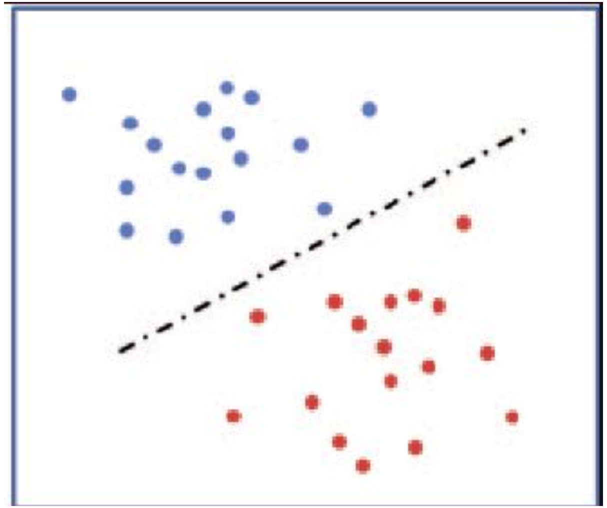
16. 심층 학습을 위한 구조적 확률 모형

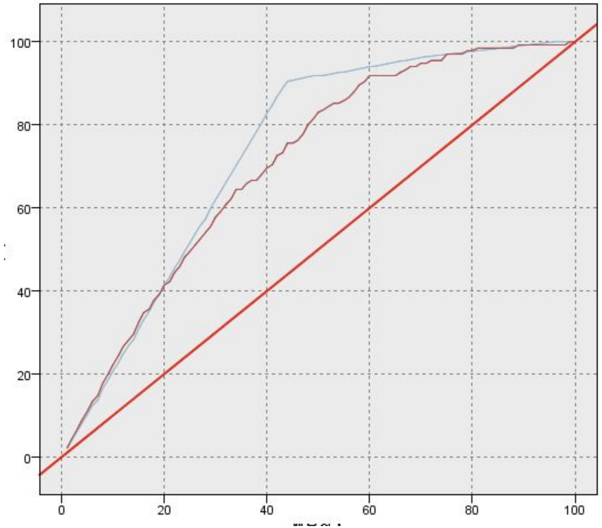
2023.05.23.
21학번 김은채

16.1 심층 학습을 위한 구조적 확률 모형

‘기계 학습의 규모를, 인공지능이 풀어야 하는 부류의 난제들을 풀 수 있을 정도로 확장하는 것’



