A Spectral Analysis of Dot-product Kernels

M. Scetbon

Z. Harchaoui

ENSAE
IP PARIS

UNIVERSITY of WASHINGTON

AISTATS

Twenty-fourth International Conference on Artificial Intelligence and Statistics

Dot-Product Kernel

Dot-product kernel on the sphere: $K(x, y) = \sum_{m \geq 0} b_m (\langle x, y \rangle)^m, \quad x, y \in S^{d-1}, \quad d \geq 2$

- K is well defined if $\sum_{m \geq 0} |b_m| < +\infty$
- K is symmetric: $K(x, y) = K(y, x)$
- K satisfies for all $N \in \mathbb{N}$, $x_1, \ldots, x_N \in S^{d-1}$ and $(a_1, \ldots, a_N) \in \mathbb{R}^N$: $\sum_{i,j=1}^{N} a_i a_j K(x_i, x_j) \geq 0$ if $b_m \geq 0$

K is a positive definite kernel

Examples:

RBF kernel: $\exp\left(\frac{\|x - y\|_2}{2\sigma^2}\right)$, *Arc-cosine kernel*: $\pi - \arccos(\langle x, y \rangle)$, *Inverse kernel*: $(2 - \langle x, y \rangle)^{-1}$

Mercer Decomposition

Integral operator: $T_K^{d\mu} : f \in L_2^{d\mu}(S^{d-1}) \longrightarrow \int_{S^{d-1}} K(x,\cdot) f(x) d\mu(x), \quad \mu \in \mathscr{P}(S^{d-1})$

- $T_K^{d\mu}$ is self-adjoint, positive semi-definite and trace-class:

$$T_K^{d\mu}(\cdot) = \sum_{m=0}^{M} \sum_{\ell_m=1}^{\alpha_m} \eta_m^{\mu} \langle \cdot, Y_{m,\ell_m}^{\mu} \rangle_{L_2^{d\mu}} Y_{m,\ell_m}^{\mu}$$

Spectral Theorem

where $M \in \mathbb{N} \cup \{+\infty\}$, (Y_{m,ℓ_m}^{μ}) is an ONS, (η_m^{μ}) a positive, non-increasing summable sequence

and (α_m) are the multiplicities of (η_m^{μ})

- If μ is the induced Lebesgue measure on S^{d-1} we have:

(Y_{m,ℓ_m}^{μ}) are the spherical harmonics

$$\eta_m^{\mu} = \frac{|S^{d-2}|\Gamma((d-1)/2)}{2^{m+1}} \sum_{s \geq 0} b_{2s+m} \frac{(2s+m)!}{(2s)!} \frac{\Gamma(s+1/2)}{\Gamma(s+m+d/2)}$$

Eigenvalue Decay

- *Polynomial decay*

Proposition

If there exists $\alpha > 1$ such that $b_m \in \mathcal{O}(m^{-\alpha})$ then we have:

$$\eta_m^\mu \in \mathcal{O}(m^{-d/(2d-2)-\alpha/(d-1)+3/(2d-2)})$$

- *Geometric decay*

Proposition

If there exists $0 < r < 1$ such that $b_m \in \mathcal{O}(r^m)$ then we have:

$$\eta_m^\mu \in \mathcal{O}\left(r^{\,c_d\, m^{\frac{1}{d-1}}}\right)$$ where c_d is a constant depending on d

- *Super-geometric decay*

Proposition

If there exists $\delta > 0$ such that $\left|\frac{b_{m+1}}{b_m}\right| \in \mathcal{O}(m^{-\delta})$ then we have:

$$\eta_m^\mu \in \mathcal{O}\left(m^{\,-\delta c_d\, m^{\frac{1}{d-1}}}\right)$$ where c_d is a constant depending on d

Approximation of the RKHS

Mercer Decomposition

$$K(x,y) = \sum_{m=0}^{M} \sum_{\ell_m=0}^{\alpha_m} \eta_m^{\mu} \, Y^{\mu}_{m,\ell_m}(x) Y^{\mu}_{m,\ell_m}(y), \qquad H_K = \left\{ \sum_{m=0}^{M} \sum_{\ell_m=0}^{\alpha_m} a_{m,\ell_m} Y^{\mu}_{m,\ell_m} \text{ s.t. } \sum_{m=0}^{M} \sum_{\ell=1}^{\alpha_m} \frac{\alpha^2_{m,\ell_m}}{\eta^{\mu}_m} < +\infty \right\}$$

n-th entropy number: $\varepsilon_n(E) := \inf\{\epsilon \colon \mathbf{N}(\epsilon, E, d) \le n\}$

where $\mathbf{N}(\epsilon, E, d)$ is the smallest number of elements of an ϵ-cover for a given set E

- *Polynomial decay:* $b_m \in \mathcal{O}(m^{-\alpha}) \implies \varepsilon_n(T_K(B)) \in \mathcal{O}(\log^{-p(\alpha,d)/2}(n))$ *where* $p(\alpha, d) = \frac{d/2+\alpha-3/2}{d-1}$
- *Geometric decay:* $b_m \in \mathcal{O}(r^m) \implies \log(\varepsilon_n(T_K(B))) \in \mathcal{O}(\log^{1/d}(n))$

Statistical Bounds for RLS

Goal: estimation of $f_\rho(\,\cdot\,) = \mathbb{E}_{(X,Y)\sim\rho}[Y \,|\, X = \cdot\,]$

RLS estimator: $\hat{f}_{n,\lambda} = \text{argmin}_{f \in H_K} \left\{ \frac{1}{\ell} \sum_{i=1}^{\ell} (f(x_i) - y_i)^2 + \lambda \|f\|^2_{H_K} \right\}$ where (x_i, y_i) are i.i.d $\sim \rho$

Rates of the RLS

- *Polynomial decay*

Proposition

If there exists $\alpha > 1$ such that $b_m \in \mathcal{O}(m^{-\alpha})$ then w.h.p we have:

$$\|f_{\ell,\lambda_\ell} - f_\rho\|_\rho^2 \in \mathcal{O}\left(\ell^{-\frac{\beta}{\beta + q(\alpha,d)}}\right) \text{ where } q(\alpha, d) = 1/p(\alpha, d) \text{ and } 2 \geq \beta > 1$$

- *Geometric decay*

Proposition

If there exists $0 < r < 1$ such that $b_m \in \mathcal{O}(r^m)$ then we have w.h.p. :

$$\|f_{\ell,\lambda_\ell} - f_\rho\|_\rho^2 \in \mathcal{O}\left(\frac{\log(\ell)^{d-1}}{\ell}\right)$$

- *Super-geometric decay*

Proposition

If there exists $\delta > 0$ such that $\left|\frac{b_{m+1}}{b_m}\right| \in \mathcal{O}(m^{-\delta})$ then we have w.h.p. :

$$\|f_{\ell,\lambda_\ell} - f_\rho\|_\rho^2 \in \mathcal{O}\left(\frac{\log(\ell)^{d-1}}{\log(\log(\ell))^{d-1}\ell}\right)$$

Examples related to deep nets

- *Neural tangent kernels*
- *Hilbertian envelope of smooth multi-layer perceptrons*
- *Link between the eigendecay and the depth of networks*

Thank you