

# 회귀분석팀

6팀

김형석

김준령

윤여원

김현우

이채은

# INDEX

---

1. Introduction

2. Subset Selection

3. Dimension Reduction

4. Shrinkage

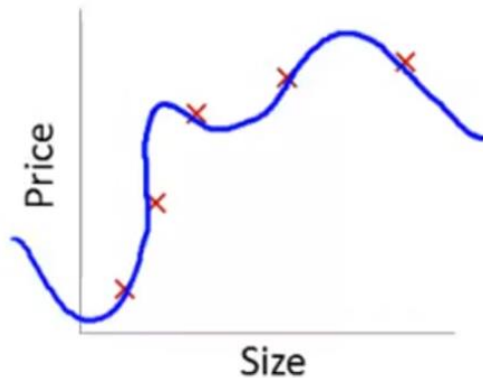
5. Regression for Spatial Data

6. Appendix

# 1

## Introduction

## 모델 선택의 필요성



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

High variance



설명변수(X)가 많은 경우

다중공선성의 발생 가능성이 높아져

학습성능 대비 추정모델의 Variance 가 커짐

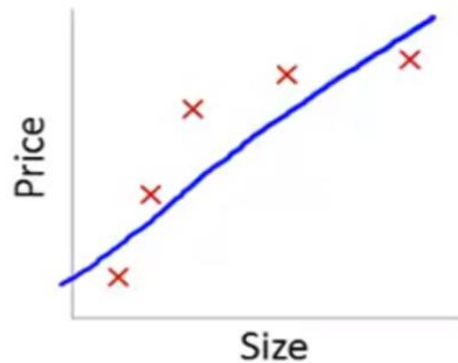


예측 성능이 크게 떨어질 가능성 존재

## 1

## Introduction

## 모델 선택의 필요성



$$\theta_0 + \theta_1 x$$

High bias



설명변수(X)가 적은 경우

추정모델의 Bias가 커져

학습성능 대비 추정모델의 Bias가 커짐

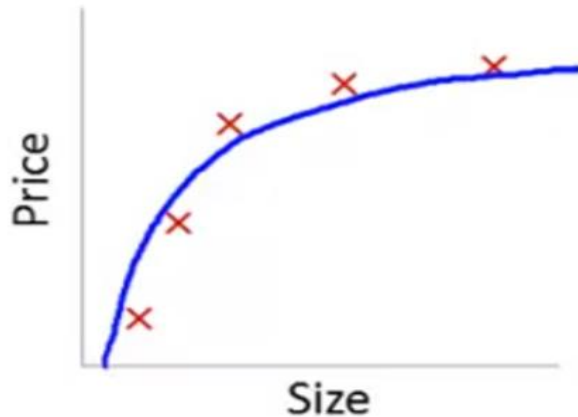


예측 성능이 크게 떨어질 가능성 존재

## 1

## Introduction

## 모델 선택의 필요성



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

"Just right"



따라서 예측오차(Prediction Error) 값을 기준으로  
적정한 개수의 설명변수( $X$ )를 갖는 최적의 회귀모델을 선택할  
필요성을 인식!

## 모델 선택의 필요성



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

"Just right"

그렇다면 예측오차(Prediction Error)란  
구체적으로 어떤 것을 의미하는 걸까?



따라서 예측오차(Prediction Error) 값을 기준으로  
적정한 개수의 설명변수( $X$ )를 갖는 최적의 회귀모델을 선택할  
필요성을 인식!

## 예측오차

예측오차 (Prediction Error)

훈련 데이터를 이용해 학습된 모델에 대해

테스트 데이터를 대입 시 발생하는 Loss 의 Expectation 값

⋮

$$\begin{aligned} E \left[ (Y - \hat{Y})^2 \right] &= \text{Bias}^2[\hat{f}(X)] + V[\hat{f}(X)] + E[\epsilon^2] \\ &= \text{Reducible Error} + \text{Irreducible Error} \end{aligned}$$

$f(X)$  : 데이터를 실제로 설명하는 모델 식

$\hat{f}(X)$  : 데이터로 추정한 모델 식



## 예측오차

예측오차 (Prediction Error)



훈련 데이터를 이용해 학습된 모델에

$$\text{Prediction Error} = \text{Bias}^2[\hat{f}(X)] + V[\hat{f}(X)] + E(\epsilon^2)$$

Irreducible Error( $E(\epsilon^2)$ ) 는 참값으로부터 발생하는

오차의 분산으로 줄일 수 없는 오차에 해당

$$E[(Y - \hat{Y})^2] = \text{Bias}^2[\hat{f}(X)] + V[\hat{f}(X)] + E[\epsilon^2]$$



따라서 Reducible Error 를

최대한 줄일 수 있는 최적의 회귀모델 선택해줘야 함

$\hat{f}(X)$  : 데이터로 추정된 모델 식

or

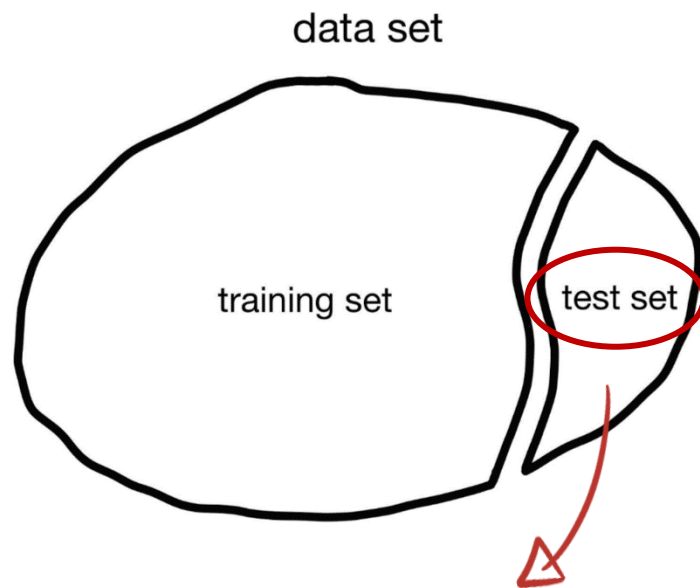
# 2

## Subset Selection

## 2

# Subset Selection

## 변수 선택 | 평가지표

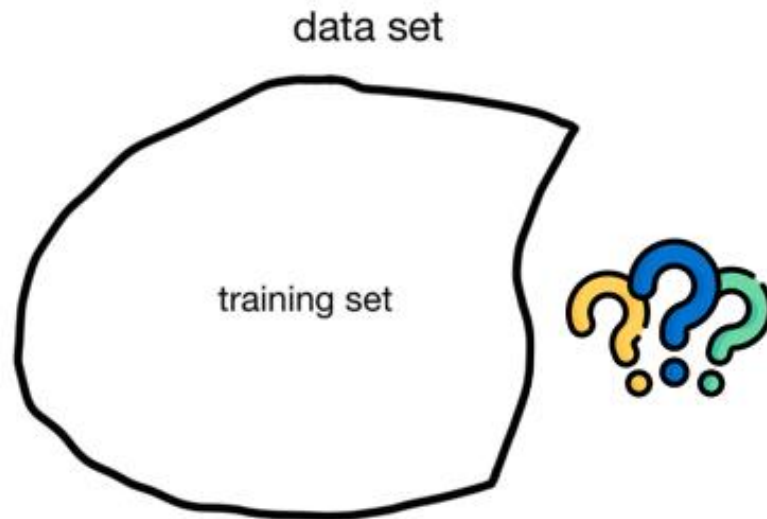


데이터가 충분히 많다면 데이터셋을 분리하여 Test Error 를 계산

## 2

# Subset Selection

## 변수 선택 | 평가지표

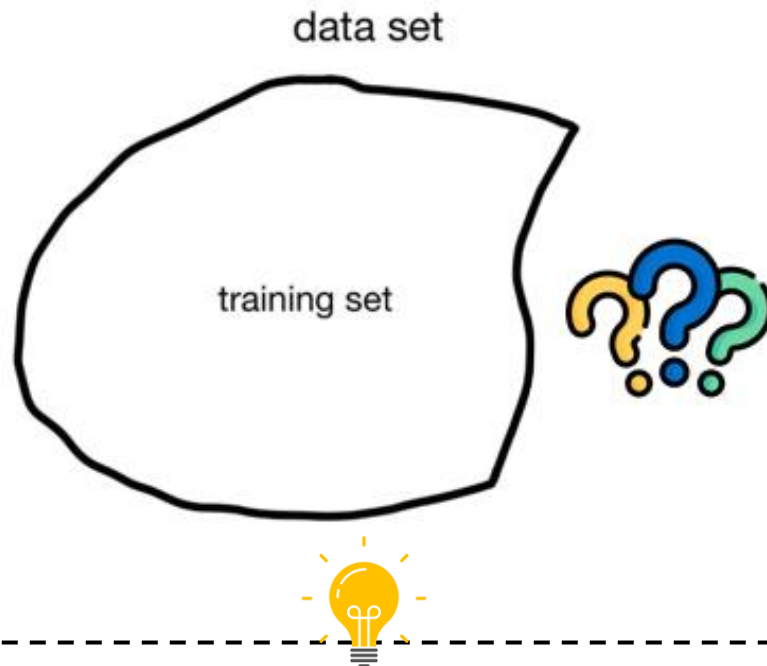


그러나 데이터의 양이 부족하다면 Test Error 계산에 어려움이 존재

## 2

## Subset Selection

### 변수 선택 | 평가지표



이 경우 현재 가지고 있는 데이터셋을 이용하여

**Test Error** 를 추정해볼 필요가 있음

## 변수 선택 | 평가지표

## Test Error 의 추정 (Estimation of Test Error)

활용할 수 있는 데이터를 통해 Test Error 의 대체 값을 계산하는 것

Analytical Methods  
(Indirect Estimation)

Adjusted  $R^2$ , AIC, BIC,  $C_p$

Subset Selection 방법에서  
기준척도로 사용됨

Sample Re-use Methods  
(Direct Estimation)

Cross-Validation, Bootstrap

데마 팀 클린업 자료 참고!

## 변수 선택 | 평가지표

## Test Error 의 추정 (Estimation of Test Error)

활용할 수 있는 데이터를 통해 Test Error 의 대체 값을 계산하는 것

Analytical Methods  
(Indirect Estimation)

Adjusted  $R^2$ , AIC, BIC,  $C_p$

Subset Selection 방법에서  
기준척도로 사용됨



이번 회귀 팀 클린업에서는  
**Analytical Methods 의  
평가지표들을 중점적으로**

다뤄볼 예정!

데마 팀 클린업 자료 참고!

평가지표 | Adjusted  $R^2$ Adjusted  $R^2$ 자유도를 감안하여 결정계수( $R^2$ )를 보정한 값

⋮

$$\text{Adjusted } R^2 = \frac{ESS/(\mathbf{p} - \mathbf{1})}{TSS/(\mathbf{n} - \mathbf{1})} = 1 - \frac{MSE}{TSS/(n - 1)}$$

분자, 분모에 각각의 제곱합에 대한 자유도를 나눠준 형태



평가지표 | Adjusted  $R^2$ 의 특징

① MSE 값을 기준으로 진행한 변수 선택 결과와 동일한 결과를 도출해줌



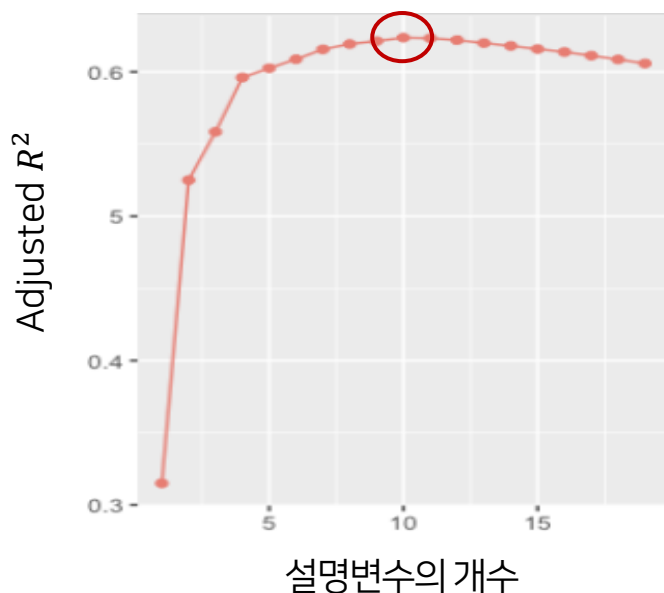
$TSS/(n - 1)$  값은 일정하므로

MSE 값에 따라 *Adjusted  $R^2$*  값이 변한다는 것을 확인 가능

$$\text{Adjusted } R^2 = \frac{ESS/(p - 1)}{TSS/(n - 1)} = 1 - \frac{\textcolor{red}{MSE}}{TSS/(n - 1)}$$

평가지표 | Adjusted  $R^2$ 의 특징

② 설명변수( $X$ ) 추가 시 Penalty 를 부과하므로 값이 항상 증가하지 않음



Adjusted  $R^2$  값의 분기점에  
해당되는 설명변수( $X$ ) 개수로 작성된  
선형회귀모델을 최적의 모델로 선정

## 2

## Subset Selection

### 평가지표 | $C_p$ , AIC, BIC

$C_p$ , AIC, BIC

Bias-Variance 특성을 고려하여 고안된 Test Error 의 대체 값으로

Likelihood 부분과 Penalty 부분으로 구성됨

Likelihood Part

모델의 Training Error 를 담당

모델 Bias 역할



Penalty Part

변수 개수에 따른 Penalty 를 담당

모델 Variance 역할

## 2

## Subset Selection

평가지표 |  $C_p$ , AIC, BIC $C_p$ , AIC, BIC

Bias-Variance 특성을 고려하여 고안된 Test Error 의 대체 값으로

Likelihood 부분과 Penalty 부분으로 구성됨

Likelihood Part

Penalty Part

각 지표별 계산식은 Appendix 참고!

모델의 Training Error 를 측정

과도하게 작은 Penalty 를 담당

모델 Bias 역할

모델 Variance 역할

## 2

## Subset Selection

### 평가지표 | $C_p$ , AIC, BIC

$C_p$ , AIC, BIC

Bias-Variance 특성을 고려하여 고안된 Test Error 의 대체 값으로

Likelihood 부분과 Penalty 부분으로 구성됨

Likelihood Part

모델의 Training Error 를 담당

모델 Bias 역할



Likelihood 부분의 경우

확률분포 식을 토대로 계산되기에

분포적 가정이 없는 모델은

$C_p$ , AIC, BIC 의 사용이 불가

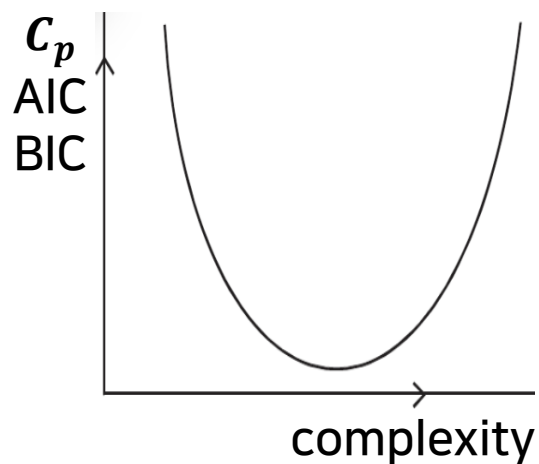
Ex) KNN Regression

## 2

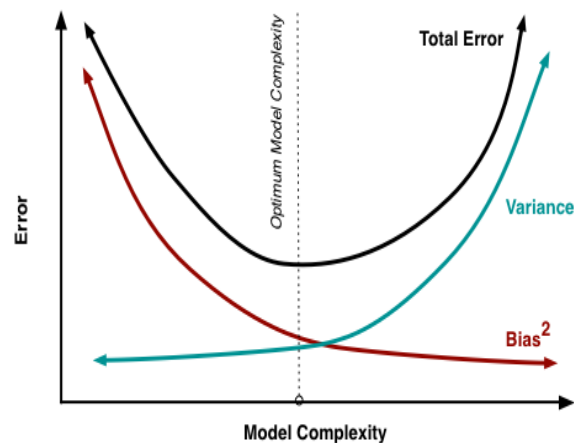
## Subset Selection

평가지표 |  $C_p$ , AIC, BIC 의 특징

Likelihood 와 Penalty 부분의 합을 통해 U 커브 형태의 그래프를 보임



$C_p$ , AIC, BIC 의 그래프



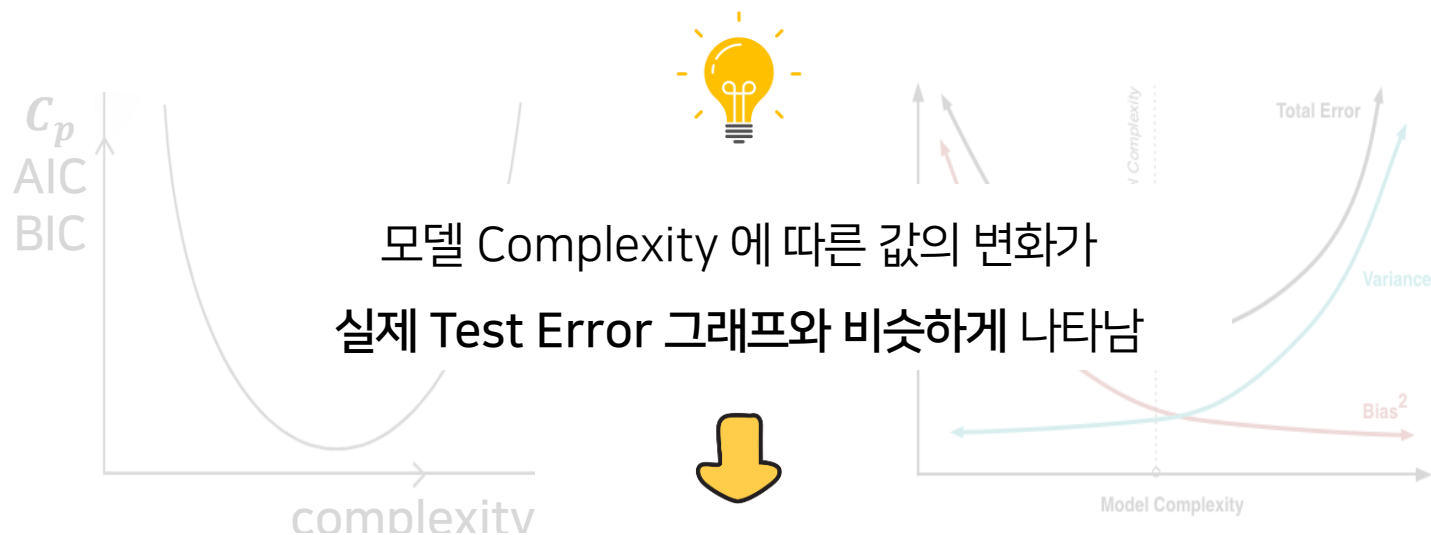
실제 Test Error 의 그래프

## 2

## Subset Selection

### 평가지표 | $C_p$ , AIC, BIC 의 특징

Likelihood 와 Penalty 부분의 합을 통해 U 커브 형태의 그래프를 보임



Test Error 의 대체 값으로  
사용하는 것에 대한 타당성을 가짐

## 변수선택법

변수선택법 (Subset Selection)

