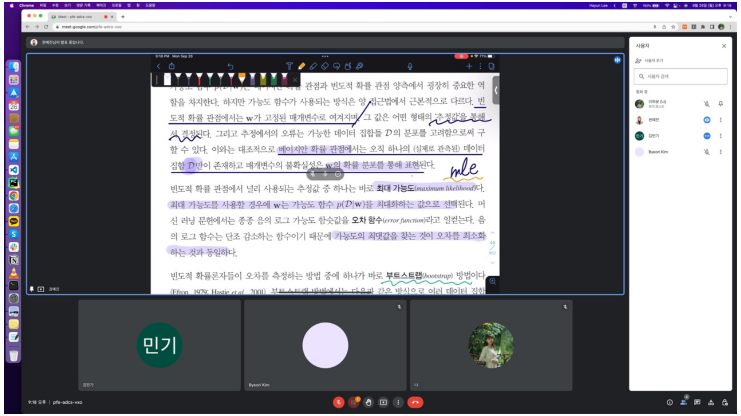
CUAI 스터디 PRML팀

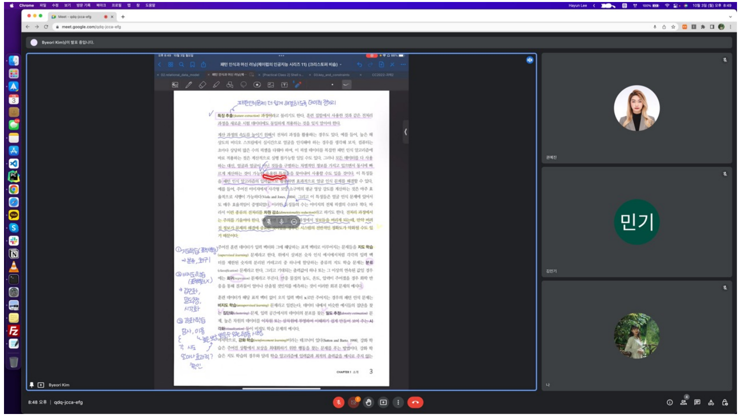
2022.11.01

발표자 : 이하윤

스터디원 소개 및 만남 인증

- 9/26, 10/3, 11/1
- 김민기, 김벼리, 권예진. 이하윤





스터디 내용

스터디 내용

스터디 내용

- 확률 밀도
  - 연속적인 변수에서의 확률

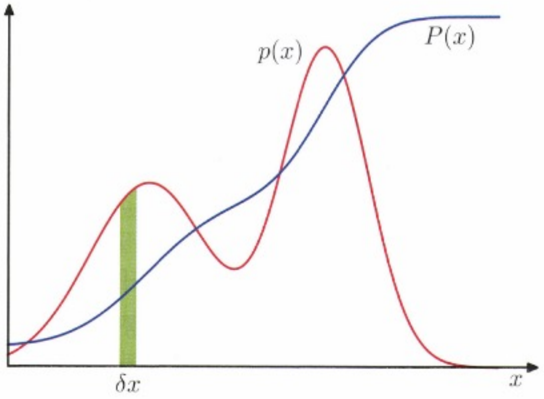


그림 1.12 다음과 같이 이산 변수에 대한 확률 개념을 연속 변수에 대해 확장할 수 있다. 변수 x가 $(x, x+\delta x)$ 구간 사이의 값을 가질 확률은 $p(x)\delta x(\delta x \to 0$일 경우)다. 확률 밀도는 누적 분포 함수 $P(x)$의 미분으로 표현할 수 있다.

스터디 내용

- 기댓값과 공분산
  - 기댓값
    - 확률 밀도 f(x) 하에서 어떤 함수 f(x)의 평균값은 f(x)의 기댓값이라고 한다.

$$\mathbb{E}[f] = \sum_{x} p(x) f(x) \qquad \mathbb{E}[f] = \int p(x) f(x)\, \mathrm{d}x$$

  - 공분산
    - X값과 y값이 얼마나 함께 같이 변동하는가에 대한 지표 (서로 독립일 경우 0)

$$\begin{aligned} \operatorname{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

스터디 내용

- 가우시안 분포(== 정규분포)

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

스터디 내용

- 곡선 피팅
- (베이지안)

가능도 함수

$$p(t|x, \mathbf{w}_{\mathrm{ML}}, \beta_{\mathrm{ML}}) = \mathcal{N}\left(t|y(x, \mathbf{w}_{\mathrm{ML}}), \beta_{\mathrm{ML}}^{-1}\right)$$

"t" 에 대한 예측 분포로 표현! 아직, Bayesian 방식은 아님!

다항 계수 w에 대한 사전 분포 도입 (가우시안)

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) = \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left\{-\frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}$$

$M+1$: M차수 다항식 벡터 w의 원소의 개수

σ : 모델의 매개변수의 분포를 제어하는 초매개변수

위 분포에서 w의 가우시안"사전분포"를 도입하여 독립적으로 추출되었다고 가정

\+

베이지안 정리

베이지안 정리 (trick ?)

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

w 의 사후 분포는 "가능도 함수"와 "사전 분포"의 곱에 비례

MAP(maximum posterior) : 사후 분포를 최대화하는 방식으로 w를 결정할 수 있다는 것

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

$$\propto \prod_{n=1}^{N} N(t_n|y(x_n, \mathbf{w}), \beta^{-1}) \times N(\mathbf{w}|0, \sigma^{-1}\mathbf{I})$$

음의 로그를 취함

$$\frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

$\lambda = \alpha/\beta$ 로 주어진 regularization된 제곱합 오차 함수와 동일

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

① $\mathbf{w}_{\mathrm{MAP}}$: 주어진 데이터에 대해 가장 가능성 높은 w를 찾는 방식(즉, 사후 분포를 최대화)

② β_{MAP} : 주어진 데이터에 대해 가장 가능성 높은 β를 찾는 방식(즉, 사후 분포를 최대화)

스터디 내용

- 결정 이론
    - 불확실성이 존재하는 상황에서 의사 결정을 내려야 하는 경우가 많음
    - => 결정 이론과 확률론을 함께 사용하면 최적의 의사 결정을 내릴 수 있음
    - Cost function / Loss function
    - 거부 옵션

- 정보이론
    - 엔트로피
    - 다중도
    - 스털링 근사식
    - 상대적 엔트로피와 상호 정보량
    - 옌센의 부등식