배이시언 결정 이론

Pattern Recognition & Machine Learning Laboratory
Ji-Hoon Park, July 06, 2021

확률과 통계(1/4)

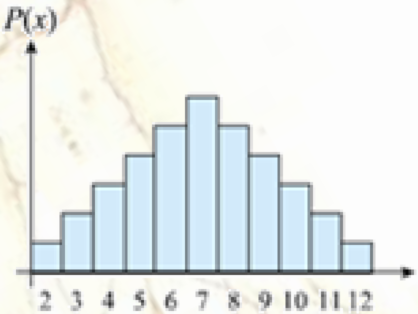
- 확률 기초
  - 확률 분포 함수
    - 예 : 이산 확률 분포, 연속 확률 분포
  - 조건부 확률
    - 선행된 조건 하에 발생하는 확률
      - 예: $P(B|A)$
  - 결합 확률
    - 두가지 서로 다른 사건이 동시에 일어날 확률
      - 예: $P(A, B)$
  - 주변 확률
    - 예: $P(C) = P(C|A)P(A) + P(C|B)P(B)$
  - 사전 확률
    - 연이어 일어나는 사건에서 두 번째 사건이 발생하기 전에 작용하는 확률
  - 우도 함수
    - 한 사건이 관찰되어 고정되어 있고 사전 사건에 따라 확률을 계산
  - 사후 확률
    - 사건 Y가 일어난 후에 따지는 조건부 확률
    - 사건 Y 가 일어나 고정되고 이전에 발생한 사건 X에 따라 확률을 계산
      - 예: P(X|Y)

확률과 통계(2/4)

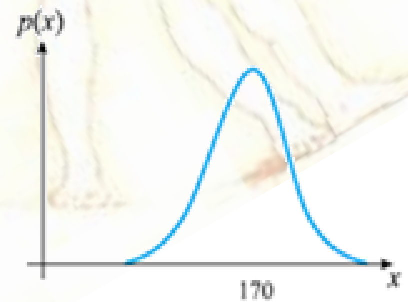
- 베이스 정리
  - 결합 확률과 곱규칙에 따라 식 도출
    - $P(X|Y) = \; = \dfrac{P(Y|X)P(X)}{P(Y)} = \dfrac{\text{우도} \times \text{사전확률}}{P(Y)}$

평균과 분산

- 이산 확률 분포
  - $\mu = \sum_x xP(x)$
    - μ: 평균
    - $P(x)$: 이산 확률
  - $\sigma^2 = \sum_x (x-\mu)^2 P(x)$
- 연속 확률 분포
  - $\mu = \int_{-\infty}^{\infty} xp(x)\,dx$
    - $p(x)$: 연속 확률
  - $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx$
- 샘플 집합
  - $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
    - N: 샘플 집합 갯수
  - $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i-\mu)^2$



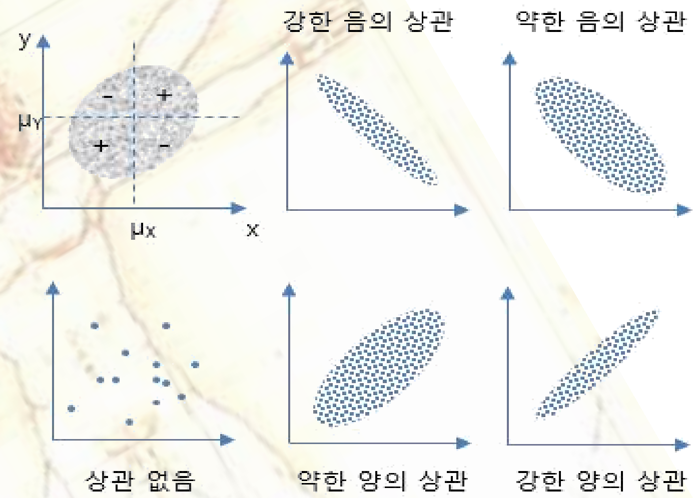
이산 확률 분포



연속 확률 분포

확률과 통계(3/4)

