Geographically Weighted Regression

$$\beta_0 + \beta_1\, Population + \beta_2\, Income = Crime$$

Supplementary Slides to Accompany the Code Walkthrough

What is Geographically Weighted Regression?

Geographically weighted regression (GWR) is a nonparametric spatial analysis technique that takes non-stationary variables into consideration…

e.g. climate; demographic factors; physical environment characteristics

…and models the local relationships between these predictors and an outcome of interest.

Standard Multiple Linear Regression Model w/ Normal Errors

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \ldots + \beta_m X_{mi} + \varepsilon_i \quad = \quad Y_i = \beta_0 + \sum_{k=1}^{m} \beta_k X_{ki} + \varepsilon_i$$

$$= \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1,m} \\ 1 & X_{21} & X_{22} & \ldots & X_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \ldots & X_{n,m} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$= \quad \boldsymbol{Y}_{n\times 1} = \underbrace{\boldsymbol{X}_{n\times p}}_{\text{(model matrix)}} \times \boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1} \quad = \quad \boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

We run OLS models to determine the global regression coefficients (β) for the independent variables with estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

The Regression Model Underlying Geographically Weighted Regression

$$Y_i(\mathbf{u}) = \beta_{0i}(\mathbf{u}) + \beta_{1i}(\mathbf{u})X_{1i} + \beta_{2i}(\mathbf{u})X_{2i} + \cdots + \beta_{mi}(\mathbf{u})X_{mi}$$

The notation $\beta_{0i}(\mathbf{u})$ indicates that the parameter describes a relationship around u, some general location in the study area. think of **u** as a vector of coordinates i.e. (X,Y) or (long,lat).

OLS Estimator

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

GWR Estimator

$$\hat{\boldsymbol{\beta}}(\boldsymbol{u}) = (\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{u})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}(\boldsymbol{u})\boldsymbol{Y}$$

W(**u**) is a square matrix of weights relative to the position of u in the study area. The weights themselves are computed from weight functions, also called kernels

(Gaussian Kernel)

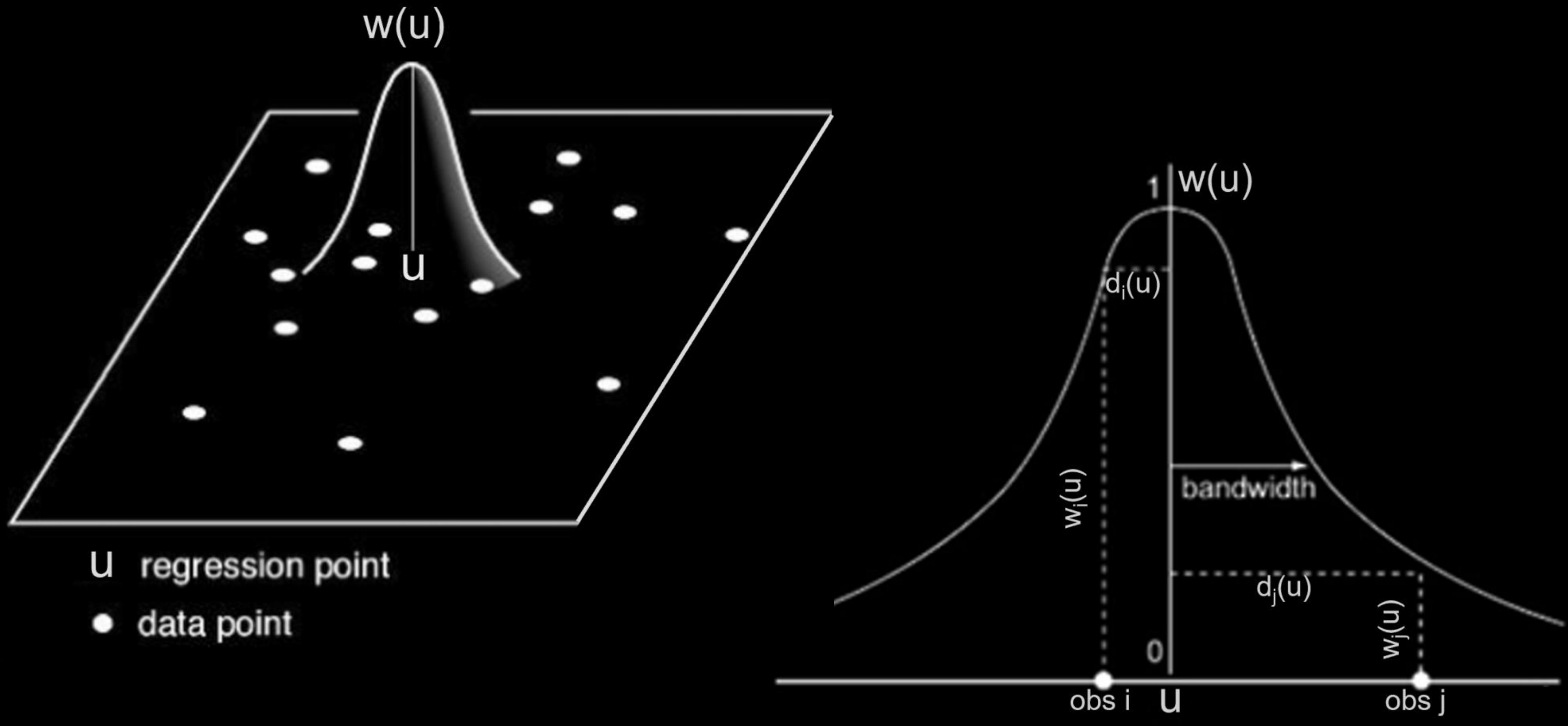
The geographical weight of the ith observation in the dataset relative to the location **u**

$$\boxed{w_i(\mathbf{u})} = e^{-0.5\boxed{(d_i(\mathbf{u})/h)^2}}$$

d_i(**u**) is a measure of the distance between the ith observation and the location **u**

h is the bandwidth

Weighting with a Gaussian Kernel, Illustrated



Pros / Strengths

- Allows visualization of stimulus-response relationships and if/how that relationship varies in space.
- Accounts for spatial autocorrelation of variables.
- Relatively simple, nonparametric approach (less assumptions to violate).
- Based on the traditional regression framework with which many are familiar.

Cons / Weaknesses

- Computationally Intensive (GWR builds a local regression equation for each feature in the dataset).
- When values for a particular explanatory variable cluster spatially or when sample sizes are small, you are likely to have problems with multicollinearity.
- Complicated approaches to calculating p-values and goodness-of-fit measures.
- Prediction efficacy contested.