20기 정규세션

ToBig's 19기 강의자
김은지

Naïve Bayes Classification

Contents

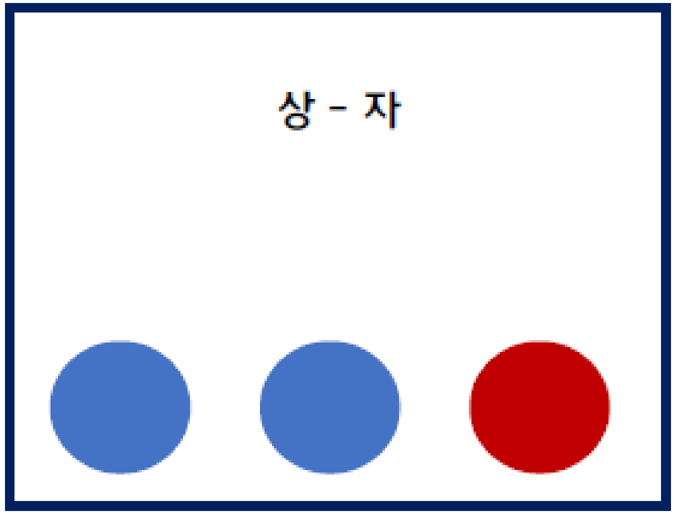
목표

1. 기본적인 확률 공식들을 바탕으로 베이즈 정리를 이해한다.

2. 베이즈 정리를 바탕으로 나이브 베이즈에 대해서 이해한다.

1. 확률이란(probability)

특정 사건이 일어날 정도를 수치로 나타낸 것



파란 공을 뽑을 확률 : 2/3
빨간 공을 뽑을 확률 : 1/3

1/3 + 2/3 = 1

2. 조건부 확률 (conditional probability)

- 어떤 사건이 일어난 조건 하에서, 다른 사건이 일어날 확률

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$ → 사건 A가 일어났을 때, 사건 B가 일어날 확률

- 곱셈 공식

$$P(B|A)P(A) = P(B\cap A) = P(A\cap B) = P(A|B)P(B)$$

3. 독립과 조건부 독립 (Independent & conditional independent)

- 독립 : 한 사건이 일어날 확률이 다른 사건이 일어날 확률에 영향을 미치지 않는다.

$$P(A \cap B) = P(A)P(B)$$

- 조건부 독립 : 한 사건이 일어났다는 가정하에서, 서로 다른 두 사건은 독립인 상황

$$P(A, B|C) = P(A|C)P(B|C)$$

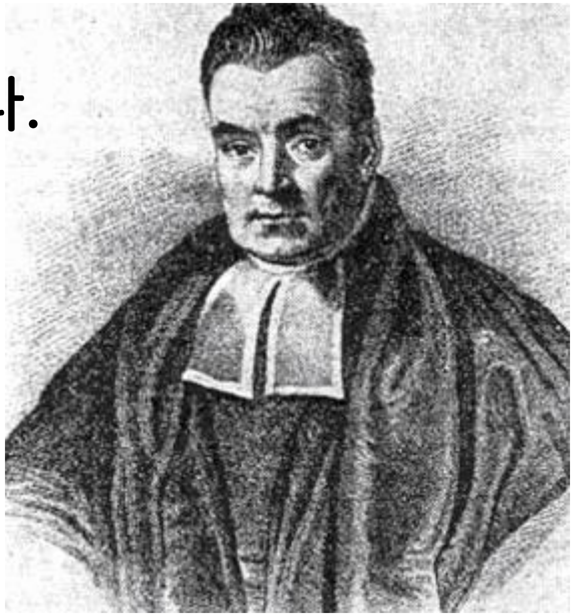
C 사건이 일어났을 때, 사건 A가 일어날 확률은
사건 B가 일어날 확률에 영향을 주지 않는다.