Analytical Methods 를 기준 평가지표로 사용하여  
회귀모델에 대한 최적의 변수 개수를 결정하는 방법

Subset Selection

Best Subset

Forward Stepwise

Backward Stepwise



## 2

# Subset Selection

## Best Subset Selection

### Best Subset Selection

가능한 모든 변수 조합을 고려하여

후보 모델들을 정의한 다음에 변수 선택을 진행하는 방법

Null model

Null

1 variable model

a

b

c

2 variables model

a, b

a, c

b, c

Full model

a, b, c

## Best Subset Selection | 장단점

가능한 후보 모델들의 총 개수

$$\binom{p}{0} + \binom{p}{1} + \dots + \binom{p}{p} = 2^p$$

$p$  : 모델의 최대 변수개수



모든 가능한 모델들을 고려하므로  
항상 Best Model 을 리턴



고려할 최대 변수 개수가 많아지면  
계산비용이 커져 효율성이 떨어짐

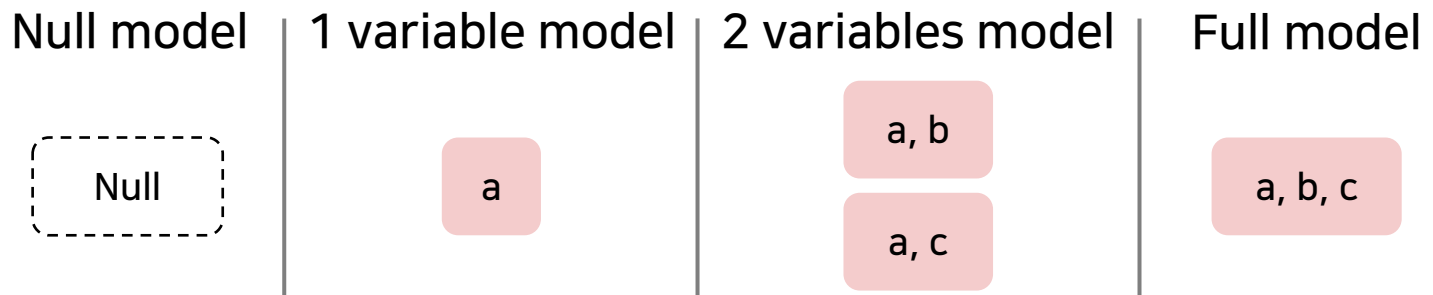
## 2

## Subset Selection

### Forward Stepwise Selection

#### Forward Stepwise Selection

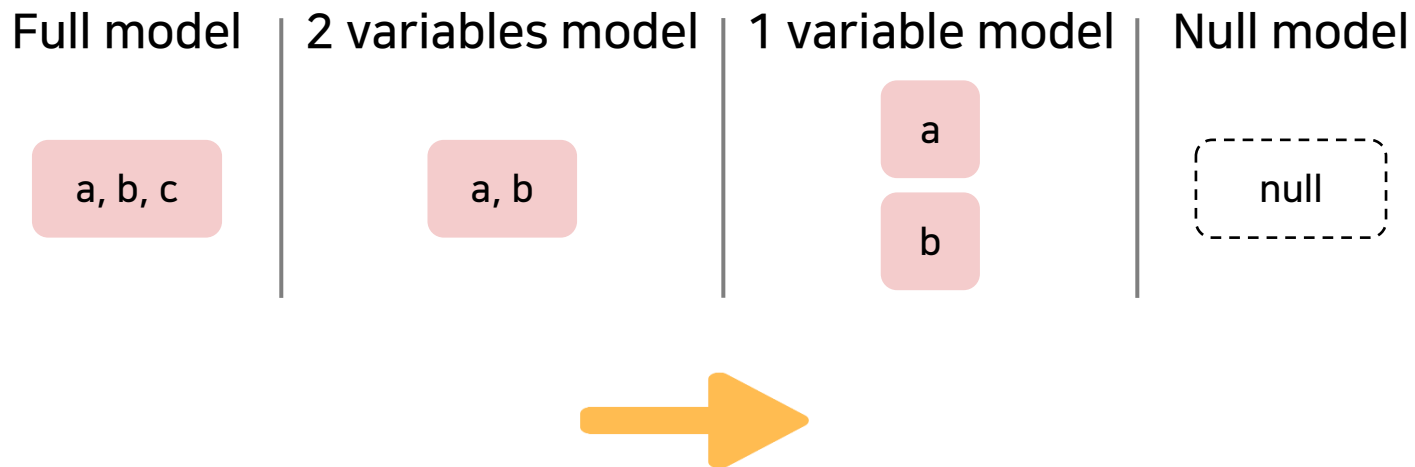
Null 모델에서부터 시작해서 가장 유의미한 특성을 선택해가며  
후보 모델들을 구축한 다음에 변수 선택 과정을 진행하는 방식



## Backward Stepwise Selection

### Backward Stepwise Selection

Full 모델에서부터 시작해서 가장 무의미한 특성을 제거해가며  
후보 모델들을 구축한 다음에 변수 선택 과정을 진행하는 방식



## Forward / Backward Stepwise Selection | 장단점

가능한 후보 모델들의 총 개수

$$1 + \sum_{k=0}^{p-1} (p - k) = 1 + p(p + 1)/2 < 2^p$$

 $p$  : 모델의 최대 변수개수

Best Subset Selection 보다  
계산적인 측면에서 더 효율적



모든 모델을 고려하지 않기에  
결과가 항상 Best Model 이라고  
장담할 수 없음

# 3

## Dimension Reduction

## 차원 축소기법

차원 축소기법 (Dimension Reduction)

고차원의 데이터를 저차원의 데이터로 변환하는 기법으로

너무 많은 설명변수( $X$ )를 사용했을 때의 문제점 해결

주성분 회귀분석  
(PCR)

부분 최소제곱법  
(PLS)

## 차원 축소기법

### 차원 축소기법 (Dimension Reduction)

고차원의 데이터를 저차원의 데이터로 변환하는 기법으로

너무 많은 설명변수( $X$ )를 사용했을 때의 문제점 해결



주성분 회귀분석

(PCR)

부분 최소제곱법

(PLS)

이번주 클린업에서는 주성분 회귀분석에 대해 다루어 보겠습니다!



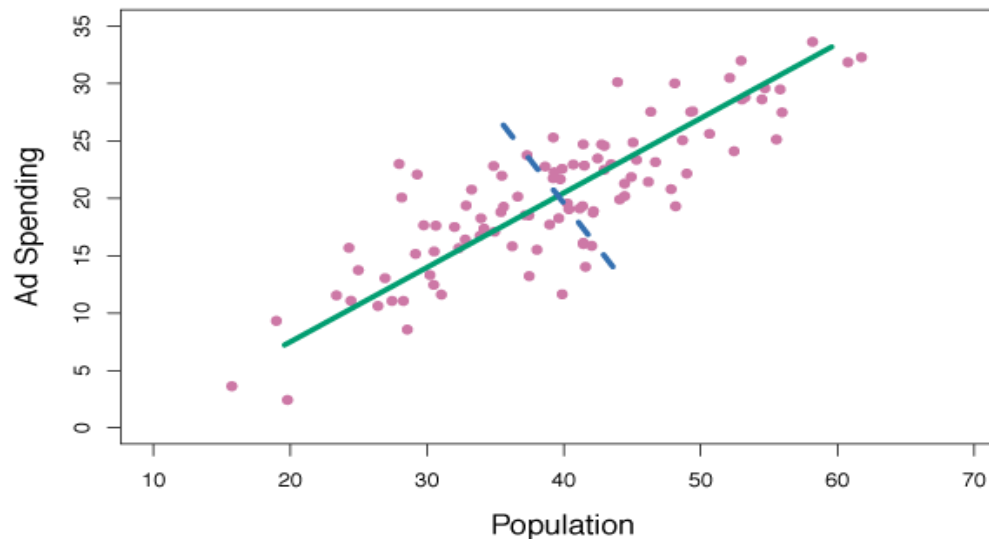
# 3 Dimension Reduction

## 주성분 회귀분석

주성분 회귀분석 (Principal Component Regression)

설명변수( $X$ ) 의 전체 분산정도를 가장 잘 설명하는

주성분( $Z$ ) 들을 새로운 설명변수로 사용하는 회귀 모델



## 주성분 회귀분석 | 모델링 과정

① 표준화된 설명변수에 대해 고유값 분해를 적용

$$X_s^T X_s = [v_1 \quad \dots \quad v_p] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} [v_1 \quad \dots \quad v_p]^{-1}$$

$X_s = [X_s^1 \quad \dots \quad X_s^p]$  : 표준화된 설명변수 벡터

데이터의 분산정도를 Fair 하게 비교하기 위해 표준화를 진행

$v_1 \quad \dots \quad , v_p$  :  $X_s^T X_s$  고유값 분해 후 고유벡터

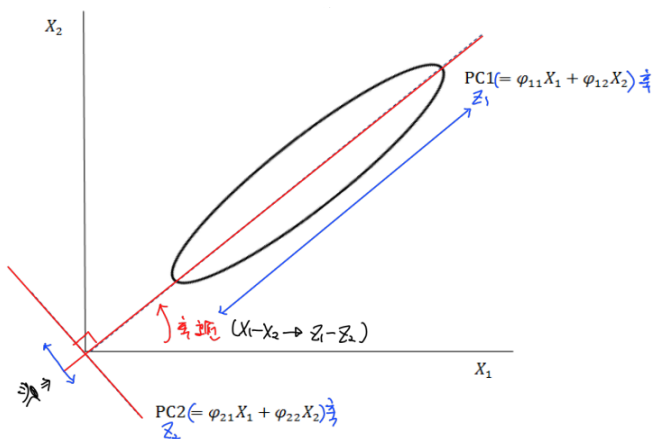
$X_s^T X_s$  는 대칭행렬이기 때문에

고유값 분해 후 고유벡터  $v_1 \quad \dots \quad , v_p$  는 서로 직교하게 됨

## 3

## Dimension Reduction

## 주성분 회귀분석 | 모델링 과정

② 선형변환을 통해 주성분( $Z$ )을 생성

$$Z = [Z_1 \quad \dots \quad Z_p]^T = X_S^T \begin{bmatrix} v_1 \\ \vdots \\ v_p \end{bmatrix}$$

각각의  $v_1 \quad \dots \quad v_p$  에 대해  $X_S$  라는 선형변환을 적용하여 새롭게 만들어진  $Z_1 \quad \dots \quad Z_p$  축들 역시 서로 직교하게 됨

## 3

## Dimension Reduction



## 주성분 회귀분석 | 모델링 과정

## 주성분(Z)의 특징에 대해 알아보자!

② 선형변환을 통해 주성분(Z)을 생성

①

 $Z_1 \dots Z_p$ 는 데이터의

분산정도를 최대한 설명할 수 있는 축임

( $\because$  각각에 대응되는  $v_1 \dots v_p$ 는

분산을 최대한 설명하는 축)

②

 $\lambda_1 > \dots > \lambda_p$ 인 경우 $Z = [Z_1 \dots Z_p]^T = X_S^T \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$   
Z 축들의 분산정도 크기는

아래 대소 관계를 따름

 $V(Z_1) > \dots > V(Z_p)$ 각각의  $v_1 \dots v_p$ 에 대해  $X_S$ 라는 선형변환을 적용하여새롭게 만들어진 Z 축들 역시 서로 직교하게 됨  
해당 내용에 대한 구체적인 증명 과정은

24-2 리드오프 3주차 전반부 교안 내용 참고

## 주성분 회귀분석 | 모델링 과정

③ 주성분( $Z$ ) 을 설명변수로 하는 선형회귀모델을 Fitting

$$y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_p Z_p$$



$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \cdots + \hat{\theta}_p Z_p$$

⋮

이 경우  $Z_1 \quad \dots \quad Z_p$  축은 서로 직교하기 때문에  
결과적으로 해당 회귀모델은 **다중공선성 문제를 해결** 가능

## 주성분 회귀분석 | 모델링 과정

주성분( $Z$ ) 을 생성할 때 전체  $X_s$  데이터를 이용하기 때문에  
주성분의 일부( $Z_1, \dots, Z_{M < P}$ ) 만으로 회귀모델을 Fitting 하더라도  
데이터의 손실은 발생하지 않음

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \dots + \hat{\theta}_P Z_P$$



Dimension Reduction( $M < P$ )

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \dots + \hat{\theta}_M Z_M$$

## 주성분 회귀분석 | 모델링 과정

주성분( $Z$ )을 생성할 때 전체  $\mathbf{X}$  데이터를 이용하기 때문에  
주성분의 일부( $Z_1, \dots, Z_{M < P}$ )  회귀모델을 Fitting 하더라도

그렇다면 축소된 차원 수인  **$M$ 의 최적값**은  
어떻게 결정해볼 수 있을까?

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \dots + \hat{\theta}_P Z_P$$



Dimension Reduction( $M < P$ )

$$\hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \dots + \hat{\theta}_M Z_M$$

## 3

## Dimension Reduction

주성분 회귀분석 | 최적의  $M$  값 결정

$$Y = \theta_0 + \varepsilon$$

⋮ **Lowest Test Error**

$$Y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_k Z_k + \varepsilon$$

⋮

$$Y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_p Z_p + \varepsilon$$



$$\therefore M = k$$

작성할 수 있는 모든 주성분 회귀 모델들의

Test Error 값을 비교해 가장 값이 작은 모델의 차원수를  $M$  으로 결정



CV, bootstrap, AIC, BIC,  $C_p$  등을 통해서도 수행 가능



## 주성분 회귀분석 | 모델링 과정 정리

## 주성분 회귀모델의 구축과정

- ① 설명변수  $X$ 를 표준화한  $X_s$  를 만듦 ( $X_{s_{ij}} = (X_{ij} - \bar{X}_j)/s_j$ )
- ②  $X_s^T X_s$  에 대해 고유값 분해를 진행하여  $Z$  축들을 정의한 뒤,  
Test Error 값을 가장 작게 하는 최적의  $Z$  축 개수(=  $M$ ) 결정  
(CV, bootstrap, AIC, BIC,  $C_p$  등을 통해서도 수행 가능)
- ③ 정의한  $Z$  축들과 원래  $Y$  데이터를 이용해, 선형회귀모델을 Fitting  
$$y = \theta_0 + \theta_1 Z_1 + \cdots + \theta_M Z_M \rightarrow \hat{y} = \hat{\theta}_0 + \hat{\theta}_1 Z_1 + \cdots + \hat{\theta}_M Z_M$$

4

Shrinkage

## Shrinkage

### Shrinkage (Regularization)

OLS 과정에서 목적식의 **회귀계수에 대한 제약조건을 추가**하여  
일부 계수의 추정치를 0으로 수축(shrink)시키는 방법

설명변수( $X$ ) 개수가 많아지면 다중공선성과 차원의 저주로 인해  
모델의 예측 성능이 저하될 수 있음



정규화(Regularization)를 통해 해당 문제점을 해결해보자!

## Shrinkage

Ex) 가상의 목적식

$$\operatorname{argmin}_{\beta} \left\{ \sum_i (y_i - \hat{y}_i)^2 + 10000\beta_3^2 + 10000\beta_4^2 \right\}$$

 $\beta_3$ 와  $\beta_4$ 의 값이 조금만 증가해도

전체 식의 값이 큰 폭으로 증가하여 모델 학습과정에 있어 지장이 생김

## Shrinkage

Ex) 가상의 목적식



이 경우  $\beta_3$ 와  $\beta_4$  회귀계수의 크기에 대해 Penalty를 부과(정규화)하여  
해당 문제를 해결할 수 있음

$\beta_3$ 와  $\beta_4$ 의 값이 조금만 증가해도



정규화를 적용해볼 수 있는 방법에는 어떤 것들이 있을까?

