15기 정규세션

ToBig's 14기 정재윤

Naïve Bayes Classifier

* Naïve : 순진하다 Bayes : 베이즈정리

Contents

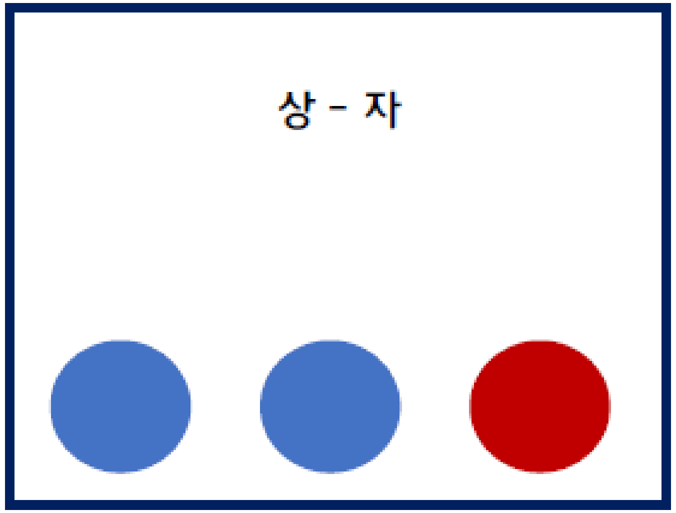
목표

[목표]

- 기본적인 확률 공식들을 바탕으로 베이즈 정리를 이해한다.
- 앞서 이해한 베이즈 정리를 바탕으로 나이브 베이즈에 대해서 이해한다.

1-1) 확률이란 (Probability)

- 특정한 사건이 일어날 가능성을 나타낸 것



파란 공을 뽑을 확률 : 2/3
빨간 공을 뽑을 확률 : 1/3

$1/3 + 2/3 = 1$

Unit 01 | 확률 기초(Probability Overview)

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

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

1-3) 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.

Unit 01 | 확률 기초 (Probability Overview)

1-4) 빈도주의통계 vs 베이지안 통계

빈도주의통계와는 다르게 베이지안 통계에서는 Parameter를 변수로 생각합니다.

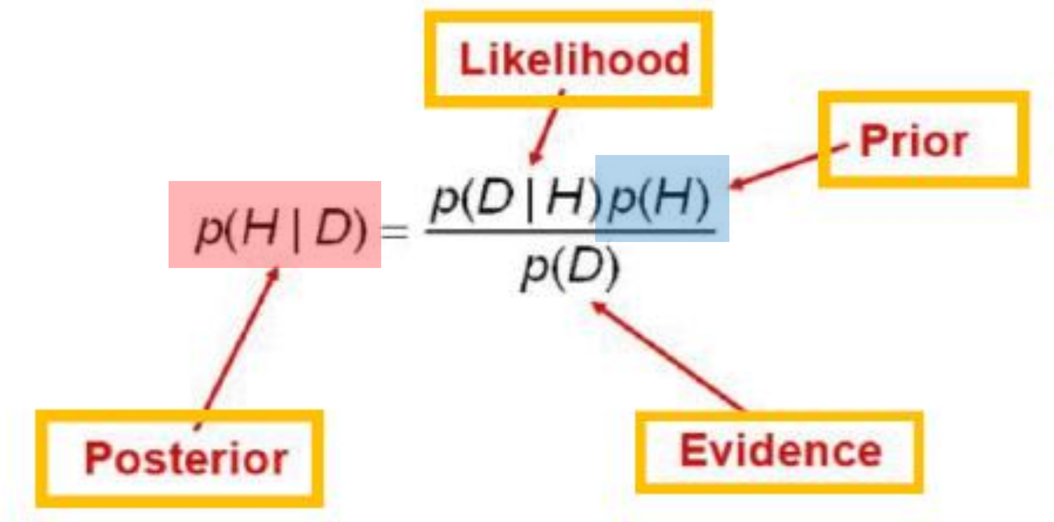
이전의 경험과 현재의 증거를 토대로 어떤 사건의 확률을 추론한다.

