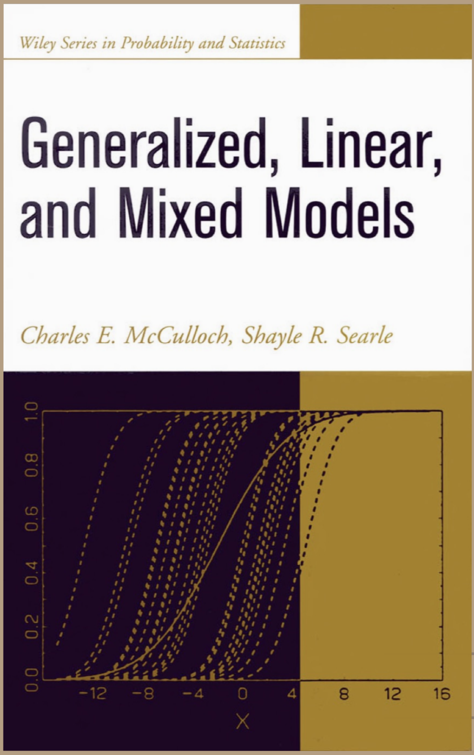
Chapter 6. LMMs



1. A General Model

a. Introduction

- Linear Model

- $E(y) = X\beta$, 단 β는 미지의 모수
- $E(y_{ij}) = \mu + \alpha_i$

μ : General mean

α_1 : 플라시보 환자의 반응변수 effect

α_2 : 약물 받은 환자의 반응변수 effect

이때, μ, α_1, α_2는 fixed effect

$E(y) = X\beta$, $\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$

- Linear Mixed Model

$E(y_{ij}) = \mu + \alpha_i + \beta_j + c_{ij}$

α_i : i 병원의 랜덤 effect

β_j : j 약물 수준의 fixed effect

c_{ij} : 상호작용의 random effect

⇒ LMM : fixed effect ⊕ random effect.

- Random model

: fixed effect 없을 때, 단 μ는 포함

- fixed effect : y의 평균 모델링

- Random effect

: y의 분산 - 공분산 구조 govern

$Var(y_{N\times1})$의 원소를 $\frac{N(N+1)}{2}$개 특정하는 작업 단순화

→ Random effect 없으면, $Var(y)$ 원소는 다양한 형태로 다뤄야 함.

→ Random effect 사용하면 데이터에 영향을 미치는 것으로 알려진 factor에 기인하는 분산·공분산 편하게 다룰수 o

b. Basic Properties

- LM : $E(y) = X\beta$, β는 fixed effect
- LMM : $E(y) = X\beta + Zu$

  Z는 알려진 행렬,

  u는 데이터 벡터 (y)에서 발생한 random effect vec.

→ u는 random variable

: 관측할수없는 실현값에 대해 조건부 모델 지정

$E(y|u) = X\beta + Zu$.

pf) $E(y|u) = E(X\beta + Zu + \varepsilon|u)$

$= E(X\beta|u) + E(Zu|u) + E(\varepsilon|u) \quad = 0 \ (\because \varepsilon \perp u)$

$= X\beta + Zu$

⇒ 확률 변수 U에 대한 실현값 u를 사용하면, $E(y|U=u)$

⇔ $E(y|u)$

$u \sim (0, D)$: $E(u) = 0$, $Var(u) = D$ 로 가정한다.

pf) Goal : $E(u) = 0$.

let, $E(u) = \tau$

$E(y|u) = X\beta + Zu = X\beta + Z\tau + Z(u - \tau)$

let, $X^* = [X \ \ Z]$, $\beta^* = [\beta' \ \ \tau']'$, $u^* = u - \tau$

$E(y|u) = X\beta + Z\tau + Z(u-\tau) = X^*\beta^* + Zu^*$

→ $E(u^*) = E(u - \tau) = E(u) - \tau = \tau - \tau = 0$

∴ $E(y|u) = X\beta + Zu$, $E(u) = 0$

cf. $Var(y|u)$ vs $Var(\varepsilon)$

- $Var(y|u)$: 일반적 표현 → 일반화 선형모형
- $Var(\varepsilon)$: 회귀 오형에서 사용하는 오차항

  일반화 선형 모형에서 정규성 가정시에만 사용.