# 4

## Shrinkage

### Shrinkage

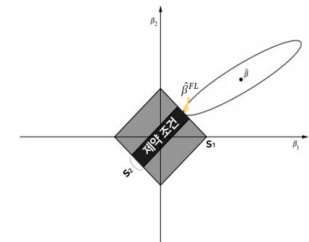
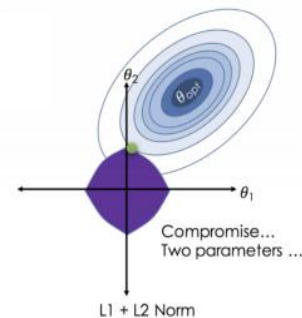
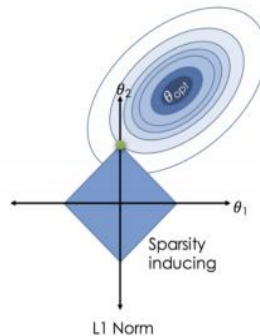
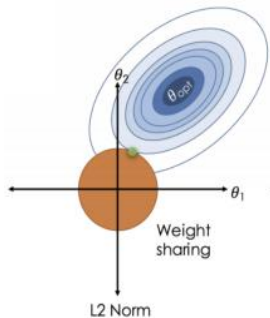
#### Regularization 의 종류

Ridge

Lasso

Elastic - Net

Fused Lasso



# 4

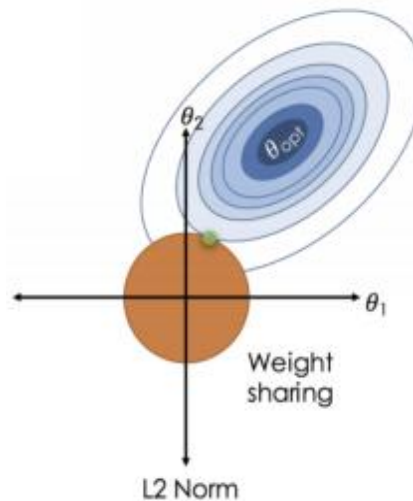
## Shrinkage

### Ridge Regression

Ridge Regression (L2 Regularization)

OLS 목적식에 대해

L2 Norm 형태의 제약조건을 회귀계수에 부여한 회귀 모델



## 4

## Shrinkage

## Ridge Regression | 목적식

Ridge Regression 의 목적식

$$\hat{\beta}_{Ridge} = \underset{\beta}{argmin} \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 \text{ subject to } \sum_j \beta_j^2 \leq s$$

$$= \underset{\beta}{argmin} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda \sum_j \beta_j^2 \right\}$$

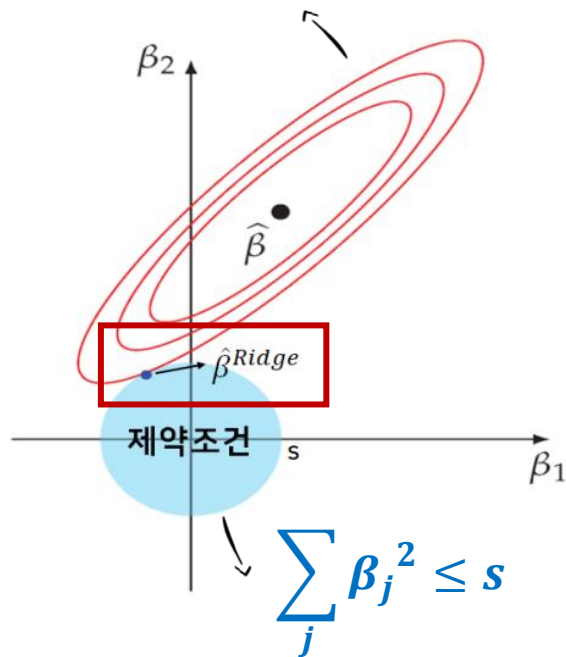


## 4

## Shrinkage

## Ridge Regression | 기하학적 해석

$$\sum_i (y_i - \beta_0 - \sum_j \beta_j x_i)^2 = RSS$$



$\hat{\beta}_{Ridge}$ 의 기하학적 위치

제약조건을 충족하기 위해

파란색 원에 존재하는 회귀계수( $\hat{\beta}$ )들 중  
RSS 값을 최소화하는 것을 찾아야 함



$\hat{\beta}_{Ridge}$ 는 타원과 원의 접점에 존재함

## Ridge Regression | 하이퍼파라미터의 영향

Ridge Regression 의 목적식

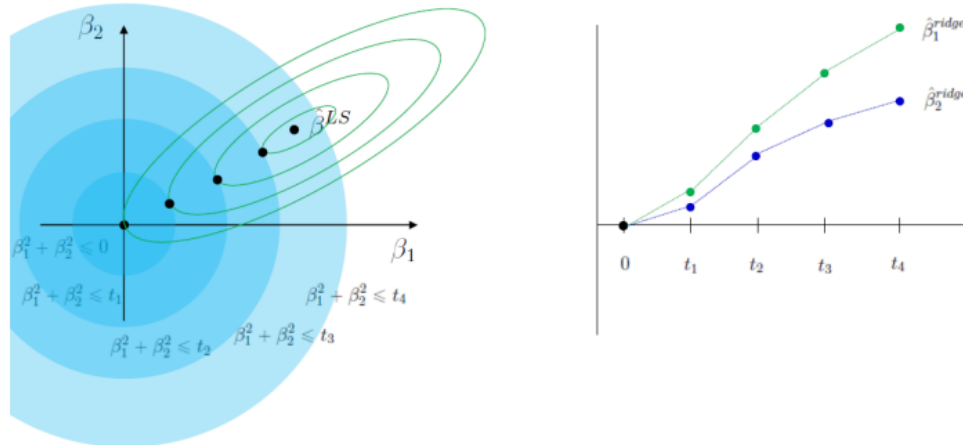
$$\hat{\beta}_{Ridge} = \underset{\beta}{argmin} \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 \text{ subject to } \sum_j \beta_j^2 \leq \textcolor{red}{s}$$
$$= \underset{\beta}{argmin} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \textcolor{red}{\lambda} \sum_j \beta_j^2 \right\}$$

목적식의 하이퍼파라미터  $\textcolor{red}{s}$ ,  $\textcolor{red}{\lambda}$ 에 주목해보자!

## 4

## Shrinkage

## Ridge Regression | 하이퍼파라미터의 영향



하이퍼파라미터  $s, \lambda$ 와 계수 추정량의 관계

**$s$  값이 작아질수록** ( $= \lambda$  의 값이 커질수록)

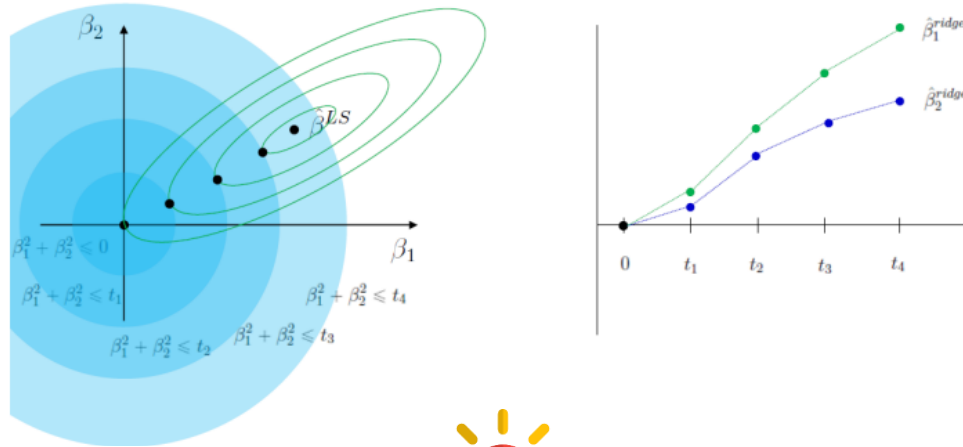
회귀계수들에 적용되는 Penalty 가 커지며,

**계수 추정량의 크기는 작아지게 됨**

## 4

## Shrinkage

## Ridge Regression | 하이퍼파라미터의 영향



하이퍼파라미터  $\lambda$  수 추정량의 관계

극단적으로  $\lambda$  값이 0이 되면 제약 조건식 부분이 사라져  
원래 OLS 목적식과 같아지게 되므로 정규화 과정의 의미가 사라짐



**Cross-Validation**을 통해 최적의  $\lambda$  값을 찾아야 함

## Ridge Regression | 특징

①  $\hat{\beta}_{Ridge}$  (Ridge Estimator)는 Unique 한 Solution이 도출됨

----- 목적식에서  $\hat{\beta}_{Ridge}$  (Ridge Estimator) 를 구하는 과정 -----

$$\begin{aligned} Q(\beta) &= (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta \\ &= Y^TY - 2\beta^TX^TY + \beta^T(X^TX + \lambda I)\beta \end{aligned}$$

$$\rightarrow \frac{\partial}{\partial \beta} Q(\beta) = 2X^TY + 2(X^TX + \lambda I)\beta = 0$$

$$\therefore \hat{\beta}_{Ridge} = (X^TX + \lambda I)^{-1}X^TY$$

## Ridge Regression | 특징

①  $\hat{\beta}_{Ridge}$  (Ridge Estimator)는 **Unique 한 Solution**이 도출됨

목적식에서  $\hat{\beta}_{Ridge}$  (Ridge Estimator)를 구하는 과정



$$Q(\beta) = (Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta$$

회귀계수들에 제약조건을 공평하게 적용하려면

반드시 설명변수들에 대한 **표준화 작업**을 먼저 진행해줘야 함

$$\rightarrow \frac{\partial Q(\beta)}{\partial \beta} = 2X^T Y + 2(X^T X + \lambda I)\beta = 0$$

$$\therefore \hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$$

## 4

## Shrinkage

## Ridge Regression | 특징

②  $\lambda I$  값을 통해 다중공선성 문제를 해결 가능

$$\hat{\beta}_{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$$

X 가 완전히 선형종속인 경우

$(X^T X + \lambda I)^{-1}$  가 존재하므로

$\hat{\beta}_{Ridge}$  계산이 가능해짐

선형종속은 아니지만  
상관관계가 높은 경우

$\det(X^T X + \lambda I)^{-1}$  의 크기를 조정하여

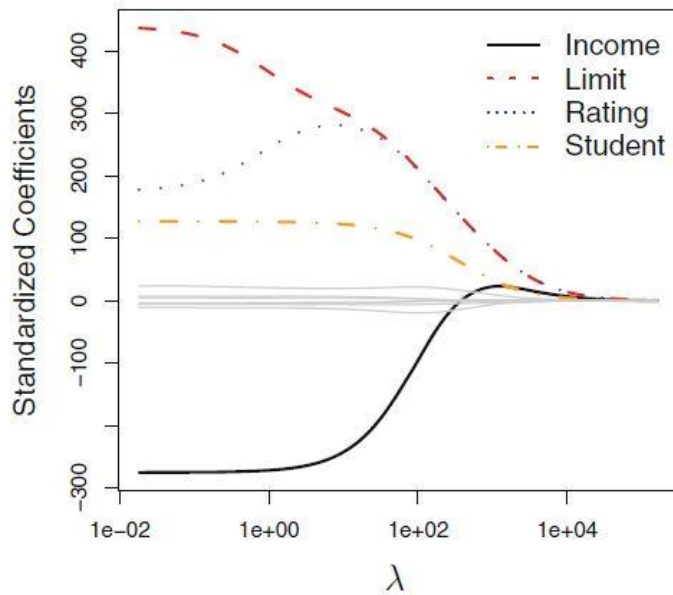
$\hat{\beta}_{Ridge}$  의 분산정도를 낮춤

## 4

## Shrinkage

## Ridge Regression | 특징

③  $\hat{\beta}_{Ridge}$  (Ridge Estimator) 를 통해선 변수선택이 불가능함



$\lambda$  값(Penalty 정도)이 커질수록  
 $\beta$  계수들은 거의 동시에  
0에 가까워지기 때문임

다중공선성의 원인이 되는 변수 역시 완전히 제거 불가능!



# 4

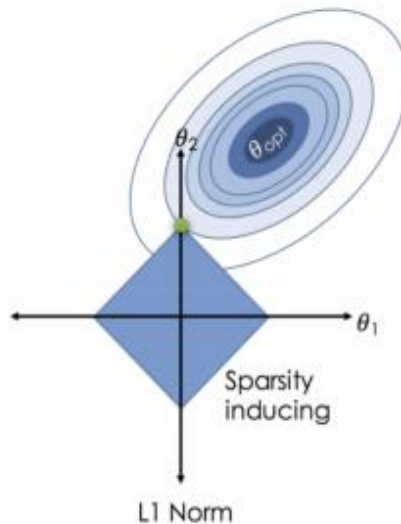
## Shrinkage

### Lasso Regression

Lasso Regression (L1 Regularization)

OLS 목적식에 대해

L1 Norm 형태의 제약조건을 회귀계수에 부여한 회귀 모델



## Lasso Regression | 목적식

Lasso Regression 의 목적식

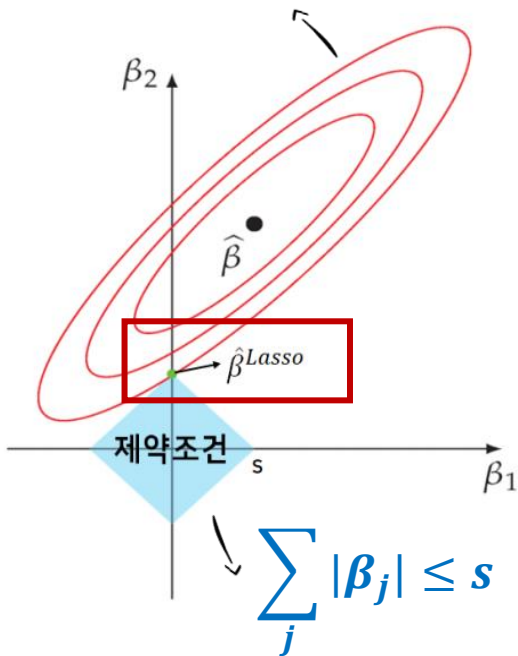
$$\begin{aligned}\hat{\beta}_{Lasso} &= \underset{\beta}{\operatorname{argmin}} \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 \text{ subject to } \sum_j |\beta_j| \leq s \\ &= \underset{\beta}{\operatorname{argmin}} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda \sum_j |\beta_j| \right\}\end{aligned}$$

## 4

## Shrinkage

## Lasso Regression | 기하학적 해석

$$\sum_i (y_i - \beta_0 - \sum_j \beta_j x_i)^2 = RSS$$



$\hat{\beta}_{Lasso}$ 의 기하학적 위치

제약조건을 충족하기 위해

파란색 마름모에 존재하는 회귀계수( $\hat{\beta}$ )들 중

RSS 값을 최소화하는 것을 찾아야 함



$\hat{\beta}_{Lasso}$ 는 타원과 마름모의 접점에 존재함

## Lasso Regression | 하이퍼파라미터의 영향

Lasso Regression 의 목적식

$$\begin{aligned}\hat{\beta}_{Lasso} &= \underset{\beta}{\operatorname{argmin}} \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 \text{ subject to } \sum_j |\beta_j| \leq s \\ &= \underset{\beta}{\operatorname{argmin}} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda \sum_j |\beta_j| \right\}\end{aligned}$$

목적식의 하이퍼파라미터  $s$ ,  $\lambda$ 에 주목해보자!

## 4

## Shrinkage

## Lasso Regression | 하이퍼파라미터의 영향

Ridge 회귀와 동일!

하이퍼파라미터  $s$ ,  $\lambda$ 와 계수 추정량의 관계 $s$  값이 작아질수록 (=  $\lambda$  의 값이 커질수록)

회귀계수들에 적용되는 Penalty가 커지며,

0으로 축소되는 계수 추정량들이 많아지게 됨

큰 $\lambda$ 값	작은 $\lambda$ 값
적은 변수 (계수가 0이 됨)	많은 변수
간단한 모델	복잡한 모델
해석 쉬움	해석 어려움
높은 학습오차 (과소적합 위험 증가)	낮은 학습오차 (과적합 위험 증가)

## Lasso Regression | 하이퍼파라미터의 영향

Ridge 회귀와 동일!

