

1 2 기 정 규 세 셴

ToBig's 11기 이영전

Naïve Bayesian Classifier

* 한줄 요약: Naïve하게(순진하게) 조건부 독립을 가정하여, 베이즈 정리와 비슷한 예측계산을 적용해 분류를 해보자.

contents

Unit 01 | MLE Review

Unit 02 | Probability Overview

Unit 03 | Bayes Theory

Unit 04 | Naive Bayesian Classification

Unit 05 | Gaussian Navie Bayes

Unit 01 | MLE Review

Unit 01 | MLE

한 페이지로 정리하는 MLE

한 마디로

“Parameter가 미지의 θ 인 확률분포에서 뽑은 표본(관측치)를 바탕으로 확률 분포의 모수인 θ 를 추정하는 기법을 말한다.”

Unit 01 | MLE Review

Unit 01 | MLE

한 페이지로 정리하는 MLE

“Parameter가 미지의 θ 인 확률분포에서 뽑은 표본(관측치)를 바탕으로 확률 분포의 모수인 θ 를 추정하는 기법을 말한다.”

- **Likelihood**란 이미 주어진 표본적 증거에 비추어 보았을 때, 모집단에 관해 어떠한 통계적 추정이 그럴듯한 정도를 말해주는 것을 가리킨다. 다시 말해 어떤 가설을 전제하였을 때, 그 전제하에 우리에게 주어진 증거가 얼마나 나타날 수 있는 가에 대한 정도이다.
- 동전 던지기를 100번 했을 때, 앞면이 56번 나왔다고 가정해보자. 반복적인 동전 던지기는 성공 확률이 P 인 베르누이 시행을 n 번 반복 시행할 때 성공횟수인 이항분포(binomial distribution)을 따른다.
- 우리가 알고 싶은 미지의 모수 θ 는 동전을 한 번 던졌을 때 앞면이 나올 확률 P 가 된다.
이를 위해 앞면이 나올 확률이 P 인 이항분포에서 뽑은 표본 x (성공횟수 = 앞면이 나온 횟수 = 56번)을 활용한다.

Unit 01 | MLE Review

Unit 01 | MLE

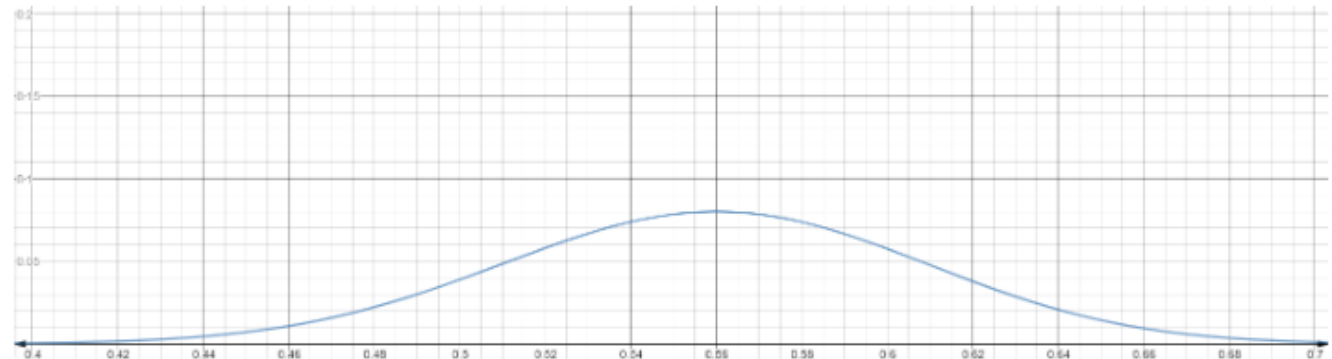
두 페이지로 정리하는 MLE

- 이항분포의 확률 함수

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- θ 값을 달리 하며 우도 계산
- 이항분포의 확률함수와 그래프는 미분가능하다
- 따라서 θ 에 대해 편미분을 해 0이 되는 지점을 구하면 우도를 최대화하는 θ 를 (MLE)를단번에 구할 수 있다.

θ	likelihood
0.48	0.0222
0.50	0.0389
0.52	0.0587
0.54	0.0739
0.56	0.0801
0.58	0.0738
0.60	0.0576
0.62	0.0378



Unit 02 | Probability Overview

Unit 01 | Overview

머신러닝의 학습 방법들

- Gradient descent based learning
- Probability theory based learning
- Information theory based learning
- Distance similarity based learning

Unit 02 | Probability Overview

Unit 01 | Overview

머신러닝의 학습 방법들

- Gradient descent based learning
- **Probability theory based learning**
- Information theory based learning
- Distance similarity based learning



오늘 배울 내용은 확률 이론 기반으로 학습하는
Naïve Bayesian Classifier입니다.

