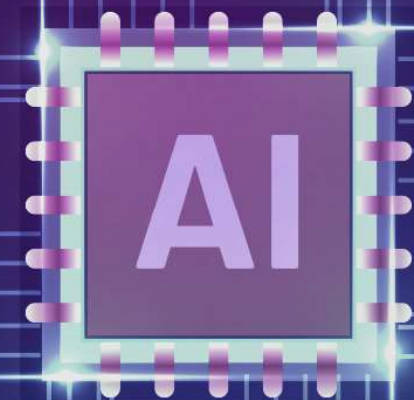


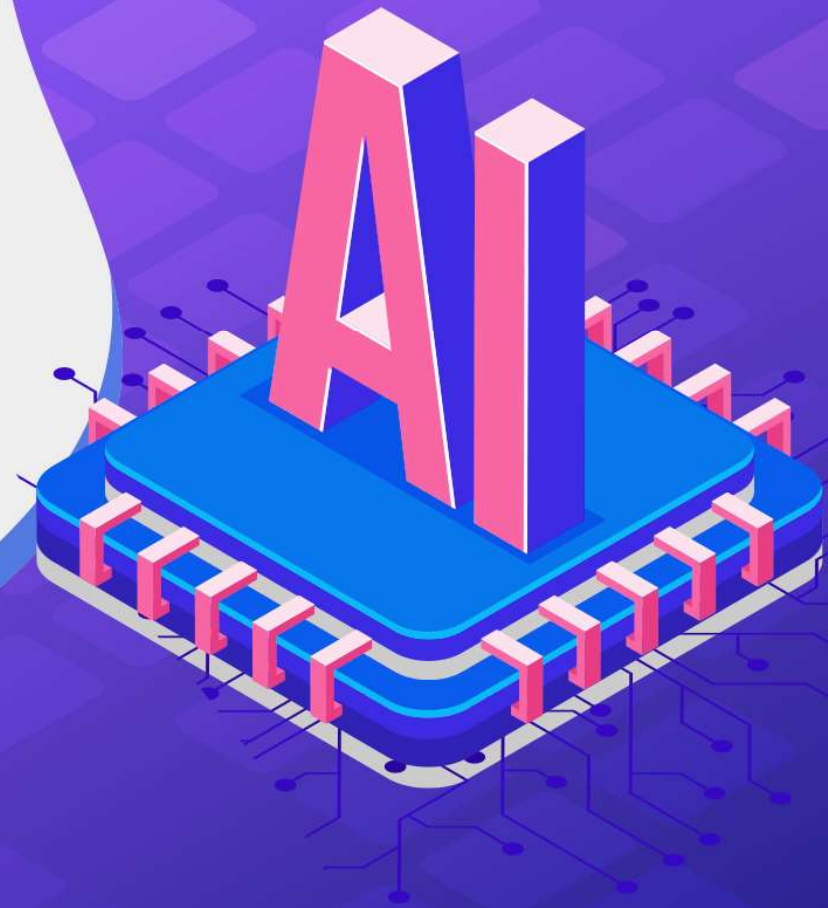
머신러닝의 학습과 분류 (2)

김재광 교수 (소프트웨어융합대학 글로벌융합학부)



Contents

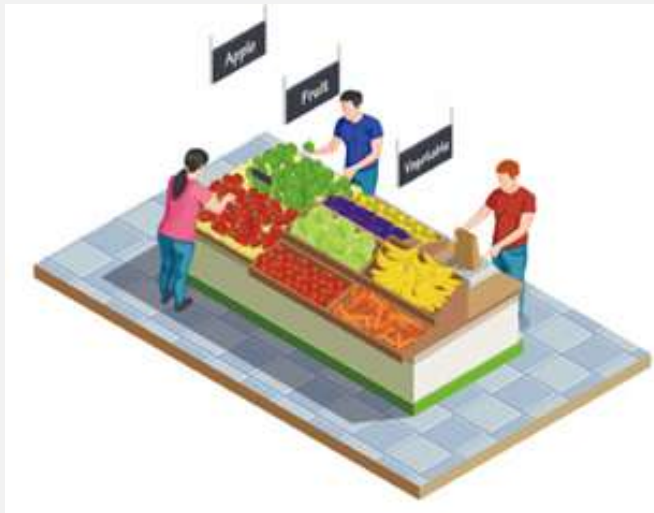
- 머신러닝의 비지도 학습
- 강화 학습
- 베이지안 네트워크와 은닉 마르코프 모델



