Discounted VAW and deterministic approximations: dynamic regret bounds in RKHS

D.B. Rokhlin, G.A. Karapetyants
Southern Federal University, Rostov-on-Don

We consider online regression with square loss in an RKHS $\mathcal{H}$ on a compact set X. Prediction quality is measured by dynamic regret with respect to a comparator sequence $f_{1:T} = (f_1, \ldots, f_T)$:

$$R_T(f_{1:T}) = \sum_{t=1}^{T} \frac{1}{2}(\widehat{y}_t - y_t)^2 - \sum_{t=1}^{T} \frac{1}{2}(f_t(x_t) - y_t)^2.$$

To approximate elements of the RKHS, we use finite-dimensional subspaces $V_m \subset \mathcal{H}$ with an orthonormal basis $g_1, \ldots, g_m$. On the feature vectors $(g_1(x), \ldots, g_m(x))$ we run the VE-DVAW algorithm [1], where discounted Vovk–Azoury–Warmuth (DVAW) forecasters [2] with different discount factors are aggregated by VAW.

Theorem 1 *Let $\mathcal{H}$ be an RKHS with kernel k, and let $\mathcal{E}(V_m) = \sup_{x \in X} \|(I - \Pi_{V_m})k(\cdot, x)\|_{\mathcal{H}}$. Assume that $|y_t|, |\widetilde{y}_t| \le Y$, $\|f_t\|_{\mathcal{H}} \le R$, $m = O(T)$, and VE-DVAW is run on the feature vectors $(g_1(x_t), \ldots, g_m(x_t))$. Then*

$$R_T^{(m)}(f_{1:T}) = O\left((\ln T)^2 + \sqrt{mT\, P_T^{\mathcal{H}}(f_{1:T})} + m \ln\left(1 + \frac{T}{m}\right) + T\mathcal{E}(V_m) \right),$$

where $P_T^{\mathcal{H}}(f_{1:T}) = \sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_{\mathcal{H}}$.

For the Gaussian kernel, truncating the Taylor expansion of the exponential gives $\mathcal{E}(V_m) \le C_1 \exp(-C_2 m^{1/d})$. With a suitable choice of $m = m(T)$, this yields

$$R_T(f_{1:T}) = O\left(\sqrt{T\, P_T^{\mathcal{H}}(f_{1:T})}\, (\ln T)^{d/2} + (\ln T)^{d+1} \right).$$

If $\mathcal{E}(V_m) \le C m^{-\beta}$, the optimal dimension depends on the unknown $P_T^{\mathcal{H}}(f_{1:T})$. Therefore, we use an additional VAW aggregation over a grid of dimensions m, which gives

$$R_T(f_{1:T}) = O\left(T^{\frac{\beta+1}{2\beta+1}} \left(P_T^{\mathcal{H}}(f_{1:T})\right)^{\frac{\beta}{2\beta+1}} + T^{\frac{1}{\beta+1}} (\ln T)^{\frac{\beta}{\beta+1}} \right).$$

REFERENCES

1. Rokhlin, D.B., Karapetyants, G.A. Ensembling discounted VAW experts with the VAW meta-learner. Global and Stochastic Analysis, vol. 12, no. 5, pp. 45–58, 2025.
2. Jacobsen, A., Cutkosky, A. Online linear regression in dynamic environments. Proc. 41st Int. Conf. Machine Learning (ICML 2024), PMLR, vol. 235, pp. 21083–21120, 2024.