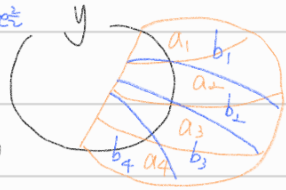
*달라 보이지만 동일한 vector space는
갖고 있으면 Data를 더 추가해도
y를 더 잘 설명할 수 없다.

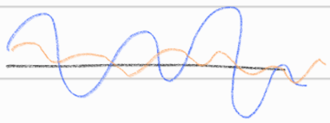


Regression Residual

: $\hat{\varepsilon} = (I-H)\varepsilon$, $H = X(X'X)^{-1}X'$

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

$\approx \dfrac{Cov(X,y)}{Var(X)}$: 상관 관계
→ 내 Data의 움직임(2nd moment, variance)로
y의 움직임을 얼마나 설명할지 보겠다는 것.



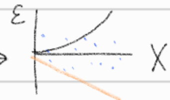
β를 잘못 구해도
Mis-specification.

Gauss Markov Theorem.

A1: $r(X)=k$, $X: S\times k$ (full rank)

A2: $y = X\beta + \varepsilon$, $E(\varepsilon)=0$ (Linear Relation)

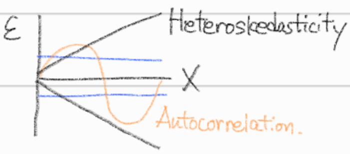
A3: X are unrelated wrt ε
$E(\varepsilon|X)=0$ or $E(\varepsilon_t|x_t)=0$. } →



: X에 따라 패턴을 따르지 말고
파란 점들처럼 Random으로
0을 중심으로 있어야 한다.

A4: $E(\varepsilon\varepsilon^T)=\sigma^2 I$, ε iid

↳ 깨지면 어떻게 될까?



A5: $\varepsilon \sim iid\ N(0, \sigma^2 I)$

*A1–A4 만족.
Linear Operator 중에서
OLS가 BLUE.
A5 까지 만족하면 BUE.

Standardization.

: Scale Issue 제거

$\varepsilon \to \frac{\varepsilon}{\sigma} \sim N$

(Studentizing)

$\hat{\varepsilon} \to \frac{\varepsilon}{\hat{\sigma}} \sim t$

정규 분포를
가정으로 함.

Mis-specification

ex) Anscombe's Quartet

Q. 내 모델이 다른 data set에서도
비슷한 performance를 보일까?

Outliers

: Model을 망칠 수도 있다 (Mis-specification)
Data를 지우기 위한 객관적인 기준으로
leverage of point, Cook's distance가 사용됨.

↓

「Distribution」

Outlier에 영향 많이 받는 분포.
Normal > Laplace > Cauchy > t-dist^n

「RANSAC」

Re-sample을 통해서 Outlier
제거와 비슷한 효과를 얻겠다는 것.

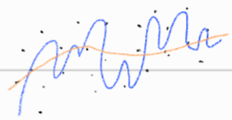
Robustness

Optimization
: LSE, LAD,
Huber, Bisquare.

• kernel의 종류가 중요한 게
아니고 optimization을 잘하는
게 중요(error가 적어야 함)

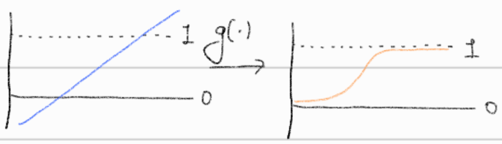
Advanced Regression : Sparsity

Data Term + Regularization Term



Ridge / Lasso

Generalized Linear Model (for real world)



Bayesian

$p(\beta|X,y) \propto p(\beta)p(X,y|\beta)$

$\ln p(\beta|X,y) \propto \ln p(\beta) + \ln p(X,y|\beta)$

우리의 가정을 나타내는 부분. Regularization 이랑
같이 설명하는 이유는 Data Dist^{n}에 맞게 설정해야하기 때문.

Normal → 정규 분포에 맞는 regularization : Ridge
Laplace → Laplace에 〃 : Lasso.
} 이 분포에서 크게 벗어난 이상치들
줄이겠다는 것.

Nonparametric statistics and Model selection.

→ Kolmogorov-Smirnov test.

: CDF 1. vs CDF 2 ← 이게 Normal 이면 Shapiro Wilk test.

(Data 1) (Data 2, Reference Distribution)

→ Wilcoxon's signed-rank test:

: CLT 못쓸때 사용 (Sample # ↓ / Skewedness or Variance ↑)
(like t-dist^n)

N variable 확장.

→ Mann-Whitney U test.

Estimating & Tests.

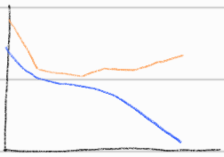
「Resampling Based Methods」

ex) Chicago Teaching Scandal. → Permutation Tests.
→ Bootstrap

: Data를 여러 번 Resample 하면서
Variance를 보는 이유?
여러 개의 distribution이 섞였을 때,
단순히 하나의 분산을 보면 잘못된 거기 때문에
일부를 여러 번 보면서 distribution이
몇 개가 있는지 각각의 분산은 어느 정도인지 보겠다는 것.
→ 모델이나 Data 하나의 분산을 계산하기 위해서가 아니다.