4. 빈도주의통계 vs 베이지안 통계

빈도주의통계와는 다르게 베이지안 통계에서는 Parameter를 변수로 생각합니다.

이전의 경험과 현재의 증거를 토대로 어떤 사건의 확률을 추론한다.

사전확률과 사후확률 사이의 관계를 조건부확률을 이용해 계산



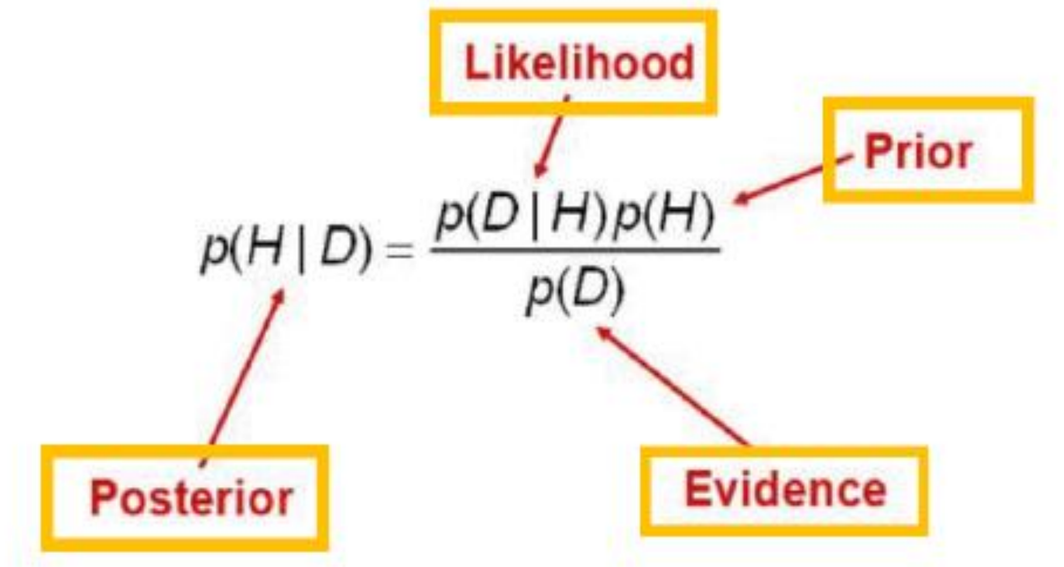
1. 베이즈 정리(Bayes' Theorem)

두 확률 변수의 사전 확률(prior)과 사후 확률(posterior) 사이의 관계를 나타내는 정리

사전 확률(prior)로부터 사후 확률(posterior)을 구하고자 한다.

Unit 02 | 베이즈 정리(Bayes' Theorem)

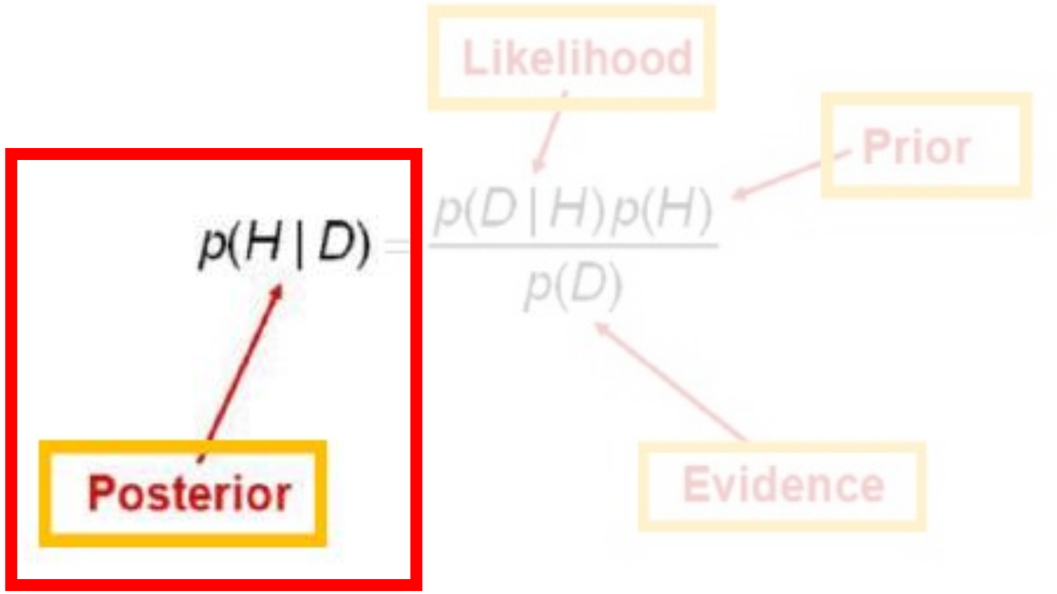
1. 베이즈 정리(Bayes' Theorem)