Unit 02 | Probability Overview

Unit 01 | Overview

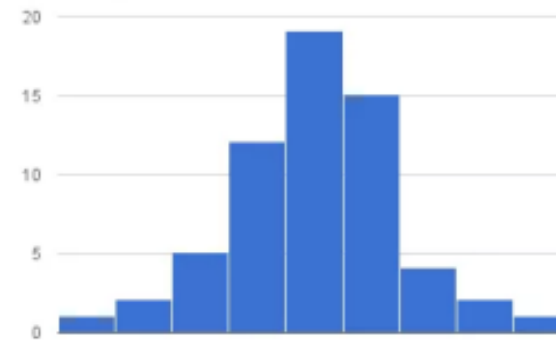
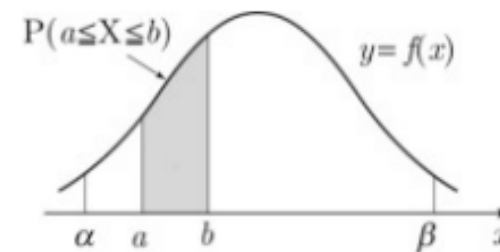
Probability

- 이산형 값

$$P(X) = \frac{\text{count}(\text{Event_}X)}{\text{count}(\text{ALL_Event})}$$

- 연속형 값

$$P(-\infty < x < \infty) \int_{-\infty}^{\infty} f(x) dx = 1$$

Source: <https://goo.gl/D9VCSL>Source: <https://goo.gl/DVNSH2>

Unit 02 | Probability Overview

Unit 01 | Overview

Basic concepts of probability

$$0 \leq P(E) \leq 1$$

$$P(S) = \sum_{i=1}^N P(E_i) = 1 \quad \text{if all } E_i \text{ are independent}$$

$$P(A \cap B) \qquad P(A \cup B) \qquad P(A^c) = 1 - P(A)$$

Unit 02 | Probability Overview

Unit 01 | Overview

Basic concepts of probability

$0 \leq P(E) \leq 1$ ← 특정 E(Event)가 일어날 확률은 0에서 1사이이다.

$P(S) = \sum_{i=1}^N P(E_i) = 1$ if all E_i are independent
← 각각의 E가 독립이라면 Probability의 Sum은 1이다.

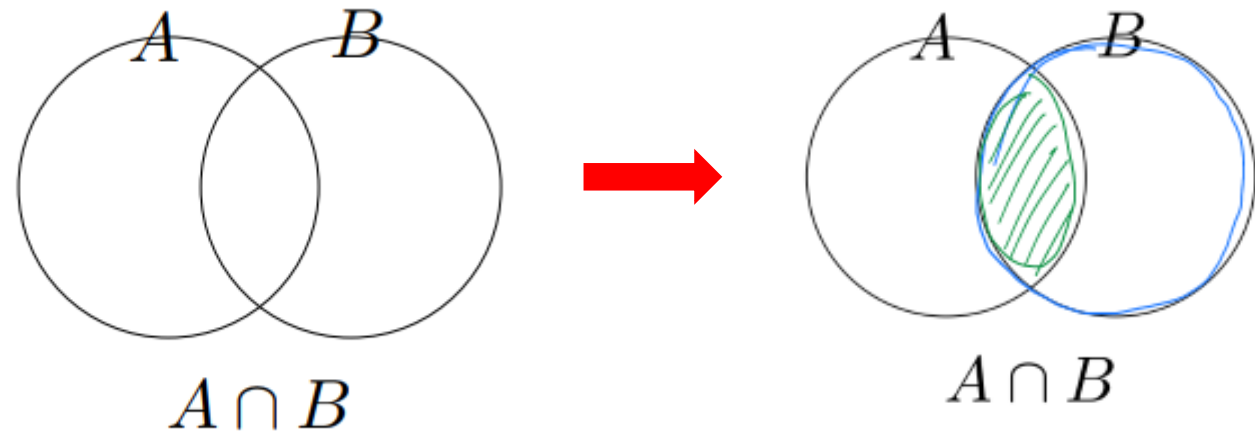
$$P(A \cap B) \quad P(A \cup B) \quad P(A^c) = 1 - P(A)$$

Unit 02 | Probability Overview

Unit 01 | Overview

Conditional Probability

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



Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

- '베이즈 정리(Bayes's Theorem)란 한마디로 조건부 확률을 '반대로 뒤집는' 것이다.