→ $Var(y|u) = Var(X\beta + Zu + \varepsilon|u) = Var(\overset{0}{Zu|u}) + Var(\varepsilon|u) = Var(\varepsilon|u)$

∵ u가 지정됨.

$= Var(\varepsilon|u) = R$

↳ R이 u의 함수가 아니면 indep.

$Var(y|u) = R$ 가정

⇒ $y \sim (X\beta, \ ZDZ' + R)$

pf) Goal : $Var(y) = ZDZ' + R$.

$Var(y) = Var(X\beta + Zu + \varepsilon)$

$= Z^2 Var(u) + Var(\varepsilon) + Z \cdot Cov(u, \varepsilon) \quad (\because X\beta$ 상수$)$

$= Z^2 \cdot D + R + 0 \quad (\because u \perp \varepsilon)$

$= ZDZ' + R$

pf2) Goal : $Var(y) = ZDZ' + R$.

$Var(y) = E[Var(y|u)] + Var[E(y|u)]$

$= E(R) + Var(X\beta + Zu)$

$= R + Z^2 Var(u) = R + Z^2 \cdot D = R^2 + ZDZ'$

⇒ fixed effect는 평균만 입력 ($X\beta$)

random effect model 행렬과 분산은 y분산만 입력

2. Attributing Structure to Var(y)

$V = Var(y) = ZDZ' + R$

a. Example

- 데이터 : 뉴욕 15개 고등학교에서 각 4개의 9학년 학급에 주어진 수학 시험 성적

→고정효과인 남/여 의 차이 이외에도 변동요인 有

① 학교 간 ② 각 학교내 학급간 ③각 학급내 학생간

→ i학교 내 j학급의 t 성별의 k학생의 성적 : y_{tijk}

$E(y_{tijk} | S_i, G_j) = \beta_t + S_i + G_j$.

β_t : 성별 t에 대한 fixed effect

S_i : $i=1, \cdots, 15$에 대한 학교 효과 (random effect)

G_{ij} : ($j=1, \cdots, 4$), i 학교에 대한 j번째 학급효과(random effect)

P_{ijk} : 개인 학생에 대해 β_i, S_i, G_j로 설명되지 않는 모든 것

↓

β : β_m (남성), β_f (여성) 두 원소 존재

u : 15개의 S_i 효과와 60개 (15×4) G_j 효과

b. Taking Covariances between factors as Zero

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \{_c S_i\}_{i=1}^{15} \\ \{_c \{_c G_{ij}\}_{j=1}^{4}\}_{i=1}^{15} \end{bmatrix}$$

$$Z = [Z_1 \quad Z_2]$$

$\therefore Zu = Z_1 u_1 + Z_2 u_2$

$$D = Var(u) = \begin{bmatrix} Var(u_1) & Cov(u_1, u_2') \\ Cov(u_2, u_1') & Var(u_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_{12} \\ D_{21} & D_2 \end{bmatrix}$$

→ $D_{21} = D_{12}'$

$\therefore E(y|u) = X\beta + Z_1 u_1 + Z_2 u_2$

$V = ZDZ' + R$

$= Z_1 D_1 Z_1' + Z_2 D_2 Z_2' + Z_1 D_{12} Z_2' + Z_2 D_{21} Z_1' + R$

→ $R = Var(y|u)$

: P_{ijk} 항의 분산-공분산 행렬

학생간 변동성 + S_i와 G_j에 기인하지 않는 변동성

∘ 2-random factor → r Random factors 확장!

