

그래피컬라소 (Graphical Lasso)를 기반으로 한
변화감지에 관한 기초연구
-생활습관병 데이터를 대상-

사회시스템 연구실
김연경

그래피컬라소에 의한 구조분석

정밀도행렬 $\Lambda = (\lambda_{ij}) \Rightarrow$ 인접행렬

$$\Lambda^* = \begin{pmatrix} * & * & * & * & * & 0 \\ * & * & * & * & * & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & 0 & * & 0 & 0 \\ * & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * \end{pmatrix}$$

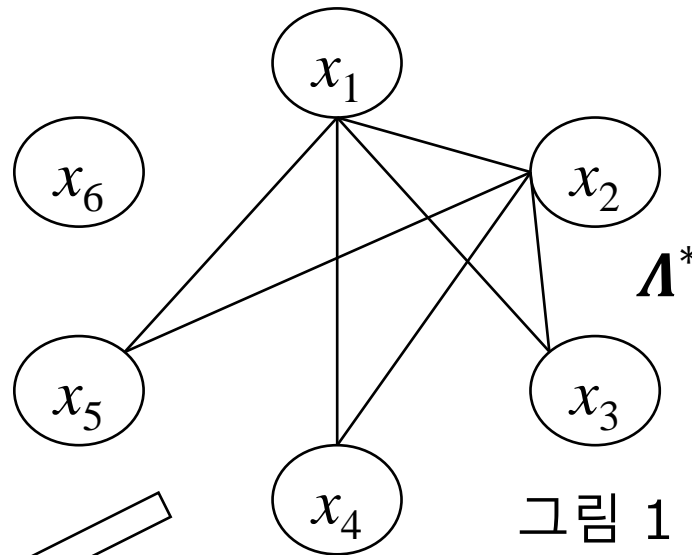
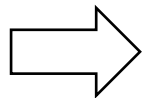


그림 1 : 변수간의 상관그래프

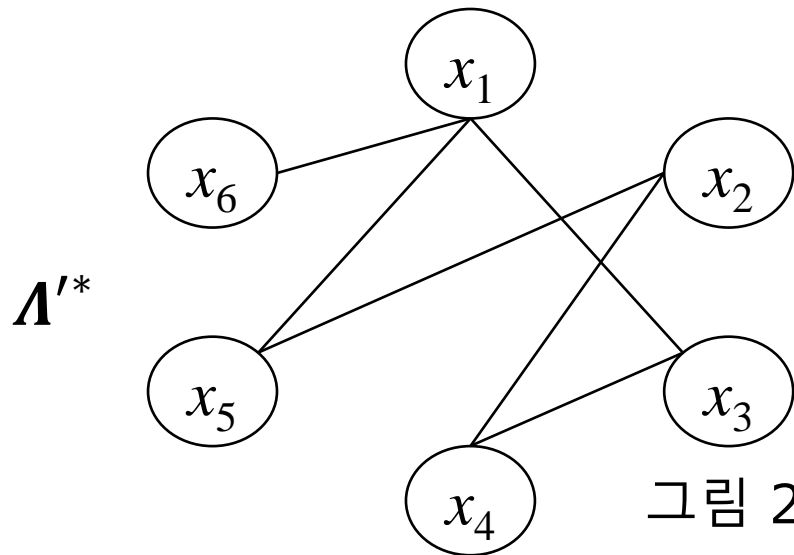


그림 2 : 구조변화 행렬

$$\Rightarrow \Lambda^* = \arg \max_{\Lambda} (\ln \det \Lambda - \text{tr}(S\Lambda) - \rho \|\Lambda\|_1)$$

S : 데이터의 표본분산공분산행렬
 $\rho > 0$: 정규화 파라미터

연구목적

- ◆ 생활습관병 데이터를 대상으로 하여,
그래피컬라소를 이용해서,
 - 데이터집합간의 구조변화
 - 요인(항목)마다의 변화값

당뇨병 데이터

◆ 당뇨병(Pima Indians Diabetes Database)

- 1990년의 미국국립당뇨병학회
- 연령 21살이상의 여자 **768명**(중복없음) 의 데이터를 사용

◆ 검사항목 (8항목)

