

13기 정규세션

ToBig's 12기 김태한

Naïve Bayes Classifier

* naive : 순진하다 Bayes : 베이즈정리

목표

[목표]

- 베이즈 정리 (Bayes' Rule)에 대하여 이해한다
- 분류 문제 (classification) 모델 중 하나인 나이브 베이즈 방법에 대하여 이해한다

예측 변수인 Y 가 0 또는 1과 같이 **두 종류로 구분**되는 분포를 가질 때,
 Y 를 예측 하는 방법 입니다

Contents

Unit 01 | MLE Review

Unit 02 | 확률 기초 (Probability Overview)

Unit 03 | 베이즈 정리 (Bayes Rule)

Unit 04 | Naïve Bayes Classification

Unit 05 | Gaussian Naïve Bayes

Unit 06 | 과제 및 데이터 설명

Unit 01 | MLE Review

1-1) 정보이론 (Information Theory)

- 정보의 양을 측정하는 응용수학의 한 갈래
- 본래 최적의 코드를 짜고 디자인하는 연구에서 출발하였다

Unit 01 | MLE Review

1-1) 정보이론 (Information Theory)

- 정보의 양을 측정하는 응용수학 분야
다음주부터 저는 투빅스를 그만두고

여러분 곁을 떠납니다 ~^

- 본래 최적의 코드를 짜고 디자인하는 연구에서 출발하였다

Unit 01 | MLE Review

1-1) 정보이론 (Information Theory)

- 뻥입니다 믿으시면 안돼요!!!

Main Idea : 잘 일어나지 않는 사건(저는 투빅스를 떠납니다~)은 자주 발생하는 사건(저는 투빅스를 계속 활동합니다~)보다 정보량이 많다(놀랍다!!)

정보의 성질 3가지

- 1) 자주 발생하는 사건은 낮은 정보량을 가진다
- 2) 덜 자주 발생하는 사건은 더 높은 정보량을 가진다
- 3) 독립사건은 추가적인 정보량을 가진다 – 정보량이 더해진다