H - 알고 싶은 정보 ex) 나가서 운동하기 적합한 날씨인가
D - 이미 알고 있는 정보 ex) 기상 정보

- Prior
  - 사전 확률
  - 과거의 경험을 토대로 내가 지정한 파라미터(H)의 확률
- Posterior
  - 사후 확률
  - 관측 결과 사건 D가 일어난 조건 하의 파라미터(H)의 확률
- Likelihood
  - 사전 확률의 과거 경험을 잘 설명하는 정도
  - 모델 파라미터(H)를 바탕으로 하는 관측결과 사건 D의 확률
- Evidence(normalizing constant)
  - 사건 D의 발생 가능성
  - 모델 파라미터(H)와 무관한 관측 결과(D) 자체의 확률
  - 우리가 관심있는 H와 무관한 값, 보통 상수로 생각하고 무시하고 계산

1. 베이즈 정리(Bayes' Theorem)

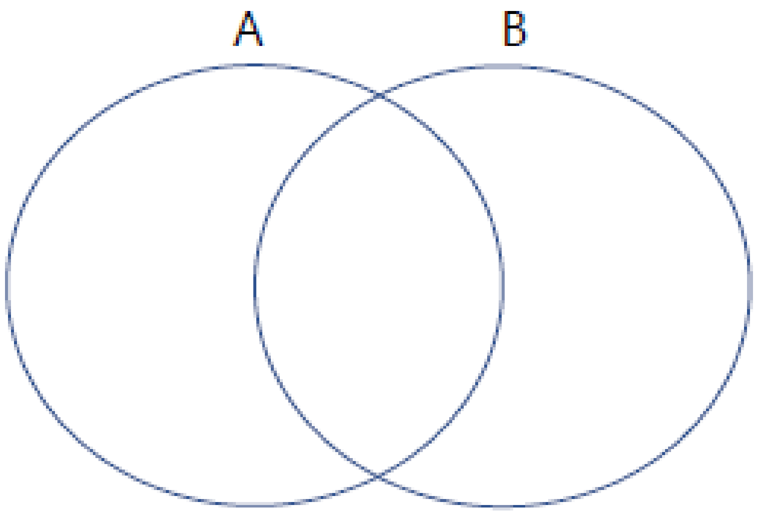


H - 알고 싶은 정보 ex) 나가서 운동하기 적합한 날씨인가
D - 이미 알고 있는 정보 ex) 기상 정보

- **Posterior**
  - 직접적(데이터, 실험 등)으로 관찰 및 표현이 불가능하지만, 계산을 통해 '사후적으로' 알 수 있는 확률

1) 사전에 예상한 가설 H에 대한 확률(사전확률)에
2) 관찰된 데이터 D로 계산한 가능성정도(likelihood)를 **곱하고**
3) 모델 파라미터와 무관한, 관측 결과가 발생할 확률 (Normalizing Constant)로 **나눠서** 구한 값