AI 머신러닝의 비지도 학습

● 클러스터(cluster)와 클러스터링(clustering)

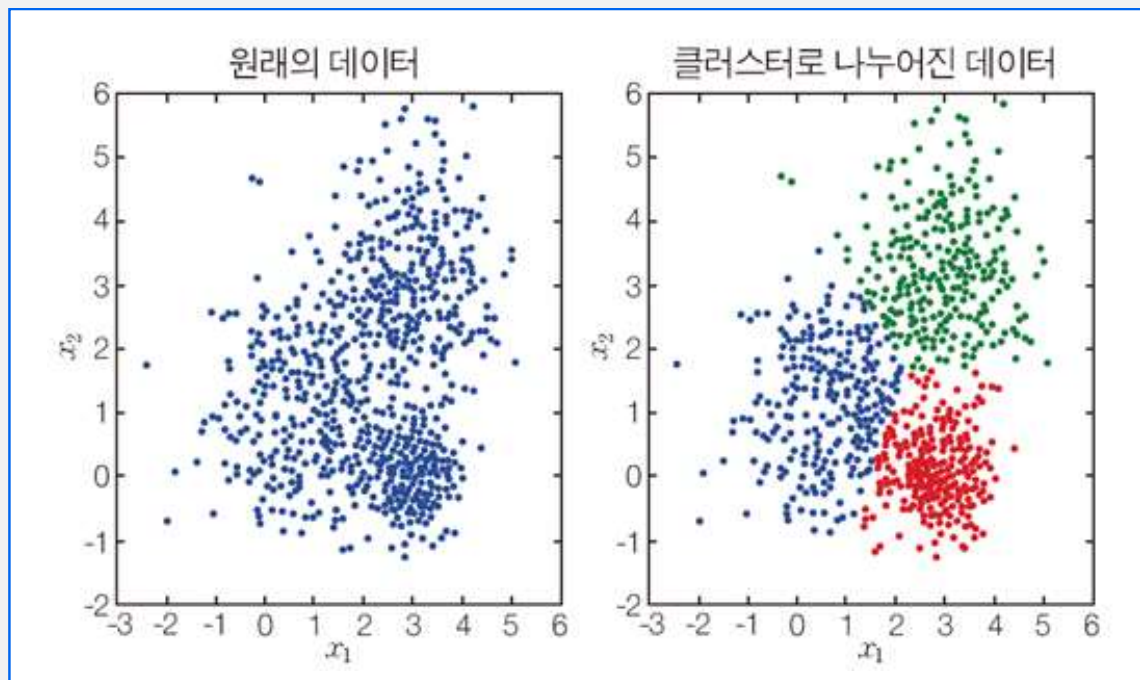
- 클러스터는 유사한 여러 개의 클래스로 나뉘어진 데이터
- 클러스터링은 유사한 특성을 가진 그룹들로 묶는 작업
- 같은 클러스터의 것은 다른 클러스터의 것보다 더 유사
- 이와 같은 유사한 것들끼리의 집합이 바로 클러스터



머신러닝의 비지도 학습 (Cont'd)

● 클러스터의 분류 예

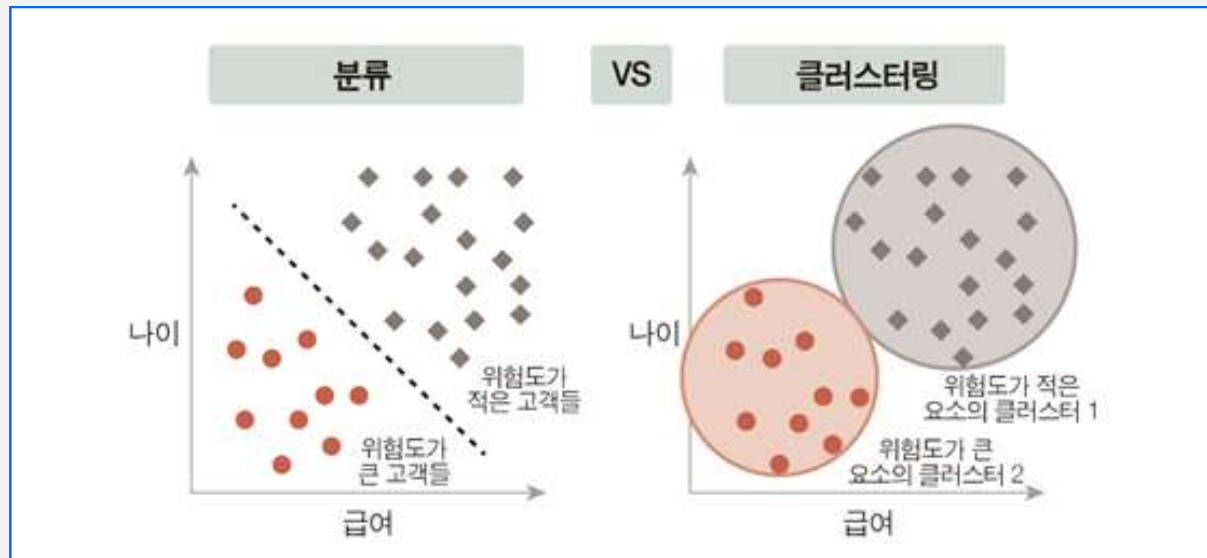
- 원래 데이터를 3개의 클러스터로 분류해 놓은 예
- 빨간색, 파란색, 녹색