사전확률과 사후확률 사이의 관계를 조건부확률을 이용해 계산

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-1) 베이즈 정리

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



베이즈 정리는 사전확률(prior)로부터 사후확률(posterior)을 구할 수 있다.
어떻게? 조건부 확률로!

prior – 사전 확률 , 과거의 경험을 토대로 내가 지정한 확률

likelihood – 사전 확률의 과거 경험을 잘 설명하는 정도

posterior – 사후 확률 , 사건 D가 일어난 조건 하의 확률

evidence – 사건 D의 발생 가능성

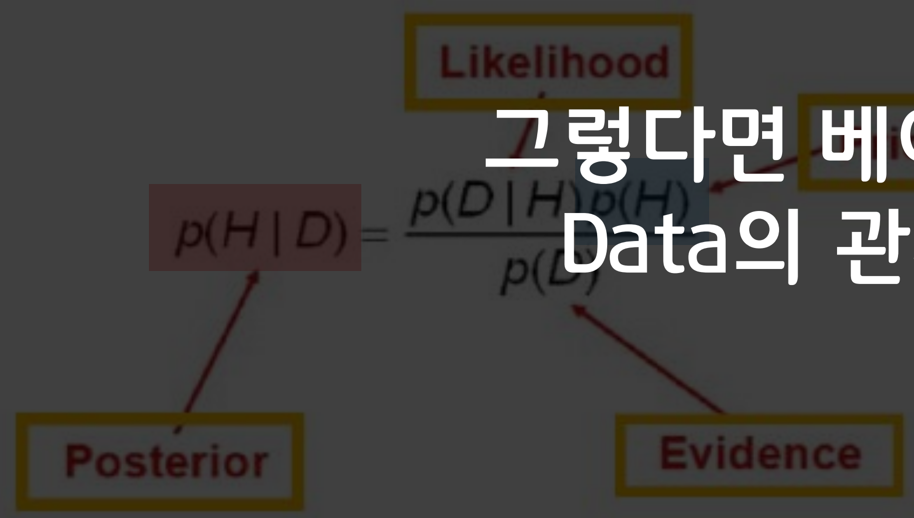
H – 알고 싶은 정보
D – 이미 알고 있는 정보

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-1) 베이즈 정리

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

'뭔 소리야 진짜…'



베이즈 정리는 사전확률(prior)로부터 사후확률(posterior)을 구할 수 있다.
어떻게? 조건부 확률로!

prior - 사전 확률 , 과거의 경험을 토대로 내가 지정한 확률

likelihood - 사전 확률의 과거 경험을 잘 설명하는 정도

posterior - 사후 확률 , 사건 D가 일어난 조건 하의 확률

evidence - 사건 D의 발생 가능성

H - 알고 싶은 정보
D - 이미 알고 있는 정보

그렇다면 베이즈 정리를 Parameter와
Data의 관계로 다시 한번 봅시다!!

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-1) 베이즈 정리

$$\underbrace{P(H|D)}_{\text{Posterior}} = \frac{\overbrace{P(D|H)}^{\text{Likelihood}}\overbrace{P(H)}^{\text{Prior}}}{\underbrace{P(D)}_{\text{Evidence}}}$$

prior - 사전 분포. Parameter의 분포를 의미
우리는 이 분포를 이미 알고 있다.