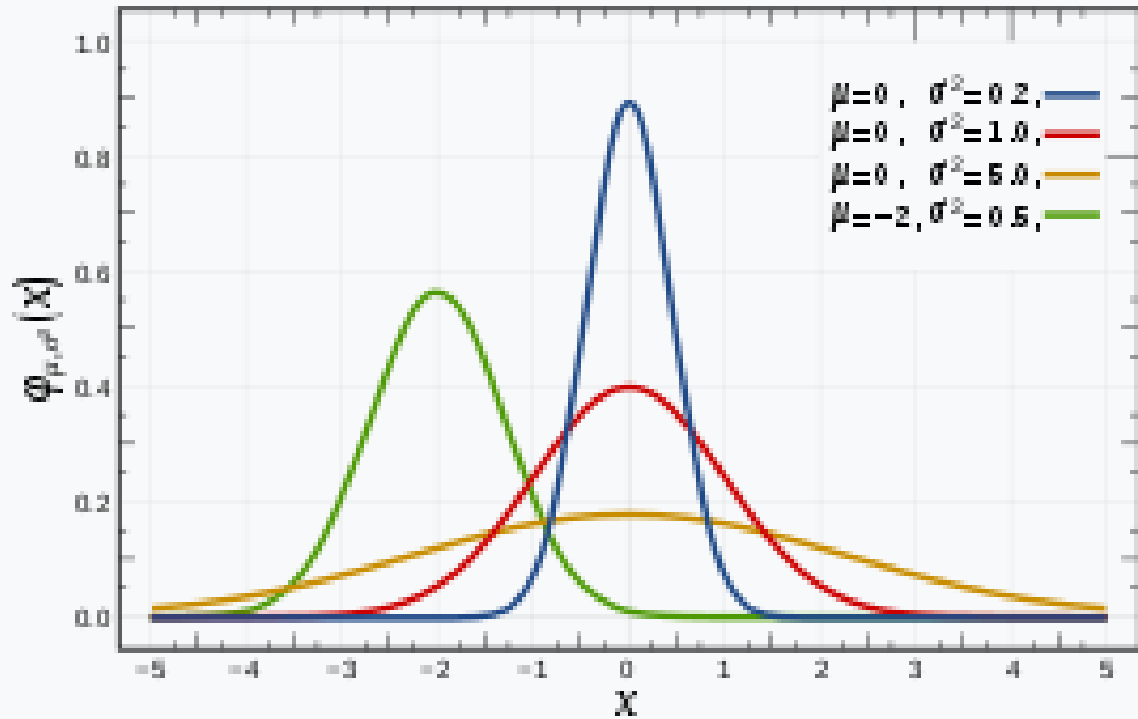
Unit 01 | MLE Review

1-1) 정보이론 (Information Theory)

- 3가지 조건을 만족하는 함수 : $I(X) = -\log P(x)$
- 여기서 log의 밑은 e가 아닌 2 !!
- Ex) $P(x_1)=0.5, P(x_2)=0.25$: $I(x_1) = 1, I(x_2) = 2$
- Ex) 독립적인 사건 x_1, x_2, x_3 가 있다 이들의 모든 정보량을 합치면
 $-\log P(x_1) - \log P(x_2) - \log P(x_3) = -\log P(x_1)P(x_2)P(x_3)$
- $P(x_1)P(x_2)P(x_3)$ 는 독립된 사건이 일어날 확률과도 일맥상통!

Unit 01 | MLE Review

1-2) 가우시안 분포 (Gaussian Distribution)



정규분포(normal distribution),
베이지안 분포(Bayesian distribution)등으로
불리는 분포

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

Unit 01 | MLE Review

1-2) 가우시안 분포 (Gaussian Distribution)

- $X = (x_1, x_2, x_3, \dots, x_n)$ 처럼 독립변수가 여러 개인 경우
- 다변수 가우시안분포 (multivariate gaussian distribution)

$$N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

Unit 01 | MLE Review

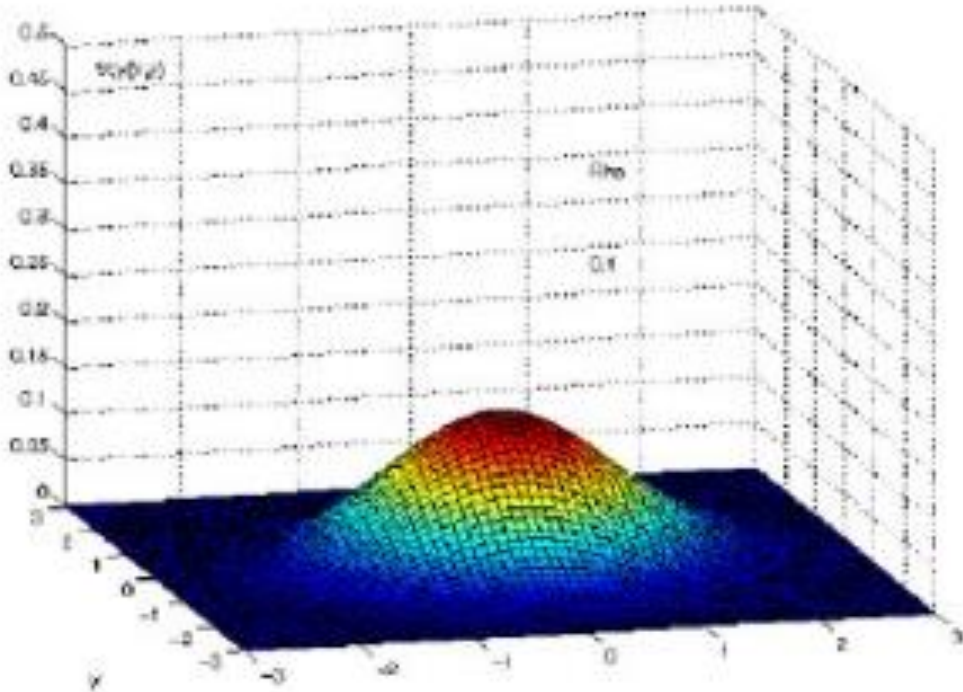
1-2) 가우시안 분포 (Gaussian Distribution)