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

Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

- 경험에 의한 확률의 업데이트
- 사전 확률(given)로 부터 사건 발생을 통해 사후 확률을 구한다.
- 빈도주의 vs. 베이즈 주의

빈도주의 관점에서 확률이란

확률을 반복적으로 선택된 표본이 사건(부분집합) A의 원소가 될 경향으로 본다.

e.g 동전 앞면이 나오는 사건의 확률값
확률값이 0.5라는 것은 빈도주의 관점에서 동전을 반복하여 던졌을 경우 동전을 던진 전체 횟수에 확률 값을 곱한 숫자만큼 해당사건이 발생한다고 본다.

10,000번을 던진다면 $10,000 \times 0.5 = 5,000$

베이지안 관점에서 확률이란

선택된 표본이 특정한 사건(부분집합)에 속한다는 가설(hypothesis), 명제(proposition), 혹은 주장(assertion)의 신뢰도(degree of belief)라고 볼 수 있다.

e.g 새는 날 수 있다.
새가 날 수 있는 가능성은 95%다....

더 많은 내용이 궁금하시다면...

<https://datascienceschool.net/view-notebook/9605664e26a0411b88f60e4ba9521dd9/>

Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

- 경험에 의한 확률의 업데이트
- 사전 확률(given)로 부터 사건 발생을 통해 사후 확률을 구한다.
- 빈도주의 vs. **베이즈 주의**

빈도주의 관점에서 확률이란
확률을 반복적으로 선택된 표본이 사
건(부분집합) A의 원소가 될 경향으로
본다.

e.g 동전 앞면이 나오는 사건의 확률값
확률값이 0.5라는 것은 빈도주의 관점에서 동전을 반복하
여 던졌을 경우 동전을 던진 전체 횟수에 확률 값을 곱한
숫자만큼 해당사건이 발생한다고 본다.
10,000번을 던진다면 $10,000 \times 0.5 = 5,000$

베이지안 관점에서 확률이란

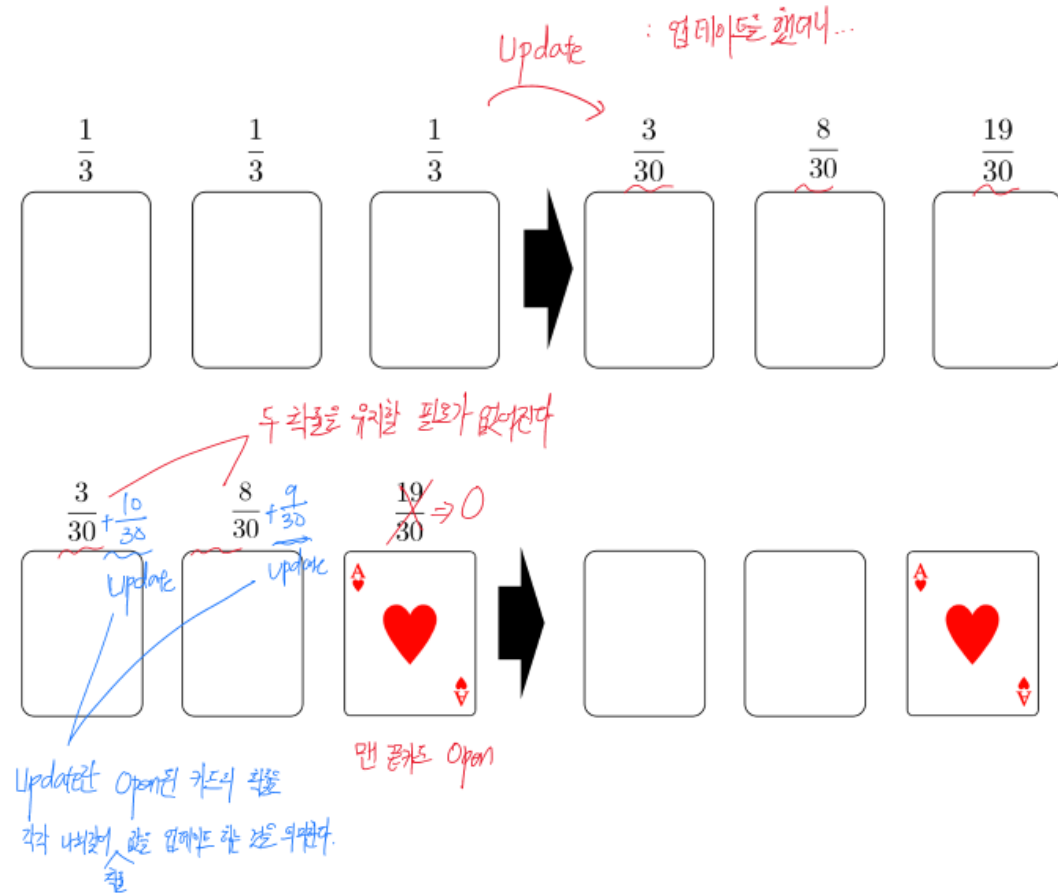
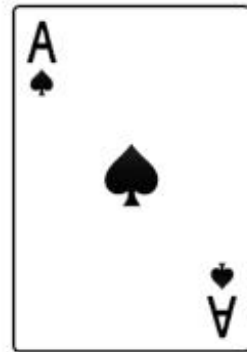
선택된 표본이 특정한 사건(부분집합)에 속한다는 가설(hypothesis),
(hypothesis) 즉 (hypothesis)의 신뢰도(degree of belief)
라고 볼 수 있다.
**객관적인 확률이란
존재하지 않는다.**

