1 통계 모델과 확률 분포

1 통계 모델과 확률 분포

- 머신 러닝
- 일반화 선형 모델과 기저함수
- 주요 기저 함수
- 기타 비선형 함수

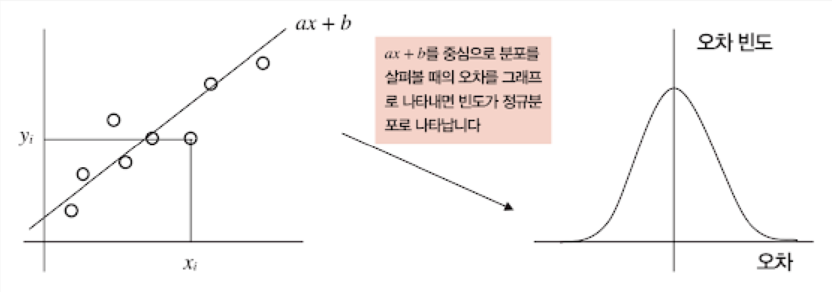
2 확률 기반

- 일반적인 사건(Event)의 발생들은 확률로 표현 가능
  - 동전을 던질 때 앞면과 뒷면 어느 방향이 나올 것인가
  - 태풍이 올 때 날씨는 어떨까?
- 회귀 분석 시 설명 변수와 목적 변수들 간에 선형함수를 통한 피팅 방법 사용
- 피팅을 위한 오차 함수 설정 시 확률이 관여됨
  - 두 변수 간의 측정 결과에 따른 측정 오차가 확률 분포를 따름
  - 측정오차는 정규 분포 등의 분포로 표현

2 확률 기반

- 확률 분포 모델
  - 설명 변수와 목적 변수가 갖는 어떤 확률에 근거한 관계
  - 임의의 분포를 갖는 오차로 표현

[측정 오차와 정규 분포]



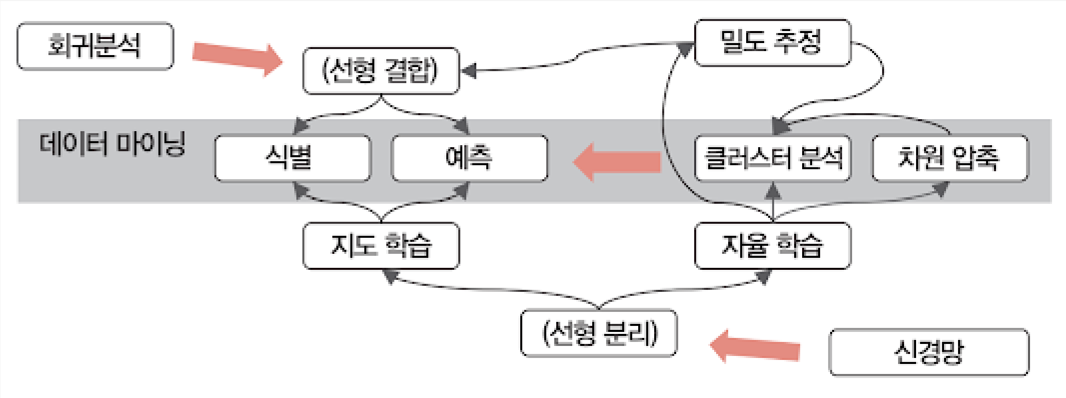
※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

3 머신 러닝

- 학습
  - 신경망에서 데이터 입력을 통해 스스로 네트워크의 가중치를 변화시키는 것
- 머신 러닝
  - 신경망의 학습처럼 '기계가 학습한다'는 개념을 의미하는 용어
  - 입력 데이터의 특성과 분포 경향 등에서 자동으로 데이터를 나누거나 재구성
  - 통계 기반 머신 러닝에서는 확률 개념이 크게 관여

3 머신 러닝

▶ 회귀 분석이 머신 러닝 및 신경망과의 연관 여부



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

3 머신 러닝

- 입력 데이터의 회귀 분석 시 주로 데이터 식별 및 예측에 목적을 둠
- 따라서 데이터를 선형 결합으로 나타내는 특성 이용
- 회귀 분석 지도학습
  - 정답 정보가 결합된 학습 데이터(또는 훈련 데이터)로 데이터의 특징을 모델링
- 신경망 학습
  - 선형 분리로 변환하는 특성 이용
  - 선형 분리가 가능하면 회귀 분석과 같은 선형 결합 방법 사용

3 머신 러닝

- 자율 학습
  - 신경망을 사용하지 않는 경우
    - ✓ 입력 데이터의 정답을 모르는 상태에서 사용
    - ✓ 클러스터 분석, 차원 압축, 밀도 추정 등에 해당
  - 데이터 마이닝
    - ✓ 식별, 예측, 클러스터 분석 등 실행 시 데이터의 새로운 특징을 찾아내는 작업

3 머신 러닝

▶ 데이터 마이닝

- 분류(Classification)
  : 일정한 집단에 대한 특징 정의를 통해
  분류 및 구분을 추론
- 군집화(Clustering)
  : 구체적인 특성을 공유하는 군집 검색
  : 군집화는 미리 정의된 특성에 대한 정보를 가지지 않는다는 점에서 분류와 다름
- 연관성(Association)
  : 동시에 발생한 사건 간의 관계를 정의