2. 베이즈 정리(Bayes' Theorem) 증명



$P(A|B) = P(A \cap B) / P(B)$
$P(A \cap B) = P(A|B) * P(B)$
$P(B \cap A) = P(B|A) * P(A)$
$P(B \cap A) = P(A \cap B)$
$P(A \cap B) = P(B|A) *P(A) = P(A|B) * P(B)$
$P(B|A) * P(A) = P(A|B) * P(B)$

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

Unit 02 | 베이즈 정리(Bayes' Theorem)

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$P(Y = yes) = 3/5$$
$$P(X = sunny \mid Y = yes) = 2/3$$
$$P(X = rainy \mid Y = yes) = 1/3$$

$$P(Y = no) = 2/5$$
$$P(X = sunny \mid Y = no) = 1/2$$
$$P(X = rainy \mid Y = no) = 1/2$$

Unit 02 | 베이즈 정리(Bayes' Theorem)

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$P(Y=yes \mid X=sunny) = \frac{P(X=sunny \mid Y=yes) \times P(Y=yes)}{P(X=sunny)}$$

$$P(Y=no \mid X=sunny) = \frac{P(X=sunny \mid Y=no) \times P(Y=no)}{P(X=sunny)}$$

by 베이즈 정리

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$P(Y=yes \mid X=sunny) = \frac{\boxed{P(X=sunny \mid Y=yes) \times P(Y=yes)}}{P(X=sunny)}$$

$$P(Y=no \mid X=sunny) = \frac{\boxed{P(X=sunny \mid Y=no) \times P(Y=no)}}{P(X=sunny)}$$

by 베이즈 정리

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$P(Y=yes|X=sunny) = \frac{2/3 \times 3/5}{\boxed{P(X=sunny)}}$$

$$P(Y=no|X=sunny) = \frac{1/2 \times 2/5}{\boxed{P(X=sunny)}}$$

?

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$P(Y = yes \mid X = sunny) = \frac{2/3 \times 3/5}{\boxed{P(X = sunny)}}$$

$$P(Y = no \mid X = sunny) = \frac{1/2 \times 2/5}{\boxed{P(X = sunny)}}$$

$$P(Y = yes \mid X = sunny) + P(Y = no \mid X = sunny) = 1$$

굳이 노란색 박스를 구할 필요가 없다!

3. 연습

Sky (X)	Enjoy Point (Y)
Sunny	Yes
Sunny	Yes
Rainy	No
Sunny	No
Rainy	Yes
Sunny	?

$$\boxed{P(Y = yes \mid X = sunny) = 2/3}$$

$$P(Y = no \mid X = sunny) = 1/3$$

1. 계산의 한계

d = 변수의 개수
k = class 개수(Yes, No)

Sky	Temp	Humid	Wind	Water	Forest	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

$$f^*(x) = argmax_{Y=y}P(X = x|Y = y)P(Y = y)$$

$P(X = x|Y = y)$
= P(x1= sunny, x2= warm, x3= normal, x4= strong, x5= warm, x6= same | y=Yes)
$P(Y = y) = (y = Yes)$
$P(X = x|Y = y) = for\ all\ \ x,y \ \ \rightarrow (2^d - 1)k$
$P(Y = y)\ for\ all\ y \ \ \rightarrow \ \ k - 1$

1. 계산의 한계

문제점 : 계산량이 많아짐

$$P(X = x|Y = y) = for\ all\ x, y \ \rightarrow (2^d - 1)k$$

변수가 늘어날수록 연산량이 기하급수적으로 증가함
어떻게 얻은 변수들인데 …. d의 개수를 줄이는 것은 피하고 싶어!

=> 해결책 : **조건부 독립**을 가정!!