머신러닝의 비지도 학습 (Cont'd)

● 분류와 클러스터링의 차이점

- 분류는 지도 학습 영역, 클러스터링은 비지도 학습 영역
- 분류는 데이터를 기준에 따라 직선으로 분류하는 것
- 클러스터링은 유사성에 따라 몇 개의 클러스터들로 묶는 것
- 급여, 나이, 위험도 상관관계의 예에서의 차이점의 예



머신러닝의 비지도 학습 (Cont'd)

● 비지도 학습

- 주어진 입력에 대응하는 출력 정보 없이 학습
- 데이터 분류에 대한 정보가 전혀 없이 패턴을 찾거나 데이터를 분류하려고 할 때 사용하는 학습 방법
- 데이터에 레이블을 전혀 사용하지 않음
- 관계를 스스로 학습한 후, 과일들을 각 그룹으로 알아서 묶기



머신러닝의 비지도 학습 (Cont'd)

● 비지도 학습

- 비슷한 성향의 고객을 그룹으로 묶기
- 블로그에서 주제별로 구분하기
- 유사한 꽃이나 동물들끼리 묶기
- 네트워크상에서의 비정상적인 접근의 탐지



머신러닝의 비지도 학습 (Cont'd)

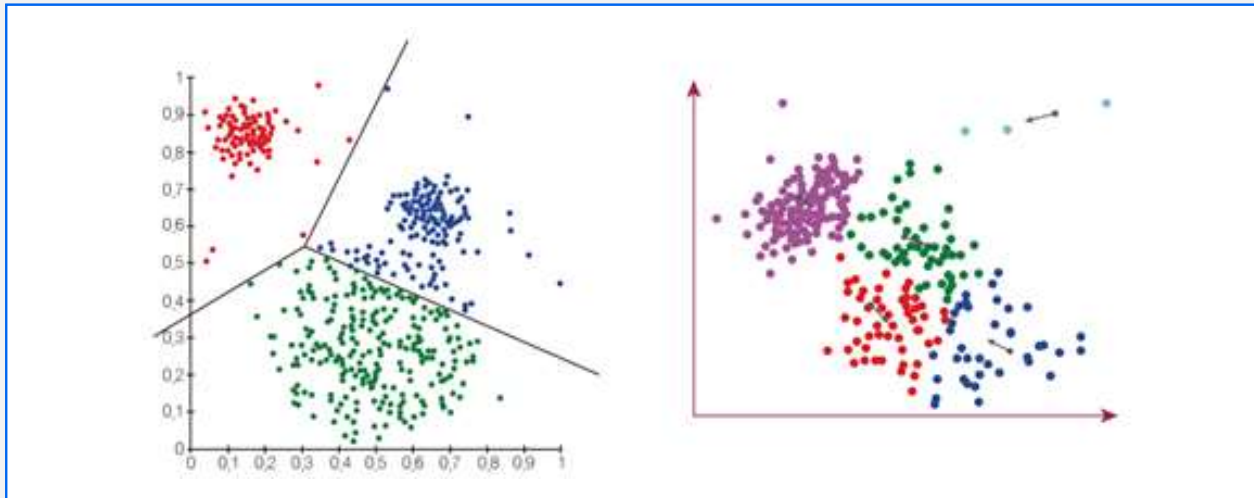
- 비지도 학습을 통한 클러스터링과 추천 시스템

- K-means 클러스터링
- 가우스 혼합 모델
- 계층적 클러스터링
- 추천 시스템 등

머신러닝의 비지도 학습 (Cont'd)

