Log-Hilbert-Schmidt distance between covariance operators and its approximations

Hà Quang Minh, Istituto Italiano di Tecnologia (IIT), Genova, ITALY

Abstract

We show conditions under which the infinite-dimensional Log-Hilbert-Schmidt distance between RKHS covariance operators can be approximated by the finite-dimensional Log-Euclidean distance.

Log-Hilbert-Schmidt distance

Log-Euclidean distance. This is the geodesic distance between two matrices $A, B \in \mathrm{Sym}^{++}(n)$ in the Log-Euclidean metric [1]

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F.$$

Log-Hilbert-Schmidt distance. This generalizes the Log-Euclidean distance to the infinite-dimensional manifold $\Sigma(\mathcal{H})$ of positive definite Hilbert-Schmidt operators on a separable Hilbert space $\mathcal{H}$ [2]. For two operators $A + \gamma I > 0$,$B + \mu I > 0$, $A, B \in \mathrm{HS}(\mathcal{H})$, $\gamma, \mu > 0$,

$$d_{\mathrm{logHS}}[(A + \gamma I), (B + \mu I)] = ||\log(A + \gamma I) - \log(B + \mu I)||_{\mathrm{eHS}},$$

with the extended Hilbert-Schmidt norm $||A + \gamma I||^2_{\mathrm{eHS}} = ||A||^2_{\mathrm{HS}} + \gamma^2$.

RKHS and covariance operators

Reproducing kernel Hilbert spaces (RKHS) and feature maps. Let $\mathcal{X}$ be any non-empty set, K a positive definite kernel on $\mathcal{X} \times \mathcal{X}$, with corresponding RKHS $\mathcal{H}_K$. Then $\exists$ a separable Hilbert space $\mathcal{H}$, which can be identified with $\mathcal{H}_K$, and a corresponding feature map $\Phi : \mathcal{X} \to \mathcal{H}$, so that

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}.$$

Covariance operators. Let $\mathbf{x} = [x_1, \ldots, x_m]$ be a data matrix randomly sampled from $\mathcal{X}$ according to some probability distribution. The feature map Φ gives the bounded linear operator $\Phi(\mathbf{x}) : \mathbb{R}^m \to \mathcal{H}$, defined by

$$\Phi(\mathbf{x})\mathbf{b} = \sum_{j=1}^{m} b_j \Phi(x_j), \quad \mathbf{b} \in \mathbb{R}^m.$$

$\Phi(\mathbf{x})$ can be viewed as a (infinite) data matrix $\Phi(\mathbf{x}) = [\Phi(x_1), \ldots, \Phi(x_m)]$ in $\mathcal{H}$, with covariance operator

$$C_{\Phi(\mathbf{x})} = \frac{1}{m}\Phi(\mathbf{x}) J_m \Phi(\mathbf{x})^* : \mathcal{H} \to \mathcal{H}, \quad J_m = I_m - \frac{1}{m}\mathbf{1}_m \mathbf{1}_m^T.$$

For two covariance operators $C_{\Phi(\mathbf{x})}$ and $C_{\Phi(\mathbf{y})}$, the Log-HS distance

$$d_{\mathrm{logHS}} = ||\log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})||_{\mathrm{eHS}} \quad (1)$$

has a closed form expressed in terms of Gram matrices.

References

[1] V. Arsigny, et al. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM journal on matrix analysis and applications 29(1):328-347, 2007.

[2] H.Q. Minh, et al. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces. NIPS 2014.

[3] H.Q. Minh, et al. Approximate Log-Hilbert-Schmidt distance between covariance operators for image classification. CVPR 2016.