2. Naïve Bayes Classification

- 가정 : 종속변수(Y)가 주어졌을 때, 입력 변수들이 모두 독립이다! (조건부 독립)
- 결과가 주어졌을 때, 예측 변수 벡터의 정확한 조건부 확률은 각 조건부 확률의 곱으로 충분히 잘 추정할 수 있다는 단순한 가정을 기초로 한다.
  -> 데이터셋을 너무 믿네 순진하긴~~ : naïve(순진하다)
  -> Naïve Bayes

$$f^*(x) = argmax_{Y=y} P(X = x|Y = y)P(Y = y)$$

$$\approx \mathrm{argmax}_{\mathrm{Y=y}} \mathrm{P}(\mathrm{Y} = \mathrm{y}) \prod_{1 \le i \le d} P(X = x_i\,|\mathrm{Y} = \mathrm{y})$$

2. Naïve Bayes Classification

d = 변수의 개수
k = class 개수(Yes, No)

Sky	Temp	Humid	Wind	Water	Forest	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

$$f^*(x) = argmax_{Y=y} P(X = x|Y = y)P(Y = y)$$

$P(X = x|Y = y)$
= P(x1= sunny, x2= warm, x3= normal, x4= strong, x5= warm, x6= same | y=Yes)

$P(Y = y) = (y = Yes)$

$P(X = x|Y = y) = for\ all\ \ x, y \ \ \rightarrow (2^d - 1)k$

$P(Y = y)\ for\ all\ y \ \ \rightarrow \ \ k - 1$

2. Naïve Bayes Classification

d = 변수의 개수
k = class 개수(Yes, No)

Sky	Temp	Humid	Wind	Water	Forest	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

$$f^*(x) = argmax_{Y=y} P(X = x|Y = y)P(Y = y)$$

$\boldsymbol{P(X = x|Y = y)}$
= P(x1= sunny, x2= warm, x3= normal, x4= strong, x5= warm, x6= same | y=Yes)

$P(Y = y) = (y = Yes)$

$P(X = x|Y = y) = for\ all\ \ x, y \ \ \rightarrow (2^d - 1)k$

$P(Y = y)\ for\ all\ y \ \ \rightarrow \ \ k - 1$

조건부 독립에 의하여!

2. Naïve Bayes Classification

$P(X = x|Y = y)$
= P(x1= sunny, x2= warm, x3= normal, x4= strong, x5= warm, x6= same | y=Yes)

조건부 독립 => $\boldsymbol{P(A \cap B|C) = P(A|C)P(B|C)}$

$\boldsymbol{P(X = x|Y = y)}$
= P(x1= sunny | Y=yes) P(x2= warm | Y=yes) P(x3= normal | Y=yes)
P(x4= strong | Y=yes), P(x5= warm | Y=yes), P(x6= same | y=Yes)

$$f^*(x) = argmax_{Y=y} P(X = x|Y = y)P(Y = y)$$

$$\approx \mathrm{argmax}_{\mathrm{Y=y}} \mathrm{P(Y = y)} \prod_{1 \le i \le d} P(X = x_i\,|\mathrm{Y = y})$$

2. Naïve Bayes Classification

1. 알아야 할 파라미터 수가 대폭 줄어들게 된다.

$$P(X = x \mid Y = y) \;\Rightarrow\; dk$$

2. Feature 들의 곱으로 바뀌면서 계산이 수월해진다.

3. Maximum Likelihood Estimation

1. 최대우도법(Maximum Likelihood Estimation, MLE)란?

모수적인 데이터 밀도 추정 방법으로써 파라미터 $\theta = (\theta_1, \dots, \theta_m)$ 으로 구성된
어떤 확률밀도함수 $P(x|\theta)$ 에서 관측된 표본 데이터 집합을 $x = (x_1, \dots, x_n)$ 이라 할 때,
이 표본들에서 파라미터 $\theta = (\theta_1, \dots, \theta_m)$ 를 추정하는 방법

3. Maximum Likelihood Estimation

1. 최대우도법(Maximum Likelihood Estimation, MLE)란?

