Probabilistic Graphic Model

확률 추론 문제

-흡연 환자의 엑스레이 진단에서 양성이 나타났을 때 폐암일 확률은?

-한국은행이 기준금리를 올렸고 S전자의 1분기 실적이 양호할 때 S전자 주식이 오를 확률?

확률 그래피컬 모델

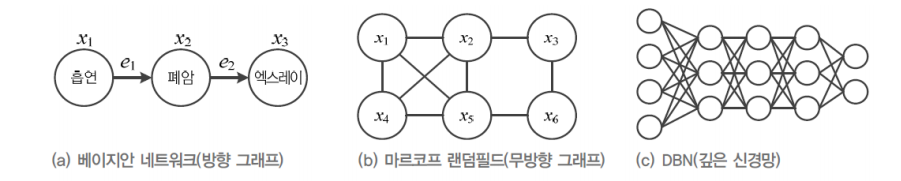
엑스레이, 흡연, 폐암을 확률변수로 정의하고 이들의 상호작용을 그래프로 표현하고, 그래프에서 확률 추론 수행

▪ 확률 이론과 그래프 이론을 결합하여 확률분포(probability distribution)를 표현하고, 관심있는 대상(확률변수)에 대한 확률을 계산할 수 있는 모델

▪ 대표적 모델 3가지

• 베이지안 네트워크 (베이지안 인프런스랑 결합하여 시험 필수)

• 마르코프 랜덤필드

• RBM과 DBN

◼ 결합확률

▪ 완벽한 확률 정보로서 모든 확률 추론 가능

• 예, 엑스레이가 양성일 때 폐암 확률은?

• 예, 흡연자의 폐암 확률은?

▪ 결합확률을 알아내는 일은 차원의 저주

• 𝑛개 확률변수가 있고, 각각 𝑞개의 값을 가진다면 총 𝑞 𝑛 − 1개의 확률을 알아내야 함

◼ 그래프 분해 (\*\*)

▪ 직접 상호작용하는 확률변수만 에지로 연결함(확률 그래피컬 모델의 핵심 아이디어)

• 베이지안 네트워크는 직접 인과관계가 있는 변수만 방향 에지로 연결

• 마르코프 랜덤필드는 이웃한 변수만 무방향 에지로 연결 • DBN은 이웃한 층 사이에만 무방향 에지로 연결

▪ 결합확률을 알아낼 필요가 없어지고, 분해된 그래프에서 부분집합의 확률분포만 알아내면 됨

▪ 에지 연결이 없는 노드는 중간 노드를 통해 상호작용을 함(예, 흡연과 엑스레이는 폐암을 통해 상호작용)

◼ 베이지안 네트워크 (Bayesian Network) 의 문제

◼ 구조 학습 (structure learning)

◼ 그래프 구조를 만드는 작업.

◼ 확률변수는 사람이 결정해야 하며, 확률변수가 정해지면 이들 간의 인과관계는 사람이 지정하 거나 데이터로부터 자동으로 알아낼 수 있다.

◼ 확률 학습 (probability learning)

◼ 노드 또는 노드와 노드 사이에 확률을 부여하는 작업이다.

◼ 부모가 없는 루트 노드는 사전 확률, 부모가 있는 노드는 조건부 확률을 알아낸다.

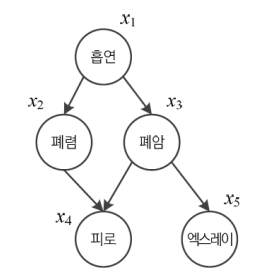
◼ 확률 추론 (probabilistic inference)

◼ 구조 학습과 확률 학습을 마친 후 테스트 단계 또는 현장 설치 후 수행하는

작업.

◼ 흡연자의 폐암 확률을 알아내는 것과 같은 각종 질문에 대한 확률을 추정하는 일이다.

