

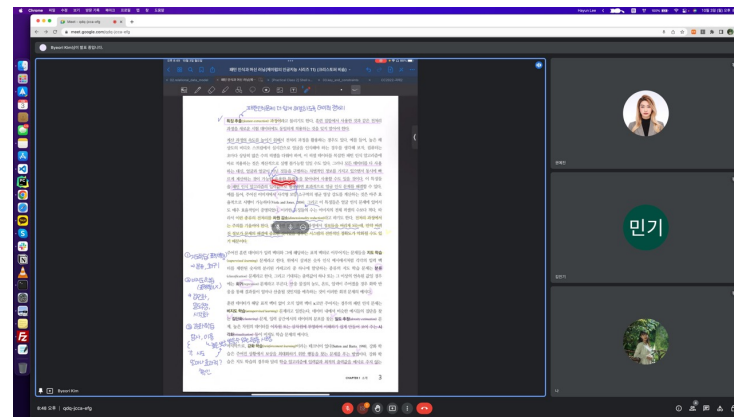
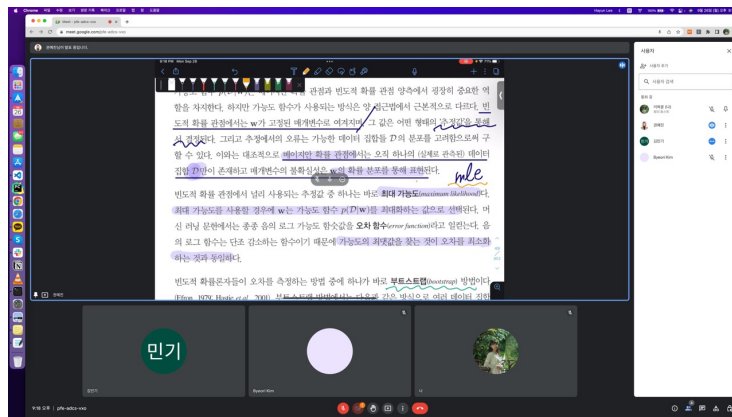
CUAI 스터디 PRML팀

2022.11.01

발표자 : 이하윤

스터디원 소개 및 만남 인증

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



스터디 내용

CHAPTER

1

소개 1

	1.1	예시: 다항식 곡선 피팅	5
	1.2	확률론	13
벼리	1.2.1	확률 밀도	19
	1.2.2	기댓값과 공분산	21
예진	1.2.3	베이지안 확률	23
	1.2.4	가우시안 분포	27
하윤	1.2.5	곡선 피팅	32
	1.2.6	베이지안 곡선 피팅	34
민기	1.3	모델 선택	36
하윤	1.4	차원의 저주	37
	1.5	결정 이론	42
벼리	1.5.1	오분류 비율의 최소화	43
	1.5.2	기대 손실의 최소화	45
	1.5.3	거부 옵션	46
예진	1.5.4	추론과 결정	47
	1.5.5	회귀에서의 손실 함수	51
민기	1.6	정보 이론	54
	1.6.1	상대적 엔트로피와 상호 정보량	61
	★ 연습문제		65

스터디 내용

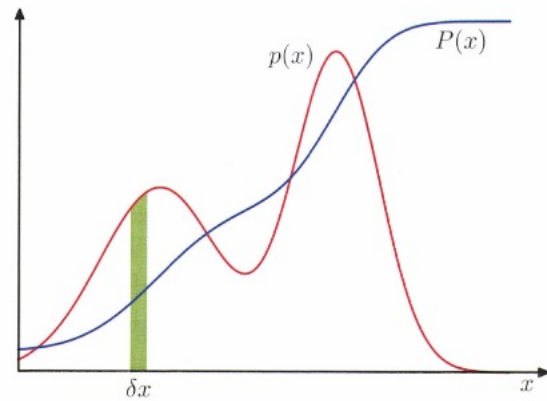
CHAPTER 2 확률 분포 75

하윤	2.1 이산 확률 변수_76	
	2.1.1 베타 분포	79
민기	2.2 다항 변수_83	
	2.2.1 디리클레 분포	85
벼리	2.3 가우시안 분포_87	
	2.3.1 조건부 가우시안 분포	95
	2.3.2 주변 가우시안 분포	98
	2.3.3 가우시안 변수에 대한 베이지안 정리	101
예진	2.3.4 가우시안 분포의 최대 가능도	104
	2.3.5 순차 추정	105
	2.3.6 가우시안 분포에서의 베이지안 추론	108
	2.3.7 스튜던트 t 분포	114
	2.3.8 주기적 변수	117
	2.3.9 가우시안 분포의 혼합	123
	2.4 지수족_126	
	2.4.1 최대 가능도와 충분 통계량	129
	2.4.2 켈레 사전 분포	130
	2.4.3 무정보적 사전 분포	131
	2.5 비매개변수적 방법_134	
	2.5.1 커널 밀도 추정	136
	2.5.2 최근접 이웃 방법론	139
	★ 연습문제	142

스터디 내용

- 확률 밀도
 - 연속적인 변수에서의 확률

그림 1.12 다음과 같이 이산 변수에 대한 확률 개념을 연속 변수에 대해 확장할 수 있다. 변수 x 가 $(x, x + \delta x)$ 구간 사이의 값을 가질 확률은 $p(x)\delta x$ ($\delta x \rightarrow 0$ 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.

$$\mathbb{E}[f] = \sum_x p(x)f(x) \quad \mathbb{E}[f] = \int p(x)f(x) dx$$

- 공분산

- x 값과 y 값이 얼마나 함께 같이 변동하는가에 대한 지표 (서로 독립일 경우 0)

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

스터디 내용

- 가우시안 분포(== 정규분포)

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

스터디 내용

- 곡선 피팅
- (베이지안)

가능도 함수

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

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

다항 계수 w 에 대한 사전 분포 도입 (가우시안)

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

$M+1$: M차수 다항식 벡터 w 의 원소의 개수
 σ : 모델의 매개변수의 분포를 제어하는 초매개변수

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 w 를 결정할 수 있다는 것

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

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

음의 로그를 취함

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

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

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

① \mathbf{w}_{MAP} : 주어진 데이터에 대해 가장 가능성 높은 w 를 찾는 방식(즉, 사후 분포를 최대화)
 ② β_{MAP} : 주어진 데이터에 대해 가장 가능성 높은 β 를 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
 - 엔센의 부등식