하이퍼파라미터  $s, \lambda$ 와 계수 추정량의 관계

$s$  값이 작아질수록 (=  $\lambda$  의 값이 커질수록)

회귀계수에 대한



추정치가 커지며

Ridge 회귀와 마찬가지로

**Cross-Validation**을 통해 **최적의  $\lambda$  값**을 찾아야 함!

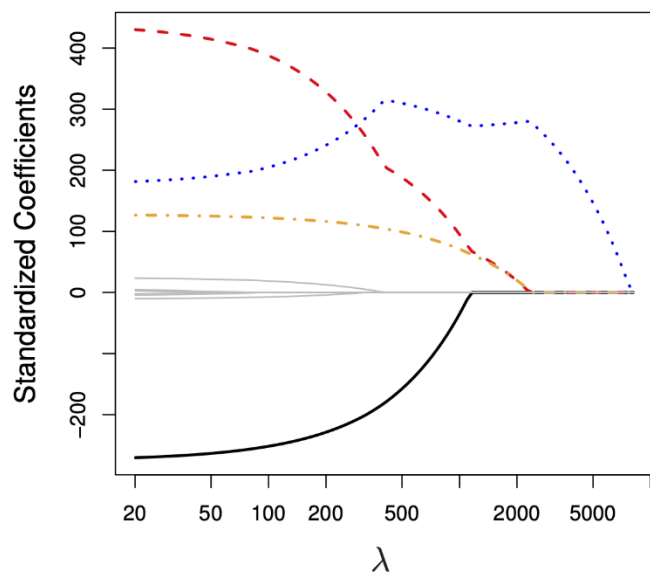
큰 $\lambda$ 값	작은 $\lambda$ 값
적은 변수 (계수가 0이 됨)	많은 변수
간단한 모델	복잡한 모델
해석 쉬움	해석 어려움
높은 학습오차 (과소적합 위험 증가)	낮은 학습오차 (과적합 위험 증가)

## 4

## Shrinkage

## Lasso Regression | 특징

①  $\hat{\beta}_{Lasso}$  (Lasso Estimator) 를 통해 변수 선택이 가능함



$\lambda$  값(Penalty 정도) 이 커짐에 따라  
 $\hat{\beta}$  값이 정확히 0으로 축소됨



변수 선택 가능!

## 4

## Shrinkage

## Lasso Regression | 특징

①  $\hat{\beta}_{Lasso}$  (Lasso Estimator) 를 통해 변수 선택이 가능함



설명변수(X)들 간에 상관관계가 높은 경우  
반응변수(Y) 예측에 유의미한 설명변수도 0으로 보낼 수 있음

따라  
림



따라서 다중공선성이 존재하는 경우  
Lasso Regression **모델의 예측 성능이 떨어지게** 됨

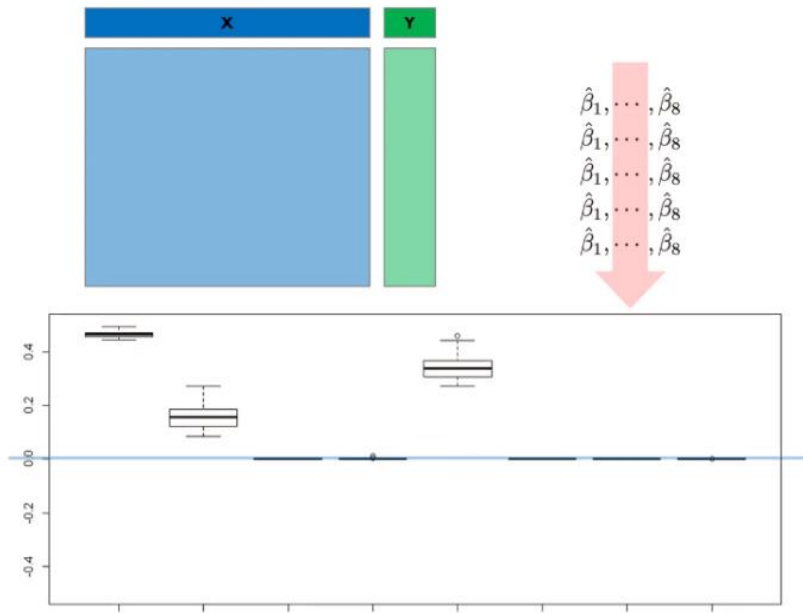


## 4

## Shrinkage

## Lasso Regression | 특징

② 데이터가 달라지더라도  $\hat{\beta}_{Lasso}$  값은 크게 변하지 않음



분할시킨 데이터셋 각각을  
모델에 넣어  $\hat{\beta}_{Lasso}$ 을 구해본 결과  
 $\hat{\beta}_{Lasso}$  값의 변동이 적음을 확인

## 4

## Shrinkage

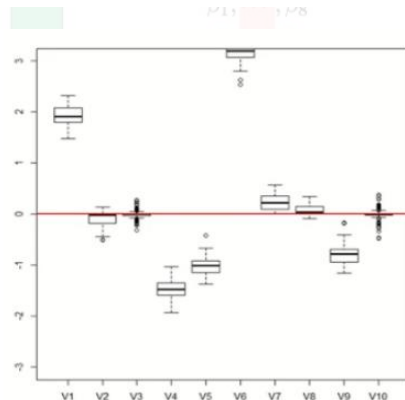
## Lasso Regression | 특징

② 데이터가 달라지더라도  $\hat{\beta}_{Lasso}$  값은 크게 변하지 않음

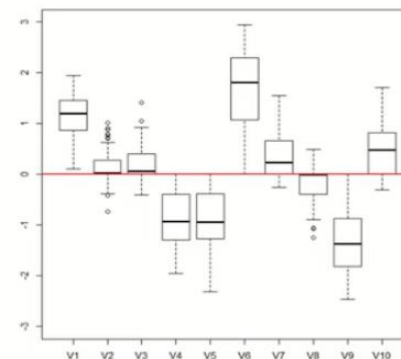


설명변수(X)들 간에 다중공선성이 존재하는 경우

$\hat{\beta}$ 의 변동성이 커져 변수선택의 결과는 일관되지 않을 수 있음



변수간 상관관계가 작을 경우



변수간 상관관계가 클 경우

근해본 결과  
음을 확인

## 4

## Shrinkage

## Lasso Regression | 특징

③  $\hat{\beta}_{Lasso}$  는 Unique 한 Solution이 도출되지 않음

$$\underset{\beta}{argmin} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda \sum_j |\beta_j| \right\}$$

제약 조건식(=  $\lambda \sum_j |\beta_j|$ ) 이 절댓값 형태로,  
 $\beta$ 에 대한 미분이 불가능하므로 유일한 해를 가지지 않음

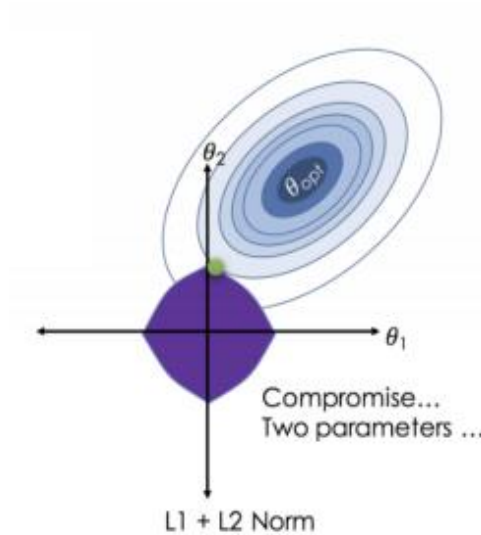
# 4

## Shrinkage

### Elastic-Net

#### Elastic-Net

Ridge 및 Lasso 의 정규화 방식을 혼합한 형태의 회귀 모델



## 4

## Shrinkage

## Elastic-Net | 목적식

Elastic-Net 의 목적식

$$\hat{\beta}_{Elastic} = \underset{\beta}{argmin} \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2$$

subject to  $(1 - \alpha) \sum_j |\beta_j| + \alpha \sum_j \beta_j^2 \leq s$  where  $\alpha = \lambda_2 / (\lambda_1 + \lambda_2)$

$$= \underset{\beta}{argmin} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda_1 \sum_j |\beta_j| + \lambda_2 \sum_j \beta_j^2 \right\}$$

## Elastic-Net | 목적식의 해석

Elastic-Net 의 제약 조건식은 **Strictly Convex** 하다는 성질을 가짐

⋮

서로 강하게 상관된 변수들이  
함께 선택되는 현상

Strictly Convex 는 Grouping Effect 를 항상 보장



따라서 Elastic-Net은 서로 상관성이 있는 변수들을  
동시에 선택 혹은 제거할 수 있음

## Elastic-Net | 목적식의 해석

Elastic-Net 의 제약 조건식은 **Strictly Convex**하다는 성질을 가짐



Elastic Net 을 통해 추정된 계수들은  
어떻게 Grouping Effect 를 보장하는 걸까?

구체적인 수식을 통해 알아보자!

따라서 Elastic-Net은 서로 상관성이 있는 변수들을  
동시에 선택 혹은 제거할 수 있음

변수들이  
현상

## Elastic-Net | 목적식의 해석

----- Elastic-Net에서의 Grouping Effect 식 -----

$$|\hat{\beta}_{Elastic_i} - \hat{\beta}_{Elastic_j}| \leq \left( \sum_i y_i / \lambda_2 \right) \sqrt{2(1 - p_{ij})}$$

 $p_{ij}$ :  $X_i$  와  $X_j$  간의 상관계수(음의 상관관계를 갖는 경우  $X_k$  대신  $-X_k$  를 사용)

$p_{ij}$  값이 커질수록  $\hat{\beta}_{Elastic_i}$  과  $\hat{\beta}_{Elastic_j}$  간의  
차이가 작아지므로 값이 서로 유사해짐을 확인할 수 있음



## Elastic-Net | 목적식의 해석

Elastic-Net에서



Shrinkage Effect 식

상관관계가 높은 설명변수들의 회귀계수 값은 유사해지므로

 $\lambda$ 의 크기를 크게 하여

해당 회귀계수 값들을 동시에 0으로 축소시킬 수 있음

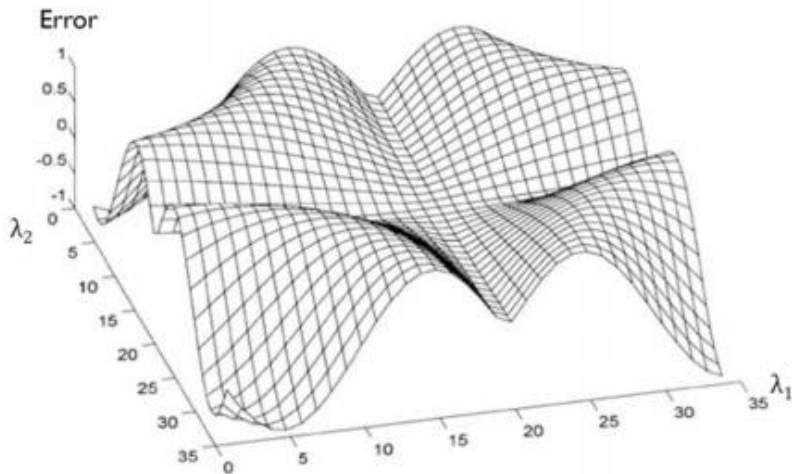
(음의 상관관계들 갖는 경우  $x_k$  대신  $-x_k$  들 사용)다중공선성 존재 시 예측성능이 떨어지는 **Lasso의 한계점을 보완 가능** $P_{ij} = \frac{P_{i,j}}{P_{i,i} P_{j,j}}$ 

차이가 작아지므로 값이 서로 유사해짐을 확인할 수 있음

# 4

## Shrinkage

### Elastic-Net | Parameter( $\lambda$ ) Selection



목적식의 파라미터  $\lambda_1, \lambda_2$  는  
Grid Search 방식을 적용하여  
최적의  $\lambda$  값들을 도출할 수 있음



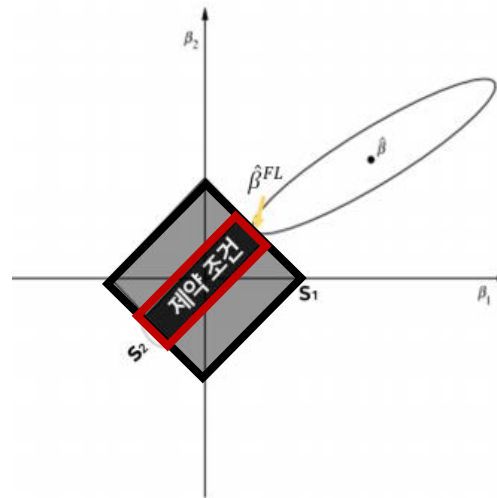
# 4

## Shrinkage

### Fused Lasso

#### Fused Lasso

Lasso보다 강한 L1 Norm 형태의 제약조건을 적용한 회귀 모델



검은 테두리 영역 : Lasso Regression의 제약조건

빨간 테두리 영역 : Fused Lasso Regression의 제약조건

## Fused Lasso | 목적식

Fused Regression 의 목적식

$$\hat{\beta}_{FL} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_i \left( y_i - \beta_0 - \sum_j \beta_j x_i \right)^2 + \lambda_1 \sum_j |\beta_j| + \lambda_2 \sum_j |\beta_i - \beta_j| \right\}$$

① 물리적으로 근접한 관측치들의 회귀계수를 비슷한 값으로 추정하게 됨

② 기존의 Lasso에 비해 Strict한 제약 공간을 형성

# 4

## Shrinkage

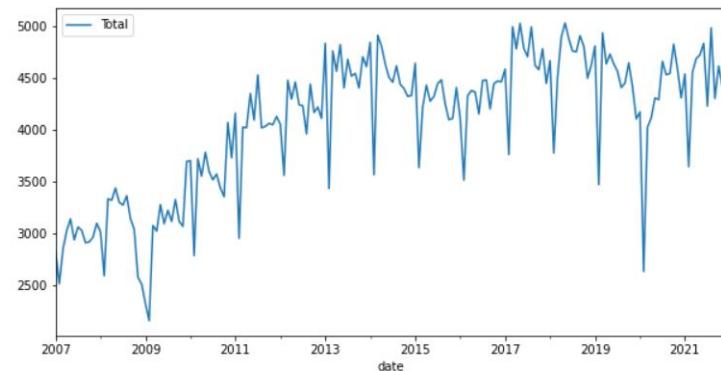
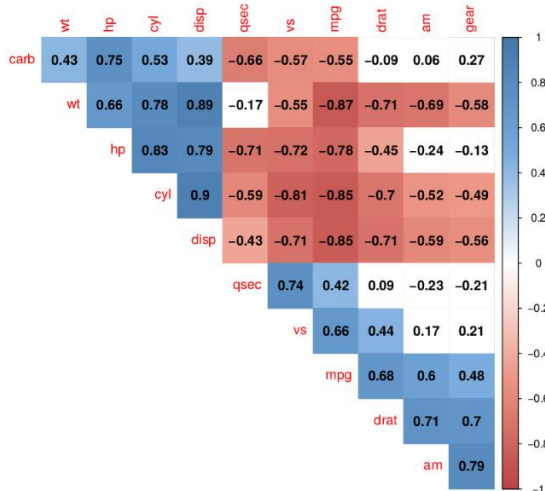
### Fused Lasso | Elastic-Net과의 비교

#### Elastic-Net

상관관계가 높은 관측치들 간의  
회귀계수 값을 비슷하게 추정

#### Fused Lasso

물리적으로 근접한 관측치들 간의  
회귀계수 값을 비슷하게 추정



# 4

## Shrinkage

### Fused Lasso | Elastic-Net과의 비교



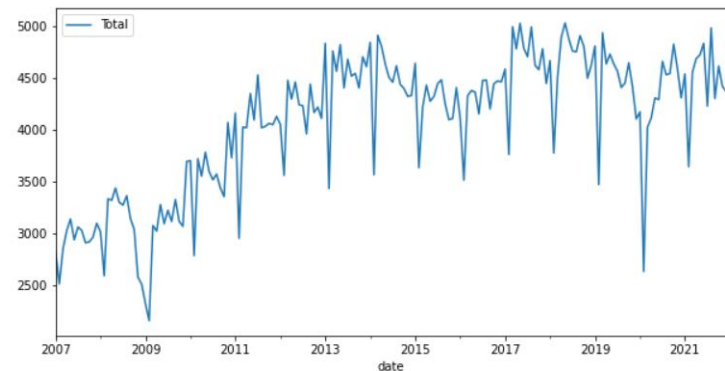
공간적, 시간적으로 **순서**가  
존재하는 데이터의 특성을  
설명하기에 용이함



