

## Contents ⚙

1 Discrete Naïve
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve
6.1 Posterior, likelihood
6.2 Bernoulli Naïve
6.2.1 Bernoulli
6.2.2 일계조각
6.3 Bernoulli Naïve
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve
8.1 Multinomial
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조각
8.3.5 Sklearn
8.4 Multinomial
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating
9 예제: Spam Filter
10 Categorical Naïve

## Naïve Bayes

- 대표적인 generative 모형으로 target, feature, parameter 모두 확률변수로 취급한다.
- Posterior의 모수는 Bayes' rule로 계산하며, 모수는 MLE나 MAP로 추정한다.
- Feature들 사이에 조건부독립성을 가정하여 추정해야 할 모수의 수가 획기적으로 감소

$$\Pr(\boldsymbol{\theta}|X, Y) = \frac{\Pr(X, Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)} = \frac{\Pr(X_1, \dots, X_K|\boldsymbol{\theta}) \times \Pr(Y) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)}$$

- Algorithm 훈련에 필요한 시간이 극히 짧다.
- Multiclass 분석도 가능하며 discrete feature인 경우 아주 훌륭한 performance를 보인다.
- 가정에 크게 민감하지 않으며 적당한 크기의 표본으로도 상대적으로 정확한 분류가 가능하다.
- 경제학에서 모든 설명변수가 이산변수인 경우는 흔치 않지만, 설문조사 자료나 자연어처리 문제에선 흔히 볼 수 있으며 image 문제에서도 간혹 사례를 볼 수 있다. 특히 자연어처리 문제에선 표본의 크기가极大로 naïve Bayes 모형의 빠른 처리속도는 큰 장점이 된다.
- Spam filter나 감성분석, 추천 algorithm과 같이 feature의 수가 많고 표본이 큰 문제에 적합하다.
- 가정이 만족되지 않는다면 algorithm의 예측확률은 '확률'로 해석하기는 어렵다.

## 1 Discrete Naïve Bayes Classifier

- Discrete : Feature가 모두 discrete 확률변수
- Naïve : feature는 같은 class 내에서 조건부 독립
- Bayes : Bayes' rule을 이용하여 추정모형을 단순화

### 1.1 모형의 구분

- Feature로 사용하는 확률변수의 특징에 따라 다음과 같이 구분한다.
  - Bernoulli Naïve Bayes: binomial distribution,  $x_k \in \{0, 1\}$
  - Multinomial Naïve Bayes: multinomial distribution,  $x_k \in \{1, 2, \dots, m\}$ ,  $m = \sum_{k=1}^K x_k$
  - Categorical Naïve Bayes: categorical distribution,  $x_k \in \{1, \dots, J_k\}$

Model	Feature Type/Encoder	Distribution	Applications	Sklearn Algorithm
Bernoulli Naïve Bayes	Bool, Binarizer	Bernoulli	spam filter	BernoulliNB
Multinomial Naïve Bayes	Counting numbers	Binomial, Multinomial distribution	spam filter	MultinomialNB
Categorical Naïve Bayes	OrdinalEncoder	Categorical distribution	survey	CategoricalNB
Gaussian Naïve Bayes	Continuous features	Gaussian distribution	logistic regression	GaussianNB
Mixed Naïve Bayes	Mixed	Mixed	Mixed feature types	None

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bernoulli
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남성
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional)
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli NB
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial NB
8.3.2 Dirichlet prior
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 1.2 주의

- 모든 모형은 하나의 자료유형만 사용하며, 여러 유형을 복합적으로 사용하려면 통계량들을 독립적으로 계산하여 최종예측에 사용한다.
- Naïve Bayes 는 모든 feature가 독립적이라고 가정하므로 기존 code의 수정이 그리 어렵지 않다.
- Categorical distribution은 generalized Bernoulli distribution이나 multinomial distribution으로 부르기도 한다.
  - 추정량 계산에선 Bernoulli의 일반화된 분포, multinomial distribution의 특수한 경우로 볼 수 있다.
  - 하지만 실제 적용에선 categorical variable과 count variable으로 구분하여 사용한다.
- Multinomial distribution은 실험횟수가 (예를 들어 문장의 길이) 고정되어 있으며 분포를 결정하는 모수 중 하나이다.

$$\frac{m!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_K^{x_K}$$

- 실제 적용에선 term frequency를 문장의 길이로 정규화하여 사용하기도 한다.
- MultinomialNB는 count 변수에 정의되어 있지만 실수를 입력으로 사용할 수 있다.