단어 그대로

모수적인 데이터 밀도 추정 방법으로써 파라미터 $\theta = (\theta_1, \ldots, \theta_m)$ 으로 구성된

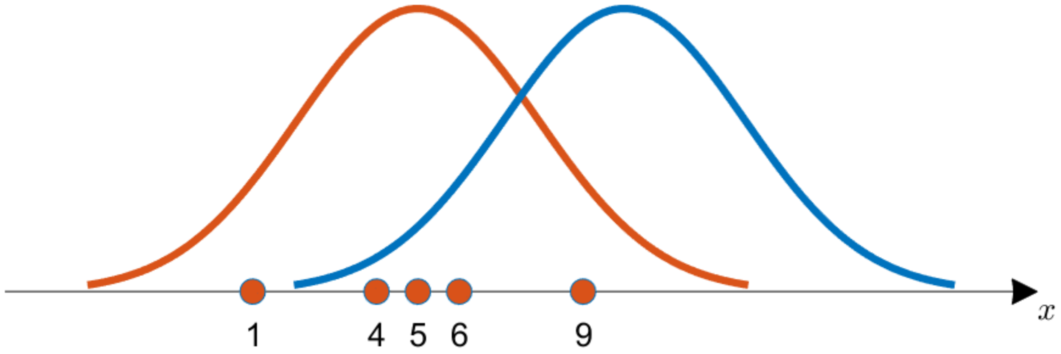
'우도(Likelihood, 가능도)'를 '최대화' 하는 지점을 찾는 것을 의미

어떤 확률밀도함수 $P(x|\theta)$ 에서 관측된 표본 데이터 집합을 $x = (x_1, \ldots, x_n)$ 이라 할 때,

이 표본들에서 파라미터 $\theta = (\theta_1, \ldots, \theta_m)$ 를 추정하는 방법

3. Maximum Likelihood Estimation

1. 최대우도법(Maximum Likelihood Estimation, MLE)란?



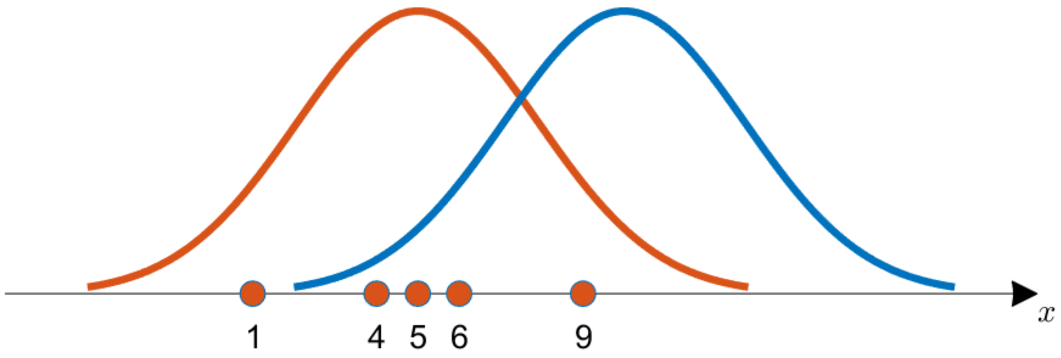
데이터 x = {1, 4, 5, 6, 9} 는

주황색 곡선과 파란색 곡선 중 어떤 곡선으로부터

추출되었을 확률이 더 높을까?

3. Maximum Likelihood Estimation

1. 최대우도법(Maximum Likelihood Estimation, MLE)란?



데이터 x = {1, 4, 5, 6, 9} 는

주황색 곡선과 파란색 곡선 중 어떤 곡선으로부터

추출되었을 확률이 더 높을까?

주황색 곡선!