- $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_n)$ 각 변수의 평균 벡터
- 공분산 (covariance) 행렬 : Σ
- 공분산 이란 : $Cov(x_1, x_2) = E[(x_1 - \mu_1)(x_2 - \mu_2)]$

$$\Sigma = \begin{bmatrix} Var[x_1] & Cov[x_1, x_2] & \dots & Cov[x_1, x_N] \\ Cov[x_2, x_1] & Var[x_2] & \dots & Cov[x_2, x_N] \\ \vdots & \vdots & \vdots & \dots \\ Cov[x_N, x_1] & Cov[x_N, x_2] & \dots & Var[x_N] \end{bmatrix}$$

Unit 01 | MLE Review

1-2) 가우시안 분포 (Gaussian Distribution)



$X = (x_1, x_2)$ 변수가 2개일 때 가우시안 분포

Unit 01 | MLE Review

1-3) 빈도주의통계 vs 베이지안 통계

- 한국에서 임의로 선택한 성인이 남성일 확률
- 빈도주의 : 성별이 여자 남자 니까 0.5
- 베이지안 통계 : 베르누이 분포 확률 p 가정후 sample 데이터를 통해 추정
- 베르누이 분포란 어떤 사건의 결과가 0 또는 1과같이 **2가지의** 경우로만 표현 될 수 있는 분포를 말합니다
- 베이지안 통계는 변수의 분포를 가정하고 시작 : 이때 해당 분포의 변수를 **모수**라 부른다
- Ex) 모수 - 베르누이 분포 : 확률 p / 가우시안 분포 : μ , Σ 보통 theta로 표기

Unit 01 | MLE Review

1-4) 최대우도추정법 (Maximum Likelihood Estimation)

- 1) 해당 변수가 어떤 분포를 가질지 정한다 ex) 베르누이 분포, 가우시안 분포
- 2) 샘플을 뽑는다 (보통 주어지죠 ㅎㅎ)
- 3) 해당 샘플을 최대한 잘 설명하는 모수를 찾는다
- 4) 해당 모수를 토대로 목표변수를 추정한다

Unit 01 | MLE Review

1-4) 최대우도추정법 (Maximum Likelihood Estimation)

- 3) 해당 샘플을 최대한 잘 설명하는 모수를 찾는다

3번의 과정에서 잘 설명하는 정도는 정보의 놀라움 $-\log P(x_1)P(x_2)P(x_3)\dots P(x_n)$ 를 기준으로 판단

이때 정보의 놀라움이 작아져야 한다 즉 $\log P(x_1)P(x_2)P(x_3)\dots P(x_n)$ 가 최대가 되어야 하고 이를 log-likelihood라 하며 $P(x_1)P(x_2)P(x_3)\dots P(x_n)$ 를 likelihood라고 부른다

그럼 최솟값을 찾는 알고리즘은? -> [gradient descent method](#)

이게 머신러닝의 전부 입니다!!!!!!

Unit 01 | MLE Review

1-5) 예시 (선형회귀)

$$p(y \mid \mathbf{x}) = \mathcal{N}(y; \hat{y}(\mathbf{x}; \mathbf{w}), \sigma^2)$$

$$\theta_{\text{ML}} = \arg \max_{\theta} \sum_{i=1}^m \log P(\mathbf{y}^{(i)} \mid \mathbf{x}^{(i)}; \theta).$$