- d차원 특징 벡터의 평균 벡터
  - 이산 확률 분포
    - $\mu = \sum_{\mathrm{x}} \mathrm{x}P(\mathrm{x})$
      - » x: d 차원 벡터
  - 연속 확률 분포
    - $\mu = \int_{R^d} \mathrm{x}p(\mathrm{x})\,dx$
      - » $R^d : d$ 차원 실수 공간
  - 샘플 집합
    - $\mu = \frac{1}{N}\sum_{i=1}^{N} \mathrm{x}_i$

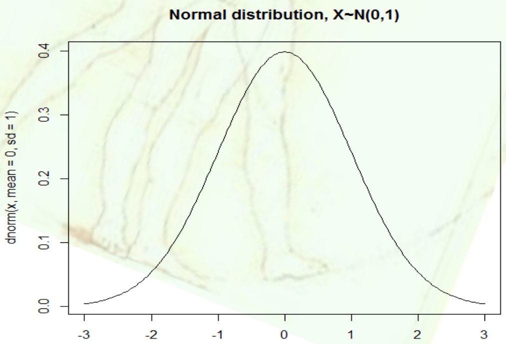


공분산 상관관계

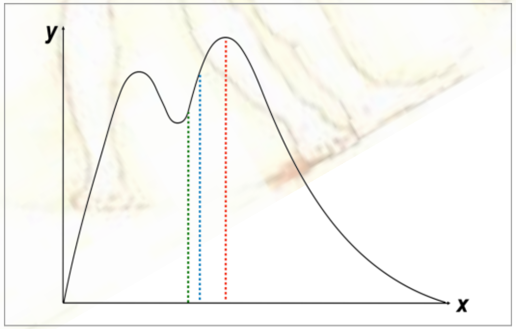
- d차원 특징 벡터의 공분산
  - 공분산은 랜덤 벡터를 구성하는 랜덤 변수간의 관계를 표현
  - 양수인 경우 두 변수는 양의 관계, 음수인 경우 음의 관계
  - 이산 확률 분포
    - $\Sigma = \sum_{\mathrm{x}} (\mathrm{x}-\mu)(\mathrm{x}-\mu)^T P(\mathrm{x})$
  - 연속 확률 분포
    - $\Sigma = \int_{R^d} (\mathrm{x}-\mu)(\mathrm{x}-\mu)^T p(x)\,dx$
  - 샘플 집합
    - $\Sigma = \frac{1}{N-1}\sum_{i=1}^{N} (\mathrm{x}_i - \mu)(\mathrm{x}_i - \mu)^T$

확률과 통계(4/4) & 베이시언 분류기(1/3)

- **확률 분포의 표현과 추정**
  - d개의 변수 증가로 d차원의 배열 사용으로 차원의 저주 발생
    - 확률 밀도함수를 정규 분포로 근사화
      - 평균 μ와 분산 σ^2 설정하여 표현
    - 여러 개의 모드를 가진 분포
      - 정규분포 표현 불가
- **최소 오류 베이시언 분류기**
  - 분류 문제의 해결
    - x를 발생시켰을 가능성이 가장 큰 부류 선택
      - $P(w_1|\mathrm{x})$ 와 $P(w_2|\mathrm{x})$ 비교
        - » $P(w_1|\mathrm{x}) > P(w_2|\mathrm{x})$이면 w_1로 분류
        - » $P(w_1|\mathrm{x}) < P(w_2|\mathrm{x})$이면 w_2로 분류



정규 분포

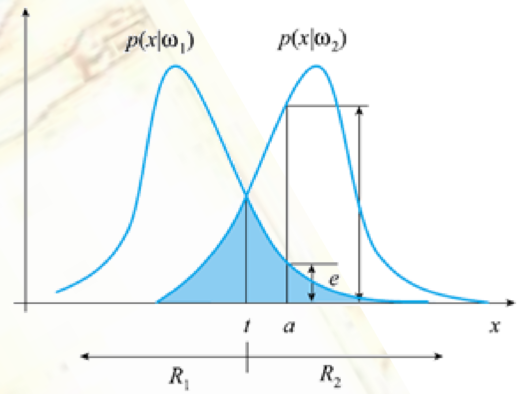


두 개의 모드를 가진 분포

베이시언 분류기(2/3)