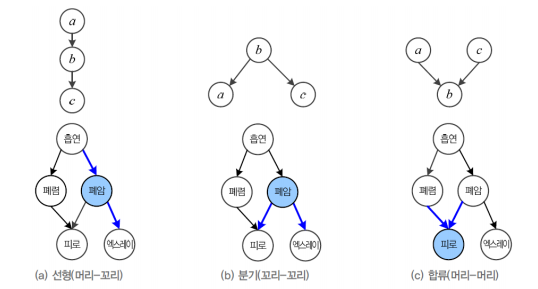
◼ 그래프 분해에 필요한 마르코프 조건

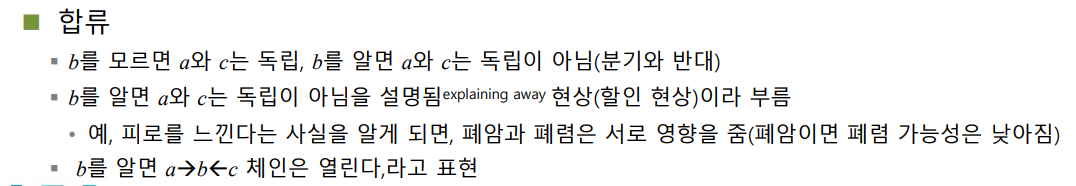
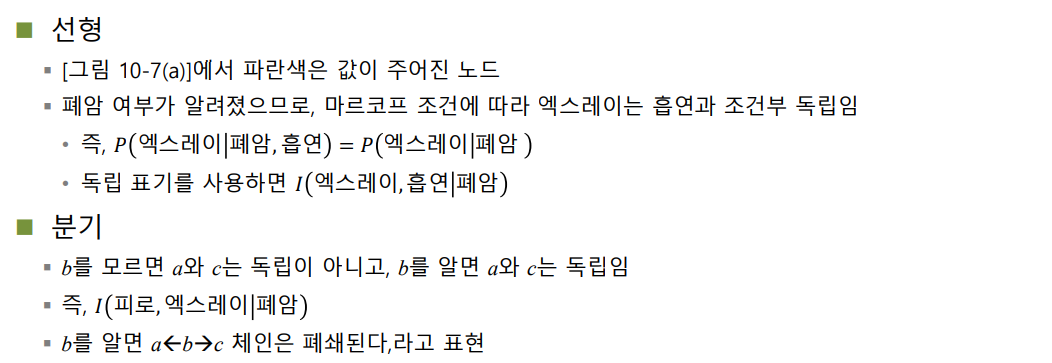
▪ 노드 𝑥의 부모를 𝑦라 할 때, 𝑦의 값이 주어지면 𝑥는 비후손(후손을 제외한 모든 노드)과 조건부 독립