$u = \{_c\, u_i\}_{i=1}^{r}$

$D = var(u) = \{_m\, D_{ii'}\}_{i,i'=1}^{r}$

$Z = [Z_1\ Z_2\ \cdots\ Z_r] = \{_r\, Z_i\}_{i=1}^{r}$

$\rightarrow E(y|u) = X\beta + \sum_{i=1}^{r} Z_i u_i$

$$V = \sum_{i=1}^{r} Z_i D_{ii} Z_i' + \sum_{\substack{i=1 \\ i \neq i'}}^{r} \sum_{i'=1}^{r} Z_i D_{ii'} Z_{i'}' + R$$

$D_{ii} = Var(u_i)$, $D_{ii'} = Cov(u_i, u_{i'}')$

→ 예제에서 D_{11} : 학교들의 분산-공분산 행렬

D_{12} : 학교와 학급의 분산 -공분산 행렬

↓

이러한 공분산=0 이 되는게 합리적

Why? 학교 효과 S_j 와 학급효과 G_{ij} 의 학교는 다른 학교

Correlation 는 없다.

D_{12} : 한 학교 효과 S_i와 같은 학교내 학급 효과 G_{ij}의 공분산

랜덤 효과 사용해 데이터의 모든 변동성 파악 → 공분산=0

∴ $D_{12} = 0$

∘ General Case로 확장

: $i \neq i'$ 에 대해 $D_{ii'} = 0$ → D_{ii} 를 D_i로 씀

$D = \{_d\, D_i\}_{i=1}^{r}$ $\quad V = \sum_{i=1}^{r} Z_i D_i Z_i' + R$

C. Traditional Variance Components model

① Customary notation

- 예제에서, 학교간 공분산 =0 , 학교들의 등분산성 가정한다면,

  15개의 학교에 대해 $D_1 = \sigma_s^2 I_{15}$ —①

- i학교내 4개의 학급에도 비슷하게 가정하면,

  $Var[C_{i1}\ C_{i2}\ C_{i3}\ C_{i4}]' = \sigma_i^2 I_4$

- 한 학교 내 학급들이 모든 다른 학교내 학급들과 독립가정 합리적.

  $D_2 = \{_d\ Var(u_i)\}_{i=1}^{15} = \{_d\ \sigma_i^2 I_4\}_{i=1}^{15}$

- 더 간단한 가정 : 한 학교내 4개 학급이 15개 학교에 대해

  동일한 분산. → $\sigma_i^2 = \sigma_c^2 \quad \forall i$

  $D_2 = \sigma_c^2 I_{60}$ —②

- ①과 ②에 의해, $R = \sigma^2 I_{1200}$

  : 전통적인 Variance Components 모형에 대한 표준 구조 만들어냄.

  ⇓

- General Case : random effects r 개

  $D = \{_d\ \sigma_i^2 I_{q_i}\}_{i=1}^{r}$ ← i의 랜덤 요인에 q_i 효과 가짐

  u_i의 order은 $q_i \times 1$

  $V = \sum_{i=1}^{r} Z_i Z_i' \sigma_i^2 + \sigma^2 I_N$ ← y의 order : N×1

② Amended notation

- $\sigma_0^2 \equiv \sigma^2$, $Z_0 \equiv I_N$, $q_0 = N$

  ↓

  $$D_* = \begin{bmatrix} \sigma_0^2 I_{q_0} & 0 \\ 0 & D \end{bmatrix} = \{_d\ \sigma_i^2 \cdot I_{q_i}\}_{i=0}^{r}$$

  $$V = \sum_{i=1}^{r} Z_i Z_i' \sigma_i^2 + \sigma^2 I_N = \sum_{i=0}^{r} Z_i Z_i' \sigma_i^2 = Z_* D_* Z_*'$$

  이 때 , $Z_* = [Z_0\ Z_1\ \cdots\ Z_r] = [Z_0\ Z]$

- Z_0에 대응되는 N×1 order 의 u_0는 분산분석 모형에서

  잔차 오차항으로 생각 !

- $i = 0, 1, \cdots, r$일 때 분산 σ_i^2은 Variance Components 라고 불림

  Why? 개별 관측치의 분산 성분이니까!

  $Var(y_{tijk}) = \sigma_s^2 + \sigma_c^2 + \sigma^2$

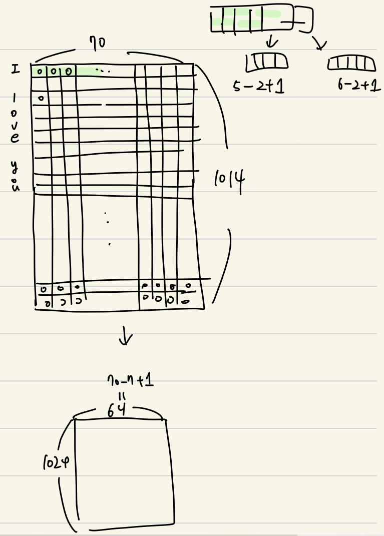
d. An LMM for Longitudinal data

- 종단자료 (longitudinal data)는 각 관측 단위 집합에 대해 연속적인 관측치.