주로 **시계열데이터**를 다룰 때  
많이 사용하는 회귀 모델

#### Fused Lasso

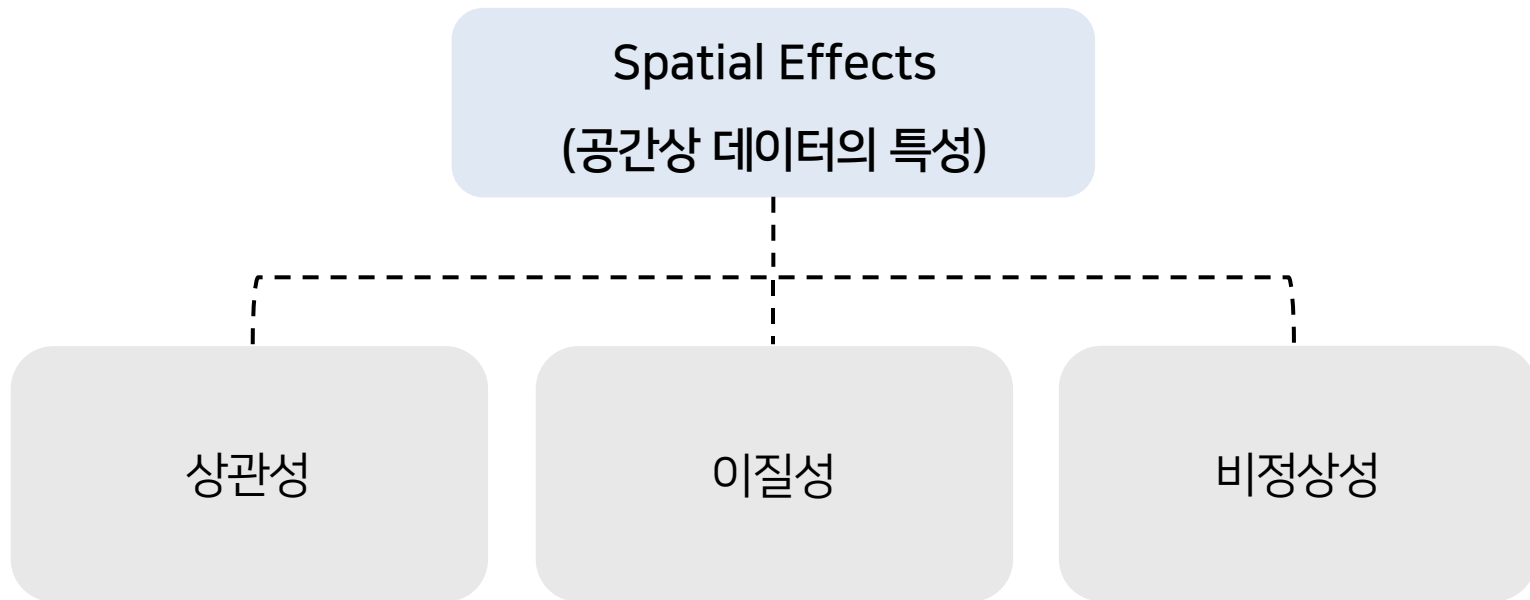
물리적으로 근접한 관측치들 간의  
회귀계수 값을 비슷하게 추정



# 5

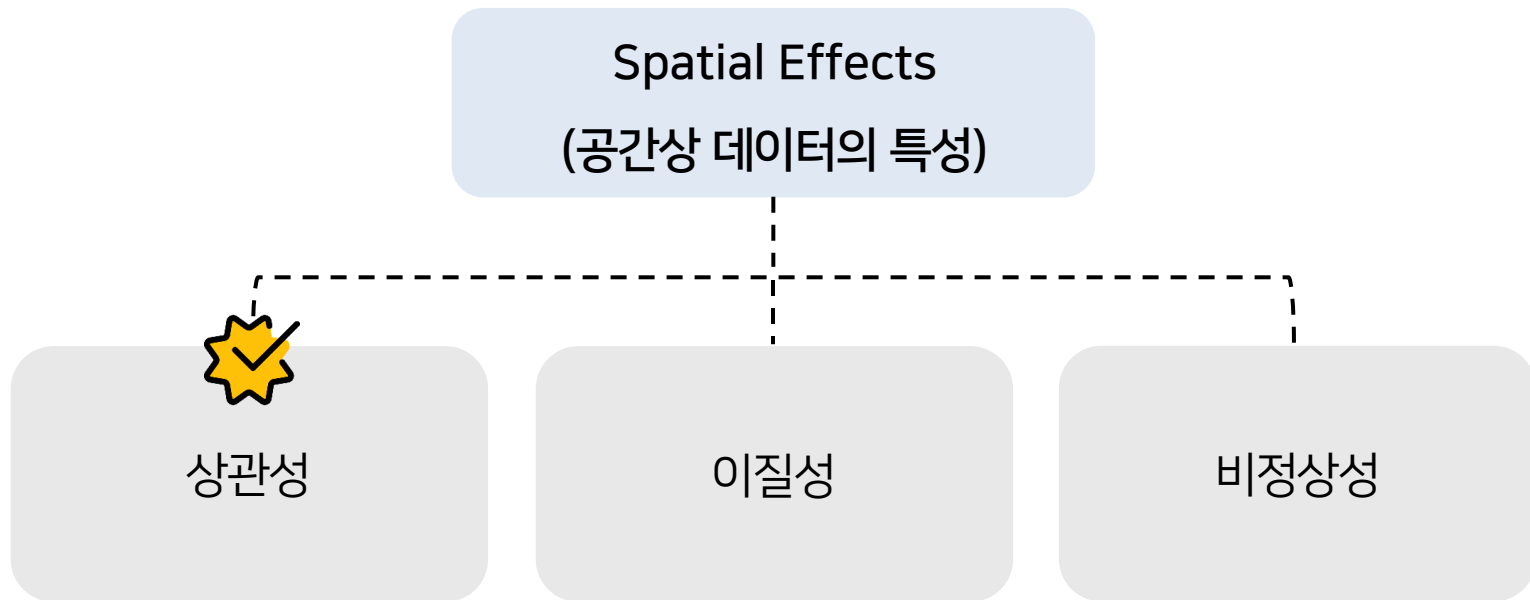
## Regression for Spatial Data

## Spatial Effects





## Spatial Effects

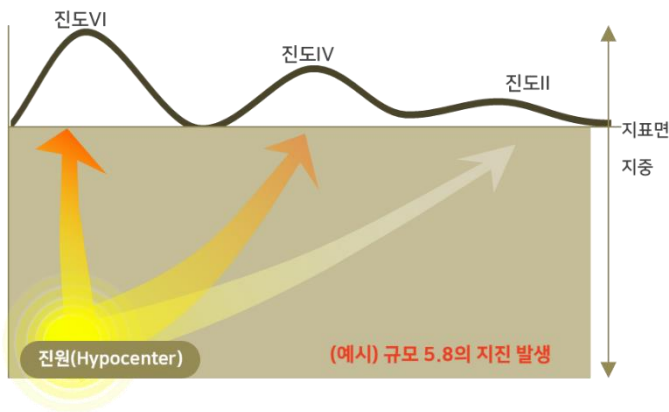


Spatial Effects 중 상관성에 대해 알아보자!

## Spatial Effects | 상관성

### 상관성 (Dependence)

한 위치에서 발생한 사건이 그 주변 지역의 사건 발생에도 영향을 미치는 것



Ex) 지진

진원 지역으로부터 가까울수록

피해의 정도가 클 것이며,

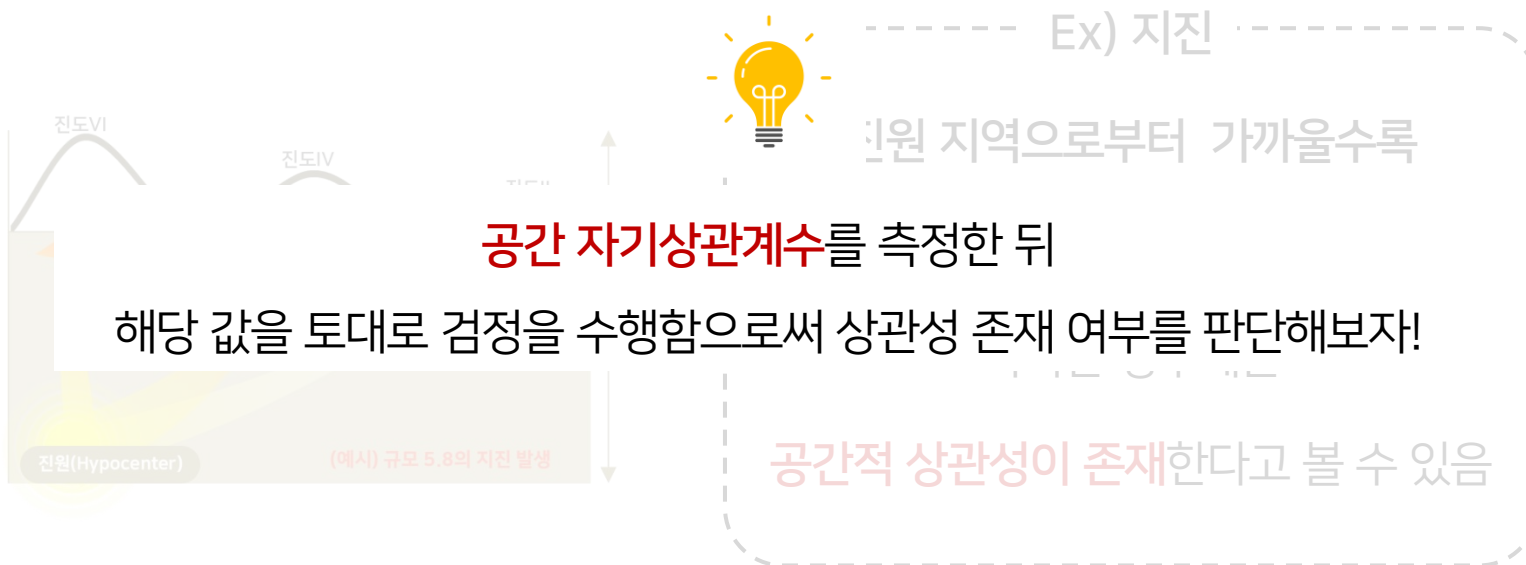
이러한 경우에는

**공간적 상관성이 존재**한다고 볼 수 있음

## Spatial Effects | 상관성

### 상관성 (Dependence)

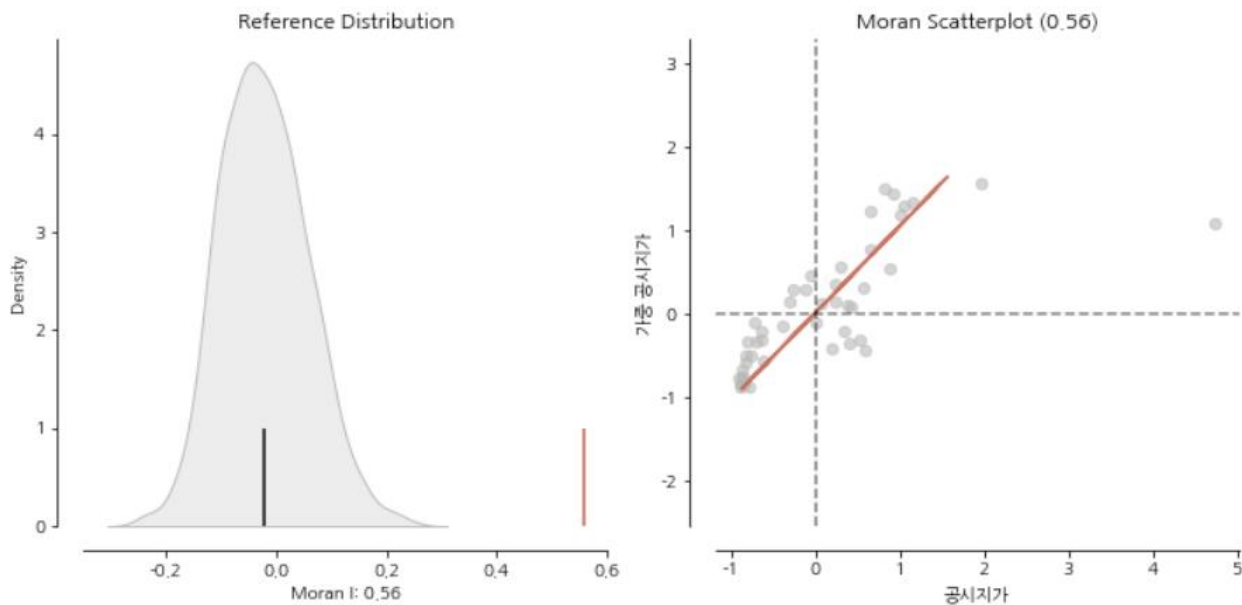
한 위치에서 발생한 사건이 그 주변 지역의 사건 발생에도 영향을 미치는 것



## 공간 자기상관계수

공간 자기상관계수 (Spatial Autocorrelation)

공간적 자기상관정도를 수치로써 나타내는 지표

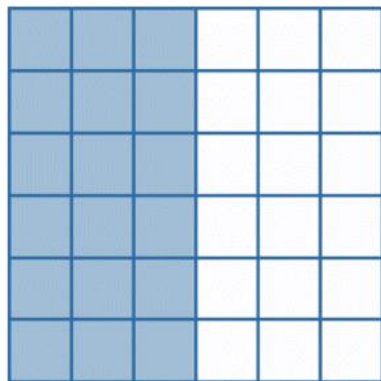


## 공간 자기상관계수

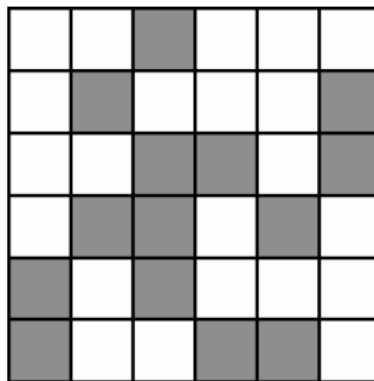
일반적으로 관측치의 자기상관성이

높은 Positive 값인 경우 근처의 관측치들과 비슷한 값을 가지며

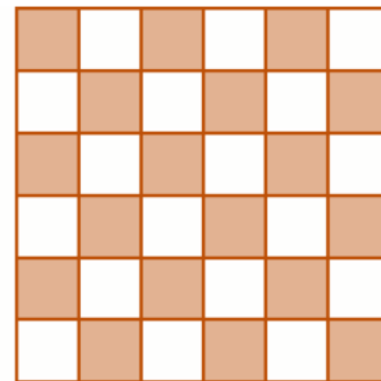
높은 Negative 값인 경우 근처의 관측치들과 상반된 값을 가짐



Positive spatial  
autocorrelation



No spatial  
autocorrelation



Negative spatial  
autocorrelation

## 공간 자기상관계수

일반적으로 관측치의 자기상관성이

높은 Positive 값인 경우 근처의 관측치들과 비슷한 값을 가지며

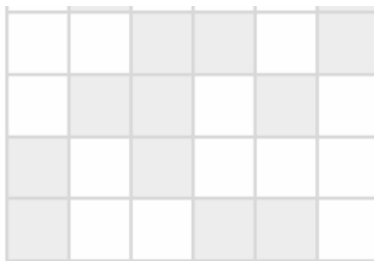
높은 Negative 값인 경우 근처의 관측치들과 상반된 값을 가짐



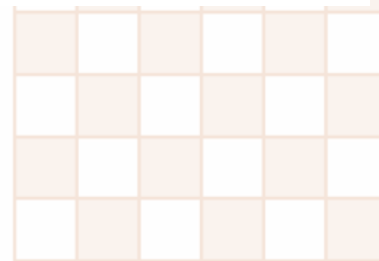
공간 자기상관계수를 측정하는 방법에는 어떤 것들이 있을까?



Positive spatial  
autocorrelation



No spatial  
autocorrelation



Negative spatial  
autocorrelation

## 공간 자기상관계수

공간 자기상관계수의 측정 방법

Moran's I

LISA

Getis-Ord

## 공간 자기상관계수 | Moran's I

Moran's I

전역(Global) 단위의 자기상관 정도를 평가할 때 사용되는 통계적 지표

Moran's I 계산식

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N \sum_{j=1}^N w_{ij}) \sum_{i=1}^N (Y_i - \bar{Y})^2}$$

$N$ : 지역단위 수,  $Y_k$ :  $k$  지역의 속성,  $w_{ij}$ : 공간 가중치



## 공간 자기상관계수 | Moran's I

①  $I$  값은  $-1 \sim 1$  사이의 값을 가짐

②  $I$  값의 절댓값이 1에 가까울수록 강한 공간 자기상관정도를 보임

③ 부호에 따라서 Positive / Negative 양상의 정도를 결정함



## 공간 자기상관계수 | Moran's I

전체 공간에 대하여 Global 하게 자기상관 정도를 측정하기 때문에  
지역 단위의 패턴(자기상관 정도) 측정이 어렵다는 한계점이 존재



LISA 지표와 Getis-Ord 지표를 사용하여 한계점을 극복해보자!