## 2 추정방법

- 추정이나 예측을 위한 분석의 궁극적인 목적은 조건부확률분포나 예측값을 구하는 것이다.

$$\Pr(Y|X) = f(X; \theta), \quad \theta = g(X, Y)$$

$$\Pr(Y|X) = \frac{\Pr(X|Y) \Pr(Y)}{\Pr(X)}$$

- Target과 feature가 모두 이산확률변수라면 확률분포의 모수는 단순한 기술통계량이 된다.
- $Y$ 와  $X$ 가 Bernoulli 분포를 따를 때 다음 세가지 방법으로 확률분포를 계산하고 각 방법의 장단점을 비교한다. 다항분포의 경우도 확률분포를 구하는 방법은 같다.

- 조건부 확률분포
- 결합확률분포
- Bayes' rule

$$\Pr(y|x) = \frac{\Pr(x,y)}{\Pr(x)} = \frac{\Pr(x|y) \Pr(y)}{\Pr(x)}, \quad x \in \{0,1\}^K, y \in \{0,1\}$$

## Contents ☰⚙️

- ▼ 1 Discrete Naïve Bayes
    - 1.1 모형의 구분
    - 1.2 주의
  - ▼ 2 추정방법
    - 2.1 Application of Bayes' rule
    - 2.2 Posterior probability
    - 2.3 조건부 확률
    - 2.4 결합확률분포
    - 2.5 Bayes' rule
  - ▼ 3 Naïve Bayes' Rule
    - 3.1 조건부 독립
    - ▼ 3.2 Bayes' Rule
      - 3.2.1 성인 남자
      - 3.2.2 여자 청소년
    - 3.3 모형의 평가
  - ▼ 4 독립과 조건부 독립
    - 4.1 조건부 독립
    - 4.2 (unconditional) independence
    - 4.3 조건부 독립
  - ▼ 5 Discrete naïve Bayesian Model
    - 5.1 Scikit-learn Implementation
    - 5.2 모형의 특징
  - ▼ 6 Bernoulli Naïve Bayes
    - 6.1 Posterior, likelihood
    - ▼ 6.2 Bernoulli Naïve Bayes
      - 6.2.1 Bernoulli distribution
      - 6.2.2 일계조각
    - 6.3 Bernoulli Naïve Bayes
  - ▼ 6.4 MLE, MAP, Bayesian Inference
    - 6.4.1 Frequency
    - 6.4.2 Bayesian Inference
  - ▼ 7 예제: Spam Filter
    - 7.1 전처리
    - 7.2 OHE vs BoW
    - 7.3 Sklearn Preprocessor
    - 7.4 모형의 평가
  - ▼ 8 Multinomial Naïve Bayes
    - 8.1 Multinomial distribution
    - 8.2 조건부 확률
    - ▼ 8.3 조건부 확률
      - 8.3.1 multinomial distribution
      - 8.3.2 Dirichlet distribution
      - 8.3.3 Loss Function
      - 8.3.4 일계조각
      - 8.3.5 Sklearn Preprocessor
    - 8.4 Multinomial distribution
  - ▼ 8.5 Updating prior
    - 8.5.1 Sklearn Preprocessor
    - 8.5.2 updating prior
  - 9 예제: Spam Filter
    - 10 Categorical Naïve Bayes

## 2.1 Application of Naïve Bayes to Titanic dataset

- Titanic 자료에서 성별과 15세를 기준으로 한 아동/성인 여부에 따른 조건부 생존 확률을 살펴보자.
  - 조건부 독립성을 가정하지 않은 모형과 naïve Bayes 를 비교한다.

\$\$ \underbrace{P(\text{survive}; \text{vert }; x)}\_{\text{posterior probability}}

\frac{ \overbrace{P(x | w; \text{survive})}^{\text{likelihood}} }{ \overbrace{P(x | w; \text{die})}^{\text{likelihood}} }

```
\quad \times  

\overbrace{\text{P}(\text{left}(\text{survive})|\text{right})}^{\text{class prior probability}}
```

Wunderbrace{PWleft(x|Wright)}\_{\text{Wtext{predictor prior probability}}}

```

} <br>
\Pr(\text{survive}) \vert \mathbf{x}) \Pr(x_1) \Pr(x_2) \cdots \Pr(x_n) \Pr(\text{survive}) \Pr(\text{survive}) $$

```

## 2.2 Posteriori probability를 계산하는 방법

- ## 1. 조건부확률

$$\Pr(\text{died} \mid \text{sex}, \text{kid}), \quad \Pr(\text{survived} \mid \text{sex}, \text{kid})$$

두 확률의 합은 1이므로 sex 와 kid에 대한 생존확률 혹은 사망확률 중 하나를 계산하면 다른 값을 알 수 있다

따라서 조건에 해당하는 4개 state 대한 확률분포를 결정하는데는 4개의 값이 필요하며, 일반적으로  $K$  개의 feature 가 있는 경우  $2^K$  개의 모수를 추정해야 한다

- ## 2. 결합확률

$$\Pr(\text{died}, \text{sex}, \text{kid}), \quad \Pr(\text{survived}, \text{sex}, \text{kid})$$

총 8개의 state가 있으며 각 state의 확률을 더하면 1이 되므로 7개의 값이 필요하다.