Likelihood – 가정한 Parameter에서 Data를 뽑을 확률

Posterior – 사후 분포 , 데이터를 바탕으로 추정한 파라미터의 분포

evidence – 해당 Data를 뽑을 확률

H - Parameter와 관련된 변수
D - Data와 관련된 변수

Unit 02 | 베이즈 정리 (Bayes' Rule)

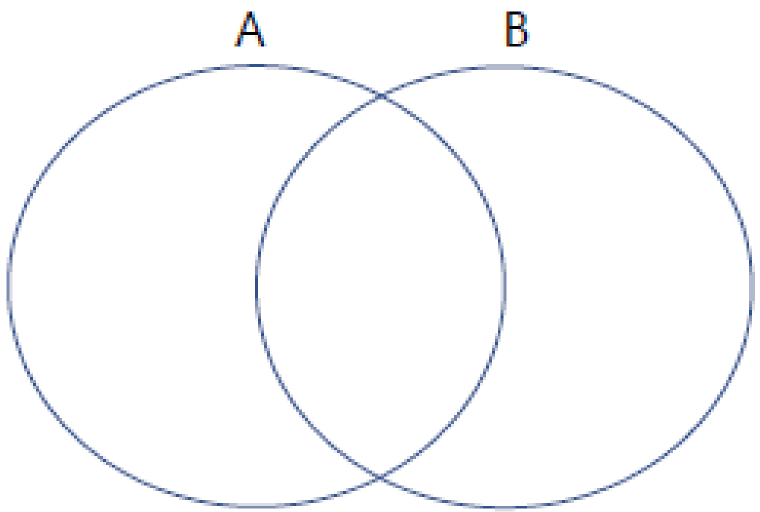
2-1) 베이즈 정리

즉, 우리는 우리가 알고 있는 Parameter의 분포와 해당 Parameter에서 뽑은 Data의 확률로부터 데이터를 바탕으로 추정한 Parameter의 분포를 구할 수 있다!!

우리는 우리가 알고 있는 정보로부터, 우리가 구하고 싶어하는 Parameter의 분포를 구할 수 있다!!!

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-2) 베이즈 정리 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' Rule)

2-3) 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' Rule)

2-3) 연습

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

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

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

by 베이즈 정리

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-3) 연습

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

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

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

?

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-3) 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 노란 박스를 구할 필요가 없다~!

Unit 02 | 베이즈 정리 (Bayes' Rule)

2-3) 연습

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

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

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

Unit 03 | Naive Bayes Classification

3-1) 계산의 한계

d=관측치 개수
K=class 개수

Sky	Temp	Humid	Wind	Water	Forecst	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

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

$\mathrm{P}(\mathrm{X} = \mathrm{x}|\mathrm{Y} = \mathrm{y})$

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

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

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

P(Y=y) for all y -> k-1

Unit 03 | Naive Bayes Classification

3-1) 계산의 한계

문제점: 계산량이 많아짐

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

변수가 늘어날 수록 기하급수적으로 연산량이 증가함

어떻게 얻은 변수들인데……. d의 개수를 줄이는 건 피하고 싶어

->>해결책: 조건부 독립을 가정!!!!

Unit 03 | Naive Bayes Classification

3-2) Naive Bayes Classification

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)

by. 조건부 독립

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

Unit 03 | Naive Bayes Classification

3-2) Naive Bayes Classification

1. 알아야할 파라미터의 수가 대폭 줄어들게 된다.

$$P(X = x|Y = y) => dk$$

2. Feature들의 곱으로 바뀌면서 계산이 수월해진다.

Unit 03 | Naive Bayes Classification

3-2) Naive Bayes Classification

- 가정 : 종속변수(Y)가 주어졌을 때, 입력 변수들이 모두 독립이다!!!!(조건부 독립 가정)

- 결과가 주어졌을 때, 예측 변수 벡터의 정확한 조건부 확률은 각 조건부 확률의 곱으로 충분히 잘 추정 할 수 있다는 단순한 가정을 기초로 한다
  -> 데이터셋을 순진하게 믿는다! -> Naïve Bayes!!

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

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

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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 :
$P(Y=1) = Py$
$P(X1=1|Y=1) = P1$
$P(X2=1|Y=1) = P2$
$P(X3=1|Y=1) = P3$
$P(X1=1|Y=0) = P4$
$P(X2=1|Y=0) = P5$
$P(X3=1|Y=0) = P6$

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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 각 행의 log likelihood를 나타내자

$\Rightarrow$ -log[P1*P2*(1-P3)*PY]

$\Rightarrow$ -log[(1-P1)*P2*(1-P3)*PY]

$\Rightarrow$ -log[P4*P5*P6*(1-PY)]

.
.
.

