Exercise 1 (Ada Boost) Let $(x_i, y_i)_{1\leqslant i\leqslant n} \in (\mathsf{X}\times\{-1,1\})^n$ be n observations and $\mathsf{H} = \{h_1,\dots,h_M\}$ be a set of M (weak) classifiers, i.e. for all $1\leqslant i\leqslant M$, : $h_i : \mathsf{X}\to\{-1,1\}$. It is assumed that for each $h\in\mathsf{H}$, $-h\in\mathsf{H}$ and that there exist $1\leqslant i\neq j\leqslant n$ such that $y_i = h(x_i)$ and $y_j\neq h(x_j)$. Let F be the set of all linear combinations of elements of H :

$$\mathsf{F} = \left\{\sum_{j=1}^{M}\beta_j h_j\,;\,\beta\in\mathbb{R}^M\right\}\;.$$

Consider the following algorithm. Set $\hat{f}_0 = 0$ and for all $1\leqslant m\leqslant M$,

$$\hat{f}_m = \hat{f}_{m-1} + \beta_m h_m \quad \text{where} \quad (\beta_m, h_m) = \underset{h\in\mathsf{H}\,,\,\beta\in\mathbb{R}}{\operatorname{argmin}}\; n^{-1}\sum_{i=1}^{n}\exp\left\{-y_i\left(\hat{f}_{m-1}(x_i) + \beta h(x_i)\right)\right\}\;.$$

Le classifieur final sera $\varphi(x) = \operatorname{sgn}\hat{f}_M(x)$.

1. Choosing $\omega_m^i = n^{-1}\exp\{-y_i\hat{f}_{m-1}(x_i)\}$, show that

$$n^{-1}\sum_{i=1}^{n}\exp\left\{-y_i\left(\hat{f}_{m-1}(x_i)+\beta h(x_i)\right)\right\} = \left(\mathrm{e}^{\beta}-\mathrm{e}^{-\beta}\right)\sum_{i=1}^{n}\omega_m^i\mathbb{1}_{h(x_i)\neq y_i} + \mathrm{e}^{-\beta}\sum_{i=1}^{n}\omega_m^i\;.$$

2. For all $1\leqslant m\leqslant M$ and $h\in\mathsf{H}$, define

$$\epsilon_m(h) = \sum_{i=1}^{n}D_m^i\mathbb{1}_{h(x_i)\neq y_i} \quad \textit{where} \quad D_m^i = \frac{\omega_m^i}{\sum_{j=1}^{n}\omega_m^j}$$

   Prove that setting

$$h_m \in \underset{h\in\mathsf{H}}{\operatorname{argmin}}\;\epsilon_m(h) \quad \text{and} \quad \beta_m = \frac{1}{2}\log\left(\frac{1-\epsilon_m(h_m)}{\epsilon_m(h_m)}\right).$$

   is licit.

3. Propose an algorithm to compute $\hat{f}_M$.