일반적으로  $K$  개의 feature 가 있는 경우  $2^{K+1} - 1$  개의 모수를 추정해야 한다.

- ### 3. Bayes' rule

$$\Pr(\text{survived or not} \mid \text{sex, kid}) = \frac{\Pr(\text{sex, kid} \mid \text{survived or not}) \Pr(\text{survived or not})}{\Pr(\text{sex, kid})}$$

$\Pr(\text{survived or not})$ 에 하나,  $\Pr(\text{sex, kid} \mid \text{survived})$ 와  $\Pr(\text{sex, kid} \mid \text{died})$ 에 각각 3개, 총 7개의 모수가 필요하다.

결합률 분포와 마찬가지로 일반적인 모형에선  $2^{K+1} - 1$  개의 모수를 추정해야 한다.

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve EM
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial Naïve Bayes
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 2.3 조건부 확률을 직접 추정

- 조건부 확률  $\Pr(Y|X)$ 에서 변수의 수가 증가하면 추정해야 할 모수(확률) 역시 기하급수적으로 증가한다.
- $X$ 와  $Y$ 가 모두 boolean이라면  $\Pr(Y|X)$ 을 추정하기 모수는 다음과 같다.

$$\Pr(Y = 1|X = 1), \quad \Pr(Y = 0|X = 1), \quad \Pr(Y = 1|X = 0), \quad \Pr(Y = 0|X = 0)$$

- 이 경우엔 두 개의 제약이 있으므로 실제 추정해야 할 모수는 2개이다.
- 일반적으로 boolean 자료  $Y$ 와  $X = (x_1, x_2, \dots, x_K)$ 를 이용하여  $\Pr(Y|X)$ 을 구할 때 추정해야 할 모수의 수는  $2^K$ 개이다.
  - $X$  조합이  $2^K$ 이고  $\Pr(Y = 1|X) + \Pr(Y = 0|X) = 1$ 이므로 각 조건에 대해 하나의 확률만 추정한다.

$\Pr(\text{survived or not} | \text{sex, kid})$

('female', 'adult') ('female', 'kid') ('male', 'adult') ('male', 'kid')

$\Pr(\text{survived}   \text{sex, kid})$	<b>0.741</b>	<b>0.641</b>	<b>0.166</b>	<b>0.490</b>
$\Pr(\text{died}   \text{sex, kid})$	0.259	0.385	0.834	0.510

survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')	('All', '')
0	0.26	0.372	0.834	0.515	0.618
1	0.74	0.628	0.166	0.485	0.382

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn 활용
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial_nb
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 2.4 결합확률분포를 추정

- 결합확률분포를 이용하여 조건부확률을 계산할 수 있다.

$$\Pr(Y = 1 \cap X = 1), \quad \Pr(Y = 0 \cap X = 1), \quad \Pr(Y = 1 \cap X = 0), \quad \Pr(Y = 0 \cap X = 0)$$

하나의 제약  $\sum_X \sum_Y \Pr(X, Y) = 1$ 과 함께 결합분포를 알기 위해 추정해야 할 모수는 총 3개이다.

- 이 방법을 사용하면  $2^{K+1} - 1$ 개의 모수를 추정해야 한다. 조건부 확률을 추정하는 것 보다 약 2배의 모수를 더 추정해야 한다.
- 이 경우는 조건부확률 뿐 아니라  $X$ 의 분포를 알 수 있다.

$$\Pr(\text{survived or not} | \text{sex, kid}) = \frac{\Pr(\text{survived or not}, \text{sex, kid})}{\Pr(\text{sex, kid})}$$

survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')
survived	<b>0.234</b>	<b>0.025</b>	<b>0.099</b>	<b>0.025</b>
died	<b>0.082</b>	<b>0.015</b>	<b>0.496</b>	0.026
All	0.316	0.039	0.595	0.051

survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')
0	0.082	0.0146	0.4958	0.0261
1	0.2337	0.0245	0.0989	0.0245
All	0.3157	0.0391	0.5946	0.0506

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) independence
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn Preprocessor
8.4 Multinomial Naïve Bayes
8.5 Updating probabilities
8.5.1 Sklearn Preprocessor
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 2.5 Bayes' rule

- 베이즈정리를 이용하면 추정해야 할 모수는 분자에 포함된 확률이므로,

$$\Pr(Y|X) = \frac{\Pr(x_1, x_2, \dots, x_K|Y) \Pr(Y)}{\Pr(X)}$$

- $\Pr(X|Y)$ 에 해당하는 모수의 수는  $2 \times (2^K - 1)$ 개, 그리고  $\Pr(Y)$ 에 해당하는 모수 1개이다. 따라서 총  $2^{K+1} - 1$ 개가 된다.