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

X1	X2	X3	Y
$\boxed{1}$	1	0	1
$\boxed{0}$	1	0	1
1	1	1	0
$\boxed{0}$	0	1	1
0	0	0	0
1	0	1	0
$\boxed{1}$	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

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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!

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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

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

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

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

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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

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

P(X1=0|Y=1) = 2/4

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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

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

P(Y=1) = 4/7

Unit 03 | Naive Bayes Classification

3-3) Maximum Likelihood Estimation

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

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

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

따라서 Y=0 으로 예상된다

Unit 03 | Naive Bayes Classification

3-4) 라플라스 스무딩 (Laplace Smoothing)

Positive or Negative		Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

P(-) = 3/5 , P(+) = 2/5

P(predictable | -) = 1/14
P(no | -) = 1/14
$\boxed{\text{P(fun | -) = 0/14}}$

$\boxed{\begin{array}{l}\text{P(predictable | +) = 0/9} \\ \text{P(no | +) = 0/9}\end{array}}$
P(fun | +) = 1/9

Unit 03 | Naive Bayes Classification

3-4) 라플라스 스무딩 (Laplace Smoothing)

Positive or Negative		Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

likelihood 부분이 둘다 0이 된다

$P(\text{predictable} \mid -) \times P(\text{no} \mid -) \times P(\text{fun} \mid -) = 0$

$P(\text{predictable} \mid +) \times P(\text{no} \mid +) \times P(\text{fun} \mid +) = 0$

Unit 03 | Naive Bayes Classification

3-4) 라플라스 스무딩 (Laplace Smoothing)

likelihood가 0이 되는 것을 방지하도록 최소한의 확률을 정해주자!!

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

v는 입력변수들의 개수

Unit 04 | Gaussian Naïve Bayes

4-1) 연속적인 입력변수 (continuous input variables)

X1	X2	X3	Y
1	0	0	True
0	1	0	True
0	0	1	False
0	1	0	True
1	1	0	False
1	0	1	?

지금까지 feature들은 베르누이 분포!

Unit 04 | Gaussian Naïve Bayes

4-1) 연속적인 입력변수 (continuous input variables)

X1	X2	X3	Y
6	0.12	152	True
1	0.2	267	True
8	0.64	363	False
3	0.97	162	True
4	0.34	667	False
5	0.5	334	?