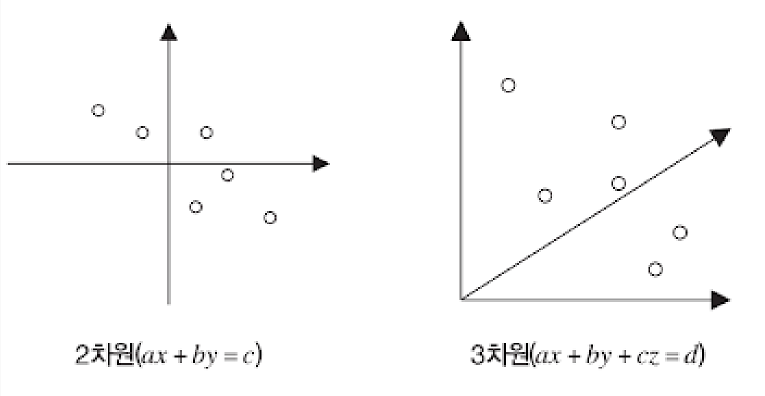
3 머신 러닝

- 데이터 마이닝
  - 연속성(Sequencing)
    : 특정 기간에 걸쳐 발생하는 관계를 규명,
    기간의 특성을 제외하면 연관성 분석과 유사
  - 예측(Forecasting)
    : 대용량 데이터 집합 내의 패턴을 기반으로 미래 예측

4 기저 함수

- 정규 직교 기저
  - 좌표축 각각이 직교하는 선형 독립 상태에서 벡터로 나타내는 좌표축
  - 차원 수 만큼 표현이 가능하여 3차원 이상도 표현

[정규 직교 기저]



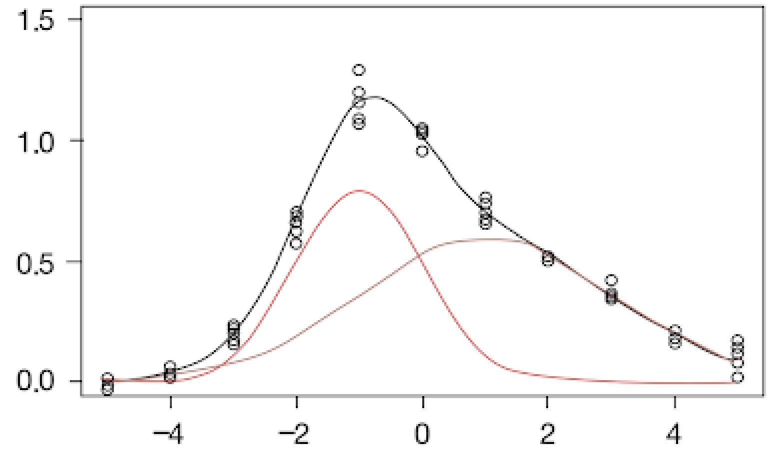
※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 기저 함수

- 회귀 분석을 정규 직교 기저로 설명하면, 선형 독립인 설명 변수의 합에 있는 목적 변수를 표현하는 것
- 설명 변수가 선형 독립인 경우만 사용
- 기저 함수
  - 설명 변수를 함수 형태로 나타낸 것
- 기저 함수 사용 시 개념
  - 일반화 선형 모델
    : 정규 분포 이외의 분포를 다루는 모델
  - 혼합 정규 모델
    : 혼합 정규 분포를 다루는 모델
    : 기저 함수와 기저 함수의 선형 결합으로 표현된 모델

4 기저 함수

[혼합 정규 분포]



- 검은색 선
  : 진한 붉은색 선과
  연한 붉은색 선의
  정규 분포 밀도
  함수의 합
- 원
  : 검은색 선에
  정규 분포를 따르는
  오차를 추가한 값

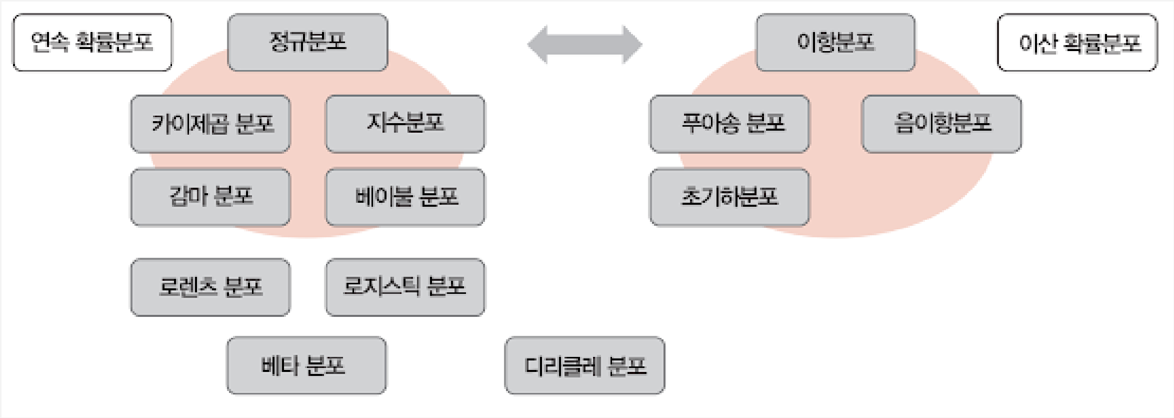
※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 주요 기저 함수

1 주요 기저 함수

▶ 기저 함수는 확률 분포에 따라 연속 확률 분포와 이산 확률 분포로 나뉨

[모델이 되는 주요 함수]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 정규 분포

- 가장 많이 사용하는 일반적인 분포 개념
- 가우스 분포라고도 함
- 실험의 측정 오차나 사회 현상 등
  자연계 현상은 정규 분포를 따른 경향이 있음
- 정규 분포를 엄격히 따르지 않아도 계산이나
  모델의 단순화를 위해 데이터 분포를 정규 분포로
  가정하는 경우가 많음