## 공간 자기상관계수 | LISA

LISA (Local Indicator of Spatial Association)

지역(Local) 단위의 자기상관정도를 평가할 때 사용되는 통계적 지표

⋮

전체 공간패턴에 대한 개별 지역의 기여도를 측정하고,  
공간적 이웃과의 상관성을 계산하여  
특정 지역에서의 공간적 클러스터링을 평가함

## 공간 자기상관계수 | LISA

LISA의 계산식

$$I_i = \frac{(Y_i - \bar{Y})}{S_Y} \sum_{j \neq i}^N w_{ij} \frac{(Y_j - \bar{Y})}{S_Y} \quad \text{where} \quad S_Y = \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2 / (N - 1)}$$

$I_i$  는 특정  $i$  지역에서의 LISA 값으로,  
해당 값이 양수인지 음수인지에 따라서 다른 클러스터링을 형성함

## 공간 자기상관계수 | LISA

LISA의 계산식

$$I_i = \frac{(Y_i - \bar{Y})}{S_Y} \sum_{j \neq i}^N w_{ij} \frac{(Y_j - \bar{Y})}{S_Y} \quad \text{where} \quad S_Y = \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2 / (N - 1)}$$

 $I_i$ 의 값이 **크고 양수**일수록

①  $i$  지역에 인접한 부분은 **비슷한 속성**을 가져서 **강한 클러스터**를 형성

② 공간의 전체적 패턴이 **양의 자기상관정도**를 가지는 데 큰 기여를 함

## 공간 자기상관계수 | LISA

LISA의 계산식

$$I_i = \frac{(Y_i - \bar{Y})}{S_Y} \sum_{j \neq i}^N w_{ij} \frac{(Y_j - \bar{Y})}{S_Y} \quad \text{where} \quad S_Y = \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2 / (N - 1)}$$

 $I_i$ 의 값이 **크고 음수**일수록

①  $i$  지역에 인접한 부분은 **상반된 속성**을 가져서 **약한 클러스터 형성**

② 공간의 전체적 패턴이 **음의 자기상관정도**를 가지는 데 큰 기여를 함

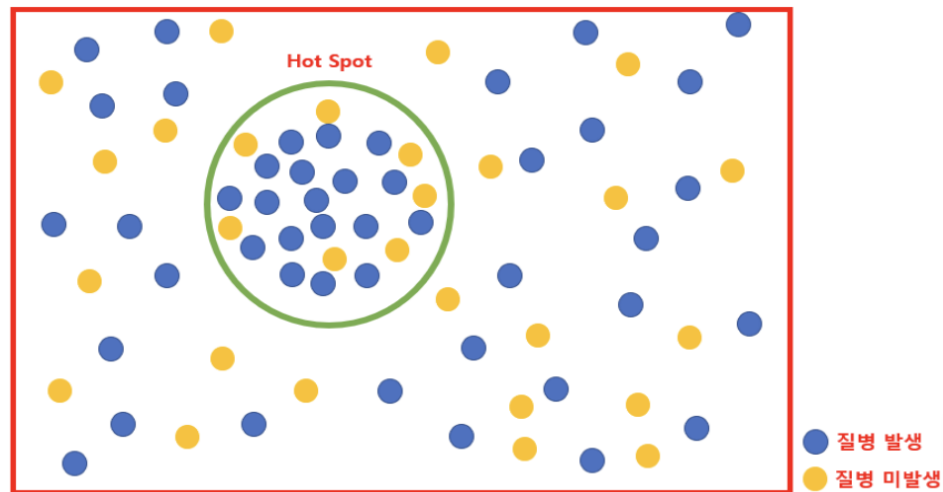
# 5

## Regression for Spatial Data

### 공간 자기상관계수 | Getis-Ord

Getis-Ord

공간 내에서 **Hot Spot** 또는 **Cold Spot** 존재 여부를 판단하는 지표



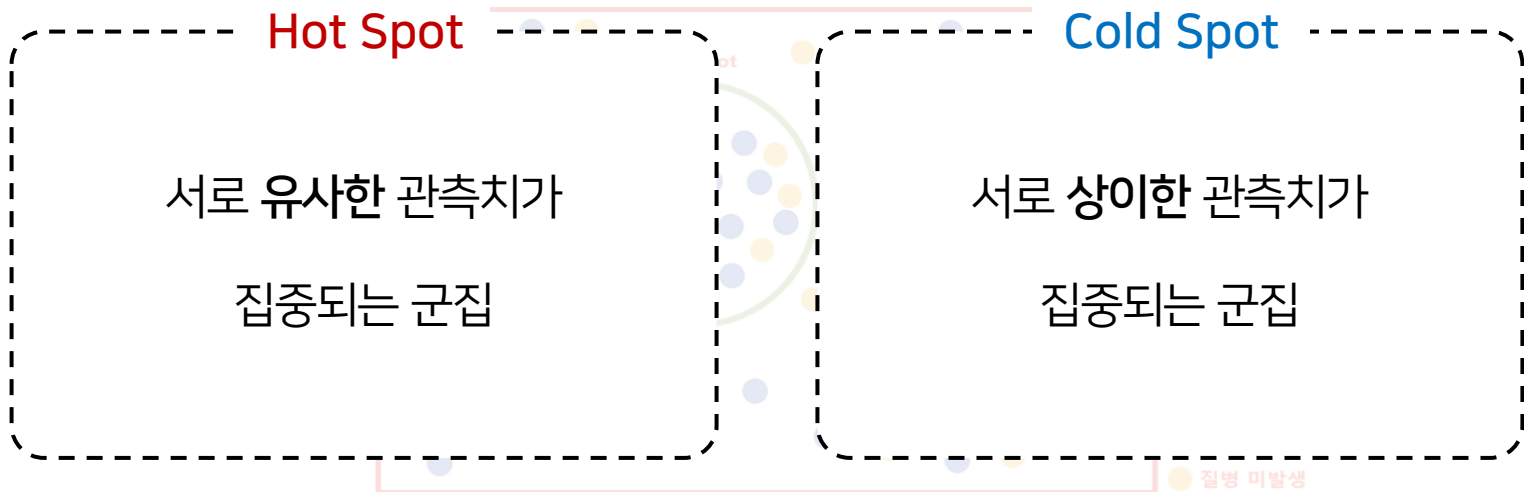
# 5

## Regression for Spatial Data

### 공간 자기상관계수 | Getis-Ord

Getis-Ord

공간 내에서 **Hot Spot** 또는 **Cold Spot** 존재 여부를 판단하는 지표





## 공간 자기상관계수 | Getis-Ord

Getis-Ord 계산식

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} Y_i Y_j}{\sum_{i=1}^N \sum_{j=1}^N Y_i Y_j} \quad (For \quad \forall i \neq j)$$


⋮

$G$  값이 양수라면 Hot Spot,  $G$  값이 음수라면 Cold Spot이

전체 공간 내에 존재한다고 판단할 수 있음

## 공간 자기상관계수 | Getis-Ord

Getis-Ord 계산식

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N \frac{Y_i Y_j}{\sum_{i=1}^N \sum_{j=1}^N Y_i Y_j} (For \forall i \neq j)$$


공간 자기상관계수를 통해  
**공간상관성 여부를 검정**할 수 있는  
 상관성 검정에 대해 알아보자!

G 값

Hot이

전체 공간 내에 존재한다고 판단할 수 있음

## 5

## Regression for Spatial Data

## 공간 자기상관계수 | 상관성 검정

공간 자기상관 지표 계산



표준화 수행



Z-검정 수행

Moran's I

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

LISA

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}}$$

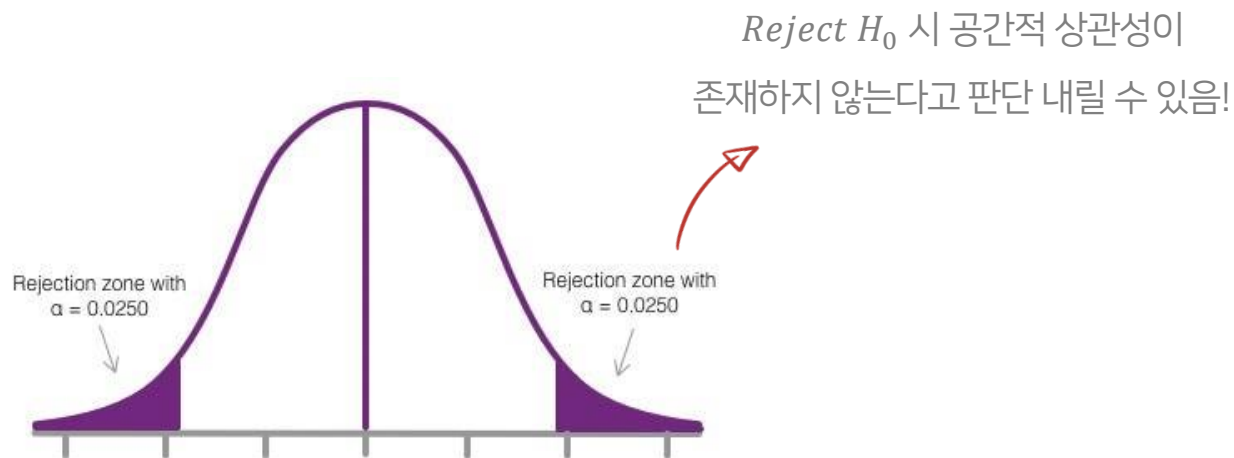
Getis-Ord

$$Z_G = \frac{G - E[G]}{\sqrt{V[G]}}$$

*Reject  $H_0$  if*

$$|z_0| > z_{1-\alpha/2, n-2}$$

## 공간 자기상관계수 | 상관성 검정



## Null Hypothesis

$H_0$ : 전역적 또는 지역적으로 공간 데이터에 상관성이 존재하지 않음

*Reject  $H_0$  if  $|z_0| > z_{1-\alpha/2, n-2}$  (Significance Level:  $\alpha$ )*

## 공간 가중치 행렬

Moran's I

$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_{i=1}^N \sum_{j=1}^N w_{ij}) \sum_{i=1}^N (Y_i - \bar{Y})^2}$$

LISA

$$I_i = \frac{(Y_i - \bar{Y})}{S_y} \sum_{j \neq i}^N w_{ij} \frac{(Y_j - \bar{Y})}{S_y} \quad \text{where} \quad S_y = \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2 / (N - 1)}$$

Getis-Ord

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N \sum_{j=1}^N w_{ij} Y_i Y_j}{\sum_{i=1}^N \sum_{j=1}^N \sum_{j=1}^N Y_i Y_j} \quad (\text{For } \forall i \neq j)$$

## 공간 가중치 행렬

Moran's I



$$I = \frac{N \sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\left( \sum_{i=1}^N \sum_{j=1}^N w_{ij} \right) \sum_{i=1}^N (Y_i - \bar{Y})^2}$$

LIS

앞서 언급한 공간 자기상관계수에  
공간 가중치행렬  $W$ 가 고려된다는 점을 확인할 수 있음

$$s^2/(N-1)$$

Getis-Ord

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N \sum_{j=1}^N w_{ij} Y_i Y_j}{\sum_{i=1}^N \sum_{i=1}^N \sum_{j=1}^N Y_i Y_j} \quad (For \forall i \neq j)$$

## 공간 가중치 행렬

공간 가중치 행렬 (Spatial Weights Matrix)

특정 관측치에 대하여 이웃 여부에 따라 1 또는 0의 값을 가지는 행렬로,  
정의에 따라 다양하게 행렬을 구성하여 공간 인접도를 표현하는 데 사용됨

Binary Contiguity  
Weights

Distance Based  
Weights

....



KNN 알고리즘 등 여러 방법들을 적용 가능하며,  
0, 1 대신 연속된 실수 값을 이용하기도 함!

## 5

## Regression for Spatial Data

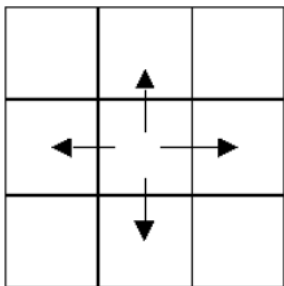
## 공간 가중치 행렬 | Binary Contiguity Weights

## Binary Contiguity Weights

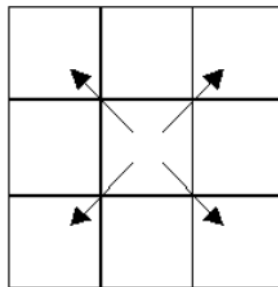
공간상의 인접 여부에 따라 위치들 간의 Neighbor 여부를 판단

$$w_{ij} = \begin{cases} 1 & (\text{if } i \text{ and } j \text{ is neighbor}) \\ 0 & (\text{Otherwise}) \end{cases}$$

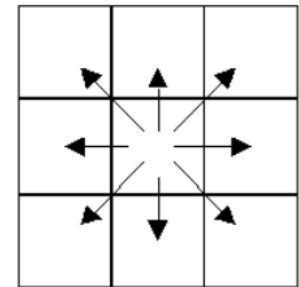
Rooks Case



Bishops Case



Queen's (Kings) Case





## 5

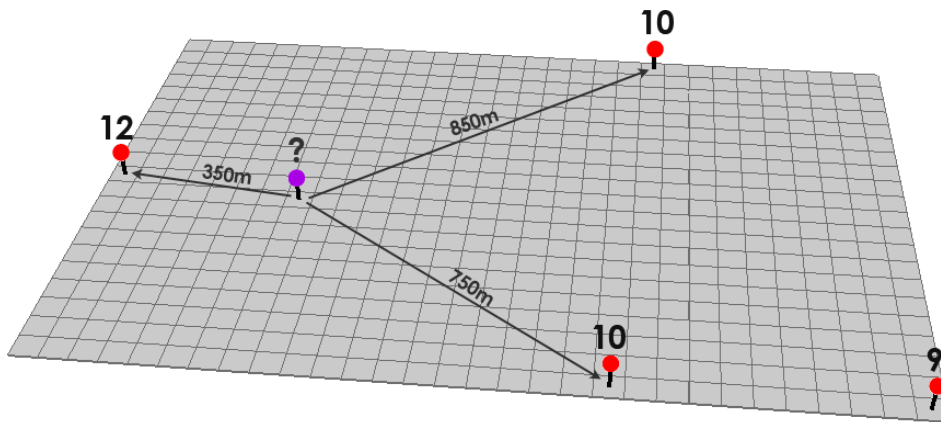
## Regression for Spatial Data

## 공간 가중치 행렬 | Distance Based Weights

## Distance Based Weights

임의의 최소 거리  $d$  를 정의한 뒤, 이를 기준으로 Neighbor 여부를 판단

$$w_{ij} = \begin{cases} 1 & (d_{ij} < d) \\ 0 & (Otherwise) \end{cases}$$



## 공간 가중치 행렬 | 정규화

일반적으로 공간 가중치 행렬은 정규화 작업을 진행시켜준 뒤 사용

## ① Row Standardized Weights

가중치 행렬의 각 행의 합이  
1이 되도록 행 단위 정규화

$$w^*_{ij} = \frac{w_{ij}}{\sum_j w_{ij}}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

행렬 전체 원소의 합이 1이 되는 것을 확인!

## 공간 가중치 행렬 | 정규화

일반적으로 공간 가중치 행렬은 정규화 작업을 진행시켜준 뒤 사용

## ② Stochastic Weights

가중치 행렬 전체 원소의 합이  
1이 되도록 정규화

$$w^*_{ij} = \frac{w_{ij}}{\sum_{i,j} w_{ij}}$$

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1/7 & 1/7 & 1/7 \\ 1/7 & 0 & 1/7 \\ 1/7 & 1/7 & 0 \end{pmatrix}$$

행렬 전체 원소의 합이 1이 되는 것을 확인!

## 공간 가중치 행렬 | 정규화

일반적으로 공간 가중치 행렬은 정규화 작업을 진행시켜준 뒤 사용

## ② Stochastic Weights



공간 가중치 행렬 값이 0, 1 이 아닌

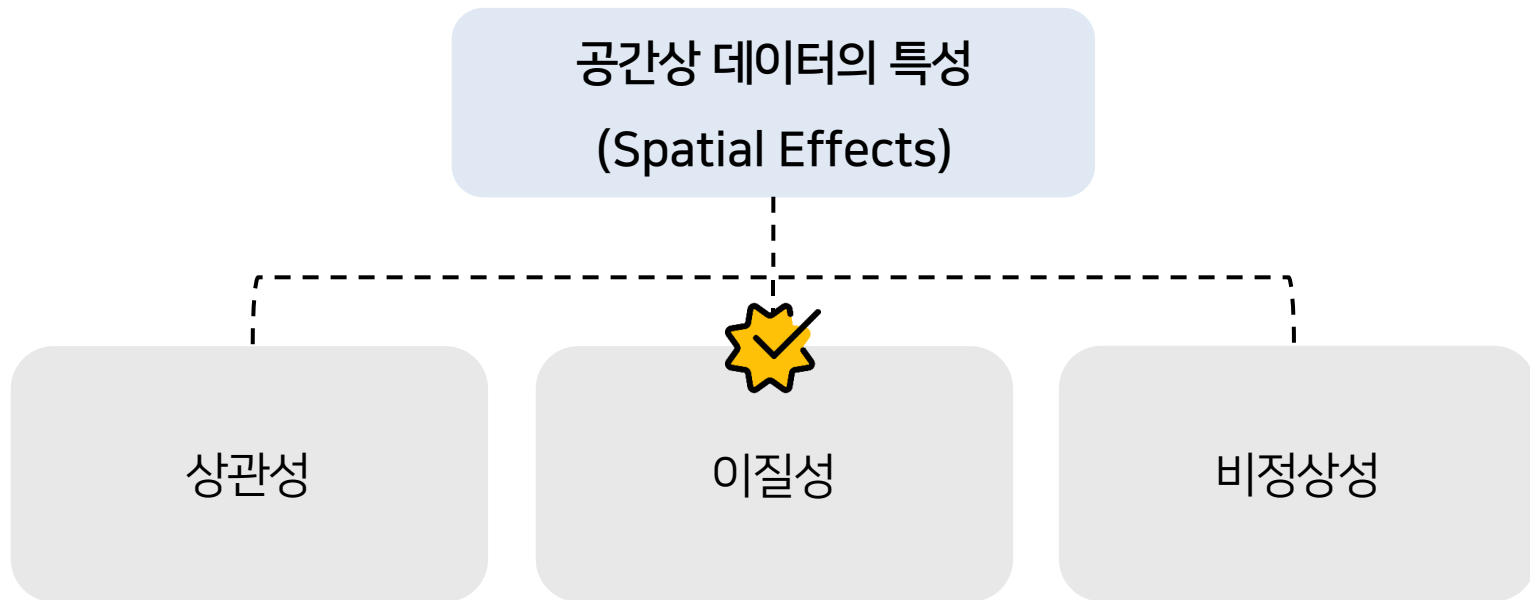
연속된 실수 값이라면 정규화를 작업을 하지 않아도 됨

$$w_{ij}^* = \frac{w_{ij}}{\sum_{i,j} w_{ij}}$$

7	1/7
1	1/7
7	0

행렬 전체 원소의 합이 1이 되는 것을 확인!