Finite-dimensional approximations

The distance in Eq. (1) can be computationally intensive on a large set of covariance operators.
Approximate feature map $\hat{\Phi}_D : \mathcal{X} \to \mathbb{R}^D$, $D << \dim(\mathcal{H})$, so that

$$\langle \hat{\Phi}_D(x), \hat{\Phi}_D(y) \rangle_{\mathbb{R}^D} = \hat{K}_D(x, y) \approx K(x, y), \quad \lim_{D \to \infty} \hat{K}_D(x, y) = K(x, y), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}.$$

Approximate covariance operator $C_{\hat{\Phi}_D(\mathbf{x})} = \frac{1}{m}\hat{\Phi}_D(\mathbf{x}) J_m \hat{\Phi}_D(\mathbf{x})^T : \mathbb{R}^D \to \mathbb{R}^D$.
Approximate Log-Hilbert-Schmidt distance

$$\left\| \log\left(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D\right) - \log\left(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D\right) \right\|_F. \quad (2)$$

Convergence. We need to determine whether (2) is truly a finite-dimensional approximation of (1), i.e.

$$\lim_{D \to \infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D) \right\|_F = ||\log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \mu I_{\mathcal{H}})||_{\mathrm{eHS}}.$$

Theorem 1. *Assume that* $\gamma \neq \mu$, $\gamma > 0$, $\mu > 0$. *Then*

$$\lim_{D \to \infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \mu I_D) \right\|_F = \infty.$$

Theorem 2. *Assume that* $\gamma = \mu > 0$. *Then*

$$\lim_{D \to \infty} \left\| \log(C_{\hat{\Phi}_D(\mathbf{x})} + \gamma I_D) - \log(C_{\hat{\Phi}_D(\mathbf{y})} + \gamma I_D) \right\|_F = ||\log(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}}) - \log(C_{\Phi(\mathbf{y})} + \gamma I_{\mathcal{H}})||_{\mathrm{eHS}}.$$

Random Fourier approximation

Let $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be of the form $K(x, y) = k(x - y)$ for some positive definite function k on $\mathbb{R}^n$. By Bochner's Theorem, $\exists$ a finite positive measure ρ on $\mathbb{R}^n$ s.t.

$$K(x, y) = \int_{\mathbb{R}^n} e^{-i\langle \omega, x - y \rangle} d\rho(\omega) = \int_{\mathbb{R}^n} \cos(\langle \omega, x - y \rangle) \rho(\omega) d\omega.$$

Without loss of generality, we can assume that ρ is a probability measure. For the Gaussian kernel $K(x, y) = e^{-\frac{||x-y||^2}{\sigma^2}}$, we have $\rho(\omega) = \frac{(\sigma\sqrt{\pi})^n}{(2\pi)^n} e^{-\frac{\sigma^2 ||\omega||^2}{4}} \sim \mathcal{N}\left(0, \frac{2}{\sigma^2} I_n\right)$. To approximate $K(x, y)$, we sample D points $\{\omega_j\}_{j=1}^D$ from the distribution ρ and compute the empirical version

$$\hat{K}_D(x, y) = \frac{1}{D} \sum_{j=1}^{D} \cos(\langle \omega_j, x - y \rangle) \xrightarrow{D \to \infty} K(x, y) \quad \text{a.s.}$$

The approximate random Fourier feature map is

$$\hat{\Phi}_D(x) = \frac{1}{\sqrt{D}} (\cos(\langle \omega_j, x \rangle), \sin(\langle \omega_j, x \rangle))_{j=1}^D \in \mathbb{R}^{2D}.$$

Example: Image classification

Method	Accuracy
Approx LogHS	**53.91**%(±4.34)
Log-HS	56.74%(±2.87)
Hilbert-Schmidt	50.17%(±2.17%)
Log-Euclidean	42.70%(±3.45)
Euclidean	26.87%(±3.52%)

The classification of fish images acquired from live underwater videos. The dataset contains 23 species of fish. At each pixel, the color values, red, green, blue, are sampled. All classifications were done by Gaussian Support Vector Machine, using the corresponding distances. Approx Log-HS, Log-HS, and Hilbert-Schmidt distances were computed with the Gaussian kernel. Approx Log-HS, using the random Fourier feature, $D = 200$, is 50 times faster to compute than Log-HS [3].