획득한 데이터들의 분포가 주황색 곡선의 중심에 더 일치하는 것처럼 보이기 때문

3. Maximum Likelihood Estimation

2. Likelihood function

- 지금 얻은 데이터가 이 분포로부터 나왔을 가능도

- 수치적으로 이 가능도를 계산하기 위해서는 각 데이터 샘플에서 후보 분포에 대한 높이(즉, likelihood 기여도)를 계산해서 다 곱한 것을 이용할 수 있다.

$$P(x|\theta) = \prod_{k=1}^{n} P(x_k|\theta)$$

likelihood function (전체 표본집합의 결합확률밀도함수)

3. Maximum Likelihood Estimation

3. Likelihood function의 최대값을 찾는 방법

- 결국 Maximum Likelihood Estimation은 Likelihood function의 최대값을 찾는 방법
- 계산의 편의를 위해 log-likelihood 의 최대값을 찾는다.
- 어떤 함수의 최대값을 찾는 방법 중 가장 보편적인 방법은 **미분**을 이용하는 방법

  => 찾고자 하는 파라미터 θ**에 대해 편미분하고 그 값이 0이 되도록 하는** θ를 찾는 과정

  을 통해 likelihood function을 최대화 시켜줄 수 있는 θ를 찾을 수 있다.

$$L(\theta|x) = \log P(x|\theta) = \sum_{i=1}^{n} \log P(x_i|\theta)$$

$$\frac{\partial}{\partial\theta} L(\theta|x) = \frac{\partial}{\partial\theta} \log P(x|\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log P(x_i|\theta) = 0$$

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

전부 0 또는 1만 나오는 베르누이 분포~!

- 모수

  P(Y=1) = Py

  P(X1=1|Y=1) = P1

  P(X2=1|Y=1) = P2

  P(X3=1|Y=1) = P3

  P(X1=1|Y=0) = P4

  P(X2=1|Y=0) = P5

  P(X3=1|Y=0) = P6

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

- 각 행의 정보를 표현 : $-\log P(X1,X2,X3,Y)$
- $-\log P(X1,X2,X3,Y) = -\log P(X1,X2,X3|Y)P(Y)$
  By 조건부 확률
- $-\log P(X1,X2,X3|Y)P(Y)$
  $= -\log P(X1|Y)P(X2|Y)P(X3|Y)P(Y)$
  By 조건부 독립

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

위에서 정한 모수를 토대로 각행의 정보를 나타내자

→ -log[P1*P2*(1-P3)*PY]

→ -log[(1-P1)*P2*(1-P3)*PY]

→ -log[P4*P5*P6*(1-PY)]

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

해당 정보들을 싹 다 더하고(로그니까 진수들의 곱이 됩니다),
P1만 보자

L = -log[P1^2*(1-P1)^2…]: log-likelihood

이제 이 값을 최소로 만들려면
편미분 했을때 0이 되는 모수들을 찾으면 된다.

dL/dP1 = -[(2/P1) – (2/(1-P1))] = 0

따라서 P1 = 1/2

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

근데 매번 이렇게 구하는 건 좀….

P1 = P(X1=1|Y=1) => Y=1인 샘플들 중

X1=1인 경우의 비율과 같다!!

따라서, 우리는 실제 적용시에는 MLE과정을 생략하고

데이터에서 개수를 세서 구한다!

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

연습

$P(Y=1|X1=0,X2=0,X3=1)$ 과

$P(Y=0|X1=0,X2=0,X3=1)$

을 비교 하면 됩니다

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

P(Y=1|X1=0,X2=0,X3=1)

P(Y=0|X1=0,X2=0,X3=1)

$$P(Y=1|X1=0,X2=0,X3=1)$$

$$=\frac{P(X1=0,X2=0,X3=1|Y=1)P(Y=1)}{P(X1=0,X2=0,X3=1)}$$

$$=\frac{P(X1=0|Y=1)P(X2=0|Y=1)P(X3=1|Y=1)P(Y=1)}{P(X1=0,X2=0,X3=1)}$$

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

P(Y=1|X1=0,X2=0,X3=1)