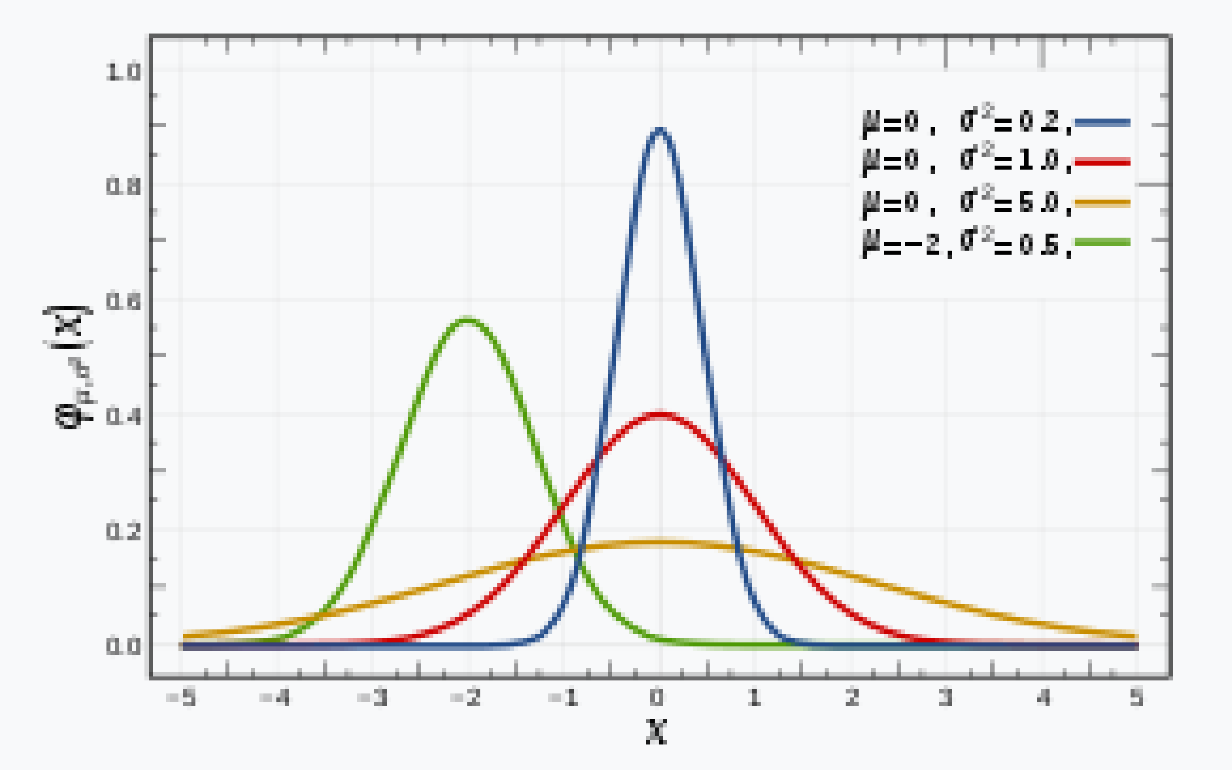
그렇다면 연속적인 경우라면??

Unit 04 | Gaussian Naïve Bayes

4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

Idea

연속적인 입력변수들을
가우시안 분포를 가진다고 가정하여
나이브 베이즈 방법을 사용한다



4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Derivation from the naïve Bayes to the logistic regression
  - $P(Y)\prod_{1\le i\le d} P(X_i|Y) = \pi_k \prod_{1\le i\le d} \frac{1}{\sigma_k^i C} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_k^i}{\sigma_k^i}\right)^2\right)$
- With naïve Bayes assumption
  - $P(Y=y|X) = \frac{P(X|Y=y)P(Y=y)}{P(X)} = \frac{P(X|Y=y)P(Y=y)}{P(X|Y=y)P(Y=y)+P(X|Y=n)P(Y=n)}$

$$= \frac{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y)}{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y) + P(Y=n)\prod_{1\le i\le d} P(X_i|Y=n)}$$

Unit 04 | Gaussian Naïve Bayes

4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- With naïve Bayes assumption
  - $P(Y=y|X) = \frac{P(X|Y=y)P(Y=y)}{P(X)} = \frac{P(X|Y=y)P(Y=y)}{P(X|Y=y)P(Y=y)+P(X|Y=n)P(Y=n)}$

$$= \frac{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y)}{P(Y=y)\prod_{1\le i\le d} P(X_i|Y=y) + P(Y=n)\prod_{1\le i\le d} P(X_i|Y=n)}$$

  - $P(Y=y|X) = \frac{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2) + \pi_2 \prod_{1\le i\le d} \frac{1}{\sigma_2^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}$

$$= \frac{1}{1 + \frac{\pi_2 \prod_{1\le i\le d} \frac{1}{\sigma_2^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2)}{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2)}}$$

