "

Final "adjustment", look at exponent:

$$\begin{aligned} -y_i \cdot \hat{b}_m(x_i) &= -I(y_i = \hat{b}_m(x_i)) + I(y_i \neq \hat{b}_m(x_i)) \\ &= -I(y_i = \hat{b}_m(x_i)) + I(y_i \neq \hat{b}_m(x_i)) - I(y_i \neq \hat{b}_m(x_i)) + I(y_i \neq \hat{b}_m(x_i)) \\ &= -1 + 2 \cdot I(y_i \neq \hat{b}_m(x_i)) \end{aligned}$$

Insert into:

$$w_i^{(m)} = w_i^{(m-1)} \cdot \exp(\hat{\beta}_m (-1 + 2I(y_i \neq \hat{b}_m(x_i))))$$

$$\begin{aligned}
&= w_i^{(m-1)} \exp\left(\overbrace{2\hat{\beta}_m}^{\hat{\alpha}_m} \cdot \mathbb{I}(y_i \neq \hat{b}_m(x_i)) - \hat{\beta}_m\right) \\
&= w_i^{(m-1)} \exp\left(\hat{\alpha}_m \cdot \mathbb{I}(y_i \neq \hat{b}_m(x_i))\right) \cdot \exp\left(-\frac{\hat{\alpha}_m}{2}\right)
\end{aligned}$$

an extra factor = the same value
 is multiplied is for each
 weight w_i = a scaling