$$\sum_{i=1}^m \log p(y^{(i)} \mid \mathbf{x}^{(i)}; \theta)$$

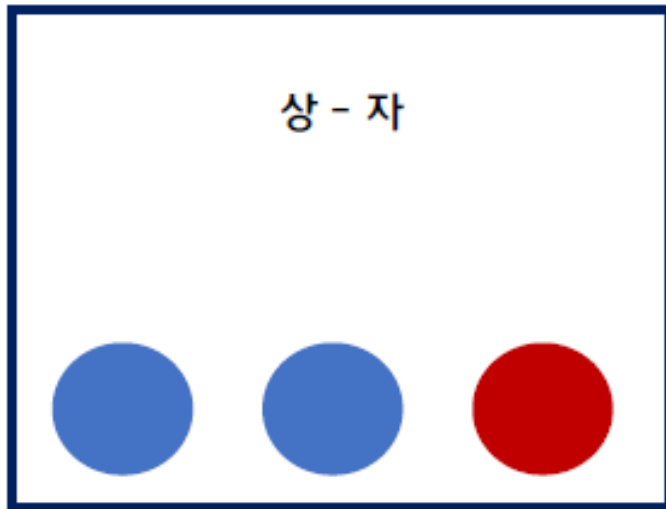
MSE는 MLE에서 유도된거다!!!!

$$= -m \log \sigma - \frac{m}{2} \log(2\pi) - \sum_{i=1}^m \frac{\|\hat{y}^{(i)} - y^{(i)}\|^2}{2\sigma^2}$$

Unit 02 | 확률 기초 (Probability Overview)

2-1) 확률이란 (probability)

- 특정 사건이 일어날 정도를 수치로 나타낸 것



파란 공을 잡을 확률 : $2/3$
빨간 공을 잡을 확률 : $1/3$

$$2/3 + 1/3 = 1$$

Unit 02 | 확률 기초 (Probability Overview)

2-2) 조건부 확률 (conditional probability)

- 어떤 사건이 일어난 조건하에서 다른 사건이 일어날 확률

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

공سم 공식

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

Unit 02 | 확률 기초 (Probability Overview)

2-3) 독립과 조건부 독립 (Independent & conditional independent)

독립 - 두 사건이 동시에 일어날 확률이 각각의 사건이 일어날 확률의 곱과 같은 경우

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

조건부 독립 - 조건부 상황에서 두 사건이 동시에 일어날 확률이 각각의 사건이 일어날 확률의 곱과 같은 경우

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

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-1) 베이즈 정리

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

The diagram shows the Bayes' Theorem formula:
$$p(H | D) = \frac{p(D | H)p(H)}{p(D)}$$
 Four yellow boxes with red text are connected to the formula by red arrows: 'Likelihood' points to $p(D | H)$, 'Prior' points to $p(H)$, 'Posterior' points to $p(H | D)$, and 'Evidence' points to $p(D)$.

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

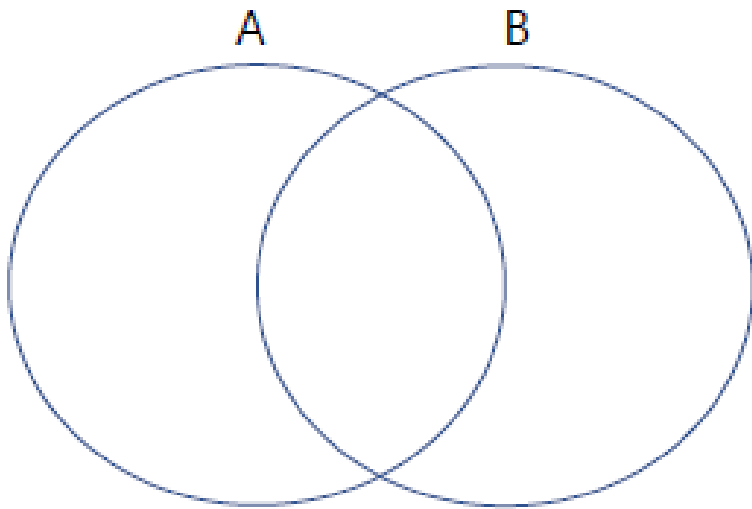
베이즈 정리는 **사전확률(prior)**로부터 **사후확률(posterior)**을 구할 수 있다.
어떻게? 조건부 확률로!