분류(classification) 알고리즘



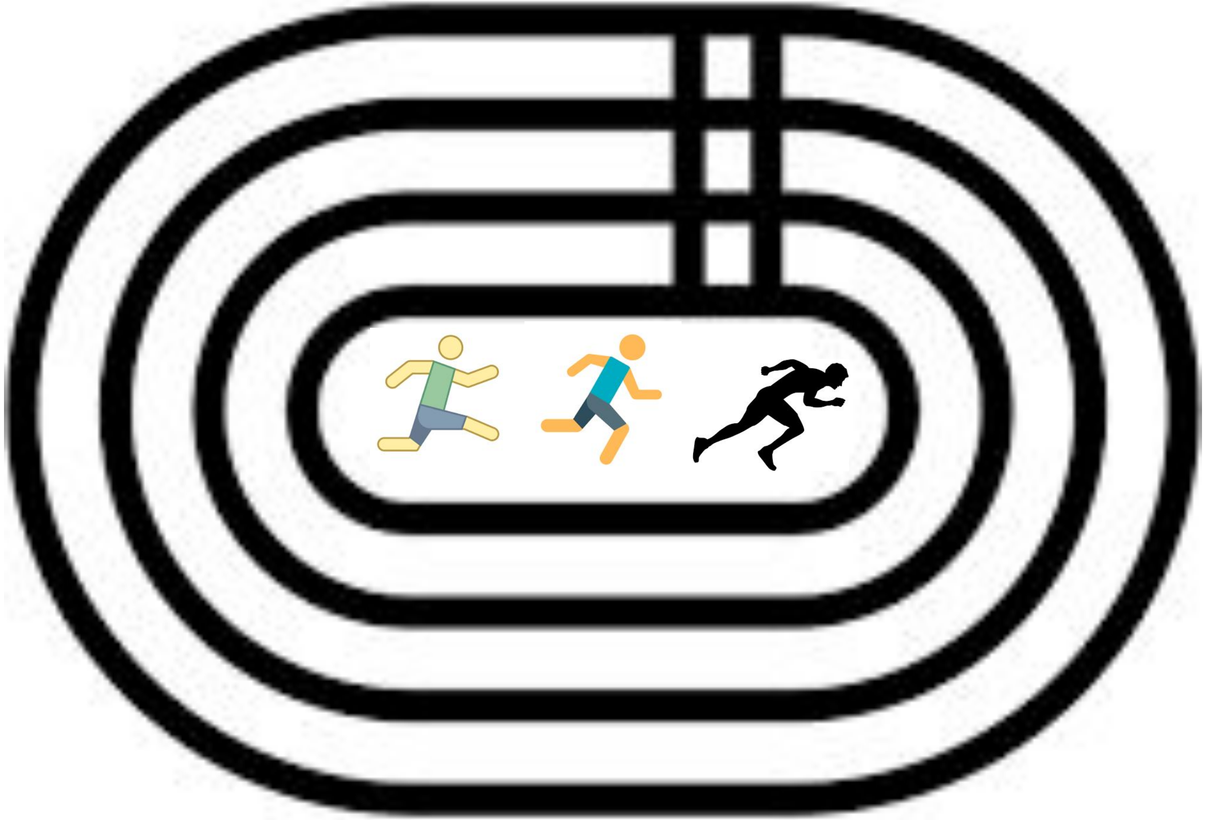
16.1 비구조적 모형화의 문제점

- 밀도 추정
- 잡음 제거
- 결측값 대체
- 표본추출(표집)