- 베이스 정리 이용하여 비교
  - » $p(\mathrm{x}|w_1)\mathrm{P}(w_1) > p(\mathrm{x}|w_2)\mathrm{P}(w_2)$ 이면 w_1로 분류
  - » $p(\mathrm{x}|w_1)\mathrm{P}(w_1) < p(\mathrm{x}|w_2)\mathrm{P}(w_2)$ 이면 w_2로 분류

• 오류 확률 예측
- 임계값 기준으로 분류
  - » $\mathrm{x} > t$ 이면 w_2로 분류
  - » $\mathrm{x} < t$ 이면 w_1로 분류
- 분류했을 경우 잘못 선택한 경우 오류 확률로 계산
  - » 그래프의 파란색 부분
  - » $E = \frac{1}{2}\left(\int_{-\infty}^{t} p(x|w_2)\,dx + \int_{t}^{\infty} p(x|w_1)\,dx\right)$ ⇐ 사전 확률을 ½로 가정했을 경우 오류율
- 임계값의 최적 위치
  - » 두 그래프 겹치는 부분이 최소의 오류
  - » 옮길 경우 오류 증가

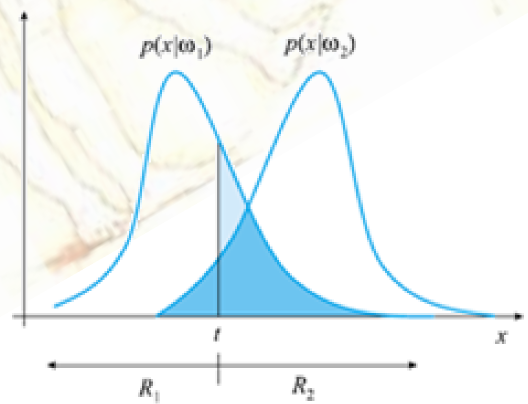


베이시언 분류기의 오류 확률

▪ 최소 위험 베이시언 분류기

➢ 분류 문제의 해결

• 손실 행렬 이용하여 D를 최소화
- $D = d_1P(w_1) + d_2P(w_2)$
  - » $P(w_1), P(w_2)$: 사전확률
  - » $d_1 = c_{11}\int_{R_1} p(\mathrm{x}|w_1)\,dx + c_{12}\int_{R_2} p(\mathrm{x}|w_1)\,dx$
  - » $d_2 = c_{21}\int_{R_1} p(\mathrm{x}|w_2)\,dx + c_{12}\int_{R_2} p(\mathrm{x}|w_2)\,dx$



베이시언 분류기 미사용 오류 확률

베이시언 분류기(3/3)& 분별 함수(1/2)

      » $P(w_1)$, $P(w_2)$: 사전확률

  – x를 $q_2 > q_1$이면 w_1으로 분류, $q_2 < q_1$이면 w_2으로 분류

      » $q_1 = c_{11}p(\mathrm{x}|w_1)P(w_1) + c_{21}p(\mathrm{x}|w_2)P(w_2)$

      » $q_2 = c_{12}p(\mathrm{x}|w_1)P(w_1) + c_{22}p(\mathrm{x}|w_2)P(w_2)$

- 우도비 결정 규칙
  – x를 $\frac{p(\mathrm{x}|w_1)}{p(\mathrm{x}|w_2)} > T$ 이면 w_1로 분류, $\frac{p(\mathrm{x}|w_1)}{p(\mathrm{x}|w_2)} < T$ 이면 w_2로 분류