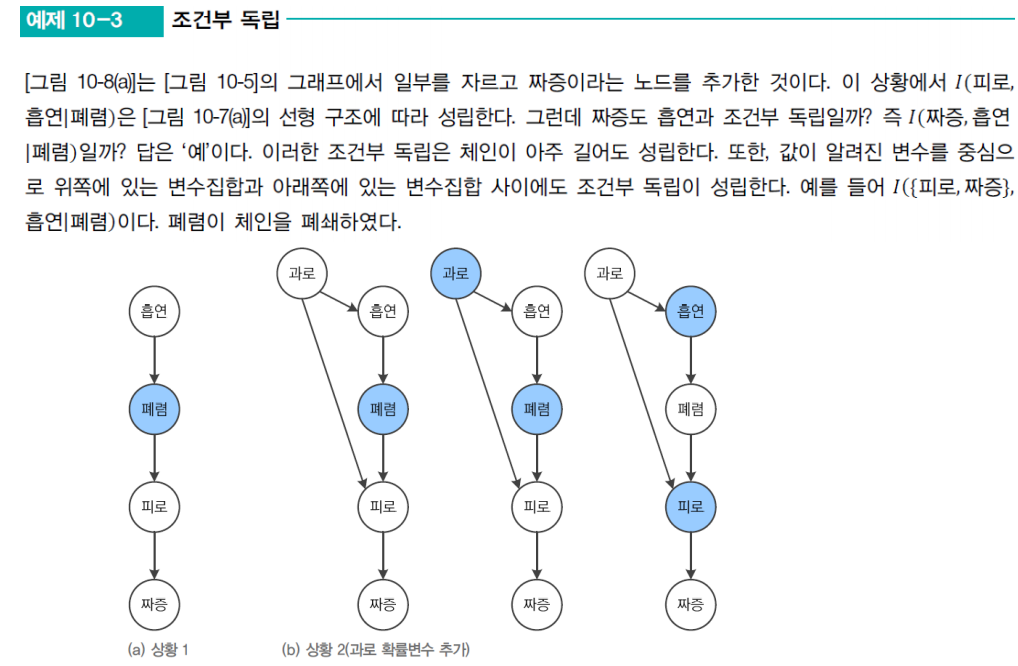
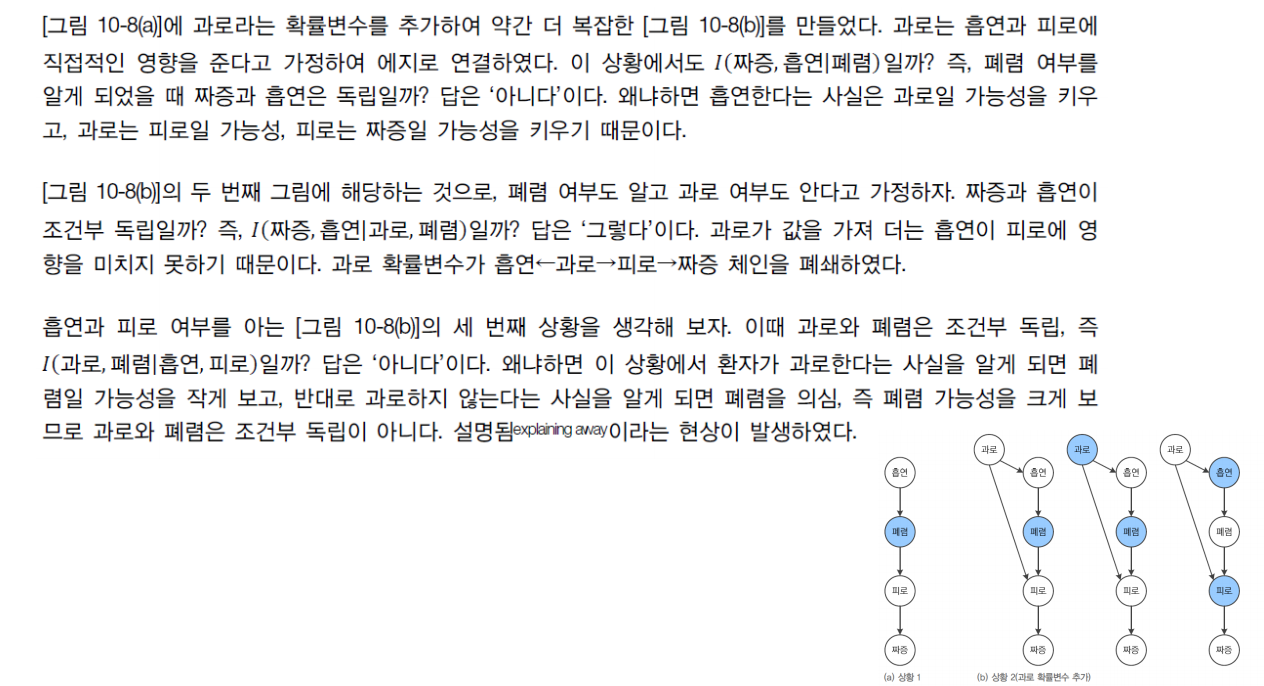
P(엑스레이 | 피로, 폐암, 폐렴, 흡연) = P(엑스레이 | 폐암)

