地统计系列

地理加权回归模型
-2-
空间权重

UP：小勇啊哈

2023年7月30日

空间权重

一、空间权重

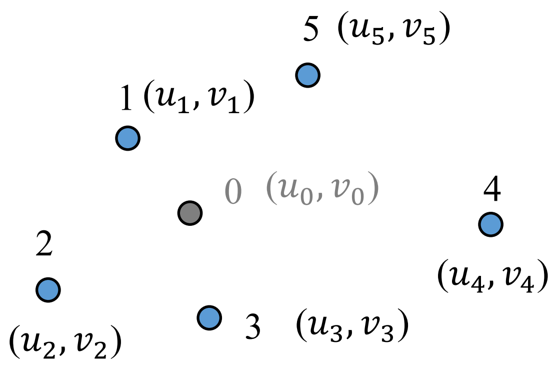
小勇啊哈 bilibili

空间权重是地理加权回归的**核心部分**，以地理学第一定律为基础，代表观测数据集中**各样本点**对估计待估计样本点y值**的贡献程度**。

$$\widehat{\beta_0} = \min_{\beta} \sum_{i=1}^{n} w_{i0}[(y_i - \boldsymbol{x}_i{}^T \boldsymbol{\beta}_0)]^2$$

上个视频该公式存在笔误，已在评论区修正

1. 数据介绍



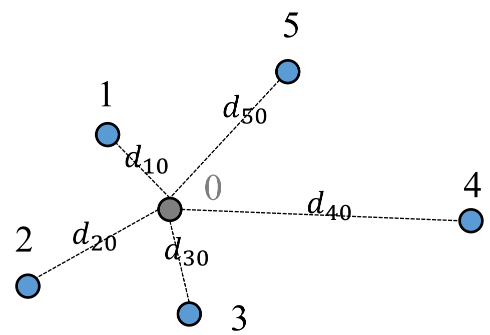
— 观测数据集中样本点

— 待估计数据集中样本点

— 0, 1-4 样本点编号

— (u_i, v_i) 样本点 i 的坐标

2. 距离度量



欧式距离:（最常用）

$$d_{i0} = \sqrt{(u_i - u_0)^2 + (v_i - v_0)^2}$$

还包括曼哈顿、闵可夫斯基、球面、余弦等距离

3. 权重计算

$$w_{i0} = f(d_{i0}, b)$$

— w_{i0}: 样本点 i 与样本点 0 之间的空间权重,值域为[0,1]

— $f(\cdot)$: 核函数

— b: 带宽

一、空间权重

小勇啊哈 bilibili

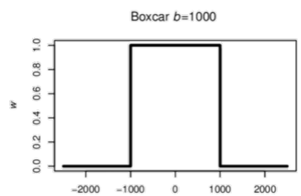
空间权重以距离为输入，**带宽**为预设参数，通过**核函数**计算而来。从直观角度理解，带宽控制空间相关范围，核函数控制空间权重衰减类型。