2 정규 분포

▶ $f(x)$: 정규 분포, μ: 평균(기대값), σ: 표준 편차, σ^2: 분산

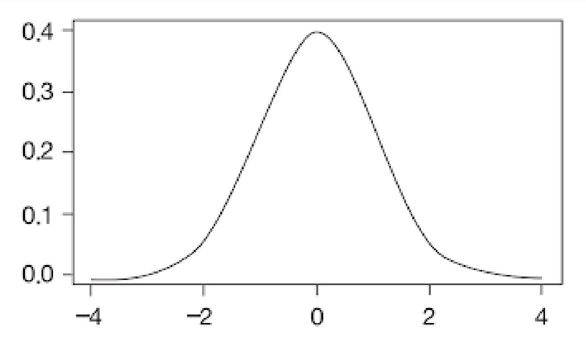
[정규 분포 식]

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

$$E(x) = \mu$$

$$\mathrm{V(x)} = \sigma^2$$

[정규 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

3 감마 분포

- 특정 수의 사건이 일어날 때까지 걸리는 시간에 관한 연속 확률 분포
- 감마 함수 Γ는 자연수 집합 N이 주어졌을 때 팩토리얼 $N!$과 동일
- 모양 매개 변수가 k 고 크기 매개 변수가 θ일 때
  - 평균 : $k\theta$
  - 분산 : $k\theta^2$

3 감마 분포

▶ $k=1$일 때 지수분포, k 가 정수일 때 얼랑 분포, k 가 반정수($^{(2n-1)}/_{2}$)고 θ가 2일 때는 카이 제곱 분포라 함

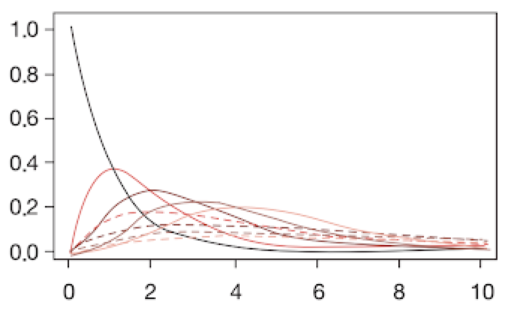
[감마 분포 식]

$$f(x)=\frac{e^{-x/\theta}}{\Gamma(k)\theta^{k}}$$

$$E(x)=k\theta$$

$$\mathrm{V(x)}=\ k\theta^{2}$$

[감마 분포 형태]



- 검은색 : k: 1 θ: 2.0
- 붉은색 실선 : k: 2 θ: 2.0
- 갈색선 : k: 3 θ: 2.0

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 지수 분포

- 감마 분포의 특별한 형태로
  사건이 일어나는 시간 간격의 확률 분포
- 푸아송 분포와도 연관
- λ : 단위 시간에 일어나는 평균 회수
- 평균 : $1/\lambda$
- 분산 : $1/\lambda^2$

4 지수 분포

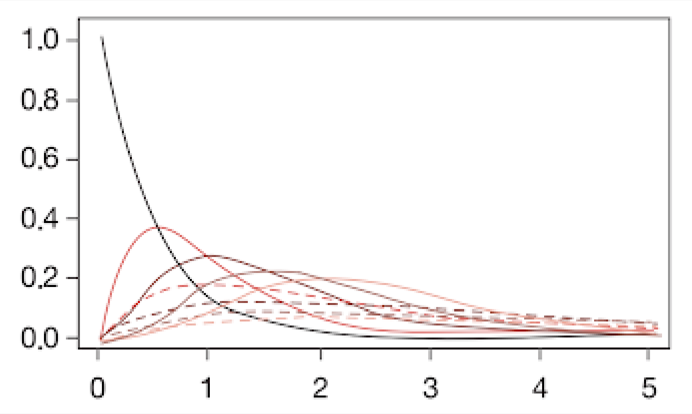
[지수 분포 식]

$$f(x) = \lambda e^{-\lambda x}$$

$$E(x) = {}^{1}/_{\lambda}$$

$$V(x) = {}^{1}/_{\lambda^2}$$

[지수 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

5 라플라스 분포

- 보통의 정규분포보다 최고점이 더 높은 분포
- 생물학, 재무, 경제 분야에서 데이터를 모형화 시 사용
- 평균 : μ
- 분산 : $2\sigma^2$

5 라플라스 분포

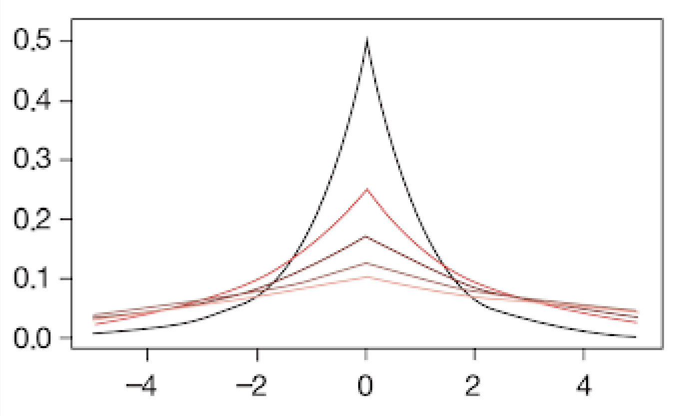
[라플라스 분포 식]

$$f(x) = \frac{1}{2\sigma} exp(-\frac{|x-\mu|}{\sigma})$$

$$E(x) = \mu$$

$$\mathrm{V}(\mathrm{x}) = 2\sigma^2$$

[라플라스 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

6 베타 분포

- 2개의 변수를 갖는 특수 함수인 베타 함수를 이용한 분포
- 베타 함수를 $B(\alpha,\beta)$로 나타낼 때,
  $\alpha>0$, $\beta>0$, $0\leq x\leq 1$ (x는 설명 변수)라는 조건 설정
- 매개변수 α, β를 바꾸면 다양한 분포로 표현 가능
- 베이즈 통계학에서는 사전 분포 모델로
  이용하는 경우가 많음

6 베타 분포

[베타 분포 식]