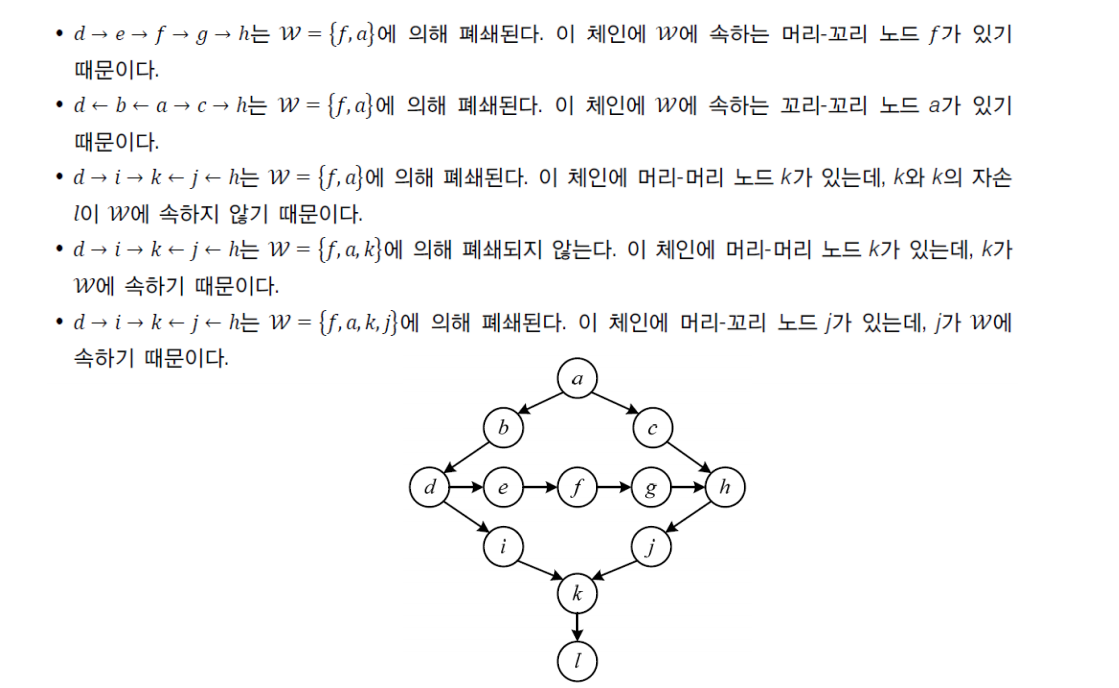
\*\*베이지안 네트워크에 확률을 부여하는 방법

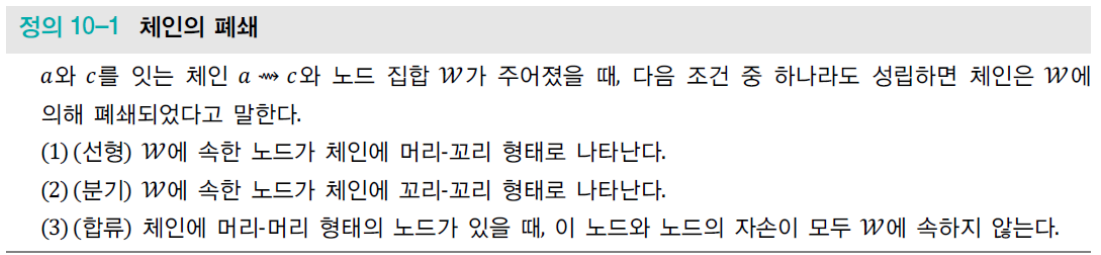
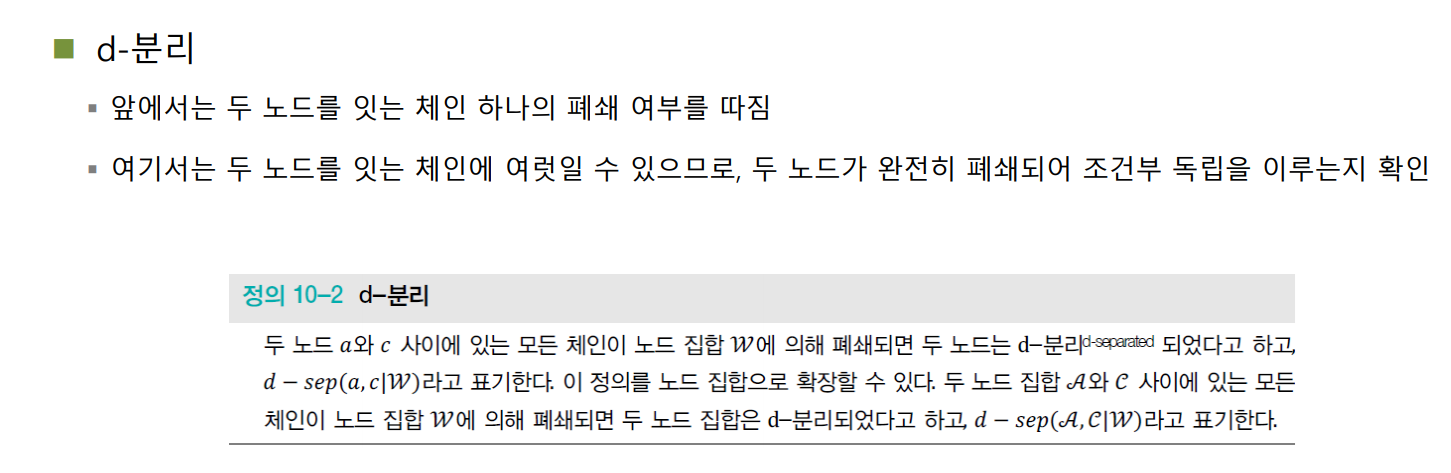
부모가 없는 루트 노드에는 사전 확률을 부여하고, 부모가 있는 노드에는 부모와 자식 사이로 한정하여 조건부 확률을 부여한다