새가 날 수 있는 가능성은 95%다....

Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem



Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

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

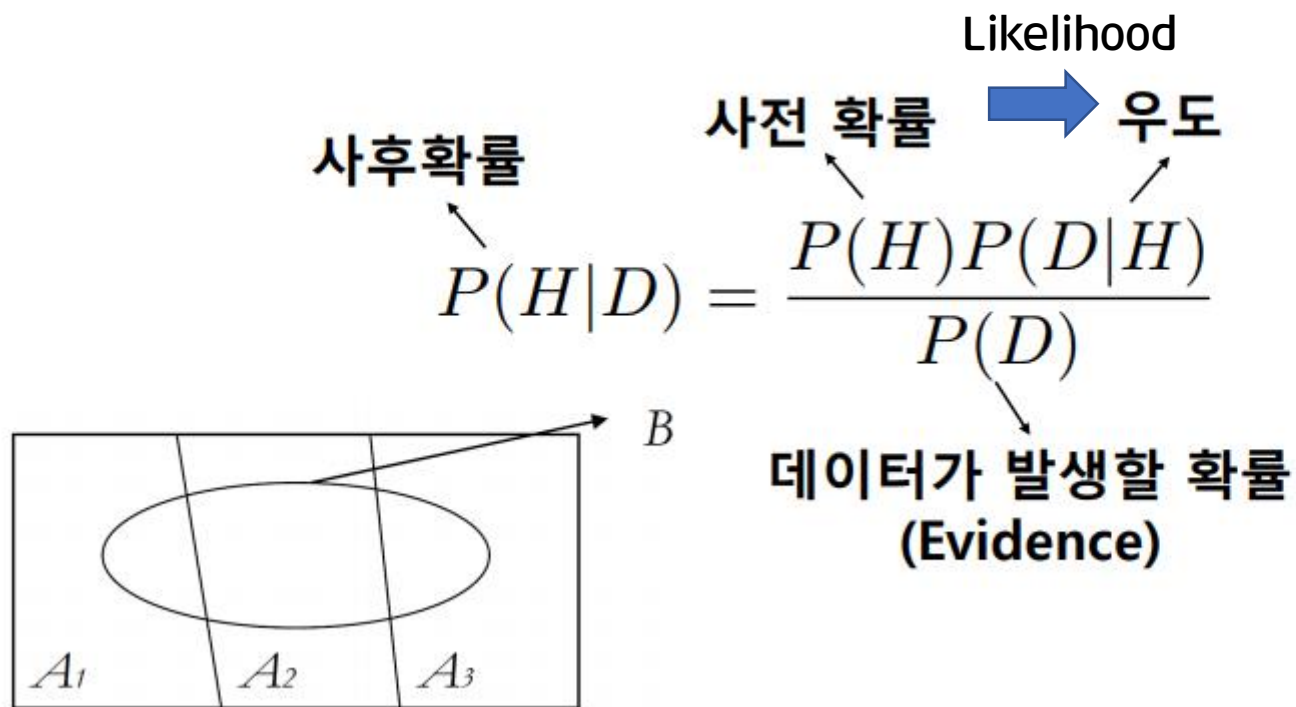
$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