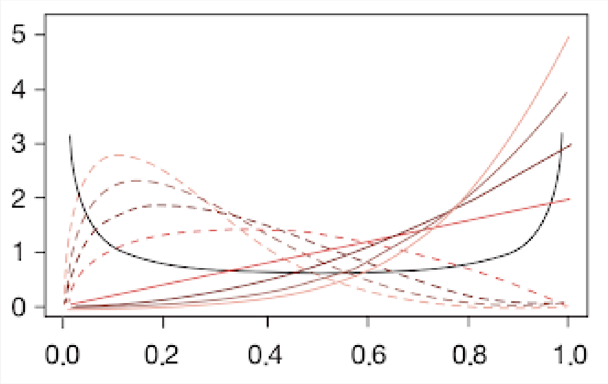
$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

$$E(x) = \frac{\alpha}{\alpha+\beta}$$

$$\mathrm{V}(\mathrm{x}) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

[베타 분포 형태]

$$B(\alpha,\beta) = \Gamma(\alpha+\beta)/[\Gamma(\alpha)\Gamma(\beta)]$$



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

7 디리클레 분포

- 베타 분포를 다변량으로 확장한 것
- 다변량 베타 분포라고도 함
- k는 2 이상의 자연수, α는 매개변수,$B(\alpha)$는 베타 함수, Γ는 감마 함수

[디리클레 분포 식]

$$f(x) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} x_i^{\alpha_i - 1} \qquad B(\alpha) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\prod_{i=1}^{K} \alpha_i)}$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

7 디리클레 분포

- 연속 함수지만 2차원 평면에서는 연속 함수로 표현 불가
- 어떤 현상이 일어나는 횟수를 확률 변수로 두는 분포가 다항 분포
- 디리클레 분포는 확률 자체를 확률 변수로 두는 분포로 자연어 처리 등에 많이 사용

8 이항 분포

- 베르누이 시행
  - 두 가지 경우 중 어느 하나가 나오는 가와 같은 실험
- 이항 분포
  - 베르누이 시행을 여러 번 반복했을 때 확률 분포
- 정규 분포나 푸아송 분포와 비슷

8 이항 분포

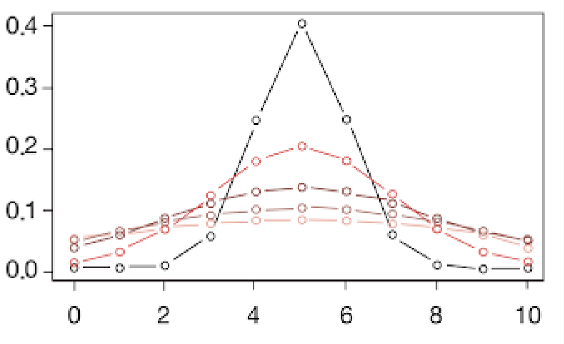
[이항 분포 식]

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$E(x)=np$$

$$V(x)=np(1-p)$$

[이항 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

8 이항 분포

- X 는 확률 분포, 확률 p 로 성공하는 실험에 있어 p^k는 k 번 성공할 확률, n 번 실행하는 경우 ${}_nC_k$ 의 조합만큼 성공할 확률이 나타날 가능성이 있음

- 실패 회수의 확률과 함께 계산 시 n 번의 독립된 시도에서 k 회 성공할 확률을 나타낼 수 있음

9 음이항 분포

- r번 성공하는데 필요한 시행 회수 k 의 분포
- 생명 과학 분야에 많이 사용

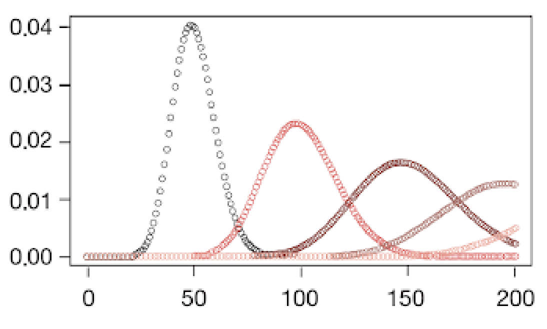
[음이항 분포 식]

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$$E(x) = r/p$$

$$\mathrm{V}(\mathrm{x}) = r(1-p)/p^2$$

[음이항 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

10 푸아송 분포

- 일정 시간 간격에 걸쳐 평균 λ번 일어나는 현상이 k번 발생할 확률 분포 X
- 푸아송 분포
  - 단위 시간에 사건이 일어날 확률
- 지수분포
  - 사건이 일어난 후 다시 발생할 때까지의 시간 간격에 관한 확률 밀도
- 푸아송 분포와 지수 분포는 같은 사건의 발생 확률을 다른 측면에서 보는 것

10 푸아송 분포

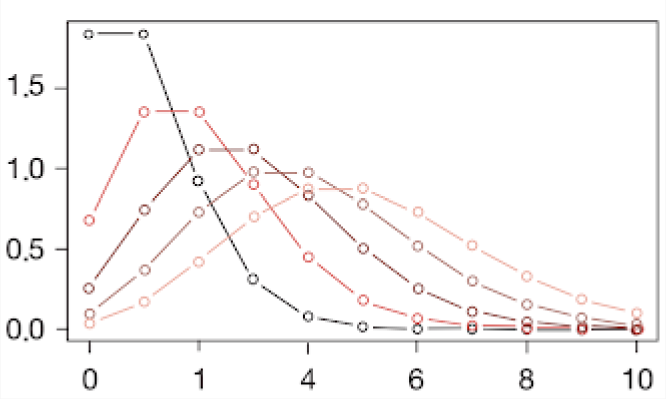
[푸아송 분포 식]

