Algorithm 10.1 *AdaBoost.M1.*

1. Initialize the observation weights $w_i = 1/N$, $i = 1, 2, \ldots, N$.
2. For $m = 1$ to M:
   (a) Fit a classifier $G_m(x)$ to the training data using weights w_i.
   (b) Compute
   $$\text{err}_m = \frac{\sum_{i=1}^N w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i}.$$
   (c) Compute $\alpha_m = \log((1 - \text{err}_m)/\text{err}_m)$.
   (d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))]$, $i = 1, 2, \ldots, N$.
3. Output $G(x) = \text{sign}\left[\sum_{m=1}^M \alpha_m G_m(x)\right]$.

a

Describe the original idea behind the AdaBoost algorithm.

b-c-d

Show that the AdaBoost algorithm reported above can be interpreted as a forward stagewise modelling procedure which minimizes the loss function $L(y, f(x)) = \exp\{-yf(x)\}$. Following this interpretation, the current estimate $f_{m-1}(x)$ is updated by adding the step-specific result of the classifier $G_m(x_i)$ to produce a new estimate $f_m(x)$. In particular, at each step m one must find G_m and α_m such that

$$(\alpha_m, G_m) = \text{argmin}_{\alpha,G} \sum_{i=1}^N \exp\{-y_i[f_{m-1}(x_i) + \frac{\alpha}{2} G(x_i)]\}.$$

Show that:

b

$$G_m(x) = \text{argmin}_G \sum_{i=1}^N w_i^{(m)} I(y_i \neq G(x_i)),$$

where $w_i^{(m)} = \exp\{-y_i f_{m-1}(x_i)\}$;

c

$$\alpha_m = \log\frac{1 - \text{err}_m}{\text{err}_m},$$

where $\text{err}_m = \frac{\sum_{i=1}^N w_i^{(m)} I(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i^{(m)}}$;

d

$$w_i^{(m+1)} \propto w_i^{(m)} \exp\{\alpha_m I(y_i \neq G_m(x_i))\}.$$