## Spatial Effects

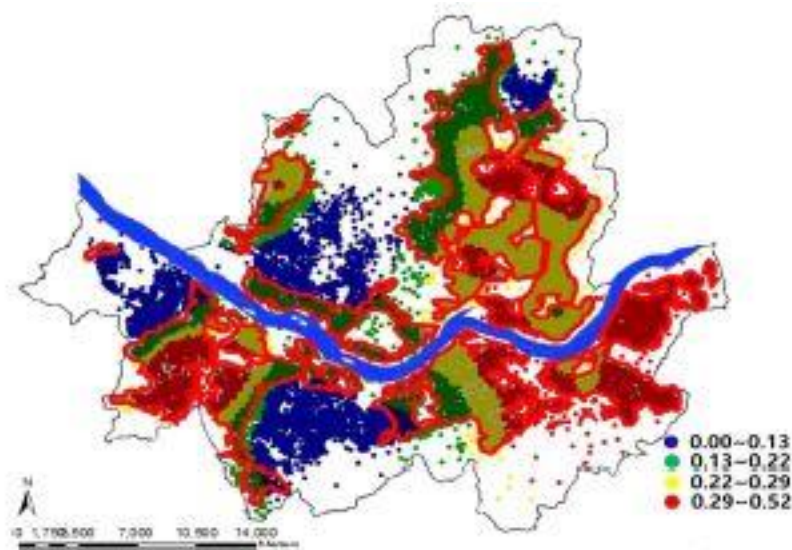


Spatial Effects 중 이질성에 대해 알아보자!

## Spatial Effects | 이질성

공간의 이질성 (Heterogeneity)

공간이 변화함에 따라 관측치의 특성이 달라지는 것



공간의 이질적 특성에 따른 소매업 매출 차이

## Spatial Effects | 이질성

Ex) 나라별 행복지수 분석

종교적 특성이 강한 나라

종교 > 돈

자본주의 국가

돈 > 종교

## Spatial Effects | 이질성



중요하게 인식하는 가치가 다른 두 나라에 대해  
변수(종교, 돈)별 균일한 회귀계수 값을 부여하는 것은 비합리적임



종교적 특성이 강한 나라

자본주의 국가

공간이 변함에 따라 차등적인 회귀계수 값을

부여해줌으로써 공간의 이질적 특성을 반영할 필요가 있음



## Spatial Effects | 이질성

단순선형 회귀모델에 대해 공간의 이질성을 반영한다면

아래와 같이 작성해볼 수 있음

이질성을 가정한 단순 선형회귀모델

$$y_i = x_i \beta_i + \varepsilon_i$$

$\beta_i$  : 관측치별 회귀계수 추정값

## Spatial Effects | 이질성

단순선형 회귀모델  해 공간의 이질성을  
반영한다면 아래와 같이 작성해볼 수 있음

그렇다면 이러한 **공간 이질성의 존재 여부를**  
**어떻게** 알아낼 수 있을까?

BP Test 를 통해  
공간에 따른 오차항의 분산 변화를 확인함으로써  
이질성 존재 여부를 파악해보자!

## Spatial Effects | 이질성 검정

## BP Test

Null Hypothesis vs Alternative Hypothesis

$$H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_p = 0 \text{ vs } H_1: \text{Not } H_0$$

데이터가 등분산성을 따라야 하므로, 귀무가설이 기각되지 않기를 원함

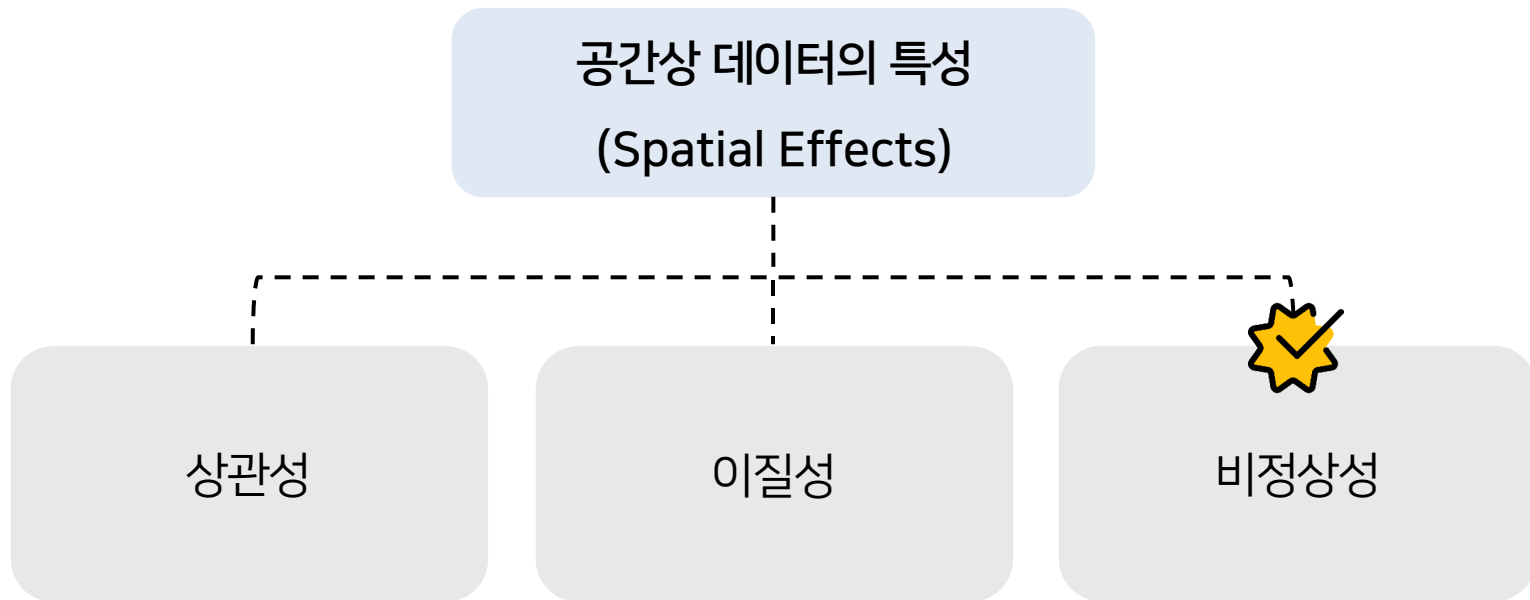
Test Statistic

$$\chi^2_{stat} = n R^2 \sim \chi^2_{p-1}$$

*Reject  $H_0$  if  $\chi^2_{stat} > \chi^2_{p-1, \alpha}$  (Significance Level:  $\alpha$ )*

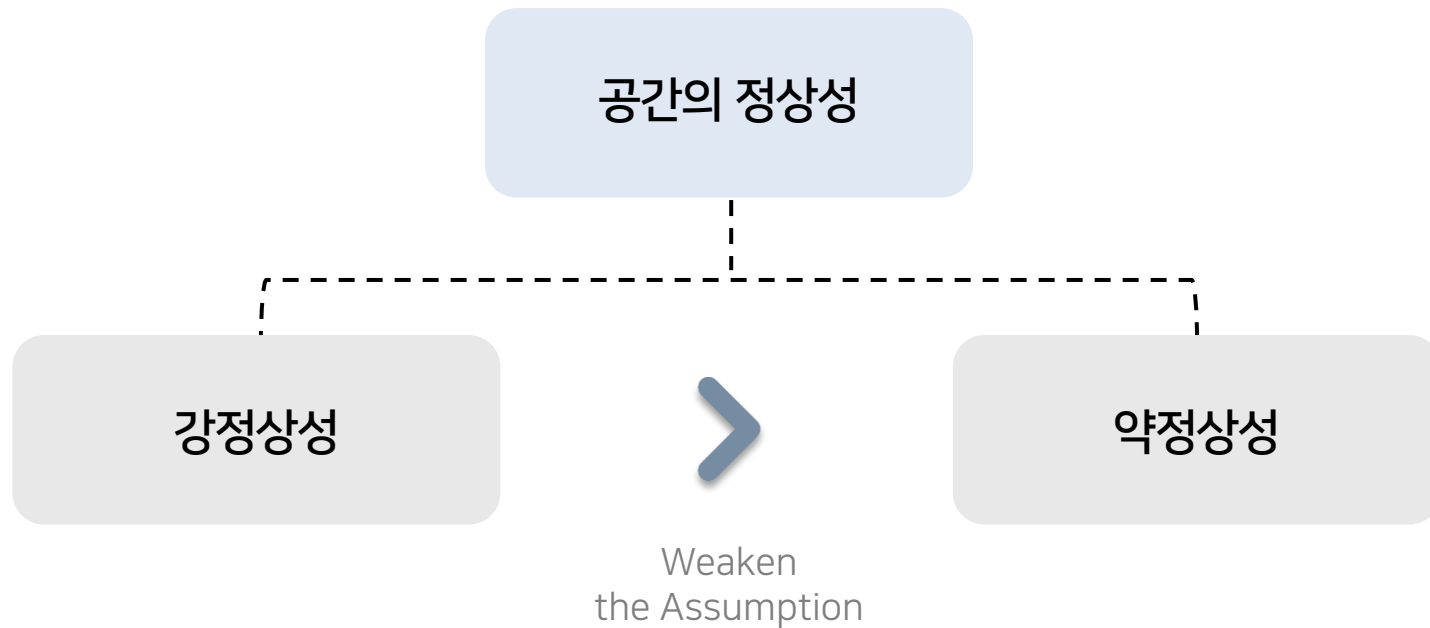
귀무가설을 Reject 하면 공간의 이질성을 지닌다고 결론을 내림!

## Spatial Effects



Spatial Effects 중 비정상성에 대해 알아보자!

## Spatial Effects | 비정상성



## Spatial Effects | 비정상성



강정상성과 약정상 간에는 어떤 차이가 있을까?

강정상성과 약정상성의 조건에 대해 알아보자!

Weaken  
the Assumption

## Spatial Effects | 비정상성

## 강정상성

종속변수의 확률 분포가 공간 lag 차(= h) 에만 의존하는 성질

강정상성을 수식으로 표현하면 아래와 같음

$$G(S_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \cdots \textcircled{1}$$

For  $\textcircled{1}$  &  $\forall_{n,h}$

$$\{G(S_i), G(S_{i+1}), \dots, G(S_{i+n})\} =^d \{G(S_{i+h}), G(S_{i+1+h}), \dots, G(S_{i+n+h})\}$$

## Spatial Effects | 비정상성

## 약정상성

강정상성의 조건을 완화한 것으로, 아래의 조건을 따름

$$\textcircled{1} E[G(S_i)] = E[G(S_j)] \text{ for } \forall i, j$$

: Constant Mean (공간 위치에 따라 평균 값은 변하지 않음)

$$\textcircled{2} V[G(S_i)] = V[G(S_j)] \text{ for } \forall i, j$$

: Constant Variance (공간 위치에 따라 분산 값은 변하지 않음)

$$\textcircled{3} \text{Cov}(G(S_i), G(S_j)) = \text{Cov}(G(S_{i+h}), G(S_{j+h})) \text{ for } \forall i, j, h$$

: Covariance 값은 반드시 lag-  $h$  에 대해서만 의존



## Spatial Effects | 비정상성

## 약정상성

강정상성의 조건을 완화한 것으로, 아래의 조건을 따름

$$\textcircled{1} E[G(S_i)] = E[G(S_j)] \text{ for } \forall i, j$$

: Constant Mean (공간 위치에 따라 평균 값은 변하지 않음)

약정상성 조건들 중 하나라도 위배되면,  
해당 공간 데이터 셋은 **Non-stationary 데이터**라고 판단

$$\textcircled{3} Cov(G(S_i), G(S_j)) = Cov(G(S_{i+h}), G(S_{j+h})) \text{ for } \forall i, j, h$$

: Covariance 값은 반드시 lag-  $h$  에 대해서만 의존

## Spatial Effects | 비정상성

비정상성

이질성

상관성



공간의 비정상성은  
확률분포가 고려되어  
상관성 및 이질성을 포괄하는 개념임



공간의 상관성 및 이질성 파악을 위한 Test 방법들은  
공간의 비정상성 여부를 파악하기 위해 사용 가능!

## 공간 자기회귀모델

공간 자기회귀모델 (Spatial Autoregressive Model)

공간 데이터에 상관성이 존재하는 경우

이를 통제하기 위해 사용하는 회귀 모형

```
graph TD; A[공간 자기회귀모델 (Spatial Autoregressive Model)] -.- B[Spatial Error Model]; A -.- C[Spatial Lag Model];
```

Spatial Error Model

Spatial Lag Model

## 공간 자기회귀모델 | Spatial Error Model

### Spatial Error Model

공간의 상관성이 오차항 부분에서 비롯되었다고 판단될 때  
사용할 수 있는 공간자기회귀모형

⋮

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + u$$

$\lambda$  : 오차항의 공간 자기상관계수

$W$  : 공간가중치행렬

$u$  : *iid* 한 분포를 따르는 오차항

## 공간 자기회귀모델 | Spatial Lag Model

### Spatial Lag Model

공간의 상관성이 반응변수 부분에서 비롯되었다고 판단될 때  
사용할 수 있는 공간자기회귀모형

⋮

$$Y = \rho WY + X\beta + \varepsilon$$

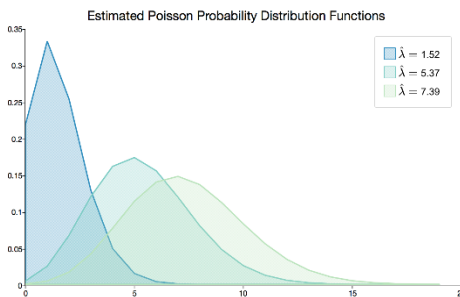
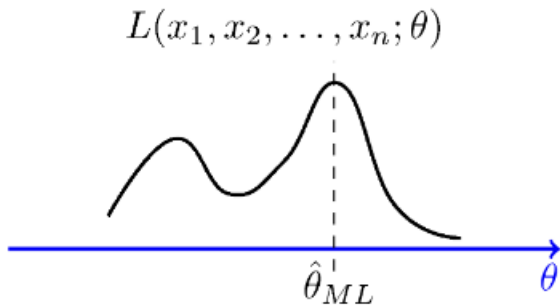
$\rho$  : 반응변수의 공간자기상관계수

$W$  : 공간 가중치행렬

$\varepsilon$  : *iid* 한 분포를 따르는 오차항

## 공간 자기회귀모델 | 모수 추정

Spatial Error Model 과 Spatial Lag Model 은  
 $\beta$  뿐만 아니라  $\rho, \lambda$  를 추가로 추정해줘야 함



OLS 를 통해 추정하게 되면  
 $\rho, \lambda$  의 불편성과 일치성을 보장하지  
 못하게 됨



MLE / GMM 기반의 최적화를 통해  
 $\beta, \rho, \lambda$  를 추정

## 5

## Regression for Spatial Data

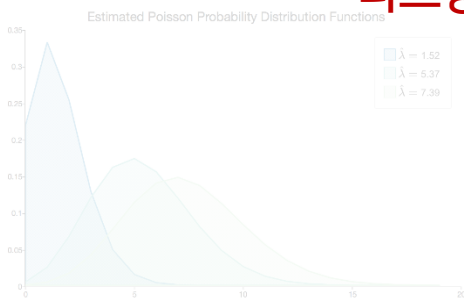
## 공간 자기회귀모델 | 모수 추정

Spatial Error Model 과 Spatial Lag Model 은

$\beta$  뿐만 아니라  $\rho, \lambda$  ? 가로 추정해줘야 함

그렇다면 Spatial Error Model과 Spatial Lag Model을  
각각 어떤 상황에서 적용해야 할까?

라그랑주 승수 검정을 통해 판단해보자!



MLE 기반의 최적화를 통해

$\beta, \rho, \lambda$  를 추정

## 공간 자기회귀모델 | 모델 선택

라그랑주 승수 검정을 통해

Spatial Error Model과 Spatial Lag Model 중 모델 선택

$$LM_{lag} = \frac{(e^T W Y / s^2)^2}{(W X \beta^T)(I - X(X^T X)^{-1} X^T)(W X \beta) / s^2 + tr(W^T W + W^2)} \sim \chi_1^2$$

$$LM_{error} = \frac{(e^T W e / s^2)^2}{tr(W^T W + W^2)} \sim \chi_1^2$$

$e$ : 잔차(Residuals) 벡터,  $s^2$ : 오차항 분산의 MLE 추정량

$\beta$ : 각각의 회귀 모형에서 추정한 회귀 계수 벡터



## 공간 자기회귀모델 | 모델 선택

## Null Hypothesis

$H_0$ : 오차항이나 설명변수에 의한 공간의 상관성이 존재하지 않음

*Reject  $H_0$  if  $|LM| > \chi^2_{1,\alpha}$  (Significance Level:  $\alpha$ )*

귀무가설을 Reject하면, 상관성이 존재한다고 결론 내림

⋮

$LM_{lag}$  이 유의하다는 결과가 나왔다면 Spatial Lag Model을,

$LM_{error}$  이 유의하다는 결과가 나왔다면 Spatial Error Model을 사용

두 Test 결과가 모두 유의하다는 결과가 나왔다면 Robust LM 통계량으로 추가적인 Test 수행

## 5

## Regression for Spatial Data

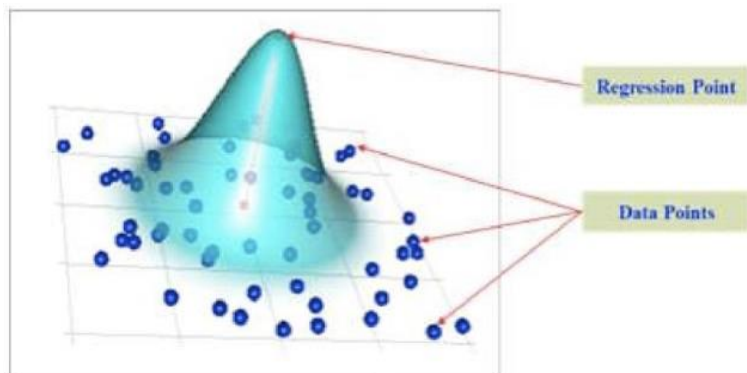
## 지리 가중회귀모델

지리 가중회귀모델 (Geographically Weighted Regression)

관측치마다 전부 다른 회귀계수 값을 추정하여

공간의 비정상성을 완전히 통제할 수 있는 국소회귀 모형

따라서 GWR을 이용하게 되면 상관계수와 이질성을 모두 통제할 수 있음!