- 결국 결합확률분포를 계산해야 하므로 필요한 정보의 양은 결합확률분포를 추정하는 것과 동일하다.
- 분모에 해당하는 확률은 전체 확률의 합을 1로 만들어주기 위한 scalar 이므로 별도로 추정할 필요가 없다.

$$\Pr(\text{survived or not} | \text{sex, kid}) = \frac{\Pr(\text{sex, kid} | \text{survived or not}) \Pr(\text{survived or not})}{\Pr(\text{sex, kid})}$$

	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')	('All', '')
$\Pr(\text{survived}) = 0.382$	<b>0.613</b>	<b>0.065</b>	<b>0.259</b>	0.065	1
$\Pr(\text{died}) = 0.618$	<b>0.133</b>	<b>0.024</b>	<b>0.803</b>	0.042	1
survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')	
:	-----:	-----:	-----:	-----:	-----:
0	0.1326	0.0235	0.8017	0.0421	
1	0.6124	0.0643	0.259	0.0643	
All	0.3157	0.0391	0.5946	0.0506	

survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')
0	0.1326	0.0235	0.8017	0.0421
1	0.6124	0.0643	0.259	0.0643
All	0.3157	0.0391	0.5946	0.0506

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남성
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian Estimation
6.4.1 Frequency
6.4.2 Bayesian Estimation
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs Bag-of-Words
7.3 Sklearn Preprocessing
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial Naïve Bayes
8.3.2 Dirichlet Prior
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn Implementation
8.4 Multinomial Naïve Bayes
8.5 Updating prior
8.5.1 Sklearn Implementation
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 모수 세어보기

- 모든 자료가 boolean,  $X_{ik}, Y \in \{0, 1\}$ , 일 경우 함수  $\Pr(Y|X)$ 를 결정하기 위해 계산해야 할 모수의 수는 얼마인가?

$$\Pr(Y|X) = \frac{\Pr(X|Y) \Pr(Y)}{\Pr(X)}$$

1.  $\Pr(X)$ 는 확률의 합이 1이 되도록 조정하기 위한 항이므로 무시한다. 필요하다면 law of total probability로 계산한다.

2.  $\sum_X \Pr(X|Y = 0) = 1$ 이므로  $\Pr(X_1, \dots, X_K|Y = 0)$ 은  $2^K - 1$ 개의 모수를 포함한다. 한편,

$$\Pr(X_1, \dots, X_K|Y) = \begin{pmatrix} \Pr(X_1, \dots, X_K|Y = 0) \\ \Pr(X_1, \dots, X_K|Y = 1) \end{pmatrix}$$

이므로  $\Pr(X_1, \dots, X_K|Y)$ 의 경우  $2 \times (2^K - 1)$ 의 모수를 추정해야 한다.

3.  $\Pr(Y = 1) = 1 - \Pr(Y = 0)$ 이므로  $\Pr(Y)$ 에는 하나의 모수만 있으면 된다.

따라서 이 모형은  $2^{K+1} - 1$ 개의 모수를 포함한다.

## 모수의 수를 세는 이유

### 3 Naïve Bayes' Rule

- 충분한 정보가 주어진다면 베이즈 규칙을 이용하여  $\Pr(Y|X)$ 을 계산하는 것은 어렵지 않다. 하지만 변수들의 수가 증가하면서 추정해야 할 모수의 수는 기하급수적으로 증가한다.
- Boolean attributes의 수가 20개라면 추정해야 할 모수의 수는 1백만개 정도, 40개라면 1조개가 넘게 된다. 현실적으로 충분한 크기의 표본을 구하기 어렵다.
- Bayes' rule을 적용할 때는  $X$ 의 조건부분포를 사용하며, 따라서  $X$ 의 조건부분포가 다음과 같은 만족한다면 추정할 모수의 수를 극단적으로 줄일 수 있다.

$$\Pr(X_1, X_2, \dots, X_K|Y) = \prod_{k=1}^K \Pr(X_k|Y)$$

- Titanic 자료에서 생존자나 사망자에 대해 청소년 여부와 성별은 서로 독립적이란 의미이다.

