How to Select the Default Parameters for Spatial and Temporal Kernel Functions?

Tsz Nam Chan[1], Pak Lon Ip[2,3], Leong Hou U[2,3], Byron Choi[1] and Jianliang Xu[1]

[1]Hong Kong Baptist University
[2]University of Macau
[3]State Key Laboratory of Internet of Things for Smart City

As stated in our experimental settings (cf. Section 6.1) of our work [2], we follow [1] and adopt the Scott's rule [3] to obtain the default parameters γ_s and γ_t. Here, we provide the details for how to obtain these two parameters.

Given a dataset $\widehat{P}$ of n spatiotemporal data points, i.e., we have $\widehat{P} = \{(\mathbf{p}_1, t_{\mathbf{p}_1}), (\mathbf{p}_2, t_{\mathbf{p}_2}), ..., (\mathbf{p}_n, t_{\mathbf{p}_n})\}$, where each spatial location $\mathbf{p}$ is denoted as $(\mathbf{p}.x, \mathbf{p}.y)$. We can find the bandwidth values h_x (cf. Equation 1), h_y (cf. Equation 2) and h_t (cf. Equation 3) for this dataset $\widehat{P}$ (based on the Scott's rule [3]).

$$h_x = n^{-\frac{1}{5}} \sigma_x \tag{1}$$
$$h_y = n^{-\frac{1}{5}} \sigma_y \tag{2}$$
$$h_t = n^{-\frac{1}{5}} \sigma_t \tag{3}$$

where σ_x, σ_y and σ_t is the standard deviation of all $\mathbf{p}.x$, $\mathbf{p}.y$ and $t_{\mathbf{p}}$ in the dataset $\widehat{P}$, respectively.

Based on these bandwidth values, we follow [1] and obtain the default parameter γ_s, where:

$$\gamma_s = \frac{1}{\sqrt{h_x^2 + h_y^2}} \tag{4}$$

Moreover, we also have:

$$\gamma_t = \frac{1}{h_t} \tag{5}$$

References

[1] T. N. Chan, R. Cheng, and M. L. Yiu. QUAD: Quadratic-bound-based kernel density visualization. In *SIGMOD*, pages 35–50, 2020.

[2] T. N. Chan, P. L. Ip, L. H. U, B. Choi, and J. Xu. SWS: A complexity-optimized solution for spatial-temporal kernel density visualization. *Proc. VLDB Endow.*, 2022 (Submitted).

[3] D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. A Wiley-interscience publication. Wiley, 1992.