PRML Chapter.1

Introduction

2017010698
수학과 오서영

목차

5. The Decision Theory

1) 오분류 최소화
(Minimizing the misclassification rate)

잘못 분류될 가능성 (오분류될 확률 값을 모두 합한 확률)
-> 이를 최소화 하는 방향으로 모델 설계

$$p(mistake) = p(x \in R_1, C_2) + p(x \in R_2, C_1) = \int_{R_1} p(\mathbf{x}, C_2) d\mathbf{x} + \int_{R_2} p(\mathbf{x}, C_1) d\mathbf{x}$$

(**EX**) cancer
case1 : 암이 아닌데 암인 것으로 진단
case2 : 암이 맞는데 암이 아닌 것으로 진단
-> case2 가 더 심각하다 -> Penalty?

5. The Decision Theory

2) 기대 손실 최소화
(Minimizing the expected loss)

Loss function (Cost function)

단순히 오 분류 개수만 세는 것이 아니라
Loss라는 개념을 정의하고 이를 최소화
-> 가능한 결정이나 행동들을
조금 더 능동적으로 조절할 수 있다.

하나의 샘플 x 가 실제로는 특정 클래스 Ck 에 속하지만,
우리가 이 샘플의 클래스를 Cj 로 선택할 때 (잘못된 선택)
들어가는 비용을 정의한다.

$$E[L] = \sum_k \sum_j \int_{R_j} L_{kj} p(\mathbf{x}, C_k)\, d\mathbf{x}$$

(L : Level of loss, loss matrix, R : decision region)

5. The Decision Theory

- x 는 반드시 하나의 Rj 에 포함되게 된다.

-> 에러 값이 최소가 되는 Rj 를 선택

-> 결국 x 에 대해 $\sum_k L_{kj} p(\mathbf{x}, C_k)$를 최소화하는 class 를 선택

$$p(\mathbf{x}, C_k) = p(C_k|\mathbf{x})p(\mathbf{x})$$

-> eliminate the common factor of p(x)

Expected loss assigns each new **x** to the class **j**
for which the quantity
(posterior class probability)

$$\sum_k L_{kj} p(C_k|\mathbf{x})$$

5. The Decision Theory

3) 추론과 판별
(Inference and decision)

- **추론(inference)** : 학습 데이터를 이용하여
p(Ck | x) 에 대한 모델을 학습

- **판별 (decision)** : 추론한 사후 확률 분포를 이용하여
실제 입력된 데이터의 클래스를 결정

5. The Decision Theory

판별 (decision)

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{p(\mathbf{x})}$$

$$p(\mathbf{x}) = \sum_k p(\mathbf{x}|C_k)p(C_k)$$

(a) Generative Models
Posterior class probability 을
class-conditional density 인 p(x | Ck) 와
Prior class probability 인 p(Ck) 로 구분하여
간접적으로 추론

- 주어진 데이터를 통해 모델링된 분포로부터 완전히 새로운 샘플들을 재 생성해 낼 수 있는 능력 -> **Generative**

- 모델이 잘못 추론되었다면 재 생성된 데이터가 원래 데이터와 유사하지 않을 가능성은 당연히 높음

- 입력 공간의 차원이 증가할 수록 더 정확한 class-conditional density 를 구하기 위한 많은 샘플이 필요
- Prior class probability 은 샘플 수를 세기만 하면 되므로 구하기 쉬움
- 새로운 데이터가 입력되었을 때 추정된 모델로부터 확률 값을 예측할 수 있으므로 낮은 확률 값을 통해 outlier 를 확인 가능

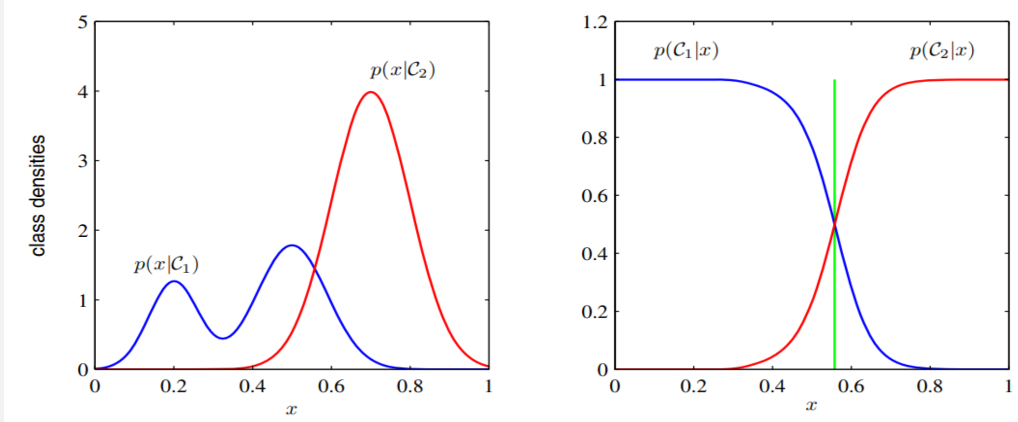
5. The Decision Theory

(b) Discriminative Models

: Posterior class probability 를 직접 근사 하는 모델
직접적인 방법 (direct) 으로 모델링

(EX) Left : Class-conditional densities for two classes
Right : posterior probabilities
- Class-conditional density p(x| C1) (blue on left)
has no effect on the posterior probabilities.
The vertical green line shows the decision boundary in x
that gives the minimum misclassification rate.