● K-means 클러스터링

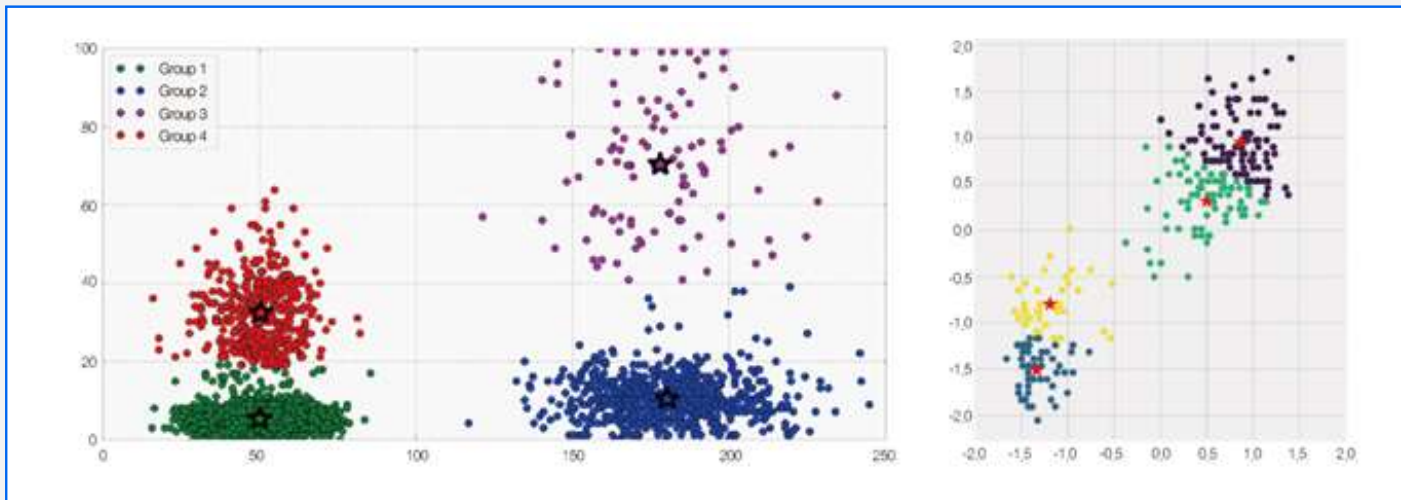
- 비지도 학습 알고리즘 중 대표적인 클러스터링 방법
- 우리말로 'K-평균 군집화'라고 함
- 간단하면서도 많이 쓰이는 클러스터링 방법 중 하나
- 유사한 특성을 가진 k개의 데이터 그룹으로 묶는 방법
- 예로 주어진 데이터 집합에서 3개와 4개의 클러스터들



머신러닝의 비지도 학습 (Cont'd)

● 클러스터와 클러스터 중심점(centroid)

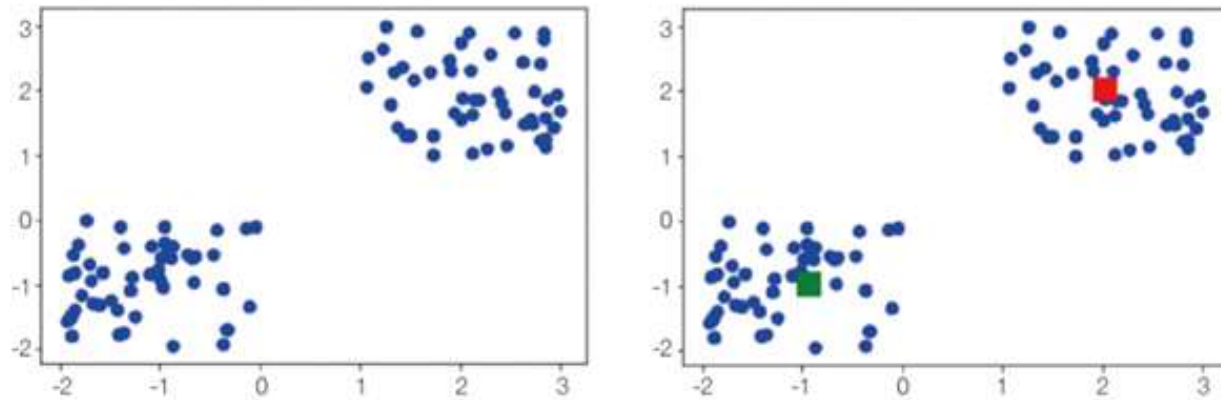
- 주어진 데이터 집합에 대해 k개의 클러스터 중심점 찾기
- 각 클러스터에는 클러스터 중심이 있음
- 각 점은 다른 중심점보다 지정된 클러스터 중심점에 더 가까움
- 4개의 클러스터로 구성된 2가지 예
- 별표는 각 클러스터의 중심점을 나타냄