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional)
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating
9 예제: Spam Filter
10 Categorical Naïve Bayes

$$\Pr(Y|X) = \frac{\Pr(X|Y) \Pr(Y)}{\Pr(X)}$$

$$\ln \Pr(Y|X) \propto \ln \Pr(X|Y) + \ln \Pr(Y)$$

$$= \sum_{k=1}^K \ln \Pr(X_k|Y) + \ln \Pr(Y)$$

- $K = 2$ 이라면  $\Pr(X = 1|Y) + \Pr(X = 0|Y) = 1$ 이므로  $\Pr(X|Y)$ 는 두 개의 모수로 정의할 수 있다.
- Feature의 수가  $K$ 개라면 각  $Y$ 에 대한 조건부 확률  $K$ 개가 필요하므로 추정해야 할 모수는  $2K + 1$ 개이다.

sex	adult	kid	adult	kid
female	0.6124	0.0643	0.1326	0.0235
male	0.259	0.0643	0.8017	0.0421

sex	adult	kid	adult	kid
female	0.59	0.087	0.146	0.01
male	0.282	0.042	0.788	0.055

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 3.1 조건부 독립성 가정

- 조건부 독립성을 만족하지 않는다면,

$$\Pr(\text{survived or not} | \text{sex}, \text{kid}) = \frac{\Pr(\text{sex}, \text{kid} | \text{survived or not}) \Pr(\text{survived or not})}{\Pr(\text{sex}, \text{kid} | \text{survived}) + \Pr(\text{sex}, \text{kid} | \text{died})}$$

	Survived(0.382)	adult	kid	All	died	adult	kid	All
female	0.612	0.064	0.677		female	0.133	0.024	0.156
male	0.259	0.064	0.323		male	0.802	0.042	0.844
All	0.871	0.129	1	All	0.934	0.066	1	

- 조건부 독립성을 만족한다면,

$$\Pr(\text{sex}, \text{kid} | \text{survived or not}) = \Pr(\text{sex} | \text{survived or not}) \times \Pr(\text{kid} | \text{survived or not})$$

	Survived			Died			
Pr(survived)=0.382	adult	kid	All	Pr(died) = 0.618	adult	kid	All
female	0.590	0.087	0.677	female	0.146	0.010	0.156
male	0.281	0.042	0.323	male	0.788	0.056	0.844
All	0.871	0.129	1	All	0.934	0.066	1

## 3.2 Bayes' Rule을 이용한 예측

- $X'$ 에 대한  $Y$ 의 예측값을 구해보자. 즉,  $\Pr(Y'|X')$ 를 극대화하는  $Y'$ 를 구하는 문제이다.

$$\begin{aligned}\Pr(Y'|X') &= \frac{\Pr(X'_1, \dots, X'_K | Y') \Pr(Y')}{\Pr(X')} \\ &= \frac{\prod_{k=1}^K \Pr(X'_k | Y') \Pr(Y')}{\Pr(X')}\end{aligned}$$

$$Y' = \arg \max_y \prod_{k=1}^K \Pr(X'_k | Y = y) \Pr(Y = y)$$

## Contents ☰

1	Discrete Naïve Bayes
1.1	모형의 구분
1.2	주의
2	추정방법
2.1	Applicaiton of Bayes' Rule
2.2	Posteriori probability
2.3	조건부 확률
2.4	결합확률분포
2.5	Bayes' rule
3	Naïve Bayes' Rule
3.1	조건부 독립
3.2	Bayes' Rule
3.2.1	성인 남자의 생존 확률
3.2.2	여자 청소년의 생존 확률
3.3	모형의 평가
4	독립과 조건부 독립
4.1	조건부 독립
4.2	(unconditional) 독립
4.3	조건부 독립
5	Discrete naïve Estimation
5.1	Scikit-learn Implementation
5.2	모형의 특징
6	Bernoulli Naïve Bayes
6.1	Posterior, likelihood
6.2	Bernoulli Naïve Bayes
6.2.1	Bernoulli
6.2.2	일계조각
6.3	Bernoulli Naïve Bayes
6.4	MLE, MAP, Bayesian
6.4.1	Frequentist
6.4.2	Bayesian
7	예제: Spam Filter
7.1	전처리
7.2	OHE vs BoW
7.3	Sklearn Preprocessor
7.4	모형의 평가
8	Multinomial Naïve Bayes
8.1	Multinomial
8.2	조건부 확률
8.3	조건부 확률
8.3.1	multinomial
8.3.2	Dirichlet
8.3.3	Loss Function
8.3.4	일계조각
8.3.5	Sklearn
8.4	Multinomial
8.5	Updating prior
8.5.1	Sklearn
8.5.2	updating prior
9	예제: Spam Filter
10	Categorical Naïve Bayes

- 조건부 독립성을 가정한 모형과 가정하지 않은 모형의 예측에 차이가 있을 수 있다.

	survived	('female', 'adult')	('female', 'kid')	('male', 'adult')	('male', 'kid')
Survived		0.6124	0.0643	0.259	0.0643
Died		0.1326	0.0235	0.8017	0.0421
All		0.3157	0.0391	0.5946	0.0506

