PC6. Ecole Polytechnique. MAP 569. Machine Learning II.

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 $\mathsf{H} = \{h_1,\ldots,h_M\}$ be a set of M classifiers, i.e. for all $1\leqslant i\leqslant M$, : $h_i : \mathsf{X}\to\{-1,1\}$. It is assumed that for each $h\in\mathsf{H}$ 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:

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

Consider the following algorithm. Set $\hat f_0 = 0$ and for all $1\leqslant m\leqslant M$,

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

1. Choosing $\omega_i^m = n^{-1}\exp\{-y_i\hat f_{m-1}(x_i)\}$, show that

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

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

$$\mathrm{err}_m(h) = \frac{\sum_{i=1}^{n}\omega_i^m\mathbb{1}_{h(x_i)\neq y_i}}{\sum_{i=1}^{n}\omega_i^m}\,.$$

   Prove that

$$h_{j_m} = \underset{h\in\mathsf{H}}{\operatorname{argmin}}\;\mathrm{err}_m(h)\quad\text{and}\quad\beta_m = \frac{1}{2}\log\left(\frac{1-\mathrm{err}_m(h_{j_m})}{\mathrm{err}_m(h_{j_m})}\right)\,.$$

3. Propose an algorithm to compute $\hat f_M$.

Exercise 2 (**Consistency of a simple random forest**) Consider a data set $\mathcal{D}_n = \{(X_i, Y_i)\in[0,1]^d\times\mathbb{R}, i = 1,\ldots,n\}$. It is assumed that the (X_i, Y_i) are i.i.d. with the same distribution as (X, Y) where

$$Y = r(X) + \varepsilon,$$

with ε a centered Gaussian noise, independent of X and r a continuous function. Define the following centered random forest estimator:

1. Grow M trees as follows:
   (a) Consider the cell $[0,1]^d$.
   (b) Select uniformly one variable $j^\star$ in $\{1,\ldots,d\}$.
   (c) Cut the cell at the middle of the $j^\star$-th side, where $j^\star$ is the coordinate chosen above.
   (d) For each of the two resulting cells, repeat $(b)-(c)$ if the cell has been cut strictly less than k_n times.
   (e) For a query point x, the m-th tree outputs the average $\hat r_n(x,\Theta_m)$ of the Y_i falling into the same cell as x, where Θ_m is the random variable encoding all selected splitting variables in each cell of the m-th tree.
2. For a query point x, the centered forest outputs the average $\hat r_{M,n}(x,\Theta_1,\ldots,\Theta_M)$ of the predictions given by the M trees.

Define the infinite random forest estimate $\hat r_{\infty,n}$ by considering the random forest estimate defined above and letting $M\to\infty$, that is

$$\hat r_{\infty,n}(x) = \mathbb{E}_\Theta[\hat r_n(x,\Theta)],$$

where $\mathbb{E}_\Theta$ is the expectation with respect to Θ only. For a tree built with the randomness Θ, we let $A_n(x,\Theta)$ be the cell containing x and $N_n(x,\Theta)$ be the number of observations falling into $A_n(x,\Theta)$. We want to prove the following theorem:

Theorem 1. *Assume that $k_n \to \infty$ is such that $2^{k_n}/n \to 0$, as $n \to \infty$. Then the random forest fulfills $\mathbb{E}[(\hat{r}_{\infty,n}(X) - r(X))^2] \to 0$, where X is independent of $(X_i, Y_i)_{i=1,\dots,n}$ with the same distribution as the X_i on $[0,1]^d$.*

1. Prove that there exists weights $W_{ni}(x,\Theta)$ and $W_{ni}^{\infty}(x)$, $1 \leqslant i \leqslant n$, such that

$$\hat{r}_n(x,\Theta) = \sum_{i=1}^{n} W_{ni}(x,\Theta) Y_i, \quad \text{and} \quad \hat{r}_{\infty,n}(x) = \sum_{i=1}^{n} W_{ni}^{\infty}(x) Y_i.$$

In this context, Stone's Theorem states that the random tree estimate $\hat{r}_n(x,\Theta)$ fulfills

$$\lim_{n\to\infty} \mathbb{E}\left[(\hat{r}_n(X,\Theta) - r(X))^2\right] = 0,$$

as soon as the two following conditions are satisfied

(i) $\mathbb{E}[\mathrm{diam}(A_n(X,\Theta))] \to 0$, as $n \to \infty$, where the diameter of any cell A is defined as

$$\mathrm{diam}(A) = \sup_{x,z\in A} \|x - z\|_2.$$

(ii) $N_n(X,\Theta) \to \infty$ in probability, as $n \to \infty$.

2. Let $x \in [0,1]^d$. What is the distribution of the number of cuts along the coordinate $j \in \{1,\dots,d\}$ in the cell $A_n(x,\Theta)$?

3. Check that, for all $x \in [0,1]^d$ and $j \in \{1,\dots,d\}$,

$$\mathbb{E}\left[\sup_{z\in A_n(x,\Theta)} z_j - \inf_{z\in A_n(x,\Theta)} z_j\right] = \left(1 - \frac{1}{2d}\right)^{k_n}.$$

4. Prove that (i) holds for a random centered tree.

5. We denote by $A_1,\dots,A_{2^{k_n}}$ the 2^{k_n} cells and by N_ℓ the number of points among $X, X_1,\dots,X_n$ which fall into A_ℓ. Show that for $\ell \in \{1,\dots,2^{k_n}\}$,

$$\mathbb{P}\left(X \in A_\ell | N_\ell, \Theta\right) = \frac{N_\ell}{n+1}.$$

Conclude that for every integer $t > 0$,

$$\mathbb{P}\left(N_n(X,\Theta) \leqslant t\right) \leqslant t 2^{k_n}/(n+1)$$

and hence that (ii) holds for a random centered tree.

6. Prove that the infinite centered random forest fulfills $\mathbb{E}[(\hat{r}_{\infty,n}(X) - r(X))^2] \to 0, \text{ as } n \to \infty$.

7. Assume that the noise ε is Gaussian with variance $\sigma^2 > 0$. Thus,

$$\mathbb{E}\left[\max_{1\leqslant i\leqslant n} \varepsilon_i^2\right] \leqslant \sigma^2(1 + 4\log n).$$

Find a condition on the number M_n of trees such that the finite centered random forest fulfills

$$\lim_{n\to\infty} \mathbb{E}[(\hat{r}_{M_n,n}(X,\Theta_1,\dots,\Theta_{M_n}) - r(X))^2] = 0.$$