Random feature-based double Vovk-Azoury-Warmuth algorithm for online multi-kernel learning

D.B. Rokhlin, O.V. Gurtovaya
Southern Federal University, Rostov-on-Don

We introduce a multi-kernel learning algorithm, VAW^2, for the online least squares regression problem in reproducing kernel Hilbert spaces (RKHS). The Vovk-Azoury-Warmuth (VAW) algorithm [1, Section 11.8] is used to construct expert strategies from random features generated for each kernel at the first level, and is then used again to combine their predictions at the second level.

Let $\mathcal{H}_j$ be RKHSs with the translation invariant reproducing kernels: $k_j(x,y) = \kappa_j(x-y)$, $\kappa_j(z) = \int_{\mathbb{R}^d} q_j(\omega)\cos\langle\omega, z\rangle\, d\omega$, where q_j are some a probability density functions. To each such kernel we associate a vector $\Phi_{\theta_j}(x) = (\sqrt{2}\cos(\langle\omega_{j,k}, x\rangle + b_{j,k}))_{k=1}^m$, $\theta_j = (\omega_{j,k}, b_{j,k})_{k=1}^m$ of random Fourier features [2], where $\omega_{j,k} \sim q_j$, $b_{j,k} \sim U(0, 2\pi)$ are i.i.d. random variables. Let $(x_t, y_t) \in \mathbb{R}^d \times \mathbb{R}$, $|y_t| \le Y$, be an arbitrary sequence. Denote by $B_R(\mathcal{H}_j)$ the R-ball in $\mathcal{H}_j$.

Theorem 1 *Let $w_{t,j} \in \mathbb{R}^m$ be generated by the VAW algorithms applied to $(\Phi_{\theta_j}(x_t), y_t)$, and $\alpha_t \in \mathbb{R}^N$ be generated by the VAW algorithm applied to (z_t, y_t), where z_t is the vector of expert predictions: $z_t = (\langle w_{t,1}, \Phi_{\theta_1}(x_t)\rangle, \ldots, \langle w_{t,N}, \Phi_{\theta_N}(x_t)\rangle)$. Then*

$$\mathsf{E}\sum_{t=1}^{T}(\langle\alpha_t, z_t\rangle - y_t)^2 \le \min_{1\le j\le N}\inf_{f_j\in B_R(\mathcal{H}_j)}\sum_{t=1}^{T}(y_t - f_j(x_t))^2 + O\left((R^2 + Y^2\ln T)\sqrt{T}\right), \quad \textit{provided} \quad m \propto \sqrt{T},$$

REFERENCES

1. Cesa-Bianchi, N., Lugosi, G. Prediction, learning, and games. Cambridge University Press, 2006.
2. Rahimi, A., Recht, B. Random features for large-scale kernel machines//Advances in neural information processing systems, vol. 20, pp. 1177–1184, 2007.