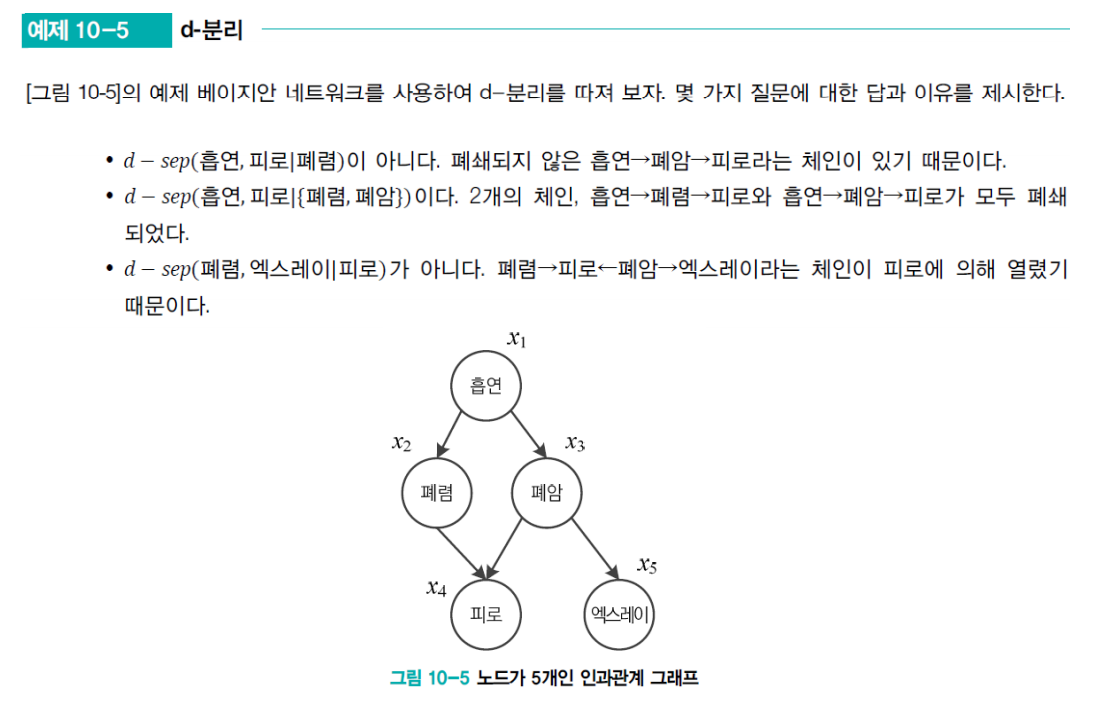
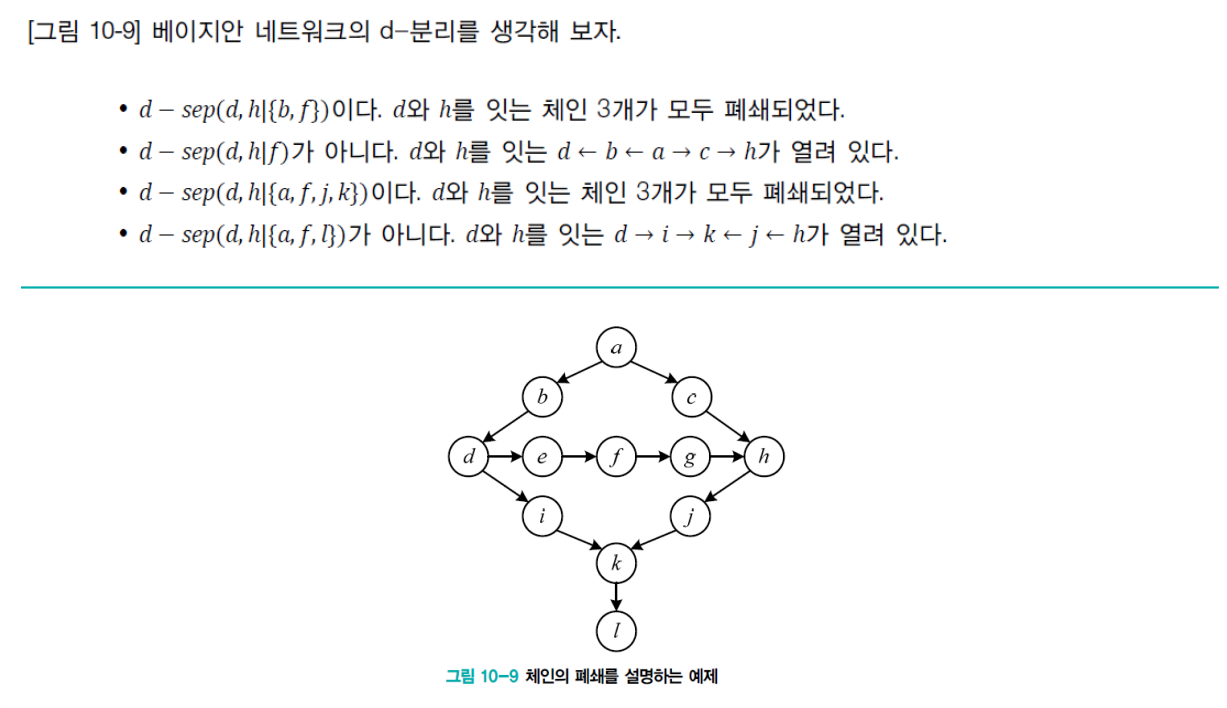
조건부 독립 – d-분리에 활용