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$E(x) = \lambda$$

$$V(x) = \lambda$$

[푸아송 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

11 카이제곱 분포

- 감마 분포의 특별한 형태
- 통계적 추론에서는 카이제곱 검정으로 자주 이용
- 집단을 몇 가지로 나눴을 때 크기가 작은 집단에 보편성이 있는지 확인
- 임상 실험이나 사회 과학 설문 조사 등에 자주 사용
- 평균 k는 2 이상의 자연수, x는 설명 변수, Γ는 감마 함수

11 카이제곱 분포

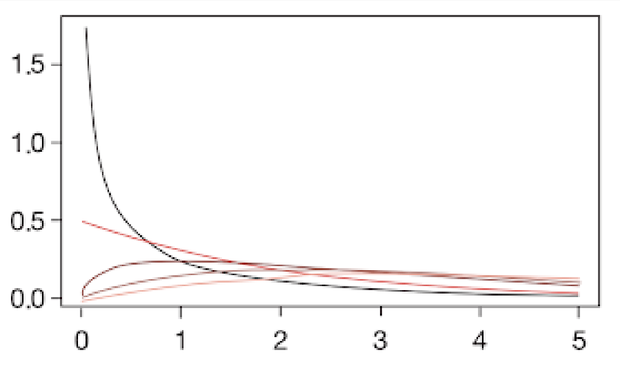
[카이제곱 분포 식]

$$f(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

$$E(x) = k$$

$$V(x) = 2k$$

[카이제곱 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

12 초기하 분포

- 반복하지 않는 시도에서 사건이 발생할 확률 분포
- 예)
  - 빨간 구슬과 흰 구슬이 들어 있는 주머니에서 구슬을 빼는 시도를 n번 할 때 빨간 구슬 k 개를 꺼낼 수 있는 확률을 구할 수 있음
  - 구슬 전체 개수는 N, 원하는 구슬을 꺼낸 전체 성공 회수는 K

12 초기하 분포

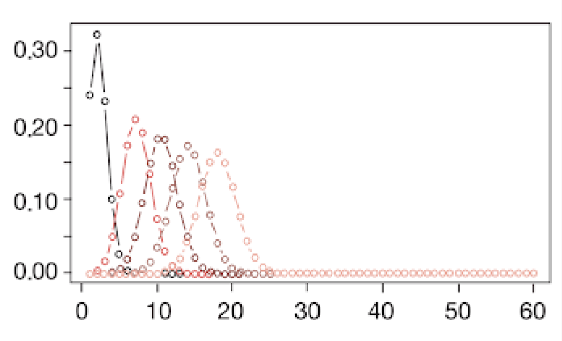
[초기하 분포 식]

$$P(X=k) = \binom{n}{k}\binom{N-n}{K-k} / \binom{N}{k}$$

$$E(x) = nK/N$$

$$V(x) = (N-n)m(N-K)K/(N-1)N^2$$

[초기하 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

13 코시 분포

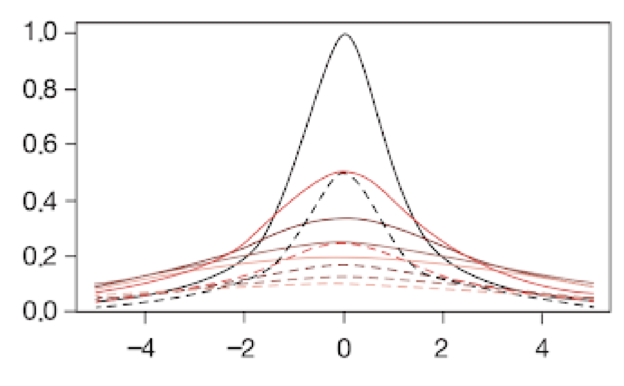
- 로렌츠 분포라고도 함
- 물리학에서는 브라이트-위그너 분포라고 함
- 분광학에서의 공명 현상 등 전자파와
  방사선의 스펙트럼 분석 등에서 자주 사용
- π는 원주율, γ는 척도 모수
  (확률 분포가 퍼져 있는 정도를 결정하는 집단의 수)
- 정규 분포와 비슷한 모양
- 그래프가 서서히 완만해 지며 중심이 그래프 아래로 잡힘

13 코시 분포

[코시 분포(로렌츠분포) 식]

$$f(x) = \frac{1}{\pi\gamma\left\{1 + (\frac{x-\mu}{\gamma})^2\right\}}$$

[코시 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

14 로지스틱 분포

- 확률 분포의 확률 변수가 특정값보다 작거나 같은 확률을 나타내는 누적 분포 함수가 로지스틱 함수인 분포
- 형태는 정규 분포와 비슷하지만 그래프의 아래가 길어 평균에서 멀어지더라도 정규 분포처럼 곡선이 내려가지 않음
- s는 표준 편차, μ는 평균, π는 원주율

14 로지스틱 분포

[로지스틱 분포 식]

$$f(x) = \frac{1}{s(1+e^{\frac{x-\mu}{s}})^2}, s > 0$$

$$E(x) = \mu$$

$$\mathrm{V(x)} = s^2\pi^2/3$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

15 베이불 분포

- 물체의 부피와 강도의 관계를 나타내는 분포
- 기기의 수명과 고장 시간 등의 신뢰성을 나타내는 지표로 이용
- $k(m)$는 실행 회수, $\lambda(\eta)$는 단위 시간에 일어나는 평균 회수