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

Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

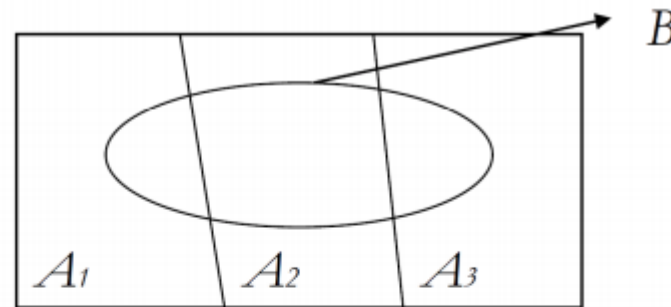
 H is Class D is Data

Unit 03 | Bayes Theory

Unit 01 | Overview

Bayes's Theorem

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$



$$P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

$$P(B) = P(A_1)P(B \cap A_1) + P(A_2)P(B \cap A_2) + P(A_3)P(B \cap A_3)$$

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

Unit 03 | Bayes Theory

Unit 01 | Overview

Quiz.

쿠키 두 그릇이 있다고 한다.

첫 번째 그릇에는 바닐라 쿠키 30개와 초콜렛 쿠키 10개가 있고,
두 번째 그릇에는 각 쿠키가 20개 씩 있다.

임의의 쿠키를 집었는데 해당 쿠키가 바닐라 쿠키이다.

이 쿠키가 그릇 1에서 나왔을 확률은?

Unit 04 | Naïve Bayesian Classifier

Unit 01 | Single Naïve Bayesian Classifier

Tutorial - Viagra Spam Filter

Viagra 스팸 필터 만들기

- Viagra라는 단어의 유무를 통해 스팸 여부 확인하는 필터 만들기
- Viagra 단어가 들어가면 무조건 스팸일까?
- 어느 정도 확률로 스팸이라고 분류할 수 있을까?

Unit 04 | Naïve Bayesian Classifier

Unit 01 | Single Naïve Bayesian Classifier

Tutorial – Viagra Spam Filter

Data Set

number	viagra	spam	number	viagra	spam
1	1	1	11	1	0
2	0	0	12	0	0
3	0	0	13	0	0
4	0	0	14	1	0
5	0	0	15	0	0
6	0	0	16	0	0
7	0	1	17	0	0
8	0	0	18	0	1
9	1	1	19	0	1
10	1	0	20	1	1

Unit 04 | Naïve Bayesian Classifier

Unit 01 | Single Naïve Bayesian Classifier

Tutorial – Viagra Spam Filter

Using Naïve Bayes

$$P(spam|viagra) = \frac{P(spam)P(viagra|spam)}{P(viagra)}$$

number	viagra	spam	number	viagra	spam
1	1	1	11	1	0
2	0	0	12	0	0
3	0	0	13	0	0
4	0	0	14	1	0
5	0	0	15	0	0
6	0	0	16	0	0
7	0	1	17	0	0
8	0	0	18	0	1
9	1	1	19	0	1
10	1	0	20	1	1

$$P(viagra \cap spam)$$

$$P(viagra) = \frac{\text{count}(viagra)}{\text{count}(ALL_{samples})}$$

$$P(H|D) = \frac{P(C)P(D|H)}{P(D)}$$

$$P(spam) = \frac{\text{count}(spam)}{\text{count}(ALL_{samples})}$$

Unit 04 | Naive Bayesian Classifier

Unit 01 | Single Naive Bayesian Classifier

Tutorial - Viagra Spam Filter

Using Naive Bayes

$$P(\text{spam}|\text{viagra}) = \frac{P(\text{spam})P(\text{viagra}|\text{spam})}{P(\text{viagra})}$$

number	viagra	spam	number	viagra	spam
1	1	1	11	1	0
2	0	0	12	0	0
3	0	0	13	0	0
4	0	0	14	1	0
5	0	0	15	0	0
6	0	0	16	0	0
7	0	1	17	0	0
8	0	0	18	0	1
9	0	0	19	0	0
10	0	0	20	0	0

필터기로 여러 단어를 검사해보고 싶다...!

$$P(\text{viagra} \cap \text{spam})$$

$$P(H|D) = \frac{P(C)P(D|H)}{P(D)}$$

$$P(\text{viagra}) = \frac{\text{count}(\text{viagra})}{\text{count}(\text{ALL}_{\text{samples}})}$$

$$P(\text{spam}) = \frac{\text{count}(\text{spam})}{\text{count}(\text{ALL}_{\text{samples}})}$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