Unit 04 | Gaussian Naïve Bayes

4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Assuming the same variable of the two classes, $\sigma_2^i = \sigma_1^i$
  - $$P(Y=y|X) = \frac{1}{1+\frac{\pi_2 \prod_{1\le i\le d} \frac{1}{\sigma_2^i C} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right)}{\pi_1 \prod_{1\le i\le d} \frac{1}{\sigma_1^i C} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right)}} = \frac{1}{1+\frac{\pi_2 \prod_{1\le i\le d} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right)}{\pi_1 \prod_{1\le i\le d} \exp\left(-\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right)}}$$

$$= \frac{1}{1+\frac{\pi_2 \exp\left(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right\}\right)}{\pi_1 \exp\left(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right\}\right)}}$$

$$= \frac{1}{1+\frac{\exp\left(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right\}+log\pi_2\right)}{\exp\left(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right\}+log\pi_1\right)}}$$

Unit 04 | Gaussian Naïve Bayes

4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Assuming the same variable of the two classes, $\sigma_2^i = \sigma_1^i$
  - $P(Y = y|X) = \dfrac{1}{1+\exp(-\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_2^i}{\sigma_2^i}\right)^2\right\}+log\pi_2+\sum_{1\le i\le d}\left\{\frac{1}{2}\left(\frac{X_i-\mu_1^i}{\sigma_1^i}\right)^2\right\}-log\pi_1)}$
  - $= \dfrac{1}{1+\exp(-\frac{1}{2(\sigma_1^i)^2}\sum_{1\le i\le d}\left\{(X_i-\mu_1^i)^2-(X_i-\mu_2^i)^2\right\}+log\pi_2-log\pi_1)}$
  - $= \dfrac{1}{1+\exp(-\frac{1}{2(\sigma_1^i)^2}\sum_{1\le i\le d}\left\{2(\mu_2^i-\mu_1^i)X_i+{\mu_2^i}^2-{\mu_2^i}^2\right\}+log\pi_2-log\pi_1)}$

Unit 04 | Gaussian Naïve Bayes

4-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Naïve Bayes classifier
  - $P(Y|X) = \frac{1}{1+\exp(-\frac{1}{2(\sigma_1^i)^2}\Sigma_{1\leq i\leq d}\{2(\mu_2^i-\mu_1^i)X_i+\mu_2^{i\,2}-\mu_2^{i\,2}\}+\log\pi_2-\log\pi_1)}$
  - Assumption to get this formula
    - Naïve Bayes assumption, Same variance assumption between classes
    - Gaussian distribution for P(X|Y)
    - Bernoulli distribution for P(Y)
  - # of parameters to estimate = 2*2*d+1=4d+1
    - With the different variances between classes
- Logistic Regression
  - $P(Y|X) = \frac{1}{1+e^{-\theta^T x}}$
  - Assumption to get this formula
    - Fitting to the logistic function
  - # of parameters to estimate = d+1
- Who is the winner?
  - Really??? There is no winner... Why?

Together, modeling joint prob.

Unit 04 | Gaussian Naïve Bayes

4-3) 나이브 베이즈 vs 로지스틱 회귀 (naive bayes vs logistic regression)

일반적인 경우 로지스틱 회귀가 훨씬 잘 맞는다 / Logistic 승!

그러나 사전 분포 확률 즉 prior를 잘 아는 경우 나이브 베이즈가 잘 맞는다 / naive bayes 승!

문제 상황을 보고 잘 판단하여 사용 할 것!!!!!!

정리

나이브 베이즈

장점

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

단점

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

Unit 06 | 과제 및 데이터 설명

6) 과제 및 데이터 설명

과제

week3_NaiveBayes_과제(200805).ipynb를 완성해주세요!!

Reference

참고자료
- 투빅스 12기 김태한님 강의
- 투빅스 13기 김미성님 강의
- 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=h09SVW6nnhM
- https://ratsgo.github.io/statistics/2017/09/23/MLE
- https://yngie-c.github.io/machine%20learning/2020/04/08/naive_bayes/
- https://bkshin.tistory.com/entry/dd?category=1042793
- https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-1%EB%82%98%EC%9D%B4%EB%B8%8C-%EB%B2%A0%EC%9D%B4%EC%A6%88-%EB%B6%84%EB%A5%98-Naive-Bayes-Classification

Q & A

들어주셔서 감사합니다.