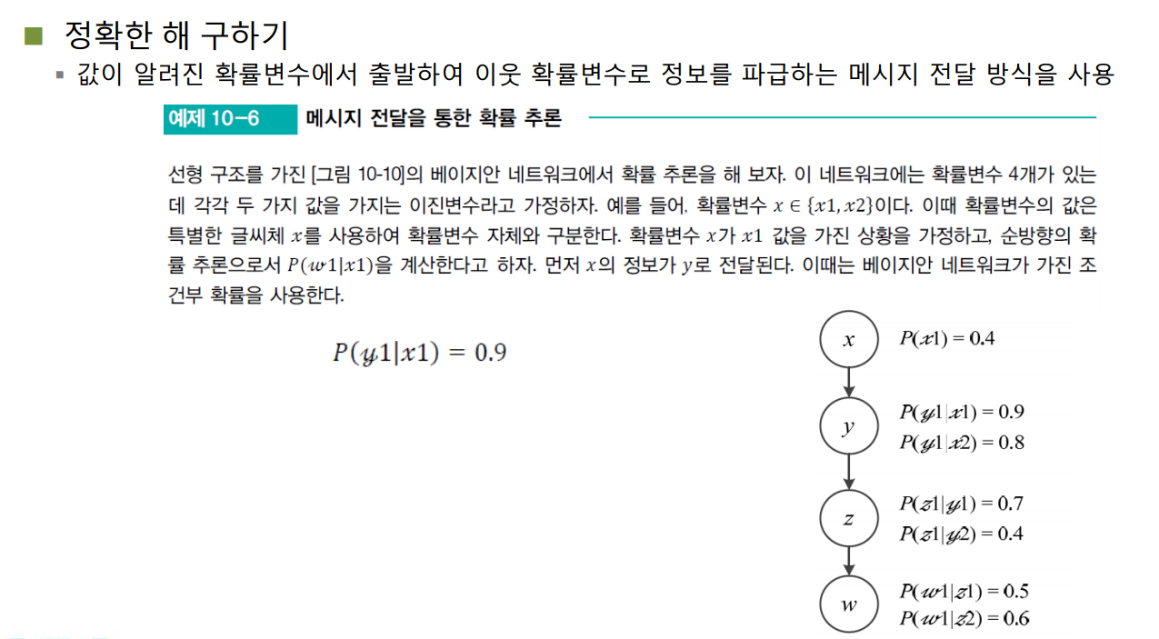










\*\*\*\*\*출제 100%