15 베이불 분포

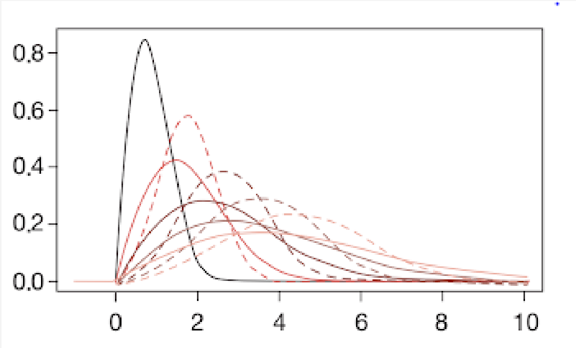
[베이불 분포 식]

$$f(x) = \frac{m}{\eta}(\frac{x}{\eta})^{m-1}exp\{-(\frac{x}{\eta})^{m}\}$$

$$E(x) = \eta\Gamma(1 + {}^{1}/_{m})$$

$$V(x) = \eta^{2}[\Gamma(1 + {}^{2}/_{m}) - (\Gamma(1 + {}^{1}/_{m})^{2}]$$

[베이불 분포 형태]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

16 손실 함수와 경사 하강법

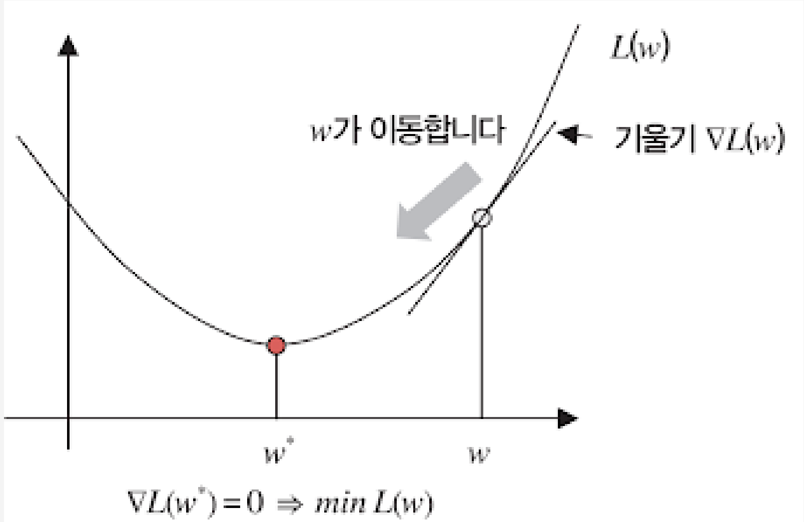
- 회귀 분석을 사용하려는 함수가 존재할 때,
  회귀 분석은 오차를 제곱한 것의 합계인 함수를
  목적 함수로 설정
- 목적 함수가 최소값이 되도록 계산
- 목적 함수와 비슷한 함수로 손실 함수 사용

16 손실 함수와 경사 하강법

- 손실 함수의 최소값 산출 식
  - 경사 하강법
    - ✓ 가중치 벡터 $\boldsymbol{w}$의 함수 $\boldsymbol{L}$ 로 작성
    - ✓ 손실 함수를 어떤 점 w_i로 편미분 한 것을 $\boldsymbol{L}$의 기울기를 $\nabla \boldsymbol{L}(\boldsymbol{w})$
    - ✓ 기울기 값이 $\nabla \boldsymbol{L}(\boldsymbol{w}^*) = 0$ 이 될 때의 $\boldsymbol{w}^*$ 가 구하려는 가중치
  - 최대 가능도 방법

16 손실 함수와 경사 하강법

[경사 하강법]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

3 베이즈 통계학

1 베이즈 통계학

- 베이즈 정리
- 로짓 함수
- 최대 가능도 추정
- EM 알고리즘
- 베이즈 판별 분석

2 베이즈 정리

- 조건부 확률에 관한 법칙
- 두 종류의 조건부 확률 사이의 관계 정의

[베이즈 정리]

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{5} P(B|A_j)P(A_j)}$$

- B라는 조건 아래 A가 일어날 조건부 확률의 정의
- 일반적인 형태의 베이즈 정리 식
- $A_1, A_2, \ldots, A_n$이 서로 관계가 없는 사건일 때 사용할 수 있는 베이즈 정리 식

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 베이즈 정리

- 승산
  - 확률 $\mathbf{p}$ 가 있을 때 $\boldsymbol{p}/(\mathbf{1}-\boldsymbol{p})$를 계산한 수치
  - 수치가 높을 수록 이길 가능성이 작아짐
  - 이 값의 로그를 취한 것을 로짓 함수
  - 로지스틱 회귀 등에서 이용
- 승산 비
  - 승산을 2개 사용해 비율을 취한 것
  - 두 집단에서 두 사건이 일어 났을 때 해당 사건이 일어나기 쉬운 정도
  - 임상 시험에서 투여한 약물의 효과가 어느 정도인가
  - 특정 두 시기에서 남녀의 인구 추세에 큰 차이가 있는지 등의 지표로 사용

2 베이즈 정리

▶ 승산 비

[승산비의 베이즈 정리 식]

$$\frac{P(A|B)}{1-P(A|B)} \bigg/ \frac{P(A)}{1-P(A)} = \frac{P(B|A)}{P(B|A^c)}$$

* B라는 조건 아래 A가 일어날 조건부 확률의 정의

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 베이즈 정리

- 예1) 질병 검사의 양성 적중율
  - 질병 검사에서 질병에 걸릴 확률이 0.01
  - 병에 걸린 사람이 검사하면 0.99의 비율로 양성
  - 건강한 사람도 0.1의 비율로 양성이 된다고 가정

[양성 적중율의 계산 식]

$$P(\text{질병}|\text{양성}) = \frac{P(\text{양성}|\text{질병})P(\text{질병})}{P(\text{양성})} = \frac{0.99 * 0.01}{0.01 * 0.99 + 0.99 * 0.1} = 0.091$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 베이즈 정리

- 예2) 뺑소니 택시의 색상
  - 뺑소니 택시 사고 발생 시 목격자 증언에 따라 택시 색이 파란색 이라고 증언
  - 거리에는 파란색과 녹색 두 종류 차량이 있으며 각각 15%, 85%의 비율 차지
  - 비슷한 상황으로 실험한 결과 목격자가 파란색이라고 했는데 실제 파란색일 비율 80%일 때 실제 녹색일 확률은?