      » $T = \frac{(c_{21}-c_{22})P(w_2)}{(c_{12}-c_{11})P(w_1)}$

  – x를 $\mathrm{p}(\mathrm{x}|w_1) > \mathrm{p}(\mathrm{x}|w_2)\frac{c_{21}}{c_{12}}$ 이면 w_1로 분류, $\mathrm{p}(\mathrm{x}|w_1) < \mathrm{p}(\mathrm{x}|w_2)\frac{c_{21}}{c_{12}}$ 이면 w_2로 분류

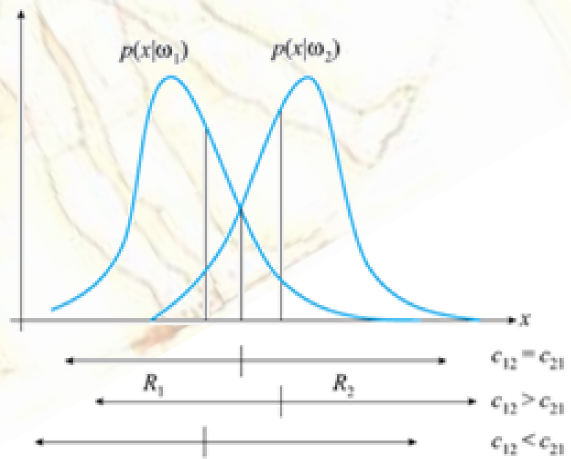
▪ M 부류로 확장

  ➢ M 부류 최소 위험 베이시언 분류기

  - x를 $k = arg\min_i q_i$일 때 w_k로 분류
    – $q_i = \sum_{j=1}^{M} p(\mathrm{x}|w_j)P(w_j)$

▪ 분별함수

  ➢ 패턴의 특징벡터 x 를 받아 M개의 함수 값 $g_i(\mathrm{x})$ 구하여 그 중 최고 값의 부류로 분류



최소 위험 베이시언 분류기

분별 함수(2/2)& 정규 분포에서 베이시언 분류기(1/3)

- $g_i(\mathrm{x}) = \begin{cases} p(\mathrm{x}|w_i)P(w_i) \\ \dfrac{1}{\sum_{j=1}^{M} c_{ji} p(\mathrm{x}|w_j)P(w_j)} \end{cases}$
- x를 $g_{12}(x) > 0$ 이면 w_1로 분류, $g_{12}(x) < 0$ 이면 w_2로 분류
  - $g_{12}(x) = g_{1(x)} - g_{2(x)}$

▪ 정규 분포의 베이시언 분류기

- ➢ 정규 분포
  - 우도가 정규 분포를 따르는 특수한 경우에 대해 분류기가 가지는 성질 분석하여 두 부류의 영역을 나누는 경계 모양 분석
    - 현실 세계에 맞는 경우가 적지 않게 존재
    - 평균과 분산의 매개 변수만으로 표현
    - 수학적 증명 용이
- ➢ 정규 분포와 분별 함수
  - 1차원 정규분포
    - $\mathrm{N}(\mu, \sigma^2) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
  - 2차원 정규분포
    - $N(\mathbf{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}-\mu_i{}^T \Sigma_i{}^{-1}(\mathbf{x}-\mathbf{\mu}_i)\right)$

정규 분포에서 베이시언 분류기(2/3)

- 우도
  - $p(\mathbf{x}|w_i) = N(\boldsymbol{\mu}_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}-\boldsymbol{\mu}_i{}^T \Sigma_i{}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right)$
- 분별 함수
  - $g_i(x) = \ln p(\mathbf{x}|w_i)P(w_i)$
  - $g_i(x) = -\frac{1}{2}\left(\mathbf{x}^T \Sigma_i{}^{-1}\mathbf{x} - 2\mu_i^T \Sigma_i^{-1}\mathbf{x} + \mu_i^T \Sigma_i^{-1}\mu_i\right) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(w_i)$
    - » $g_i(x) = g_2(x)$ 만족 해야 결정 경계 상에 x 존재
    - » $g_{ij}(x) = g_i(x) - g_2(x)$: 결정 경계 함수