16.1 비구조적 모형화의 문제점

- 메모리 - 표현 저장 비용
- 통계적 효율성
- 실행 시간-추론 비용
- 실행 시간- 표본추출 비용

16.1 비구조적 모형화의 문제점



16.2 그래프를 이용한 모형 구조의 서술

- 유향 모형들
- 무향 그래프 모형
- 분배함수
- 분리와 종속 분리(d- 분리)
- 무향 그래프와 유향 그래프의 변환
- 인수 그래프

유향 모형들

- 간선에 방향이 있다는 뜻

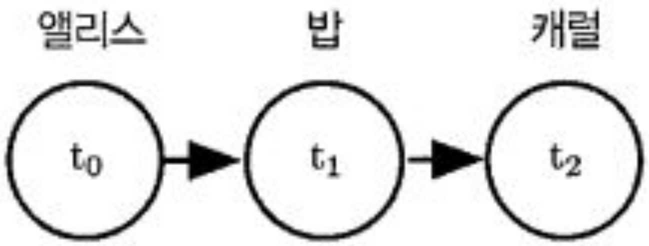


그림 1

$$p(t_0, t_1, t_2) = p(t_0)p(t_1 | t_0)p(t_2 | t_1).$$

그림 2

0분~10분까지 6초 단위로 나누어 이산화를 했을 시,

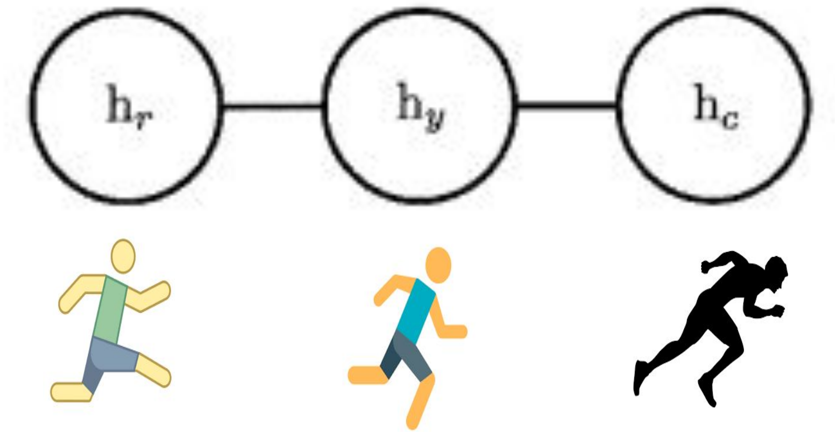
비구조적: 999,999개(t0, t1, t2는 100가지 값을 가짐)

구조적(유향): 19,899개(t0→ t1은 9900(99*100)개 + t1→ t2도 9900(99*100)개)

무향 그래프 모형

유향그래프: 각 화살표를 특정 방향으로 그릴 만한 명확한 이유가 존재하는 상황

무향그래프: 상호작용들의 방향이 그렇게 명확하지 않은 것들
(방향이 존재 하지 않음, 양방향)



$$\tilde{p}(\mathbf{x}) = \prod_{C \in \mathcal{G}} \phi(C).$$

분배 함수

비정규화 확률분포: $p(\mathbf{x}) = \frac{1}{Z}\tilde{p}(\mathbf{x}).$ $Z = \int \tilde{p}(\mathbf{x})d\mathbf{x}.$

하나의 스칼라 변수 $x \in \mathbb{R}$를 하나의 파벌 포텐셜 $\phi(x) = x^2$으로 모형화한다

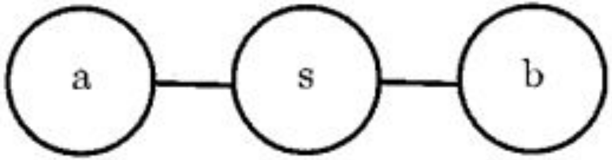
$$Z = \int x^2 dx$$

$$\phi(x;\beta) = \exp(-\beta x^2)$$

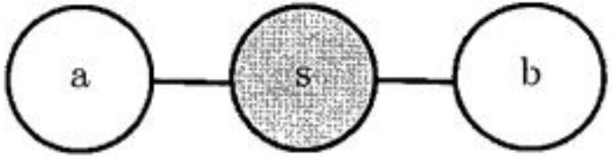
분리와 종속 분리(d-분리)

무향: 분리

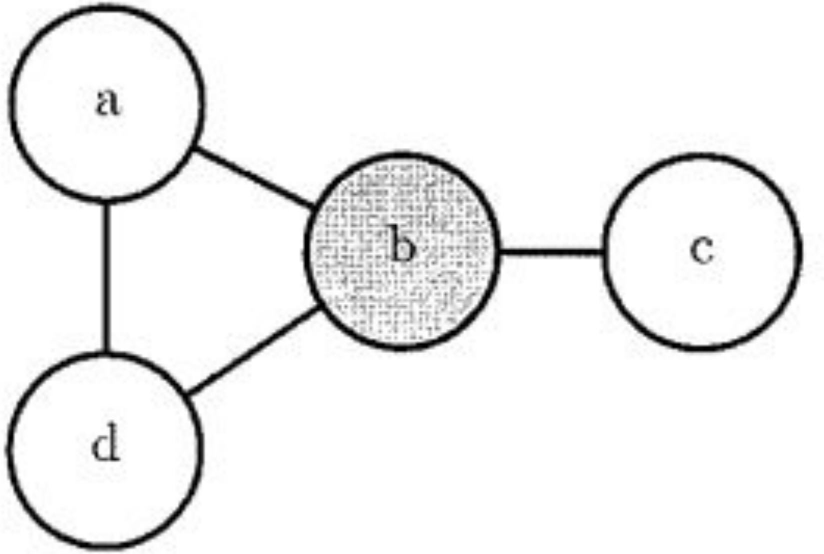
유향: 종속 분리(d-분리)