-> Class-conditional density 가 posterior probabilities 에
영향을 주지 않는 경우 바로 사후 분포를 찾는게 더 편할 수 있다.



5. The Decision Theory

4) 회귀를 위한 손실 함수 (Loss functions for regression)

: 회귀 문제는 분류를 하는 것이 아니라
실수인 target 을 예측하는 것

expected loss : 주어진 데이터로부터 얻어진 손실 함수의 평균값

$$E[L] = \iint L(t, y(\mathbf{x}))p(\mathbf{x}, t)\ d\mathbf{x}dt$$

손실함수를 $L(t_y(\mathbf{x})) = \{y(\mathbf{x}) - t\}^2$ 로 정의
목표 : E[L] 을 최소화하는 y(x) 를 찾는 것

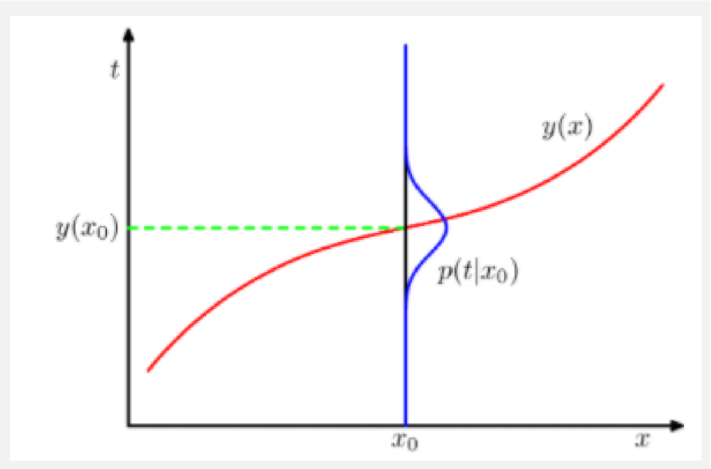
- 조건부 x 에 대한 t 값의 평균 -> regression 의 결과와 동일

$$E[L] = \iint \{y(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\ d\mathbf{x}dt$$

$$\frac{\delta E[L]}{\delta y(\mathbf{x})} = 2\int \{y(\mathbf{x}) - t\} p(\mathbf{x}, t) dt = 0$$

$$y(\mathbf{x}) = \frac{\int t p(\mathbf{x}, t) dt}{p(\mathbf{x})} = \int t p(t|\mathbf{x}) dt = E_t[t|\mathbf{x}]$$

5. The Decision Theory



$y(\mathbf{x}) = E_t[\mathbf{t}|\mathbf{x}]$: 수학적으로는 최적의 결과를 의미

현실적으로는 실제 가능한 모든 데이터 중 일부의 관찰 데이터만 얻는 것이므로 전체 데이터에 대한 실제 평균 값은 구하기 어렵다.
-> 오로지 관찰 데이터의 평균 값만을 얻을 수 있음.

5. The Decision Theory

Insight

- **y(x)** : 샘플 데이터로부터 우리가 예측한 근사 모델 식
- **E [t|x]** : 수학적으로 정답인 평균 값이다.

-> 존재 가능한 모든 경우의 데이터를 확보하여 평균 값을 구하면 실제로 최적의 함수를 만들 수 있다.

하지만 현실적으로는 샘플 데이터만 주어지게 되고,
샘플 데이터만으로 수학적으로 정답인 값은 추정되기 어렵다.
만약 샘플 데이터로부터 추정된 y(x) 식이 E[t|x] 와 동일하다면
매우 훌륭하게 식을 추정한 것이 된다.

$$\begin{aligned}\{y(\mathbf{x})-t\}^2 &= \{y(\mathbf{x})-\mathbb{E}[t|\mathbf{x}]+\mathbb{E}[t|\mathbf{x}]-t\}^2 \\ &= \{y(\mathbf{x})-\mathbb{E}[t|\mathbf{x}]\}^2+2\{y(\mathbf{x})-\mathbb{E}[t|\mathbf{x}]\}\{\mathbb{E}[t|\mathbf{x}]-t\}+\{\mathbb{E}[t|\mathbf{x}]-t\}^2\end{aligned}$$

$$\mathbb{E}[L]=\int\{y(\mathbf{x})-\mathbb{E}[t|\mathbf{x}]\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x}+\int\{\mathbb{E}[t|\mathbf{x}]-t\}^2 p(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

5. The Decision Theory

$$E[L] = \int \{y(\mathbf{x}) - E[t|\mathbf{x}]\}^2 p(\mathbf{x}) d\mathbf{x} + \int var[t|\mathbf{x}] p(\mathbf{x}) d\mathbf{x}$$

에러 를 구성하는 요소를 파악해 보기
: expected loss 값은 크게 2가지 요소로 나누어 볼 수 있다

샘플 데이터로부터 추정된 y(x) 가 E[t|x] 와
동일한 결과를 가진다면 두번째 term 만 남게 된다.

에러를 최소화하는 방향으로 식을 근사
-> y(x) 를 최대한 E[t|x] 와 동일하게 만드는 방향

첫번째 term : 모델 y(x) 와 관련된 요소로 조건부 평균을 통해
최소 제곱 방식을 사용하는 방식
두번째 term : 분산으로서 샘플이 포함하고 있는 noise를 의미

5. The Decision Theory

$$\mathbb{E}[L_q] = \iint |y(\mathbf{x}) - t|^q p(\mathbf{x}, t) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t$$

다른 함수를 도입 -> q 값에 따른 함수의 변화 모양