🧠 머신러닝의 비지도 학습 (Cont'd)

● K-means 클러스터링 알고리즘 작동의 예

- K-means 클러스터링 알고리즘을 작동시킨 예
- 왼쪽은 원래의 데이터
- 오른쪽은 $k = 2$ 인 K-means 알고리즘을 작동시킨 결과
- 빨강과 녹색인 사각형은 각 클러스터의 중심점



머신러닝의 비지도 학습 (Cont'd)

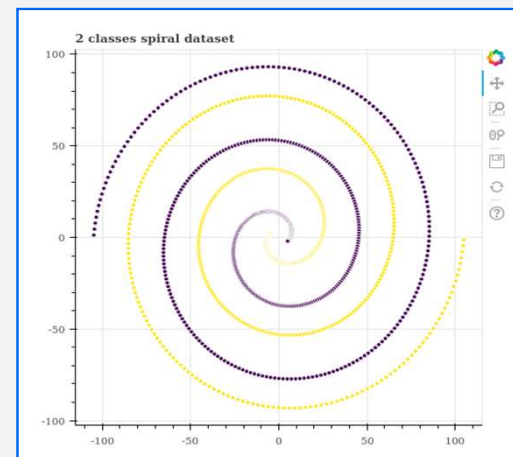
● K-means 클러스터링의 장단점

- 장점

- ✓ 알고리즘이 비교적 간단하고, 수행 속도가 빠르다는 점
- ✓ 주어진 데이터에 대한 사전 정보 없이 클러스터링을 함
- ✓ 데이터를 분류하는 머신러닝과 데이터 마이닝의 도구

- 단점

- ✓ 클러스터링의 개수 k 와 최초로 지정하는 중심점들에 따라 결과가 다소 달라질 수 있는 점
- ✓ 데이터가 convex 하다고 가정하고 있음



머신러닝의 비지도 학습 (Cont'd)

● K-means 클러스터링의 활용 분야

- 통계 : 주어진 데이터의 분류나 성향 분석
- 전자상거래 : 고객의 구매 이력으로 고객 분류
- 건강 관리 : 질병과 치료를 위한 패턴 탐지
- 패턴 : 유사한 이미지를 그룹화
- 재무 : 신용카드 사기 탐지
- 회사 : 매출 등을 토대로 회사의 등급 분류
- 기술 : 네트워크 침입과 악의적 활동 탐지
- 기상 예보 : 폭풍 예측

머신러닝의 비지도 학습 (Cont'd)

● 추천 시스템(Recommender System)

- 추천을 위해 연관 데이터 정의에 도움 주는 클러스터링 방법
- 사용자의 '선호도'를 예측하는 정보 필터링의 일종
- 네이버나 구글 등에서 상업적으로 활용 중
- 현재 검색해본 책이나 동영상 등의 추천
- 또 인기 있는 식당, 연구 관련 기사, 금융 서비스 등 추천



머신러닝의 비지도 학습 (Cont'd)

● 추천 시스템의 활용과 기타 비지도 학습 방법

- 가령 교보문고에서 책을 검색하면 그 사람이 이전에 검색했던 도서나 관련 도서를 알려줌
- 사용자의 검색 경험 정보 파악, 적절한 광고 내보내기



- 그 외 비지도 학습 방법에는 가우스 혼합 모델, 계층적 클러스터링, PCA/T-SNE 등이 있음

머신러닝의 비지도 학습 (Cont'd)

● 지도 학습과 비지도 학습의 특징 비교

기반	지도 학습	비지도 학습
입력 데이터	입력과 출력(값 또는 레이블)이 지정된 데이터를 사용하여 학습함	출력값이나 레이블이 전혀 없는 데이터를 사용하여 학습함
주요 기능	분류, 회귀	클러스터링, 추천 시스템
계산의 복잡성	비교적 간단함	상당히 복잡함
정확성	매우 정확함	다소 덜 정확함