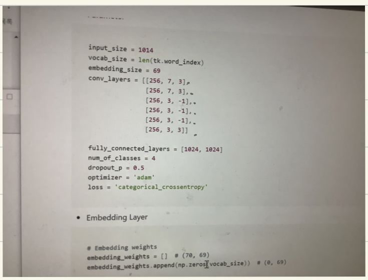
  ex) 환자 그룹에 대해 주차별 혈압 측정 자료

- y_i : i 환자에 대한 n_i 측정값에 관한 vec.

  $E(y_i|u_i) = X_i\beta + Z_i u_i$



$6 \times 2 = 12$

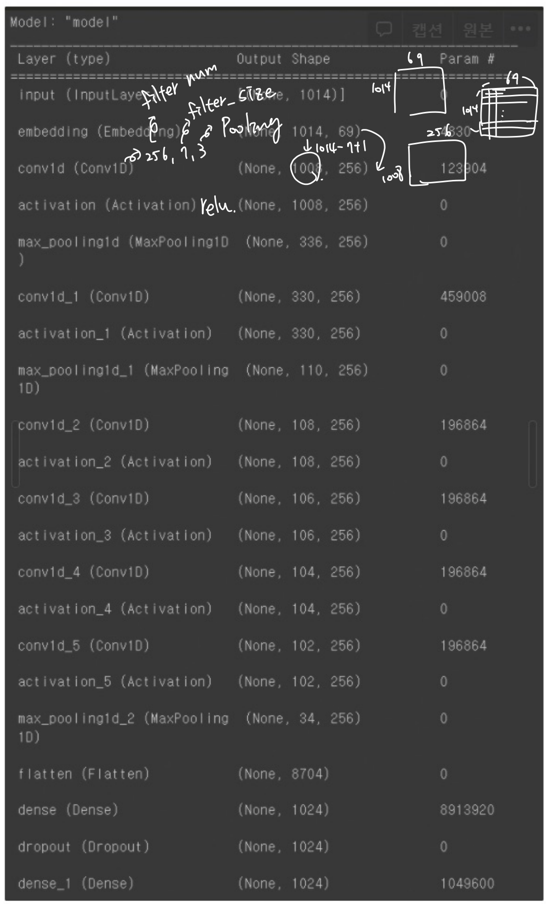


```
input_size = 1014
vocab_size = len(tk.word_index)
embedding_size = 69
conv_layers = [[256, 7, 3],
               [256, 7, 3],
               [256, 3, -1],
               [256, 3, -1],
               [256, 3, -1],
               [256, 3, 3]]

fully_connected_layers = [1024, 1024]
num_of_classes = 4
dropout_p = 0.5
optimizer = 'adam'
loss = 'categorical_crossentropy'
```

- Embedding Layer

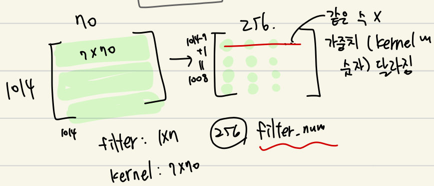
```
# Embedding weights
embedding_weights = []  # (70, 69)
embedding_weights.append(np.zeros(vocab_size))  # (0, 69)
```

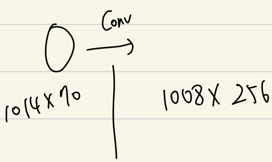


Model: "model"