제대로된 스팸 필터 만들기

- Viagra라는 단어 외에 영향을 주는 단어들은 무엇이 있을까?
- 오히려 스팸을 제외해주는 단어는 어떻게 찾을까?
- 한 번에 여러 단어들을 고려하는 필터기를 만들어보자

Unit 04 | Naïve Bayesian Classifier

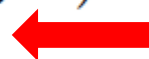
Unit 02 | Naïve Bayesian Classifier

Feature의 확장

$$P(spam|viagra)$$



$$P(spam|viagra, hello, lucky, marketing...)$$



변수가 많을 때, 조건부 확률의 변화

Q. Feature(변수)가 늘어날 때
조건부 확률값은
어떻게 변할까?

$X_1, X_2, X_3, \dots, X_n$ 일 때,

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Multivariate Multiplication Rule

$$P(Y|X_1, X_2) = \frac{P(Y \cap X_1 \cap X_2)}{P(X_1 \cap X_2)}$$

$$P(Y \cap X_1 \cap X_2) = P(Y|X_1, X_2)P(X_1 \cap X_2)$$

$$\begin{aligned} &P(X_1, X_2, X_3, \dots X_n) \\ &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots P(X_n|X_1 \dots X_{n-1}) \end{aligned}$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Multivariate Multiplication Rule

$$P(Y|X_1, X_2) = \frac{P(Y \cap X_1 \cap X_2)}{P(X_1 \cap X_2)}$$

$$P(Y \cap X_1 \cap X_2) = P(Y|X_1, X_2)P(X_1 \cap X_2)$$

$$P(X_1, X_2, X_3, \dots, X_n)$$

$$= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots P(X_n|X_1 \dots X_{n-1})$$



$X_1, X_2, X_3, \dots, X_n$ 일 때,
(Feature의 개수가 늘어날 때)

- 계산이 어려워진다.
- Feature 차원이 증가하면
Sparse Vector가 생성
즉 확률이 0이 되는 값이
늘어난다.

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Naïve Bayes Classifier 드디어 등장했다!

- 복잡하게 해결하지 말고 단순(Naïve)하게 해결하자
- 각 변수의 관계를 독립으로 가정한다.
- 계산이 용이해지고, 성능이 생각보다 좋다

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Joint Probability

$P(A \cap B) = P(A)P(B)$ if A and B are independent

$$P(Y|X_1 \cap X_2) = \frac{P(Y)P(X_1 \cap X_2|Y)}{P(X_1 \cap X_2)}$$
$$\frac{P(Y)P(X_1|Y)P(X_2|Y)}{P(X_1)P(X_2)}$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Joint Probability

$$P(Y|X_1 \cap X_2) = \frac{P(Y)P(X_1 \cap X_2|Y)}{P(X_1 \cap X_2)} = \frac{P(Y)P(X_1|Y)P(X_2|Y)}{P(X_1)P(X_2)}$$

$$P(Y_c|X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i|Y_c)}{\prod_{i=1}^n P(X_i)}$$

Y_c is a label

 $X_1, X_2, X_3, \dots, X_n$ 일 때,
(Feature의 개수가 늘어날 때)

Unit 04 | Naive Bayesian Classifier

Unit 02 | Multinomial Naive Bayesian Classifier

Joint Probability

$$P(Y|X_1 \cap X_2) = \frac{P(Y)P(X_1 \cap X_2|Y)}{P(X_1 \cap X_2)} = \frac{P(Y)P(X_1|Y)P(X_2|Y)}{P(X_1)P(X_2)}$$

하지만 Naive Bayes 방법 또한
문제가 있다.

$$P(Y_c|X_1, \dots, X_n) = \frac{\prod_{i=1}^n P(X_i|Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

← X1, X2, X3, ... Xn일 때,
(Feature의 개수가 늘어날 때)

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Issue

- 너무 많은 확률 값을 가졌을 때 발생하는 문제
→ 확률값이 0에 수렴하게 되는 문제가 발생한다.

- 1. 곱하지 말고 더하자!
→ Log를 취해주자!

$$P(Y_c|X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i|Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Issue

- 너무 많은 확률 값을 가졌을 때 발생하는 문제
→ 확률값이 0에 수렴하게 되는 문제가 발생한다.

- 1. 곱하지 말고 더하자!
→ Log를 취해주자!

$$P(Y_c|X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i|Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

$$\rightarrow \log\{P(Y_c) \prod_{i=1}^n P(X_i|Y_c)\} = \log P(Y_c) + \sum_{i=1}^n \log P(X_i|Y_c)$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Issue

- 너무 많은 확률 값을 가졌을 때 발생하는 문제
→ 확률값이 0에 수렴하게 되는 문제가 발생한다.