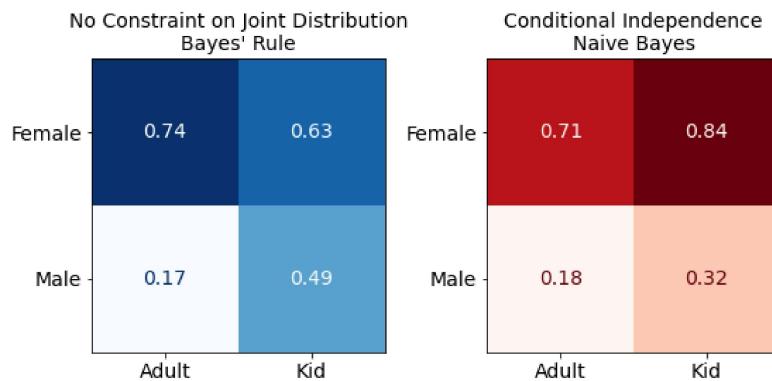
### 3.2.1 성인 남자의 생존 확률

$$\begin{aligned} \Pr(\text{survived} | \text{male} \cap \text{adult}) &= \frac{\Pr(\text{male} \cap \text{adult} \mid \text{survived}) \times \Pr(\text{survived})}{\Pr(\text{male} \cap \text{adult})} \\ &= \frac{0.259036 \times 0.381609}{0.594636} = \frac{0.098851}{0.594636} \\ &= 0.16623 \end{aligned}$$

### 3.2.2 여자 청소년의 생존 확률

$$\begin{aligned} \Pr(\text{survived} | \text{female} \cap \text{kid}) &= \frac{\Pr(\text{female} \cap \text{kid} \mid \text{survived}) \times \Pr(\text{survived})}{\Pr(\text{female} \cap \text{kid})} \\ &= \frac{0.064257 \times 0.381609}{0.039080} = \frac{0.024521}{0.039080} \\ &= 0.62745 \end{aligned}$$

Figure 1. Pr(Survived) in Titanic Dataset



## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional)
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial NB
8.3.2 Dirichlet prior
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

Figure 2. Prediction of BernoulliNB

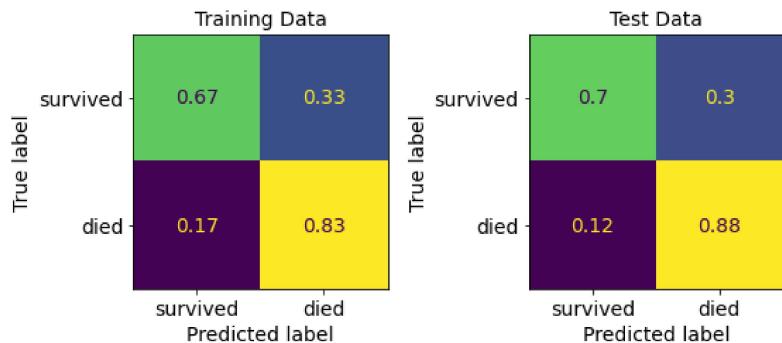


### 3.3 모형의 평가

- 다음은 BernoulliNB의 30% out-of-sample 검증결과이다.
- 이항변수로 변환한 나이와 성별만을 사용했음에도 선형확률모형이나 logistic regression에 비해 오히려 나은 점이 있다.

	precision	recall	f1-score	support
Survived	0.76	0.70	0.73	137
	0.85	0.88	0.86	255
accuracy			0.82	392
macro avg	0.80	0.79	0.79	392
weighted avg	0.81	0.82	0.81	392

Figure 3. BernoulliNB predictions



## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남성
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve EM algorithm
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial Naïve Bayes
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 4 독립과 조건부 독립

- $P(Y|X) = P(Y)$ 이라면  $X$ 는  $Y$ 에 대한 정보를 얻는데 아무런 도움이 되지 않는다.  
이때 두 사건  $X$ 와  $Y$ 는 **독립** **independent**이라고 한다.

$$P(X \cap Y) = P(Y|X)P(X) = P(Y)P(X)$$

- 다음 조건을 만족하면  $X$ 는  $Y$ 는  $Z$ 에 대해 **조건부 독립**이라고 한다.  
 $X$  is **conditionally independent** of  $Y$  given  $Z$ .

$$\Pr(X|Y \cap Z) = \Pr(X|Z)$$

- 다음도 동일한 조건이다.

$$P(X \cap Y \cap Z) = P(X|Z)P(Y|Z)\Pr(Z)$$
$$P(X \cap Y|Z) = P(X|Z)P(Y|Z)$$

### 4.1 조건부 독립 - 흔히 사용하는 복잡한 예

- 조건부 독립은 Bayes 분류모형의 핵심적인 조건이다.
- 교재나 온라인에서 많이 사용하는 조건부독립적 사건의 예이다.
  - 두 개의 동전 중 하나는 정상적인 동전으로  $\Pr(H) = 1/2$ 이다.
  - 다른 한 동전은 두 면 모두 앞면 문양만 있는 동전으로  $\Pr(H) = 1$ 이다.
  - 이 두 동전 중 한 동전을 무작위로 선택해 두 번을 던지는 실험이다.
- 사건 A: 처음 던지기에서 앞면이 나오는 사건
- 사건 B: 두번째 던지기에서 앞면이 나오는 사건
- 사건 C: 선택한 동전이 정상적인 동전인 사건