▪ 선형 분별

➢ 공분산 행렬이 같은 경우

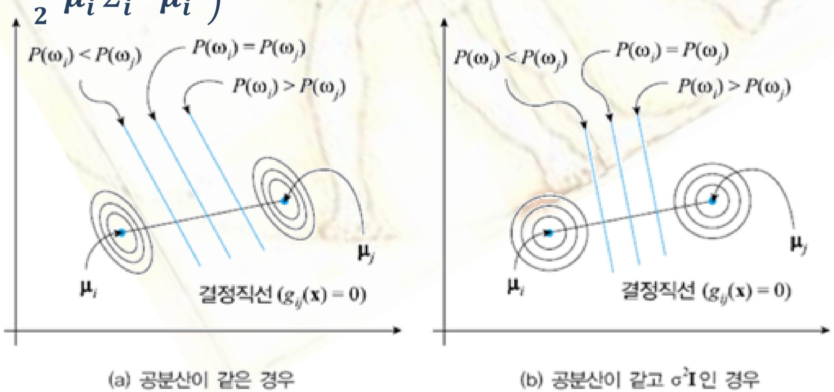
- $g_i(x)$에서 i와 무관한 항 제거
  - $g_i(x) = \Sigma_i^{-1}\mu_i{}^T\mathbf{x} + \left(\ln P(w_i) - \frac{1}{2}\mu_i^T \Sigma_i^{-1}\mu_i\right)$
  - $g_i(x) = \mathbf{w}_i^T\mathbf{x} + b_i$ 꼴

➢ 선형 분류기

- $g_{ij}(x) = \mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)$
- x_0지나고, $\Sigma^{-1}(\mu_i - \mu_j)$에 수직

➢ 공분산 행렬 $\Sigma = \sigma^2\mathrm{I}$

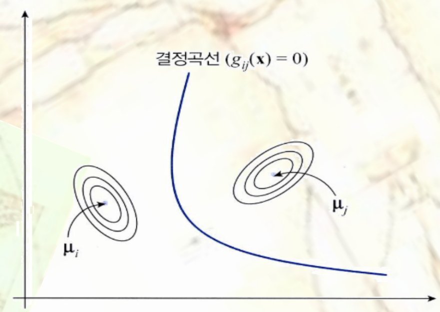
- $\mu_i - \mu_j$에 수직
- 등고선이 원형



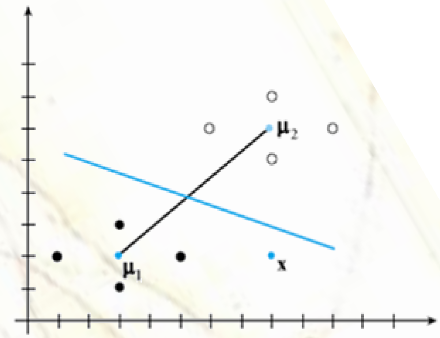
선형 분류기

정규 분포에서 베이시언 분류기(3/3) & 베이시언 분류의 특성(1/2)

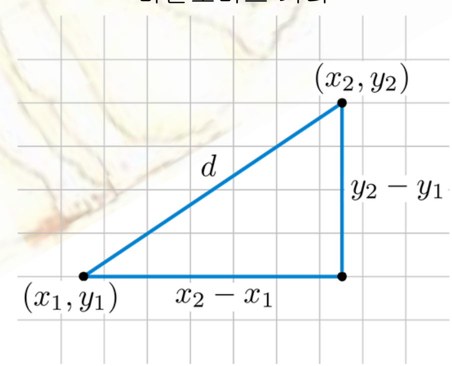
- **2차 분별**
  - ➢ 다양한 공분석 행렬
    - 2차 분별 분석
      - $g_i(x) = -\frac{1}{2}\mathbf{x}^T \Sigma_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^T \Sigma_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T \Sigma_i^{-1}\boldsymbol{\mu}_i - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(w_i)$
- **최소 거리 분류기**
  - ➢ 마할라노비스 거리
    - $\left((\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)\right)^{1/2}$
  - ➢ 유클리디언 거리
    - $\left((\mathbf{x}-\boldsymbol{\mu}_i)^T (\mathbf{x}-\boldsymbol{\mu}_i)\right)^{1/2}$
  - ➢ 베이시언분류기 최소 거리 분류기 형태
    - $k = arg\min_i \left((\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)\right)^{1/2}$
- **베이시언 분류 특징**
  - ➢ 비현실성 내포
  - ➢ 일반적인 확률 분포사용 시 차원의 저주 발생
  - ➢ 오류율 측면에서 최적
  - ➢ 확률을 신뢰도 값으로 삼아 후처리 등에 유용하게 활용 가능