Layer (type)	Output Shape	Param #
input (InputLayer)	[(None, 1014)]	0
embedding (Embedding)	(None, 1014, 69)	256830
conv1d (Conv1D)	(None, 1008, 256)	123904
activation (Activation)	(None, 1008, 256)	0
max_pooling1d (MaxPooling1D)	(None, 336, 256)	0
conv1d_1 (Conv1D)	(None, 330, 256)	459008
activation_1 (Activation)	(None, 330, 256)	0
max_pooling1d_1 (MaxPooling1D)	(None, 110, 256)	0
conv1d_2 (Conv1D)	(None, 108, 256)	196864
activation_2 (Activation)	(None, 108, 256)	0
conv1d_3 (Conv1D)	(None, 106, 256)	196864
activation_3 (Activation)	(None, 106, 256)	0
conv1d_4 (Conv1D)	(None, 104, 256)	196864
activation_4 (Activation)	(None, 104, 256)	0
conv1d_5 (Conv1D)	(None, 102, 256)	196864
activation_5 (Activation)	(None, 102, 256)	0
max_pooling1d_2 (MaxPooling1D)	(None, 34, 256)	0
flatten (Flatten)	(None, 8704)	0
dense (Dense)	(None, 1024)	8913920
dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 1024)	1049600

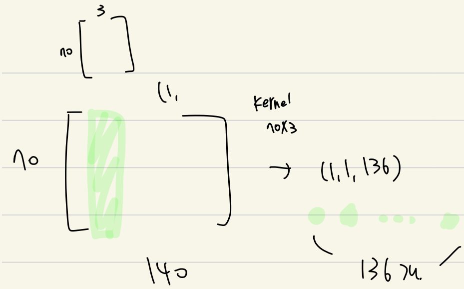
conv_layer = [filter_num, filter size, Pooling_size]

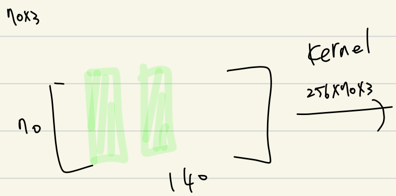




filter_num = 256

filter_size = 7





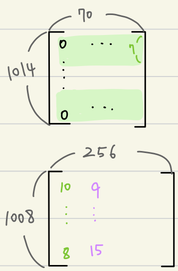
Input $\xrightarrow[\text{conv 시행}]{\text{kernel } 256\times(7\times70)}$ Feature map

[1047 x 70] → [1008 x 256]

Input

kernel
: 256 x (7x70)

Feature Map

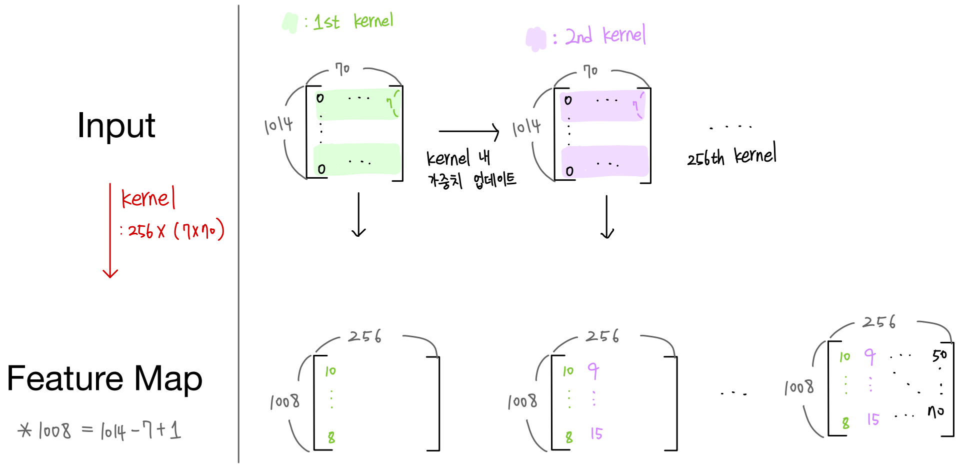


: 1st kernel

kernel 내 가중치 업데이트

: 2nd kernel

* 1008 = 1014 − 7 + 1



Input
kernel
: 256 x (7x70)
Feature Map
* 1008 = 1014 - 7 + 1
: 1st kernel
: 2nd kernel
70
1014
kernel 내
가중치 업데이트
256th kernel
256
1008
10
8
9
15
50
70