H - 알고 싶은 정보

D - 이미 알고 있는 정보

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-2) 베이즈 정리 증명



$$\begin{aligned}P(A|B) &= P(A \cap B) / P(B) \\P(A \cap B) &= P(A|B) * P(B) \\P(B \cap A) &= P(B|A) * P(A) \\P(B \cap A) &= P(A \cap B) \\P(A \cap B) &= P(B|A) * P(A) = P(A|B) * P(B) \\P(B|A) * P(A) &= P(A|B) * P(B)\end{aligned}$$



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

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

$$P(Y = yes) = 3/5$$

$$P(X = sunny | Y = yes) = 2/3$$

$$P(X = rainy | Y = yes) = 1/3$$

$$P(Y = no) = 2/5$$

$$P(X = sunny | Y = no) = 1/2$$

$$P(X = rainy | Y = no) = 1/2$$

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

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

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

by 베이즈 정리

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

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

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

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

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

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

↑
?

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

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

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

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

굳이 노란 박스를 구할 필요가 없다~!

Unit 03 | 베이즈 정리 (Bayes' Rule)

3-3) 연습

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

$$P(Y = \text{yes} | X = \text{sunny}) = 2/3$$

$$P(Y = \text{no} | X = \text{sunny}) = 1/3$$

Unit 04 | Naive Bayes Classification

4-1) 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

$$f^*(x) = \operatorname{argmax}_{Y=y} P(X = x|Y = y)P(Y = y)$$

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

$$= P(x_1 = \text{sunny}, x_2 = \text{warm}, x_3 = \text{normal}, x_4 = \text{strong}, x_5 = \text{warm}, x_6 = \text{same} | y = \text{Yes})$$

$$P(Y=y) = (y=\text{Yes})$$

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

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

Unit 04 | Naive Bayes Classification

4-1) 계산의 한계

문제점: 계산량이 많아짐

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

변수가 늘어날 수록 기하급수적으로 연산량이 증가함

어떻게 얻은 변수들인데..... d의 개수를 줄이는 건 피하고 싶어

->>해결책: 조건부 독립을 가정!!!!

Unit 04 | Naive Bayes Classification

4-2) Naive Bayes Classification

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

$$f^*(x) = \operatorname{argmax}_{Y=y} P(X = x|Y = y)P(Y = y)$$

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

$$= P(x_1 = \text{sunny}, x_2 = \text{warm}, x_3 = \text{normal}, x_4 = \text{strong}, x_5 = \text{warm}, x_6 = \text{same} | y = \text{Yes})$$

$$P(Y=y) = (y=\text{Yes})$$

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

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

Unit 04 | Naive Bayes Classification

4-2) Naive Bayes Classification

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

$$f^*(x) = \operatorname{argmax}_{Y=y} P(X = x|Y = y)P(Y = y)$$

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

$$=P(x_1=\text{sunny}|Y=\text{yes})P(x_2=\text{warm}|Y=\text{yes})P(x_3=\text{normal}|Y=\text{yes})P(x_4=\text{strong}|Y=\text{yes})P(x_5=\text{warm}|Y=\text{yes})P(x_6=\text{same}|Y=\text{yes})$$

$$P(Y=y) = (y=\text{Yes})$$

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

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

조건부 독립에 의하여 ~ !!

Unit 04 | Naive Bayes Classification

4-2) Naive Bayes Classification

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

1. 알아야할 파라미터 개수의 변화: 길이2개 -> 1개로 변함

$$P(X, Y) = P(X)(Y)$$

$$P(X = x_i | Y = y) \rightarrow (2 - 1)dk$$

2. 식의 변화

개별 피쳐 간의 곱셈으로 변함(독립이니까)

$$f^*(x) = \operatorname{argmax}_{Y=y} P(X = x | Y = y) P(Y = y) \\ \approx \operatorname{argmax}_{Y=y} P(Y = y) \prod_{1 \leq i \leq d} P(X = x_i | Y = y)$$

Unit 04 | Naive Bayes Classification

4-2) Naive 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 추정 할 수 있다는 단순한 가정을 기초로 한다
 - > 으휴 데이터 셋을 너무 믿네 순진하긴 ~ naive : 순진하다
 - > Naive Bayes !!

$$f^*(x) = \operatorname{argmax}_{Y=y} P(X = x|Y = y)P(Y = y)$$

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

Unit 04 | Naive Bayes Classification

4-3) 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 :

$$P(Y=1) = P_Y$$

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

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

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

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

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

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

Unit 04 | Naive Bayes Classification

4-3) 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	?