일반회귀모형(OLS)

공간단위1	공간단위2
$\beta_1$	$\beta_2$
공간단위3	공간단위4
$\beta_3$	$\beta_4$

$$\beta_1 = \beta_2 = \beta_3 = \beta_4$$

공간가중회귀모형(GWR)

공간단위1	공간단위2
$\beta_1$	$\beta_2$
공간단위3	공간단위4
$\beta_3$	$\beta_4$

$$\beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4$$

## 지리 가중회귀모델

----- GWR의 모델 식 구조 -----

$$Y = X \otimes \beta + \varepsilon \quad (\otimes: \text{Componentwise Multiplication})$$

$$\text{where } \beta = \begin{bmatrix} \beta_0(u_1, v_1) & \cdots & \beta_p(u_1, v_1) \\ \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \cdots & \beta_p(u_n, v_n) \end{bmatrix}, \varepsilon \sim iid N(0, \sigma^2 I)$$

 $(u_i, v_i)$ :  $i$  번째 공간좌표

## 지리 가중회귀모델

## GWR Fitting 결과 예시

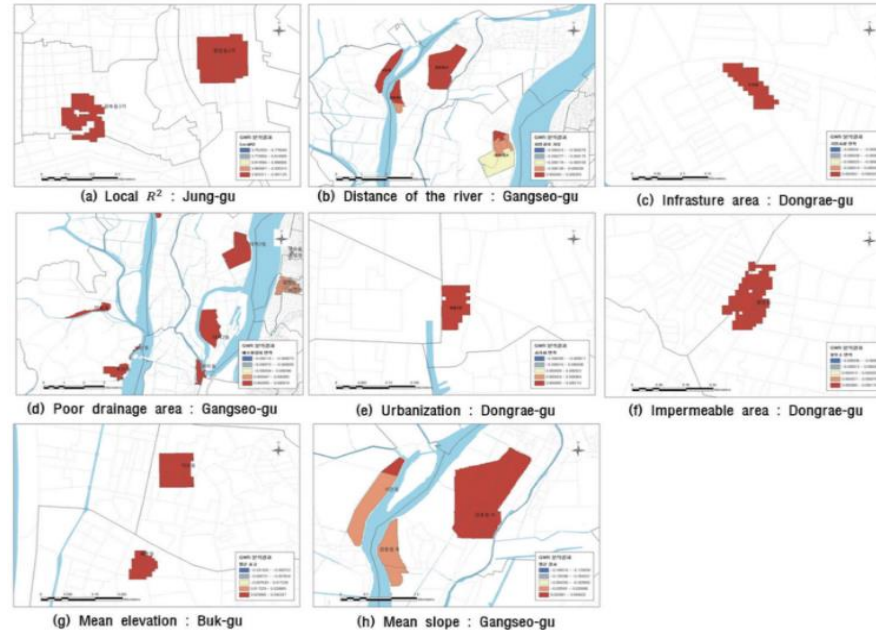


Fig 2. The distribution of the GWR coefficients

## 지리 가중회귀모델

GWR Fitting 결과 예시

Variables	Minimum value	Maximum value	Average value
Intercept	-0.06691	2.22807	0.30683
Infrastructure area	-0.00004	0.00003	-0.00001
Poor drainage area	-0.00012	0.00007	0.00002
Urbanization area	-0.00005	0.00012	0.00002
Impervious area	-0.00004	0.00017	0.00003
Mean elevation	-0.20183	0.4225	-0.00471
Mean slope	-0.16852	0.06463	-0.03617

Table 4. Estimation of the GWR model coefficients

## 지리 가중회귀모델

### GWR Fitting 결과 예시

Variables	Minimum value	Maximum value	Average value
Intercept	-0.06691	2.22807	0.30683
Infrastructure area	-0.00004	0.00003	-0.00001
Poc			
Ur			
In			
M			
Mean slope	-0.16852	0.06463	-0.03617

그렇다면 지리 가중회귀모델을 사용하기 전에  
비정상성 여부를 어떻게 검정할 수 있을까?

통계적 검정 과정에 대해 알아보자!

Table 4. Estimation of the GWR model coefficients

## 지리 가중회귀모델 | 비정상성 진단

## Null Hypothesis

$H_0$ : 일반선형회귀와 GWR 간에 유의미한 차이가 없음 (= 정상성을 가짐)

## Test Statistic

일반선형회귀와 GWR 로부터 계산된 RSS 값들을 기반으로 F-test를 진행

$$F_0 = \frac{RSS_{GWR} / df \text{ of } RSS_{GWR}(= m)}{RSS_{Origin} / df \text{ of } RSS_{Origin}(= n - p - 1)}$$

*Accept  $H_0$  if  $|F_0| < F_{1-\alpha/2, m, n-p-1}$  (Significance Level:  $\alpha$ )*

귀무가설을 Reject 하면 해당 공간 데이터는 비정상성 데이터라고 결론을 내림!

# 6

## Appendix



평가지표 |  $C_p$  계산 식 $C_p$  (Mallow's  $C_p$ )

$$\frac{1}{n} (RSS + 2d\hat{\sigma}^2)$$

$n$ : # of Training data set

$d$ : # of Parameters

(= # of 설명변수( $X$ ) + intercept +  $\sigma^2$ )

$\hat{\sigma}^2$ : MLE of the Error's Variance

$$(\hat{\sigma}^2 = RSS/n)$$

파란색 글씨 부분은 Likelihood 를

빨간색 글씨 부분은 변수에 따른 Penalty 정도를 나타냄

## 평가지표 | AIC 계산 식

----- **AIC(Akaike Information Criterion)** -----

$$n \log(2\pi\hat{\sigma}^2) + \textcolor{blue}{RSS/\hat{\sigma}^2} + \textcolor{red}{2d}$$

$n$ : # of Training data set

$d$ : # of Parameters

(= # of 설명변수( $X$ ) + intercept +  $\sigma^2$ )

$\hat{\sigma}^2$ : MLE of the Error's Variance

$$(\hat{\sigma}^2 = \textcolor{blue}{RSS}/n)$$

파란색 글씨 부분은 **Likelihood** 를

빨간색 글씨 부분은 **변수에 따른 Penalty 정도**를 나타냄

## 평가지표 | BIC 계산 식

--- **BIC**(Bayesian Information Criterion) ---

$$n \log(2\pi\hat{\sigma}^2) + RSS/\hat{\sigma}^2 + (\log n)d$$

$n$ : # of Training data set

$d$ : # of Parameters

(= # of 설명변수( $X$ ) + intercept +  $\sigma^2$ )

$\hat{\sigma}^2$ : MLE of the Error's Variance

$$(\hat{\sigma}^2 = RSS/n)$$

파란색 글씨 부분은 Likelihood 를

빨간색 글씨 부분은 변수에 따른 Penalty 정도를 나타냄

## Best Subset Selection | 작동원리

① 같은 개수의 변수를 사용한 후보군 중 RSS 값이 가장 낮은 모델 선별

feature :  $X_1, X_2, X_3, X_4$  인 경우

$\mu_0 : Y = \beta_0 + \varepsilon$  (Null Model)

$k = 1$

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$\vdots$

$\mu_1$

$$Y = \beta_0 + \beta_4 X_4 + \varepsilon$$

$k = 3$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

$\vdots$

$\mu_3$

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$k = 2$

$\mu_2$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$\vdots$

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

$k = 4$

$\mu_4$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

## Best Subset Selection | 작동원리

② 선별된  $p$  개의 최종 후보 모델들에 대하여

Analytical Methods 평가지표에 따라 최적의 모델 결정

$\mu_k$  : 각 group에서 training RSS가 가장 작은 Model

$$\mu_0 : Y = \beta_0 + \varepsilon$$

$$\mu_1 : Y = \beta_0 + \beta_4 X_4 + \varepsilon$$

$$\mu_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\mu_3 : Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$\mu_4 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$



$C_p, AIC, BIC, \text{Adjusted } R^2$  중 1개를 기준으로 최적의 Model 을 선별

## Forward Stepwise Selection | 작동원리

- ① 최대 변수 개수가  $p$  이고, 변수 개수가 1개인 모델 후보군들 중  
RSS 값이 가장 낮았던 모델이  $X_k$  를 포함하는 것을 확인

*feature* :  $X_1, X_2, X_3, X_4$  인 경우

$$\mu_0 : Y = \beta_0 + \varepsilon \text{ (Null Model)}$$

$$k = 1$$

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

$$\vdots$$

$$\mu_1 : Y = \beta_0 + \beta_4 X_4 + \varepsilon$$

## Forward Stepwise Selection | 작동원리

② 다음 후보군인 변수 개수가 2개인 모델들을 작성 시  
무조건 변수 하나는  $X_k$  로 고정

$$k = 2$$

$$\mu_2 \quad Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \varepsilon$$

$$\vdots$$

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

## Forward Stepwise Selection | 작동원리

③ 변수 개수가 커짐에 따라 같은 방식을 적용하여 모델 후보군들을 작성

$$k = 3$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \varepsilon$$

$$\vdots$$

$$\mu_3 \quad Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$k = 4$$

$$\mu_4 \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$



## Forward Stepwise Selection | 작동원리

④ 선별된 최종 후보 모델들에 대해

Analytical Methods 를 적용해 최적의 모델 결정

$$\mu_0 : Y = \beta_0 + \varepsilon$$

$$\mu_1 : Y = \beta_0 + \beta_4 X_4 + \varepsilon$$

$$\mu_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\mu_3 : Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$\mu_4 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$



$C_p, AIC, BIC, \text{ Adjusted } R^2$  중 1개를 기준으로 최적의 Model 을 선별

## Backward Stepwise Selection | 작동원리

- ① 최대 변수개수가  $p$ 이고, 변수 개수가 1개인 모델 후보군들 중 RSS 값이 가장 낮았던 모델이  $X_k$  를 포함하지 않는 것을 확인

*feature* :  $X_1, X_2, X_3, X_4$  인 경우

$$\mu_4 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \text{ (Full Model)}$$

$$k = 3$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \varepsilon$$

$\vdots$

$$\mu_3 : Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$X_1$  을 포함하지 않음을 확인 가능!

## Backward Stepwise Selection | 작동원리

② 다음 후보군인 변수 개수가  $p-2$  개인 모델들을 작성 시  
변수  $X_k$  를 제외한 나머지 변수들을 이용해 작성

$$k = 2$$

$$\mu_2 \quad Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

$$\vdots$$

$$Y = \beta_0 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

이 경우  $X_1$  을 제외한 나머지인  $X_2, X_3, \dots, X_p$  중  
 $p-2$  개 변수들을 골라 가능한 모든 모델들을 작성하는 식으로 후보군을 생성

## Backward Stepwise Selection | 작동원리

③ 변수 개수가 작아짐에 따라 같은 방식을 적용하여 모델 후보군들을 작성

$$k = 1$$

$$Y = \beta_0 + \beta_2 X_2 + \varepsilon$$

$\mu_1$

$$Y = \beta_0 + \beta_3 X_3 + \varepsilon$$

$$\mu_0 : Y = \beta_0 + \varepsilon \text{ (Null Model)}$$

## Backward Stepwise Selection | 작동원리

④ 선별된 최종 후보 모델들에 대해

Analytical Methods 를 적용해 최적의 모델 결정

$$\mu_0: Y = \beta_0 + \varepsilon$$

$$\mu_1: Y = \beta_0 + \beta_3 X_3 + \varepsilon$$

$$\mu_2: Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

$$\mu_3: Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

$$\mu_4: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$



$C_p, AIC, BIC, \text{ Adjusted } R^2$  중 1개를 기준으로 최적의 Model 을 선별

## Ridge Regression vs Lasso Regression

Ridge Regression	Lasso Regression
L2 - norm	L1 - norm
변수 선택 불가능	변수 선택 가능
Unique Solution (미분을 통해)	Closed form solution 존재 X (numerical optimization 이용)
상관관계가 높은 상황에서 좋은 예측 성능 (상관성이 있는 변수들에 대하여 적절한 가중치를 배분)	변수 간 상관관계가 높은 상황에서 Ridge에 비해 상대적으로 예측 성능이 떨어짐
제약 범위가 원	제약 범위가 마름모

## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

①  $i$  번째 관측치(장소)에 대해 오차항  $\varepsilon(u_i, v_i)$  을 계산

⋮

$$Y = X \otimes \beta + \varepsilon$$

$$\varepsilon(u_i, v_i) = Y(u_i, v_i) - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)X_{1i} - \cdots - \beta_p(u_i, v_i)X_{pi}$$

## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

② 구해본 오차항  $\varepsilon(u_i, v_i)$  을 기준으로 공간 가중치 행렬( $W_i$ ) 을 정의

⋮

$$W_i \ (i = 1 \sim n) = \begin{bmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{bmatrix} \quad \text{where } w_{ik} = \exp\left[-\frac{1}{2}\{d_i(u_k, v_k)/h\}^2\right]$$

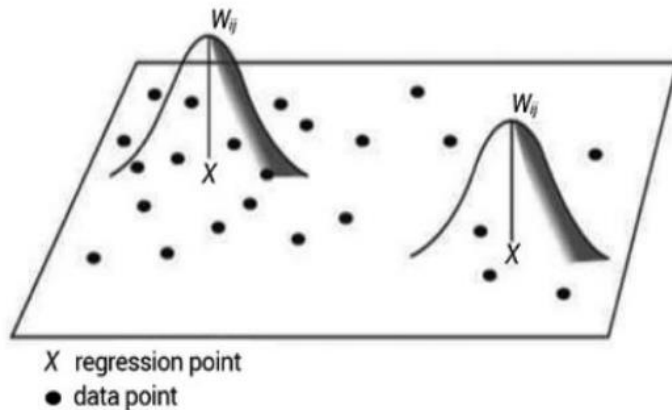
$d_i(u_k, v_k)$ :  $(u_i, v_i)$  와  $(u_k, v_k)$  사이의 거리,  $h$ : 대역폭



## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

② 구해본 오차항  $\varepsilon(u_i, v_i)$  을 기준으로 공간 가중치 행렬( $W_i$ ) 을 정의



$w_{ik}$  의 경우

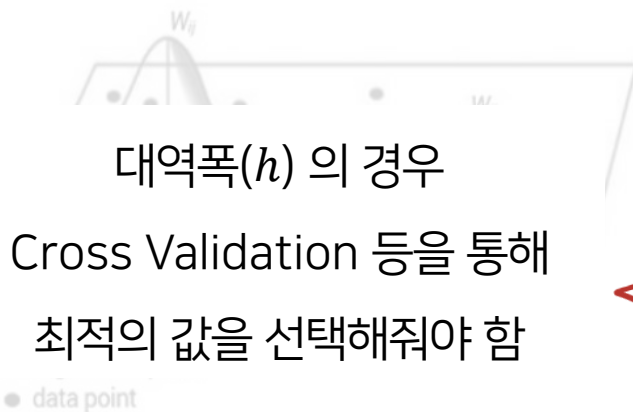
Gaussian 커널 형태의 함수를  
이용해 거리 정도를 계산하여 정의

$$\exp\left[-\frac{1}{2}\{d_i(u_k, v_k)/h\}^2\right]$$

## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

② 구해본 오차항  $\varepsilon(u_i, v_i)$  을 기준으로 공간 가중치 행렬( $W_i$ ) 을 정의



$w_{ik}$ 의 경우

Gaussian 커널 형태의 함수를  
이용해 거리 정도를 계산하여 정의

$$\exp\left[-\frac{1}{2}\{d_i(u_k, v_k)/h\}^2\right]$$

## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

- ③  $W_i$  를  $\{\varepsilon(u_i, v_i)\}^2$  에 곱하여 목적식을 정의해준 뒤  
편미분하여 추정량 ( $= \hat{\beta}(u_i, v_i)$ ) 도출

⋮

$$\arg \min_{\beta_j(u_i, v_i)} \sum_{i=1}^n W_i \{\varepsilon(u_i, v_i)\}^2, \quad j = 1 \sim p$$



$\beta_j(u_i, v_i)$  별로 편미분을 진행

$$\therefore \hat{\beta}(u_i, v_i) = [\hat{\beta}_0(u_i, v_i), \dots, \hat{\beta}_p(u_i, v_i)]^T = (X^T W_i X)^{-1} X^T W_i Y$$

## 지리 가중회귀모델

지리가중회귀모델의  $\beta$  추정 과정

- ③  $W_i$  를  $\{\varepsilon(u_i, v_i)\}^2$  에 곱하여 목적식을 정의해준 뒤  
편미분하여 추정량 ( $= \hat{\beta}(u_i, v_i)$ ) 도출



추가로  $\hat{\sigma}^2$  추정량 또한 아래와 같이 도출 가능

$$\hat{\sigma}^2 = RSS_{GWR}/m$$

where  $m = \text{tr}((I - L)^T(I - L))$ ,  $L = X(X^T W X)^{-1} X^T W$

$$\therefore \hat{\beta}(u_i, v_i) = [\hat{\beta}_0(u_i, v_i), \dots, \hat{\beta}_p(u_i, v_i)]^T = (X^T W_i X)^{-1} X^T W_i Y$$

---

**감사합니다**

---