## Contents

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial
8.3.2 Dirichlet
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

- 독립적은 아니지만 조건부 독립적인 사건임을 보이기 위해서는 다음 관계를 확인한다.

$$\Pr(A \cap B|C) = \Pr(A|C) \Pr(B|C), \quad \Pr(A \cap B) \neq \Pr(A) \Pr(B)$$

- 정상적인 동전의 경우 두 번의 시행에서 앞면이 나올 확률은 독립적이므로  $\Pr(A \cap B|C) = \Pr(A|C) \Pr(B|C)$  이다.
- 두 번째 조건을 확인하기 위해 전확률의 법칙을 이용하면

$$\Pr(A) = \Pr(A|C) \Pr(C) + \Pr(A|C^C) \Pr(C^C) = \frac{1}{2} \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{3}{4} = \Pr(B)$$

- 이는 사전 정보가 없을 때 앞면이 3개, 뒷면이 1개인 '동전'에서 앞면이 나올 확률이다. 마찬가지로

$$\Pr(A \cap B) = \Pr(A \cap B|C) \Pr(C) + \Pr(A \cap B|C^C) \Pr(C^C)$$

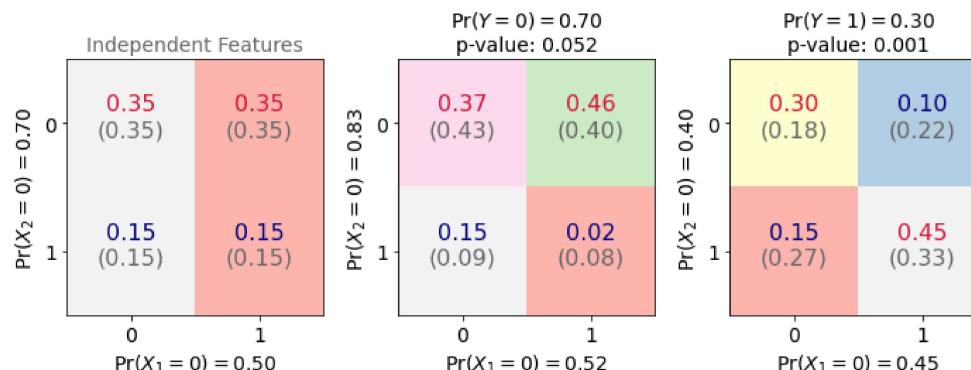
- 선택한 동전이 정상적인 동전이라면 앞면이 나올 확률이  $1/2$  이므로  $\Pr(A \cap B|C) = 1/4$
- 선택한 동전이 이상한 동전이라면 앞면이 나올 확률이 1이고 따라서  $\Pr(A \cap B|C^C) = 1$

$$\Pr(A \cap B) = \frac{1}{4} \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{5}{8}$$

## 4.2 (unconditional) independence vs. conditional independence

- 하나의 조건이 다른 조건을 의미하지 않는다.
- Naïve Bayes 모형에서 필요한 조건은 조건부 독립성이다.
- 독립성 검증  
이산변수의 경우 독립성 검증에는 Chi-square 검증방법을 주로 사용한다.  
반면 연속변수의 경우 마땅한 독립성 검증방법이 없으므로 확인이 어렵다.

Figure 4. Independent but Not conditionally Independent, n=50



## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자 여자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) independence
4.3 조건부 독립
5 Discrete naïve EM algorithm
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial Naïve Bayes
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 4.3 조건부 독립성의 검증

- 이산화된 변수의 독립성 검증에는 Chi-Square Test를 주로 사용한다.

$H_0 : X \text{ and } Y \text{ are independent.}$

$$\sum_{i,j} \frac{(observed - expected)^2}{expected} \sim \chi^2_{(df)}$$

여기서  $df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$

- Titanic dataset에선 성인 여부와 성별은 독립적이지만 생존여부에 대한 조건부 독립으로 보기 어렵다.