2차 분별 분석으로 만든 2차 분류기



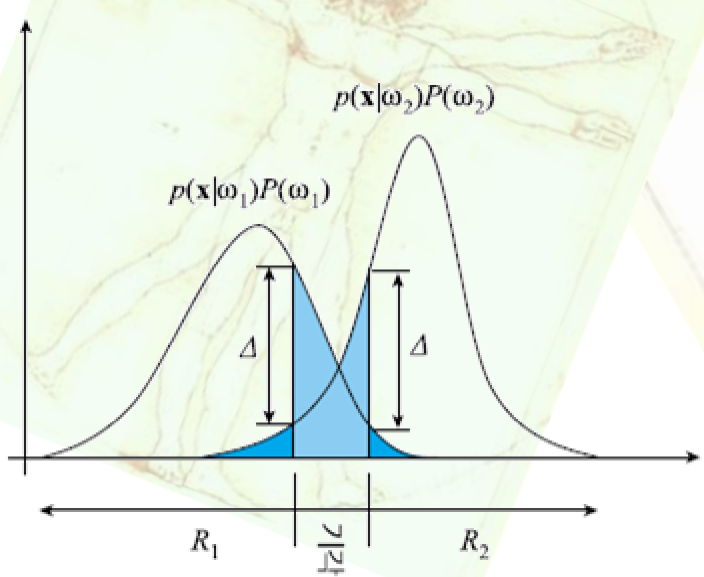
마할노비스 거리



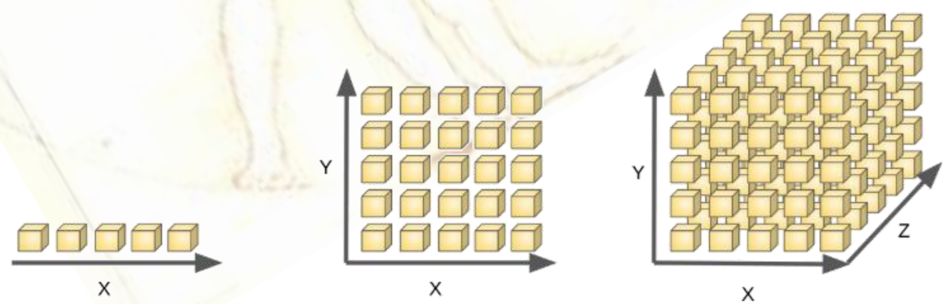
유클리디안 거리

베이시언 분류의 특성(2/2)& 기각 처리

- 나이브 베이시언 분류기
  - 차원의 저주 해결
  - $p(\mathrm{x}|w_i) = \prod_{j=1}^{d} p(\mathrm{x}_j|w_i)$
- 기각 처리와 기각률
  - 최소 오류 베이시언 분류기에서 두 부류의 확률 값 차이가 Δ 이내 영역 기각
  - 기각률: 기각으로 제외할 연한 파란색 영역
    - $p(\mathrm{x}|w_i)P(w_i)$를 $\sum_{j=1}^{M} c_{ji}\, p(\mathrm{x}|w_j)P(w_j)$로 교체



기각 처리



차원의 저주