

Kernel Smoothing And Heat Equation

October 26, 2020

1 Introduction

It is universally acknowledged that observed data with respect to the underlying patterns we are seeking for in the physical world is, to some extent, contaminated with random noise. Commonly, the observed data would be expressed into two parts: the systematic component (i.e.: a true underlying oracle function $u(x)$) and a random component (i.e.: noise ϵ). Yet, due to the fundamental sorrow of the limited observations in an infinite world, we may fail to achieve the oracle function. In fact, we can only estimate it based on our finite noise infested observations.

For decades, statisticians have been proposed various methods to estimate the oracle function. One of the non-parametric ways is to use direct diffusion to smooth away all the noise. In an absolute non-rigorous sense, the direction diffusion process achieves the oracle risk by (weighted) averaging each observation neighbor. Each time we do this, our estimated function will be closer to the true function until at a finite time T we achieve our goal. After that the estimated function will be bounded away from the truth. People coin the term "over-smoothing" to indicate this phenomenon. Our goal is to discover the optimal T where we are closest to the truth given an observed data set.

However, the diffusion process is not widely used among statisticians. Instead, they use kernel smoothing, a nonparametric method proved to be equivalent to the heat equation in physics world. Therefore, finding optimal time T can be translated to the problem of finding the optimal sigma. In statistics literature (?? reference needed), the optimal sigma is found to be a constant for every point x . This is a significant constraint. Hence, in this paper, we will allow the spatial bandwidth to be different for each point. In the next section, we will derive the optimal sigma for each observation point. We will do simulations to show that our "adaptive" bandwidth performs better than the best-fixed bandwidth.

Some literature review...

Goal of this paper...

Structure of the paper...

2 Methodology

The connection between Heat Equation and Kernel Smoothing can be expressed (??reference for Equation (1)) as

$$T = \frac{1}{2}\sigma^2 \quad (1)$$

where T denotes the total smoothing time in the direction process, and σ denotes the proportion of the spatial width of the Gaussian smoother in paper ?? (Equation (2)).

$$\sigma = \text{bandwidth} \times 0.3706506 \quad (2)$$

Given $\mathbf{x}_i = (x_1, x_2, \dots, x_p)^\top \subseteq \mathbb{R}^p$, $Y_i \subseteq R$, we have the data set

$$\mathcal{D}_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

Denote $u(\mathbf{x})$ as the oracle function, and $u(\mathbf{x}_i)$ is the first order Taylor expansion

$$u(\mathbf{x}_i) \approx u(\mathbf{x}) + \frac{\partial u}{\partial \mathbf{x}}(\mathbf{x}_i - \mathbf{x}) \quad (3)$$

And given a Gaussian kernel with b as the bandwidth parameter

$$\phi_b(x) = \frac{1}{\sqrt{2\pi b}} \exp\left(\frac{-x^2}{2b}\right) \quad (4)$$

We have MSE at \mathbf{x}_i with respect to b

$$\text{MSE}_j = \frac{\sigma^2}{2\sqrt{\pi} \cdot b} + \left[\frac{\partial^2 u}{2\partial x^2} \cdot b \right]^2 \quad (5)$$

Take the derivative with respect to b , we can solve the equation

$$b = \left(\frac{\sigma^2}{2\sqrt{\pi} u''^2} \right)^{\frac{2}{5}} \quad (6)$$

(Detailed calculation in Appendix)

3 Appendix

We let

$$\phi_b(x) = \frac{1}{\sqrt{2\pi b}} \exp\left(\frac{-x^2}{2b}\right)$$

where b is the band width parameter. Then we have our MSE at x_i

$$\begin{aligned} (\hat{u}(x_i) - u(x_i)) &= \sigma^2 \sum_{i=1}^n w_i^2 + \left[\frac{\partial u}{\partial x} \cdot \sum_{i=1}^n w_i(x_i - x) \right] + \left[\frac{\partial^2 u}{2\partial x^2} \cdot \sum_{i=1}^n w_i(x_i - x)^2 \right] + \sum_{i=1}^n \epsilon_i w_i \\ E \left[(\hat{u}(x_i) - u(x_i))^2 \right] &= \text{Var}[(\hat{u}(x_i) - u(x_i))] + (E[(\hat{u}(x_i) - u(x_i))])^2 \\ &= \sigma^2 \sum_{i=1}^n w_i^2 + \left[\frac{\partial u}{\partial x} \cdot \sum_{i=1}^n w_i(x_i - x) + \frac{\partial^2 u}{2\partial x^2} \cdot \sum_{i=1}^n w_i(x_i - x)^2 \right]^2 \\ MSE_j &= \sigma^2 \sum_{i=1}^n \phi_b^2(x_j - y_i) + \left[\frac{\partial u}{\partial x} \cdot \sum_{i=1}^n \phi_b(x_j - y_i) \cdot (x_j - y_i) + \frac{\partial^2 u}{2\partial x^2} \cdot \sum_{i=1}^n \phi_b(x_j - y_i) \cdot (x_j - y_i)^2 \right]^2 \end{aligned}$$

Now we replace summation with integration

$$\begin{aligned} MSE_j &= \sigma^2 \int_{\omega} \phi_b^2(x_j - y) dy + \left[\frac{\partial u}{\partial x} \cdot \int_{\omega} \phi_b(x_j - y) \cdot (x_j - y) dy + \frac{\partial^2 u}{2\partial x^2} \cdot \int_{\omega} \phi_b(x_j - y) \cdot (x_j - y)^2 dy \right]^2 \\ &= \frac{\sigma^2}{2\sqrt{\pi} \cdot b} + 0 + \left[\frac{\partial^2 u}{2\partial x^2} \cdot \int_{\omega} \phi_b(x_j - y) \cdot (x_j^2 - 2x_j \cdot y + y^2) dy \right]^2 \\ &= \frac{\sigma^2}{2\sqrt{\pi} \cdot b} + \left[\frac{\partial^2 u}{2\partial x^2} \cdot \int_{\omega} \phi_b(x_j - y)(x_j^2) dy - 2 \int_{\omega} \phi_b(x_j - y) \cdot (x_j \cdot y) dy + \int_{\omega} \phi_b(x_j - y) \cdot y^2 dy \right]^2 \\ &= \frac{\sigma^2}{2\sqrt{\pi} \cdot b} + \left[\frac{\partial^2 u}{2\partial x^2} \cdot (x_j^2 - 2x_j^2 + b + x_j^2) \right]^2 \\ &= \frac{\sigma^2}{2\sqrt{\pi} \cdot b} + \left[\frac{\partial^2 u}{2\partial x^2} \cdot b \right]^2 \end{aligned}$$

Set $h(b) = \frac{\sigma^2}{2\sqrt{\pi}} \cdot \frac{1}{\sqrt{b}} + \frac{1}{4} u''^2 \cdot b^2$, and take derivative, then set $h'(b) = 0$

$$\begin{aligned} h'(b) &= -\frac{\sigma^2}{4\sqrt{\pi}} \cdot b^{-\frac{3}{2}} + \frac{1}{2} u''^2 b = 0 \\ b &= \left(\frac{\sigma^2}{2\sqrt{\pi} u''^2} \right)^{\frac{2}{5}} \end{aligned}$$