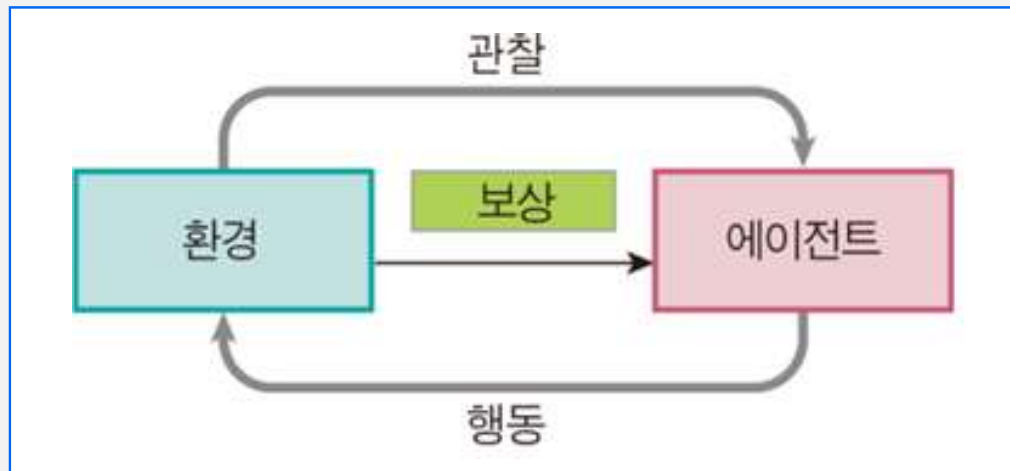
● 강화 학습이란?

- 강화 학습은 시행착오를 통해 보상하는 행동 학습
- 최적의 값을 추구하기 위해 당근과 채찍을 사용
- 로봇이 미로에서 옳은 방향으로 진입하면 +2점, 막힌 길로 들어가면 -3점 등
- 입출력이 쌍으로 된 훈련 집합으로 제시되지 않는다는 점에서 일반적인 지도 학습과는 다름

강화 학습 (Cont'd)

● 강화 학습의 응용 분야

- 보상(reward)이 주어지는 문제 해결에 매우 효과적
- 통신망, 로봇 제어, 엘리베이터 제어, 그리고 체스와 바둑 같은 게임에 주로 응용됨
- 알파고도 강화 학습을 통해 실력 향상
- 최근 게임에서는 거의 필수적으로 강화 학습이 사용됨



베이저안 네트워크와 은닉 마르코프 모델

● 베이즈의 정리(Bayesian theorem)

- 과거의 데이터들을 기반으로 미래를 예측하는 모델
- 머신러닝, 통계학, 경제학에 널리 적용되고 있음
- 검색 엔진, 스팸 메일 차단, 금융 이론, 승부 예측, 기상 예측, 의료 분야, 인공지능 등에 폭넓게 활용됨
- 베이즈(Thomas Bayes)는 확률에 대한 연구로 유명
- 베이즈의 정리는 확률적 추론에 이용되는 정리

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

베이지안 네트워크와 은닉 마르코프 모델 (Cont'd)

● '베이즈의 정리'의 응용 예

- $P(Y|X)$ 는 'X가 주어졌을 때 Y가 발생할 조건부 확률'
- 비교적 구하기 쉬운 확률을 통해 어려운 확률을 추정
- 나이브 베이지안과 은닉 마르코프 모델 등에 적용

- 증상과 의학 진단에 활용
- $P(X)$ 는 환자 중에 열이 많이 나는 환자가 있을 확률
- $P(Y)$ 는 환자 중에 독감에 걸린 환자가 있을 확률
- $P(Y|X)$ 는 열이 많이 나는 환자가 독감 환자일 확률
- $P(X|Y)$ 는 독감 환자가 열이 많이 나는 환자일 확률

베이지안 네트워크와 은닉 마르코프 모델 (Cont'd)

- 베이지안 네트워크(Bayesian network)

- '빌리프 네트워크(Belief network)'라고도 불림
- 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델
- 추론과 학습을 수행하기 위한 효과적인 알고리즘이 존재
- 예를 들어, 질환과 증상 사이의 확률 관계를 나타낼 수 있음
- 증상이 주어지면 다양한 질병의 존재 확률 계산 가능

베이지안 네트워크와 은닉 마르코프 모델 (Cont'd)

- 은닉 마르코프 모델(Hidden Markov Model, HMM)

- HMM은 마르코프(Markov) 모델의 일종
- 은닉된 상태와 관찰 가능한 결과로 이루어진 확률형 모델
- 동적 베이지안 네트워크로 간단히 나타낼 수 있음
- 대량의 데이터를 통계적으로 분석하여 추론에 응용
- 음성인식, 자연어 처리 등에 활용