核函数类型

原则：距离越近，权重越大

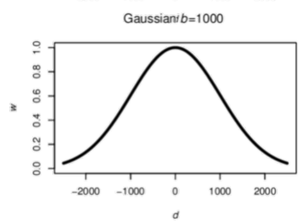
- 盒状（Box-car）函数

$$w_{ij} = \begin{cases} 1, d_{ij} \le b \\ 0, d_{ij} > b \end{cases}$$



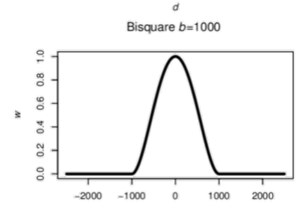
- 高斯（Gaussian）函数

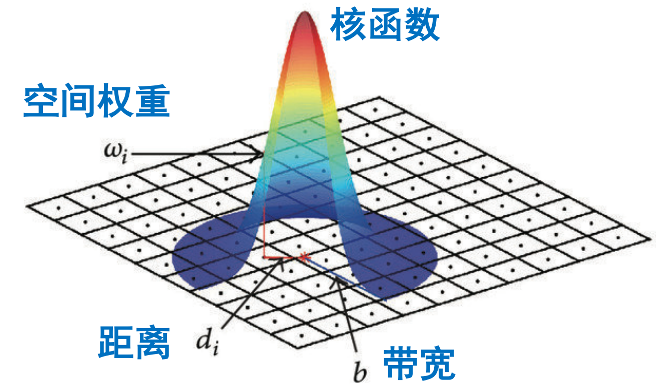
$$w_{ij} = e^{-(\frac{d_{ij}}{b})^2}$$

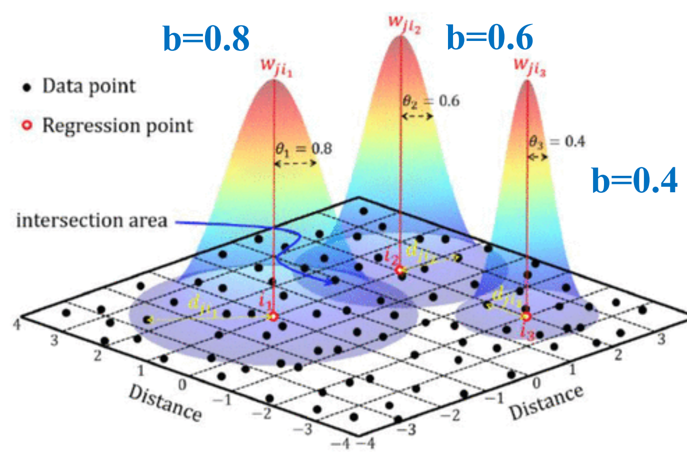


- 二次平方（Bi-square）函数

$$w_{ij} = \begin{cases} (1 - (\frac{d_{ij}}{b})^2)^2, d_{ij} \le b \\ 0, d_{ij} > b \end{cases}$$



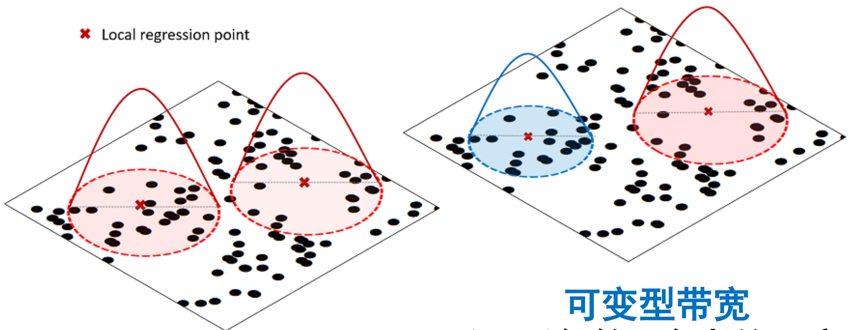




一、空间权重

小勇啊哈 bilibili

带宽是核函数的关键参数，包括**固定型**带宽和**可变型**带宽。当观测样本点分布较为**均匀**时，一般选择固定型带宽；若样本分布**不均**，选择可变型带宽。



可变型带宽
（最近邻第K个点的距离）

固定型带宽（固定值b）

如何选取合适的带宽值（b / K）呢？

评价方法(**取最小值**)：交叉验证法（CV），赤池信息量准则（AIC），贝叶斯信息准则（BIC）

$$CV = \frac{1}{n}\sum_{i=1}^{n}[y_i - \widehat{y_{\neq i}}(b)]^2$$

— n: 观测数据集中样本数

— y_i: 观测数据集中第i个样本因变量值

— $\widehat{y_{\neq i}}(b)$: 基于带宽b和观测数据集中**除第i个样本外其他样本**对第i个样本的估计值

$$AIC = 2nln(\hat{\sigma}) + nln(2\pi) + n[\frac{n + tr(\boldsymbol{S})}{n - 2 - tr(\boldsymbol{S})}]$$

— $\underset{n\times n}{\boldsymbol{S}}$: 帽子矩阵，$\widehat{\boldsymbol{Y}} = \boldsymbol{SY} = \begin{bmatrix} \boldsymbol{X}_1(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\ \boldsymbol{X}^T\boldsymbol{W}_1 \\ \boldsymbol{X}_2(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\ \boldsymbol{X}^T\boldsymbol{W}_2 \\ \cdots \\ \boldsymbol{X}_n(\boldsymbol{X}^T\boldsymbol{W}_n\boldsymbol{X})^{-1}\ \boldsymbol{X}^T\boldsymbol{W}_n \end{bmatrix}\boldsymbol{Y}$

— $tr(\boldsymbol{S})$:矩阵$\boldsymbol{S}$的迹

— $\hat{\sigma}^2$:随机误差项方差的极大似然估计

$\hat{\sigma}^2 = \sum_{i=1}^{n}[y_i - \hat{y}_i]^2/(n - 2tr(\boldsymbol{S}) + tr(\boldsymbol{S}^T\boldsymbol{S}))$

Thanks