(a)



(b)



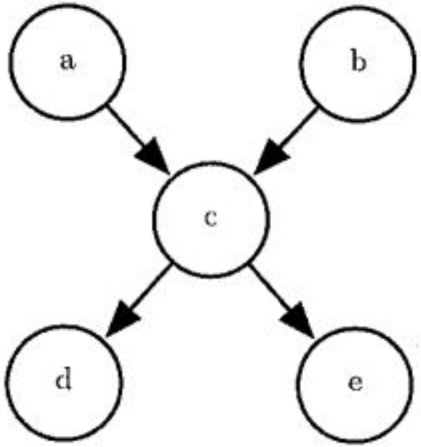


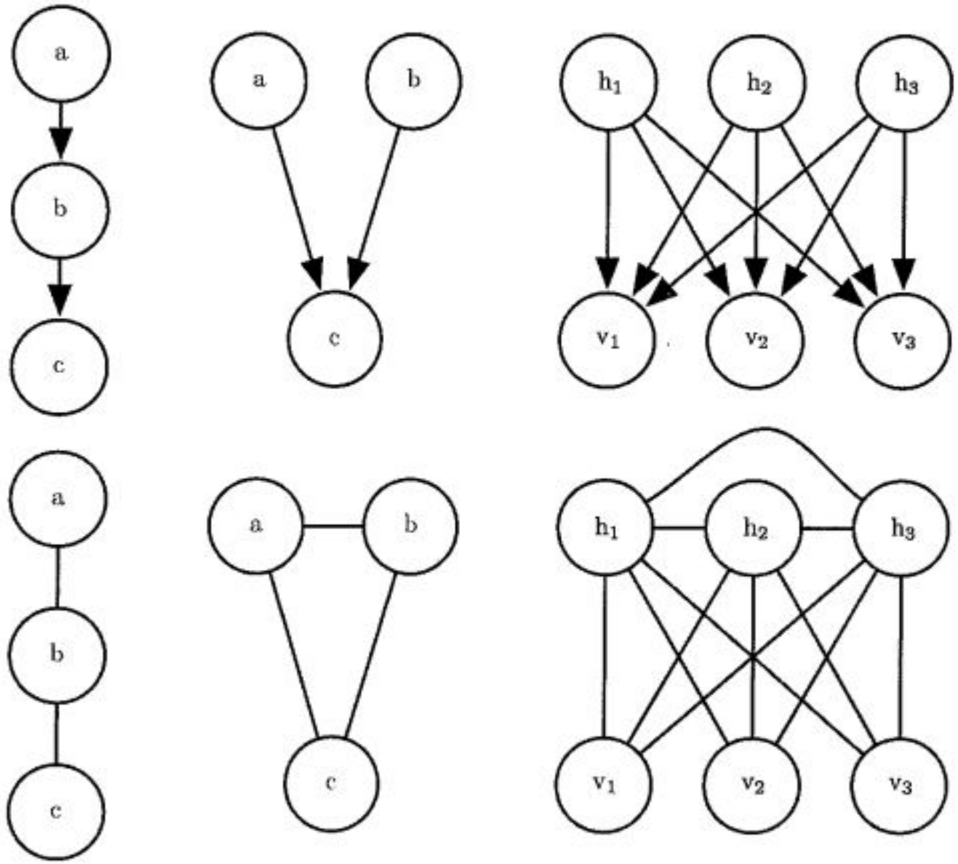
그림 16.9: 이 그래프에서 여러 종속 분리 성질들을 파악할 수 있다. 이를테면:

- a와 b는 공집합이 주어졌을 때 종속 분리이다.
- a와 e는 c가 주어졌을 때 종속 분리이다.
- d와 e는 c가 주어졌을 때 종속 분리이다.