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

Unit 04 | Naive Bayes Classification

4-3) 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 각행의 정보를 나타내자

→ $-\log[P1*P2*(1-P3)*PY]$

→ $-\log[(1-P1)*P2*(1-P3)*PY]$

→ $-\log[P4*P5*P6*(1-PY)]$

.

.

.

Unit 04 | Naive Bayes Classification

4-3) 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	?

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

$L = -\log[P1^2 * (1-P1)^2 \dots]$: 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 04 | Naive Bayes Classification

4-3) 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...

$P1 = P(X1=1|Y=1) \Rightarrow Y=1$ 인 샘플들 중 $X1=1$ 인 경우의 비율과 같다!!

따라서 우리는 실제 적용시에는 MLE과정을 생략하고 데이터에서 개수를 세서 구한다!

Unit 04 | Naive Bayes Classification

4-3) 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 04 | Naive Bayes Classification

4-3) 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

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

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

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

Unit 04 | Naive Bayes Classification

4-3) 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

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

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

Unit 04 | Naive Bayes Classification

4-3) 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

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

$$P(Y=1) = 3/7$$

Unit 04 | Naive Bayes Classification

4-3) 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

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

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

따라서 Y=0 으로 예상된다

Unit 04 | Naive Bayes Classification

4-4) 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

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

$$P(\text{predictable} | -) = 1/14$$

$$P(\text{no} | -) = 1/14$$

$$P(\text{fun} | -) = 0/14$$

$$P(\text{predictable} | +) = 0/9$$

$$P(\text{no} | +) = 0/9$$

$$P(\text{fun} | +) = 1/9$$

Unit 04 | Naive Bayes Classification

4-4) 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 04 | Naive Bayes Classification

4-4) 라플라스 스무딩 (Laplace Smoothing)

likelihood가 0이 되는 것을 방지하도록 최소한의 확률을 정해주자!!

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

v는 입력변수들의 개수

Unit 05 | Gaussian Naïve Bayes

5-1) 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	?

지금까지는 입력변수들이 베르누이 분포!

베르누이 분포란 어떤 사건의 결과가 0 또는 1과같이 2가지의 경우로만 표현 될 수 있는 분포를 말합니다

ex) 주사위에서 홀수가 나오는 사건
홀수가 나오거나 : 0
홀수가 안나오거나 : 1

Unit 05 | Gaussian Naïve Bayes

5-1) 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 05 | Gaussian Naïve Bayes

5-1) 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??

Logistic Regression을 써도 되기는 하다만...

뭔가 Naive Bayes를 쓰고싶단 말이지...

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

WARNING

- 여기부터는 난이도가 확 증가합니다!!
- 심호흡 크게 한번 쉬세요~!

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

Idea

연속적인 입력변수들을 가우시안 분포를 가진다고 가정하여 나이브 베이즈 방법을 사용한다

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Derivation from the naïve Bayes to the logistic regression

- $P(Y) \prod_{1 \leq i \leq d} P(X_i|Y) = \pi_k \prod_{1 \leq i \leq d} \frac{1}{\sigma_k^i \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{X_i - \mu_k^i}{\sigma_k^i}\right)^2\right)$

- With naïve Bayes assumption

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

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- With naïve Bayes assumption

$$\begin{aligned}
 P(Y = y|X) &= \frac{P(X|Y = y)P(Y=y)}{P(X)} = \frac{P(X|Y = y)P(Y=y)}{P(X|Y = y)P(Y=y) + P(X|Y = n)P(Y=n)} \\
 &= \frac{P(Y = y) \prod_{1 \leq i \leq d} P(X_i|Y = y)}{P(Y = y) \prod_{1 \leq i \leq d} P(X_i|Y = y) + P(Y = n) \prod_{1 \leq i \leq d} P(X_i|Y = n)}
 \end{aligned}$$