* Model Selection.

: 가성비를 보자!



Assignment 1

Question 4.

- Scale Issue : 결국 X로 Y를 설명하기 위해서는 percent, ratio로 설명한다는 것 이해하기.
  - ↳ $y = k$로 값이 결정돼 있는 상황에서 $X_i = 5$나 50,500으로 변해도 상관 없다. $\hat{\beta}_1$이 반영돼서 조정이 되기 때문이다.
- Logarithm? Scale을 줄여 (간격) Data 간 편차를 줄여 준다. → Skewness↓ Kurtosis↓ → 복잡한 계산 하지 말고 쉬운 계산 하자!

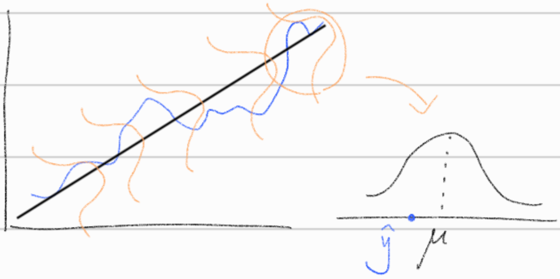
Question 5.

- 변수를 추가했을 때 유의미한 변화가 있을까? Model Selection을 다시 보자.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS\downarrow}{TSS} \rightarrow R^2_{adj} = 1 - \frac{RSS\downarrow / (n-g\uparrow)}{TSS \;/\; (n-1)}$$

: 변수를 추가해서 RSS가 크게 감소하는 것이 아니라면 큰 의미가 없다는 것.

- 왜 $R^2 = 1.0$일 때 예측값에 대한 검증을 해야할까? ex)



이어서 Machine Learning은 왜 분산을 못 구할까? 모든 sample들을 이어서 만든 머신러닝 모델이 피팅선이라고 생각하자. 오차가 있더라도 매우 작아 유의미한 분산 X. (그리고 X값에 따라 달라져 이상)

*Random이기에 당연히 μ를 중심으로 다양한 Data가 나올 것이다. But model은 $\hat{y}$로 고정은 했으니 μ로 예측할 때 보다 편차가 더 커질 것.

↓

t, z tests 불가.

↓

우리의 model에 대한 검정 및 관리 X.

↓

Ocum's razor 측면에서도 Data에 따라 매번 새로 계산해야 하는 (머신러닝) 모델보단 단순히 검정하고 추가 / 제거가 쉬운 Regression 선호.

*머신러닝 모델은 오차를 계산 과정에 반영해 Bias를 거의 0으로 만든다.

→ 머신러닝은 Data가 거의 그대로 반복되거나 분산이 없는 (Random이 아닌) 분야에서 쓰여야 한다.

QnA. A3Rsru or A3Rmi $\rightarrow$ Unbiased ①, Consistent ② :

A3Rmi $E[\varepsilon_t | X] = 0 \ \forall t \rightarrow E[\hat{\beta}] = E[(X'X)^{-1}X'y]$

$$= E[\cancel{(X'X)^{-1}X'X}\beta] + E[(X'X)^{-1}X'\varepsilon]$$

$$= \beta + E[E[(X'X)^{-1}X'\varepsilon | X]]$$

$$= \beta + E[(X'X)^{-1}X'E(\varepsilon | X)]$$

$$= \beta.$$

$\rightarrow E[E[(X'X)^{-1}X'\varepsilon | X]]$

$= E[(X'X)^{-1}X'\underbrace{E[\varepsilon | x_1, x_2, \cdots, x_s]}_{\neq 0}]$

x_1 → Column.

Because Same Row 일 때만 0이라서.

$(E[\varepsilon_t | x_{1t}, x_{2t}, \cdots, x_{st}] = 0.)$

$E[X'\varepsilon] = 0.$

A3Rsru $E[\varepsilon_t | x_s] = 0$ if $t = s$.

ε_t: t번째 원소.

* plim : 확률 변수의 변화를 sample을 늘려가며 보는 것.

E( ) : 모든 가능성의 확률을 고려한 가중 평균.

Consistency : $\underbrace{\text{plim}}_{\text{확률 변수.}} \hat{\beta} = \text{plim}(X'X)^{-1}X'y$

Sample # ↑

$$= \beta + \text{plim}(X'X)^{-1}X'\varepsilon$$

$$= \beta + \text{plim}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}$$

$$= \beta + Var(X)^{-1}Cov(X, \varepsilon)$$

$$Cov(X, \varepsilon) = \underbrace{E(X'\varepsilon)}_{=0} - E(X')\underbrace{E(\varepsilon)}_{=0}$$