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

2 베이즈 정리

▶ 예2) 뺑소니 택시의 색상

[뺑소니 택시 색의 계산 식]

$$\frac{P(\text{파란색}|\text{목격})}{P(\text{녹색}|\text{목격})} = \frac{P(\text{목격}|\text{파란색})}{P(\text{목격}|\text{녹색})} * \frac{P(\text{파란색})}{P(\text{녹색})} = \frac{0.8}{0.2} * \frac{0.15}{0.85} = \frac{12}{17}$$

$$P(\text{파란색}|\text{목격}) + P(\text{녹색}|\text{목격}) = 1\text{에서}$$

$$P(\text{녹색}|\text{목격}) = \frac{12}{12 + 17} \cong 0.59$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

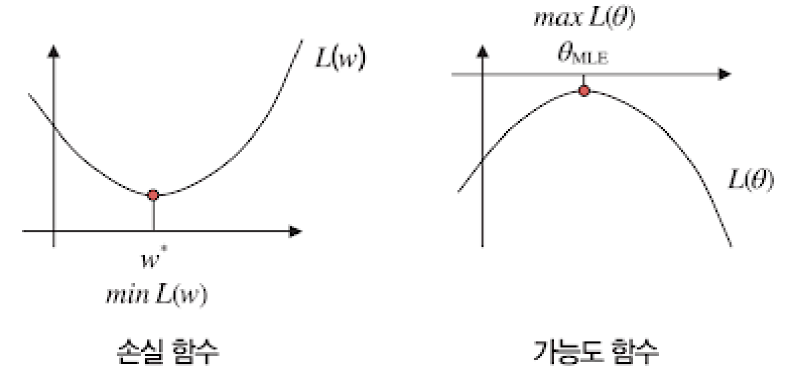
3 최대 가능도 방법과 EM 알고리즘

- 관측 데이터가 의미 있는 값과 노이즈로 구성된 경우
- 의미 있는 값을 구할 때 최소제곱법 사용 시 손실 함수 설정
- 손실 함수 대신 가능도 함수 설정
- 가능도 함수가 최대가 되는 θ 설정 시 θ는 최대 가능도 추정량이라 함
- 엔트로피가 높은 상태, 즉 노이즈가 가장 균형 있게 흩어져 있는 상태를 최대가능도 추정량으로 정해서 의미 있는 데이터를 구할 수 있음

3 최대 가능도 방법과 EM 알고리즘

- 가능도 함수는 적분 형태가 나타낼 때가 많아 로그 가능도 방정식 형태로 답을 구함

[손실 함수와 가능도 함수]



[로그 가능도 방정식]

$$\frac{\partial \log L(\theta)}{\partial \theta_1} = \frac{\partial \log L(\theta)}{\partial \theta_2} = \cdots = \frac{\partial \log L(\theta)}{\partial \theta_k} = 0$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

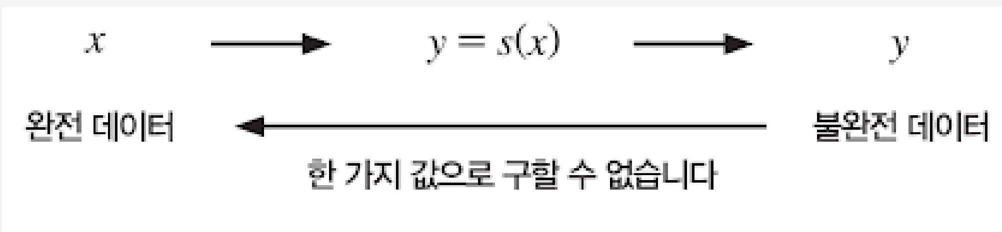
3 최대 가능도 방법과 EM 알고리즘

- 복잡한 가능도 함수에서 최대가능도 추정량을 직접 구할 수 없을 수 있음
- 반복 계산으로 구하는 방법 선택
- 의미 있는 값 x를 완전 데이터라고 하며 관찰할 수 있는 데이터 y 는 불완전 데이터라고 함

3 최대 가능도 방법과 EM 알고리즘

- x에 추가되는 어떤 작용 s(x)의 정보가 없으면 데이터 x를 한 가지 값으로 구할 수 없음

[완전 데이터와 불완전 데이터의 관계도]



※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

3 최대 가능도 방법과 EM 알고리즘

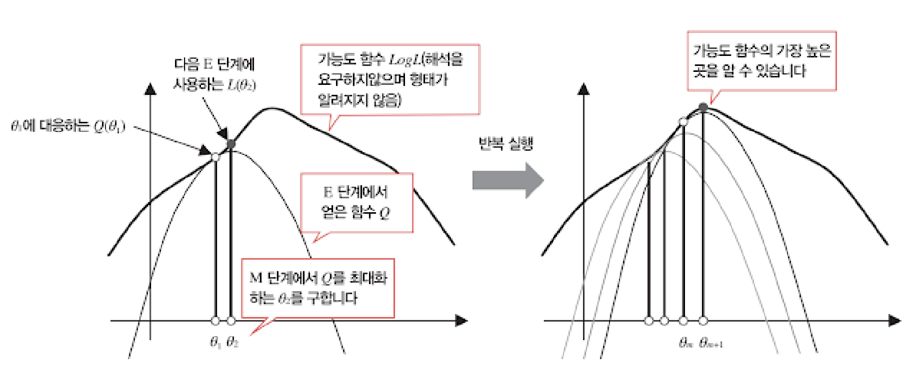
- EM(Expectation Maximization) 알고리즘
  - 숨겨진 변수를 포함하는 모델에 사용하는 가상 데이터 전체의 가능도 함수를 바탕으로 불완전한 데이터에서 최대가능도 추정량을 구하는 방법
  - 아래쪽 경계를 결정하기 위한 θ에 의존하는 볼록 함수 Q를 결정하는 E 단계와 Q에서의 θ를 극대화하는 M 단계를 반복 실행해 로그 가능도 함수의 최대값을 탐색
  - E 단계에서 구하는 Q는 사후 분포