또한, 일부 변수들이 관측되었을 때 일부 변수들이 더 이상 종속 분리가 아니라는 점도 알 수 있다. 구체적으로 말하면:

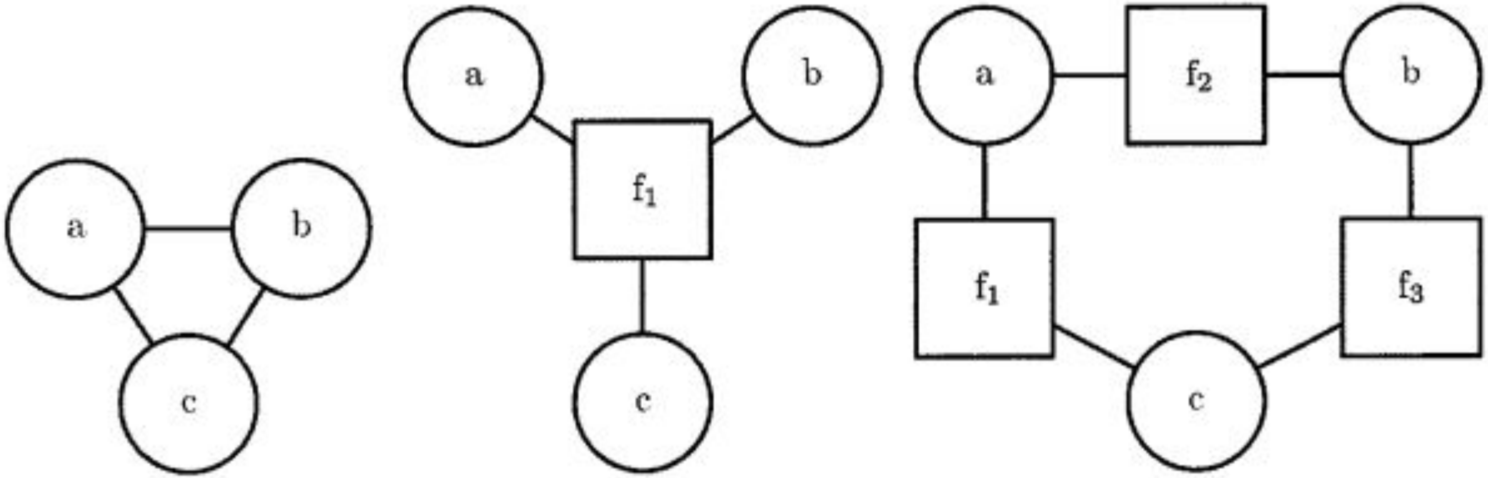
- a와 b는 c가 주어졌을 때 종속 분리가 아니다.
- a와 b는 d가 주어졌을 때 종속 분리가 아니다.

무향 그래프와 유향 그래프의 변환



인수 그래프

무향 모형의 그래프를 서술하는 표준적인 문법에 존재하는 중의성을 해소하는 방법



그래프 모형의 표본 추출

조상표집:

위상 정렬 연산을 하고 그 순서에 따라 추출한다. 위상 정렬이 없다면 부모 변수를 추출하기도 전에 자식 변수를 추출하는 일이 발생 할 수 있음

- 일반적으로 조상 표집은 아주 빠르고 편리
- 유향 모형에만 적용할 수 있다

16.4 구조적 모형화의 장점

- 확률 분포의 표현 비용은 물론 학습과 추론 비용도 극적으로 줄어듦
- 유향 모형의 경우 표본 추출 빨라짐(단, 무형 모형에서는 상황이 복잡해질 수 있음)
- 학습으로 얻은 지식의 표현과 기존 지식에 기초한 추론으로 얻은 지식의 표현을 명시적으로 분리 가능→ 덕분에 디버깅 하기 쉬워짐
- 다양한 부류의 그래프들에 적용할 수 있음
- 다수의 알고리즘과 다수의 구조를 마치 두 집합의 곱집합처럼 조합함으로써 다양한 가능성을 얻게 됨