- | | | |
|----------|----------|---------|
| (1) 임신횟수 | (2) 글루코스 | (3) 혈압 |
| (4) 피부두께 | (5) 인슐린 | (6) BMI |
| (7) 유전지수 | (8) 연령 | |

알츠하이머 데이터

◆ 알츠하이머(Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults)

- 2010년 워싱턴대학 연구소
- **150명** (중복있음) 이 373번의 MRI촬영

◆ 검사항목 (11항목)

- (1) 방문횟수 (2) MR 지연시간 (3) 연령 (4) 교육기간
- (5) 사회적 지위 (6) 정신 단기간 검사 (7) 행동검사 (8) 두상내부검사
- (9) 뇌면적 (10) ASF (11) 성별

알츠하이머 데이터 기초분석 : t-검정

항목	t 값	p 값
방문일수	-1.85	0.03
MR 지연시간	-2.35	0.01
연령	-0.11	0.45
교육기간	-3.79	0.00
사회적 지위	1.16	0.12
정신 단기간 검사	-11.58	0.00

항목	t 값	p 값
행동검사	23.43	0.00
두상내부검사	-0.82	0.21
뇌면적	-6.33	0.00
ASF	0.63	0.27
성별	-4.38	0.00

그래피컬라소를 이용한 구조분석

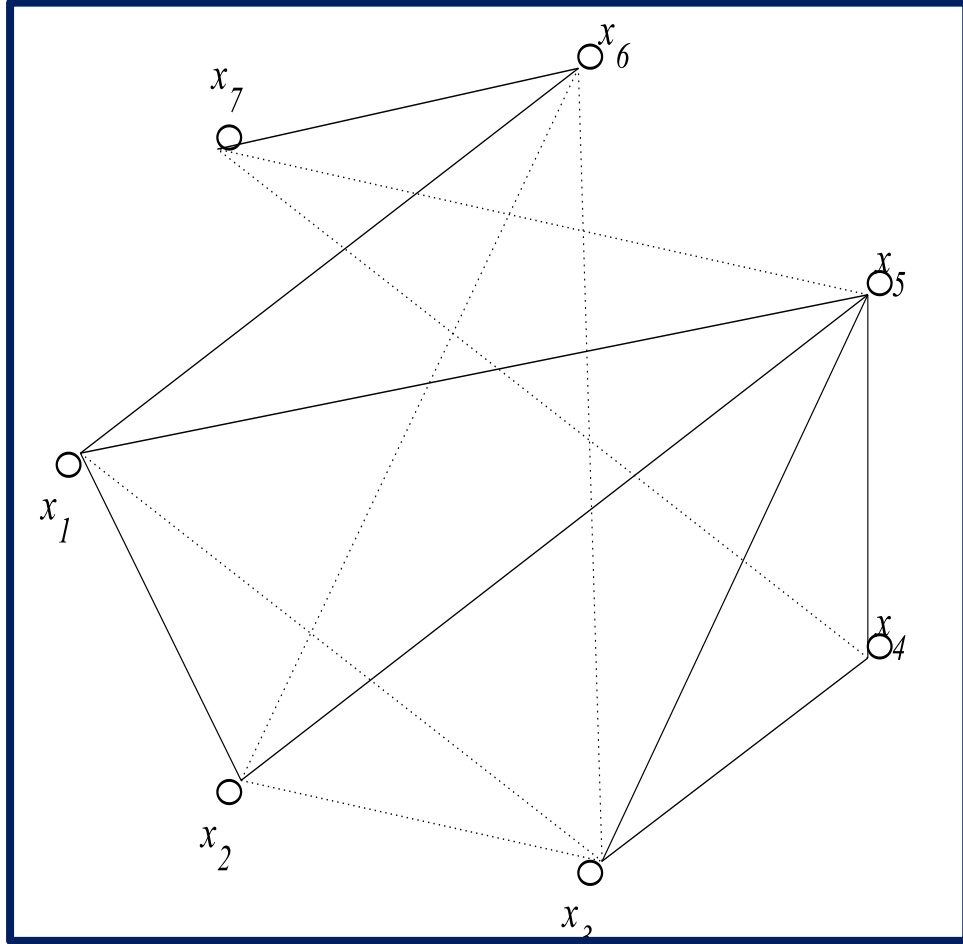


그림 3 : 정상인

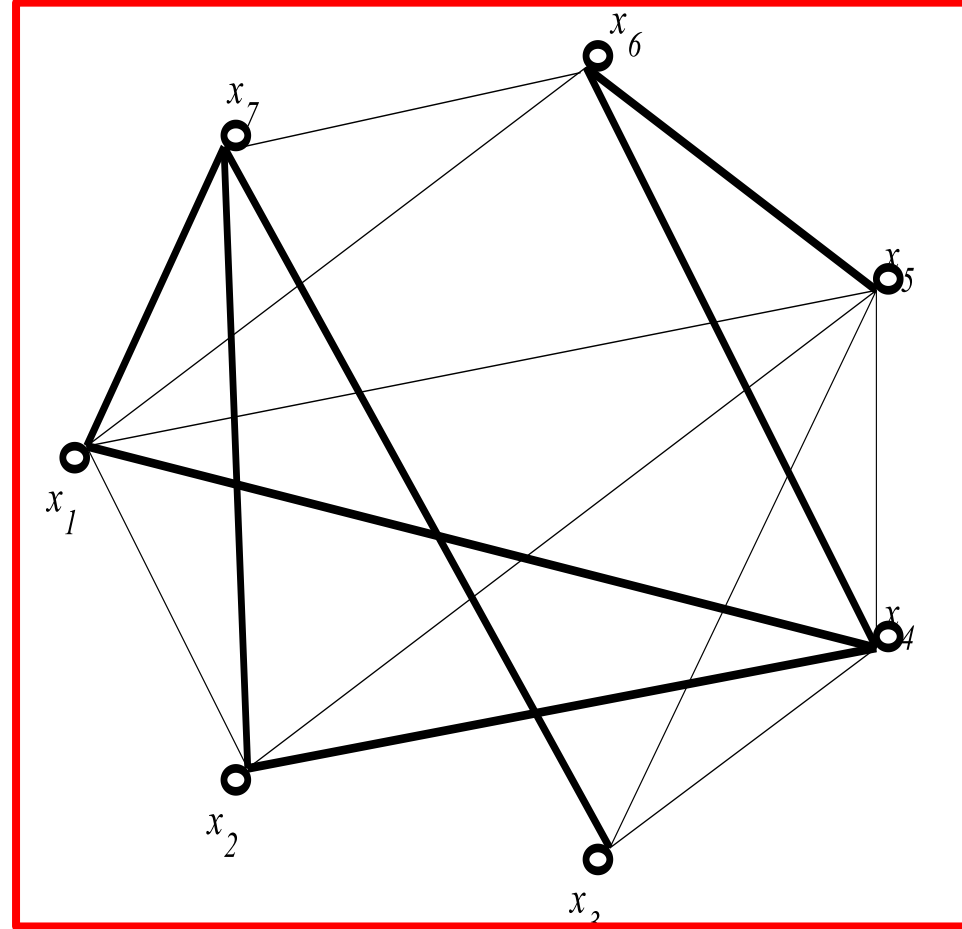


그림 4 : 알츠하이머를 앓고 있는 환자

	항목
X_1	방문일수
X_2	MR 지연시간
X_3	연령
X_4	정신 단기간 검사
X_5	행동검사
X_6	뇌면적
X_7	성별

요약

- T검정에 의하면, 정신단기간검사와 행동검사의 평균차가 크다
- 행동검사 0인 사람은 뇌면적도 크다
- 여자 정상인은 병원에 자주 방문하고 MRI시간이 길어 진다
- 남자 알츠하이머 환자는 병원에 자주 방문하고 있고, 정신단기간검사는 여자 알츠하이머 환자가 보다 높다.

정신단기간검사는 알츠하이머진단의 검사항목과 영향성이 높다고 생각할 수 있다

앞으로 분석한 항목이 진단에 중요한 항목이 되는지 검토할 필요가 있다

参考文献

- データヘルス計画HP 厚生労働者
- 井出剛：疎な相関グラフの学習による相関異常の検出, 2009
http://ide-research.net/papers/2009_DMSM_Ide.pdf
- 糖尿病(1990年)と認知症(2010年)のデータ Kaggle
<https://www.kaggle.com/>