3 최대 가능도 방법과 EM 알고리즘

[EM 알고리즘 식]

- E(Expectation) 단계
  : $Q(\cdot \mid \theta_m)$을 산출
  θ_m 은 $\hat{\theta}_{MLE}$ 의 m 번째 근사치
- M(Maximization) 단계
  : $\theta_{m+1} = \llbracket argmax \rrbracket \_\theta \; Q(\theta | \theta_m)$

[EM 알고리즘 형태]

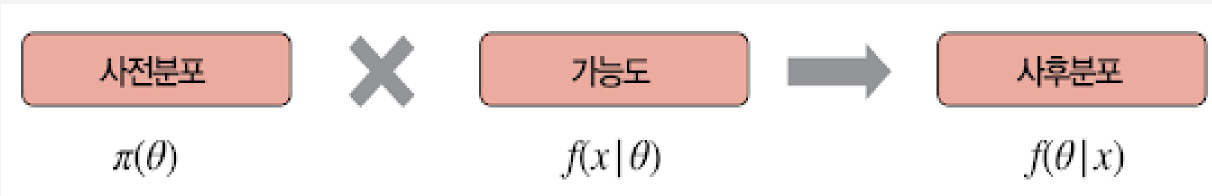


※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

- 데이터의 모집단 분포는 유일하지 않음
- 관측하지 않은 데이터를 다룬다는 의미
- 사전 분포(Prior Distribution) 혹은 주관 분포(Subjective Distribution)
  - 추론해야 하는 대상이 매개변수 $\boldsymbol{\theta}$ 일 때 확률 밀도 함수 $\boldsymbol{\pi}(\boldsymbol{\theta})$

4 베이즈 추론



$$f(\theta|x) = \frac{1}{f(x)} * f(x|\theta)\pi(\theta) \propto f(x|\theta)\pi(\theta)$$

$$f(x) = \int f(x|\theta)\pi(\theta)\, d\theta$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

- 베이즈 추정량
  - 사후 분포의 평균 제곱 오차를 최소화하는 값
  - 사후 분포의 평균값

[베이즈 추정량 식]

$$\hat{\theta}(x) = arg_i min \int_\theta |t - \theta|^2 f(\theta|x)\, d\theta$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

- 사후 메디안 추정량
  - 사후 분포의 평균 절대 오차를 최소화하는 값
  - 사후 분포의 중간값

[사후 메디안 추정량의 식]

$$\hat{\theta}(x) = arg_i min \int_{-\infty}^{\infty} |t-\theta| f(\theta|x)\, d\theta$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

- MAP 추정량
  - 사후밀도를 최대화하는 θ값
  - 사후 최빈값 추정량이라고도 함
  - MAP은 최대 사후 확률이라고도 함

[MAP 추정량 식]

$$\hat{\theta}(x) = arg_{\theta}max \frac{f(x|\theta)\pi(\theta)}{f(x)}$$

[MAP 추정량 형태]

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

- 켤레 사전 분포
  - 사전 분포 $\pi(\theta)$와 사후 분포 $f(\theta|x)$가 같은 유형의 분포를 보일 때를 켤레 사전 분포라 함
  - 사후 분포를 다루기 쉬운 특징
  - $\pi(\theta|\alpha)$처럼 추가 매개변수가 들어가면 사후 분포에서는 $f(\theta|x,\alpha)$ 같은 형태로 표현
  - α를 하이퍼 파라미터라고 함

[사후 예측분포 식]

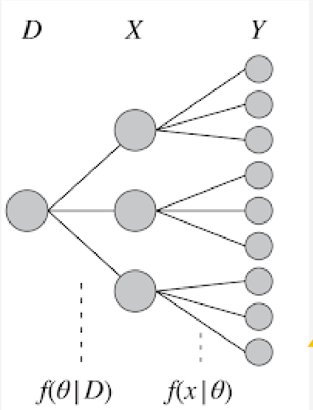
$$\boldsymbol{f(x|D) = \int f(x|\theta)f(\theta|D)\, d\theta}$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

4 베이즈 추론

▶ 켤레 사전 분포
- 사후 예측분포 $f(x \mid D)$는 진정한 밀도 함수 $f(x)$에 가까운 것 이라는 개념에 기반을 둔 것
- 따라서 데이터 D 를 기준으로 나타나는 다음 데이터 x 값을 예측하는 형태가 됨

[사후 예측 분포의 형태]



매개 변수 X값을 사후분포에서 선택한 후 이를 이용해 Y값을 계산해서 예측분포를 얻음

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017

5 베이즈 판별 분석

- 베이즈 추론의 예
  - 데이터 x가 어떤 모집단분포에 유래했는지 파악하는데 사용
  - 선형 판별 분석
    : 지도 학습의 훈련 데이터를 바탕으로 결정하는 경우가 많음
- 베이즈 통계학 개념이 포함된 요소를 추가한 것이 베이즈 판별 분석

5 베이즈 판별 분석

- N가지 종류의 데이터 x와 추론 대상 매개변수 i가 있는 모집단분포 $f(x|i)$와 사전 분포 $\pi(i)$로부터 사후 확률이 최대가 되는 모집단 분포 $f(x|i)$의 유래를 판정

[베이즈 판별 분석 식]

$$f(i|x) = \frac{f(x|i)\pi(i)}{\sum_{j=1}^{n} f(x|j)\pi(j)} \propto f(x|i)\pi(i)$$

※ 출처 : 처음 배우는 인공지능, 다다 사토시, 한빛미디어, 2017