Likelihood



- 1. 곱하지 말고 더하자!
→ Log를 취해주자!

$$P(Y_c | X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i | Y_c)}{\prod_{i=1}^n P(X_i)}$$

Y_c is a label

← Ignore! ($\because Y_0, Y_1$)

Condisder Likelihood →

$$\log \{ P(Y_c) \prod_{i=1}^n P(X_i | Y_c) \} = \log P(Y_c) + \sum_{i=1}^n \log P(X_i | Y_c)$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Naïve Bayesian Classifier

Issue

- 너무 많은 확률 값을 가졌을 때 발생하는 문제
→ 확률값이 0에 수렴하게 되는 문제가 발생한다.

- 2. 작게나마 확률이 나올 수 있도록 변경!
→ 스무딩

스무딩을 통해 작게나마 확률 값을 얻을 수 있게 됨.
If. $K = 2$, $\text{num_class} = 2 \rightarrow \underline{1/4}$

$K = \text{parameter}$



$$P(X|Y) = \frac{\text{count}(X \cap Y) + k}{\text{count}(Y) + (k * |\text{number of class}|)}$$

← Num of class

Condisder Likelihood →

$$\log\{P(Y_c) \prod_{i=1} P(X_i|Y_c)\} = \log P(Y_c) + \sum_{i=1} \log P(X_i|Y_c)$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Multinomial Naïve Bayes

- X값이 Binary가 아니라 1이상의 값을 가지는 문제
- 일반적으로 Text 문제를 분류할 때 많이 쓰인다.
- 단어의 존재 유무가 아닌 단어의 출현 횟수를 Feature로 사용한다.

Text의 Feature 표현

문자 → Feature 변환

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

One-hot Encoding

- 하나의 단어를 Vector의 Index로 인식, 단어 존재시 1 없으면 0

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

France = [0, 0, 0, 1, 0, 0, ..., 0]

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Bag of words

- 단어별로 인덱스를 부여해서 한 문장(또는 문서)의 단어 개수를 Vector로 표현한다.

the dog is on the table

are cat dog is now on table the

	are	call	from	hello	home	how	me	money	now	tomorrow	win	you
0	1	0	0	1	0	1	0	0	0	0	0	1
1	0	0	1	0	1	0	0	1	0	0	2	0
2	0	1	0	0	0	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0	0	1	0	1

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Bag of words

- 단어별로 인덱스를 부여해서 한 문장(또는 문서)의 단어 개수를 Vector로 표현한다.

the dog is on the table

are cat dog is now on table the

	are	call	from	hello	home	how	me	money	now	tomorrow	win	you
0	1	0	0	1	0	1	0	0	0	0	0	1
1	0	0	1	0	1	0	0	1	0	0	2	0
2	0	1	0	0	0	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0	0	1	0	1

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Multinomial Naïve Bayes

$$P(Y_c|X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i|Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

- ★ 앞에서 배운 기본식을 기억하자!
- ★ Likelihood만 바뀐다.

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Multinomial Naïve Bayes

$$P(Y_c | X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i | Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

★ 앞에서 배운 기본식을 기억하자!

★ Likelihood만 바뀐다.

Likelihood



$\prod_{i=1}^n$

$$P(X_i | Y_c) = \frac{\sum t f(x_i, d \in Y_c) + \alpha}{\sum N_{d \in Y_c} + \alpha \cdot V}$$

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Multinomial Naïve Bayes

$$\prod_{i=1}^n P(X_i | Y_c) \xrightarrow{\text{Likelihood}} P(X_i | Y_c) = \frac{\sum tf(x_i, d \in Y_c) + \alpha}{\sum N_{d \in Y_c} + \alpha \cdot V}$$

- x_i : A word from the feature vector \mathbf{x} of a particular sample.
- $\sum tf(x_i, d \in y_c)$: The sum of raw term frequencies of word x_i from all documents in the training sample that belong to class y_c .
- $\sum N_{d \in y_c}$: The sum of all term frequencies in the training dataset for class y_c .
- α : An additive smoothing parameter ($\alpha=1$ for Laplace smoothing).
- V : The size of the vocabulary (number of different words in the training set).