P(Y=0|X1=0,X2=0,X3=1)

$$\frac{\boxed{P(X1=0|Y=1)}P(X2=0|Y=1)P(X3=1|Y=1)P(Y=1)}{P(X1=0,X2=0,X3=1)}$$

$P(\boxed{X1=0}\boxed{|Y=1)} = 2/4$

3. Maximum Likelihood Estimation

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

P(Y=1|X1=0,X2=0,X3=1)

P(Y=0|X1=0,X2=0,X3=1)

$$\frac{P(X1=0|Y=1)P(X2=0|Y=1)P(X3=1|Y=1\,\boxed{P(Y=1)}}{P(X1=0,X2=0,X3=1)}$$

$\boxed{P(Y=1)}$ = 3/7

3. Maximum Likelihood Estimation

P(Y=1|X1=0,X2=0,X3=1)

P(Y=0|X1=0,X2=0,X3=1)

4. 연습

X1	X2	X3	Y
1	1	0	1
0	1	0	1
1	1	1	0
0	0	1	1
0	0	0	0
1	0	1	0
1	0	0	1
0	0	1	?

$$P(Y=1|X1=0,X2=0,X3=1) = \frac{\boxed{2/4*2/4*1/4*4/7}}{P(X1=0,X2=0,X3=1)}$$

$$P(Y=0|X1=0,X2=0,X3=1) = \frac{\boxed{1/3*2/3*2/3*3/7}}{P(X1=0,X2=0,X3=1)}$$

따라서 Y=0 으로 예상된다

4. 라플라스 스무딩(Laplace Smoothing)

Likelihood가 0이 되는 것을 방지하도록 최소한의 확률을 정해주자!!

$$P_{LAP} = \frac{c(x) + 1}{\sum x \, [c(x)+1]}$$

실제보다 한 번씩 더 관찰되었다고 가정하기

$$P(x|c) = \frac{count(x,c) + 1}{\sum_{x \in v} count(x,c) + v}$$

분자에 1 더하고
분모에 입력변수들의 개수 v 더하는 꼴

4. 라플라스 스무딩(Laplace Smoothing)

Ex) “단어가 스포츠를 나타내는 단어인가, 아닌가?”에 대한 NB 케이스

14개의 단어로 이루어진 데이터

Word	P(word | Sports)	P(word | Not Sports)
a	$\frac{2+1}{11+14}$	$\frac{1+1}{9+14}$
very	$\frac{1+1}{11+14}$	$\frac{0+1}{9+14}$
close	$\frac{0+1}{11+14}$	$\frac{1+1}{9+14}$
game	$\frac{2+1}{11+14}$	$\frac{0+1}{9+14}$

정리

Naïve Bayes

장점

- 입력 공간의 차원이 높을 때 유리
- 텍스트에서 강점
- 가우시안 나이브베이즈를 활용하면 input이 연속형일때도 사용 가능

단점

- 희귀한 확률이 나왔을 때 (라플라스 스무딩)
- 조건부 독립이라는 가정 자체가 비현실적

Unit 04 | 과제 설명

과제

week3_NaiveBayes_assignment.ipynb 를 완성해주세요!

참고자료

참고자료
-투빅스 12기 김태한님 강의자료
-투빅스 14기 정재윤님 강의자료
-투빅스 13기 김미성님 강의자료
-투빅스 18기 김성훈님 강의자료
- https://ratsgo.github.io/machine%20learning/2017/05/18/naive/
- https://datascienceschool.net/view-notebook/c19b48e3c7b048668f2bb0a113bd25f7/
- https://medium.com/@LSchultebraucks/gaussian-naive-bayes-19156306079b
- https://www.edwith.org/machinelearning1_17/joinLectures/9738
-https://www.youtube.com/watch?v=hO9SVW6nnhM
-https://ratsgo.github.io/statistics/2017/09/23/MLE/
-https://ebbnflow.tistory.com/330
-https://angeloyeo.github.io/2020/07/17/MLE.html

Q & A

들어주셔서 감사합니다.