16.5 종속관계의 학습

- 좋은 생성 모형은 관측변수에 관한 분포를 정확하게 포착할 수 있어야 함

- 직접 연결 된 가시 변수들 사이의 의존성을 모형으로 포착하려는 경우, 모든 변수를 연결하는 것은 일반적으로 비현실적이므로 서로 밀접하게 연관된 변수들만 연결하고 그 외의 변수들 사이의 간선을 생략한 그래프를 만들어야 함 → 구조 학습: 지정된 특정한 구조를 가진 모형을 훈련해서 점수를 부여함

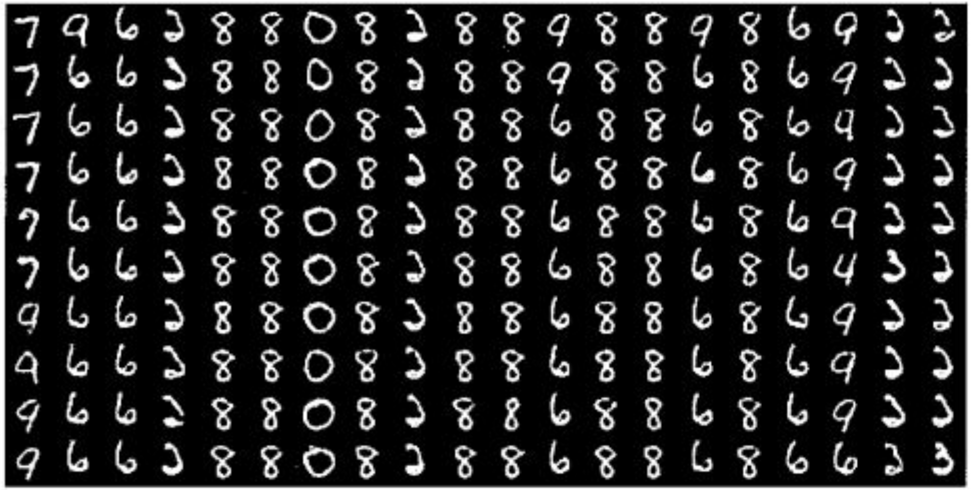
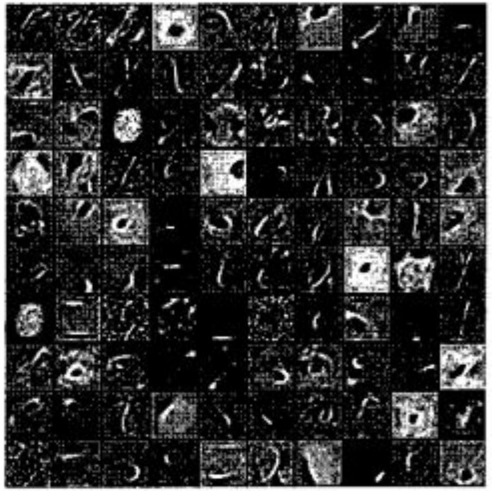
16.6 추론과 근사 추론

확률 모형의 주된 용도 중 하나는 변수들 사이의 관계를 질의하는 것

$$\log p(\boldsymbol{v}) = \mathbb{E}_{\mathbf{h} \sim p(\mathbf{h} \mid \boldsymbol{v})} \left[\log p(\boldsymbol{h}, \boldsymbol{v}) - \log p(\boldsymbol{h} \mid \boldsymbol{v}) \right]$$

16.7 구조적 확률 모형에 대한 심층 학습 접근 방식

- 잠재 변수를 포함한 층이 아예 없거나 하나뿐이지만 모형 안의 조건부 분포들을 정의 하는 데에는 깊은 계산 그래프를 사용하는 것들이 많다.
- 본질적으로 심층학습은 항상 분산 표현 개념을 활용한다.
- 전통적인 그래프 모형의 변수들을 설계하는 방식도 전통적인 분야와 다르다.
- 미지수에 대한 허용 오차를 명확하게 적용한다.

감사합니다