Unit 04 | Naïve Bayesian Classifier

Unit 02 | Multinomial Naïve Bayesian Classifier

Multinomial Naïve Bayes

	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

$$P(Y_c | X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i | Y_c)}{\prod_{i=1}^n P(X_i)}$$

Likelihood

$$\rightarrow P(X_i | Y_c) = \frac{\sum tf(x_i, d \in Y_c) + \alpha}{\sum N_{d \in Y_c} + \alpha \cdot V}$$

계산해 주세요!

Test	5	Chinese Chinese Chinese Tokyo Japan	?
------	---	-------------------------------------	---

Unit 05 | Gaussian Naïve Bayes

Unit 01 | Gaussian Naïve Bayes

Gaussian Naïve Bayes

- Category 데이터가 아닌경우 NB에 적용
- Continuous 데이터의 적용을 위해 y 의 정규분포(Gaussian)으로 가정한다.
- 확률 밀도 함수 상의 해당 값 X 가 나올 확률로 NB를 구현한다.

Unit 05 | Gaussian Naïve Bayes

Unit 01 | Gaussian Naïve Bayes

Gaussian Naïve Bayes

- 마찬가지로 앞에서 배운 기본식에서 출발한다.
Likelihood만 바뀐다.

$$P(Y_c | X_1, \dots, X_n) = \frac{P(Y_c) \prod_{i=1}^n P(X_i | Y_c)}{\prod_{i=1}^n P(X_i)} \quad Y_c \text{ is a label}$$

$$P(x_i | Y_c) = \frac{1}{\sqrt{2\pi\sigma_{Y_c}^2}} \exp\left(-\frac{(x_i - \mu_{Y_c})^2}{2\sigma_{Y_c}^2}\right) \quad \leftarrow \text{특정 } Y_c \text{에 대한 } X_i \text{의 평균, 표준편차를 } \mu, \sigma \text{에 대입한다.}$$

Unit 05 | Gaussian Naïve Bayes

Unit 01 | Gaussian Naïve Bayes

Gaussian Naïve Bayes

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

← Sample은 남자일까 여자일까?

Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033*10-02	176.25	1.2292*10+02	11.25	9.1667*10-01
female	5.4175	9.7225*10-02	132.5	5.5833*10+02	7.5	1.6667

$$P(x_i | Y_c) = \frac{1}{\sqrt{2\pi\sigma_{Y_c}^2}} \exp \left(-\frac{(x_i - \mu_{Y_c})^2}{2\sigma_{Y_c}^2} \right)$$

← 특정 Yc에 대한 Xi의 평균, 표준편차를 μ, σ 에 대입한다.

Unit 05 | Gaussian Naïve Bayes

Unit 01 | Gaussian Naïve Bayes

Gaussian Naïve Bayes

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

← Sample은 남자일까 여자일까?

Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033*10-02	176.25	1.2292*10+02	11.25	9.1667*10-01
female	5.4175	9.7225*10-02	132.5	5.5833*10+02	7.5	1.6667

$$P(\text{male}) = 0.5$$

$$p(\text{height} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6 - \mu)^2}{2\sigma^2}\right) \approx 1.5789,$$

Unit 00 | 과제 공지

과제 01 | Multinomial Naïve Bayes

Multinomial Naïve Bayes 예제로 데이터에서 Test값은 몇 % 확률로
어떤 클래스로 분류될까요?
(개념을 바탕으로 직접 풀어보세요)
(제출하지 않으셔도 됩니다. 이해가 안가시면 여쭙보세요)

Unit 00 | 과제 공지

과제 01 | Naïve Bayes from scratch

실습코드 2번째 코드는 미완성의 코드입니다.
나이브 베이즈 클래시파이어를 구현 했지만 몇 가지 문제점을 가지고 있습니다.

첫 째, 너무 많은 확률값을 가지고 있습니다.
둘 째, 값이 0에 수렴할 수 있습니다.
(어떤 기법을 적용해서 해결하면 좋을까요?)

해당 문제를 코딩으로 구현하고 설명해주세요.
e.g 너무 많은 확률 값을 가지고 있기에 00을 적용했습니다.
Code ...

Q & A

들어주셔서 감사합니다.

- 참고자료
 - 본 강의는 Teamlab 가천대학교 최성철 교수님의 Naïve Bayes강의를 기반으로 만들어 졌습니다.
 - 그 외의 자료들은...
 - Rat's go blog
 - Data Science from scratch
 - http://databaser.net/moniwiki/pds/BayesianStatistic/%EB%B2%A0%EC%9D%B4%EC%A6%88_%EC%A0%95%EB%A6%AC%EC%99%80_MLE.pdf