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

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Assuming the same variable of the two classes, $\sigma_2^i = \sigma_1^i$

$$\begin{aligned}
 P(Y = y|X) &= \frac{1}{\frac{\pi_2 \prod_{1 \leq i \leq d} \frac{1}{\sigma_2^i} \exp(-\frac{1}{2}(\frac{X_i - \mu_2^i}{\sigma_2^i})^2)}{1 + \frac{\pi_1 \prod_{1 \leq i \leq d} \frac{1}{\sigma_1^i} \exp(-\frac{1}{2}(\frac{X_i - \mu_1^i}{\sigma_1^i})^2)}}} = \frac{1}{1 + \frac{\pi_1 \prod_{1 \leq i \leq d} \frac{1}{\sigma_1^i} \exp(-\frac{1}{2}(\frac{X_i - \mu_1^i}{\sigma_1^i})^2)}{\pi_2 \prod_{1 \leq i \leq d} \frac{1}{\sigma_2^i} \exp(-\frac{1}{2}(\frac{X_i - \mu_2^i}{\sigma_2^i})^2)}}} \\
 &= \frac{1}{1 + \frac{\pi_1 \exp(-\sum_{1 \leq i \leq d} \{\frac{1}{2}(\frac{X_i - \mu_1^i}{\sigma_1^i})^2\})}{\pi_2 \exp(-\sum_{1 \leq i \leq d} \{\frac{1}{2}(\frac{X_i - \mu_2^i}{\sigma_2^i})^2\})}}} \\
 &= \frac{1}{1 + \frac{\exp(-\sum_{1 \leq i \leq d} \{\frac{1}{2}(\frac{X_i - \mu_1^i}{\sigma_1^i})^2\} + \log \pi_1)}{\exp(-\sum_{1 \leq i \leq d} \{\frac{1}{2}(\frac{X_i - \mu_2^i}{\sigma_2^i})^2\} + \log \pi_2)}}}
 \end{aligned}$$

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

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

Unit 05 | Gaussian Naïve Bayes

5-2) 가우시안 나이브 베이즈 (gaussian naive bayes classification)

- Naïve Bayes classifier

- $$P(Y|X) = \frac{1}{1 + \exp(-\frac{1}{2(\sigma_1^i)^2} \sum_{1 \leq i \leq d} \{2(\mu_2^i - \mu_1^i)X_i + \mu_2^{i^2} - \mu_1^{i^2}\} + \log \pi_2 - \log \pi_1)}$$

- Assumption to get this formula

- Naïve Bayes assumption, Same variance assumption between classes
 - Gaussian distribution for $P(X|Y)$
 - Bernoulli distribution for $P(Y)$

Together, modeling joint prob.

- # of parameters to estimate = $2 \times 2 \times d + 1 = 4d + 1$
 - 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?

Unit 05 | Gaussian Naïve Bayes

5-3) 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!!!!!!

Unit 06 | 과제 및 데이터 설명

6-1) naive bayes 코드 실습

naive_bayes_code_practice.ipynb 파일을 켜주세요

Unit 06 | 과제 및 데이터 설명

6-2) 과제 및 데이터 설명

과제

`naive_bayes_code_assignment.ipynb`를 완성해주세요!!

정리

key word로 정리해 봅시다 - 이것들만 이해했으면 오늘 성공 :)

- Bayes' Rule : 사전확률을 통해 사후확률을 맞추겠다!
- 조건부 독립 : 경우의 수가 너무 많으니 줄여버리겠다!
- Naive Bayes : 조건부 독립 + Bayes' Rule
- 우도의 모수 형태에 따라 바뀐다 :
베르누이 나이브 베이즈 , 가우시안 나이브 베이즈 , 다항 나이브 베이즈

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

Reference

참고자료

- 투빅스 10기 황이은님 강의
 - 투빅스 11기 이영전님 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/>
 - 데이터 과학을 위한 통계
 - 데이터 과학 입문
 - 패턴 인식과 머신러닝
- + 내 머리

Q & A

들어주셔서 감사합니다.