=====

unconditional independence test

kid adult kid

sex

female 412 51

male 776 66

p-value 0.0578

=====

conditional independence test for the survived

kid adult kid

sex

female 305 32

male 129 32

p-value 0.0016

=====

conditional independence test for the died

kid adult kid

sex

female 107 19

male 647 34

p-value 0.0002

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남성
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) independence
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli distribution
6.2.2 일계조각
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial distribution
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조각
8.3.5 Sklearn Preprocessor
8.4 Multinomial distribution
8.5 Updating prior
8.5.1 Sklearn Preprocessor
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 5 Discrete naïve Bayes algorithms

- 다양한 discrete naïve Bayes algorithms이 존재하며 feature의 형태 혹은 분포함수에 대한 가정에 따라 구분한다.
- Feature의 분포는 Bernoulli, Multinomial, Categorical 등을 가정하며 ComplementNB는 Multinomial distribution 모형을 수정한 것이다.
- Bernoulli distribution의 prior로는 Beta distribution, 나머지 분포들은 Dirichlet distribution을 사용한다.

### 5.1 Scikit-learn Naïve Bayes Algorithms

```
sklearn.naive_bayes.BernoulliNB(*, alpha=1.0, binarize=0.0, fit_prior=True, class_prior=None)
sklearn.naive_bayes.MultinomialNB(*, alpha=1.0, fit_prior=True, class_prior=None)
sklearn.naive_bayes.ComplementNB(*, alpha=1.0, fit_prior=True, class_prior=None, norm=False)
sklearn.naive_bayes.CategoricalNB(*, alpha=1.0, fit_prior=True, class_prior=None, min_categories=None)
```

- alpha: float, default=1.0  
Additive (Laplace/Lidstone) smoothing parameter (0 for no smoothing).
- binarize: float or None, default=0.0  
Threshold for binarizing (mapping to booleans) of sample features
- alpha=0, fit\_prior=False, 그리고 class\_prior 를 target의 표본분포로 지정하면 모수의 계산에 모수에 대한 prior 정보를 사용하지 않는다.

### 5.2 모형의 특징

- MultinomialNB 는 가장 기본이 되는 모형으로 모든 유형의 discrete features 뿐 아니라 연속변수도 사용할 수 있다.

The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. For non-integer counts, practice, **fractional counts** such as tf-idf may also work.

- 다른 모형들은 MultinomialNB의 변형으로 볼 수 있으며 documentation에선 다른 모형을 사용할 때도 MultinomialNB를 같이 추정하여 비교할 것을 권하고 있다.
- BernoulliNB 은 짧은 문장 분석에 효과가 제일 좋을 수도 있다.
- BernoulliNB에선 training set에 포함되지 않은 feature들이 있을 경우 해당 표본은 fit을 나쁘게 하지만, MultinomialNB에선 그런 표본들을 그냥 무시해 버린다.
- ComplementNB 는 자연어 분석에서 공통적으로 나타나는 단어의 비대칭적인 빈도수, 길이, imbalance를 고려한 모형으로 MultinomialNB와 유사하지만 unbalanced datasets에 상대적으로 더 나은 성과를 냈다.
- CategoricalNB 는 비교적 최근에 포함된 분포로 상대적인 performance에 대해선 그리 알려져 있지 않다. Categorical distribution (a generalized Bernoulli distribution, multinomial distribution)을 사용한다.
- Categorical distribution은 multinomial distribution  $(x_1, \dots, x_k; p_1, \dots, p_k)$ 에서 시행회수가 1인 분포

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Estimation
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood, prior
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli 분포
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian Estimation
6.4.1 Frequency Estimation
6.4.2 Bayesian Estimation
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessing
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial Naïve Bayes
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn Implementation
8.4 Multinomial Naïve Bayes
8.5 Updating prior
8.5.1 Sklearn Implementation
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 6 Bernoulli Naïve Bayes classifier

### 6.1 Posterior, likelihood, prior

- 모수  $\theta$ 의 posterior는 다음과 같다.

$$\begin{aligned}\Pr(\boldsymbol{\theta}|X, Y) &= \frac{\Pr(X, Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)} = \frac{\Pr(X|\boldsymbol{\theta}) \times \Pr(Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)} \\ \Pr(\boldsymbol{\theta}|X, Y) &\propto \Pr(X, Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta}) \\ &= \underbrace{\Pr(X|Y, \boldsymbol{\theta}) \times \Pr(Y)}_{\text{Likelihood}} \times \Pr(\boldsymbol{\theta})\end{aligned}$$

### 6.2 Bernoulli Naïve Bayes estimator

- Naïve Bayes estimator는 MAP이다.
- $\Pr(Y)$ 에 대한 정보는  $\boldsymbol{\theta}$  추정과정에서 사용하지 않으므로  $\hat{\boldsymbol{\theta}}_{\text{MAP}}$ 의 유도과정에선  $Y = c$ 에 대한 조건부분포를 사용하고 하침자  $c$ 로 표시한다.

$$\begin{aligned}\arg \max_{\boldsymbol{\theta}} \Pr(\boldsymbol{\theta}|X, Y) &= \arg \max_{\boldsymbol{\theta}} \Pr(X|Y, \boldsymbol{\theta}) \times \Pr(Y) \times \Pr(\boldsymbol{\theta}) \\ &= \arg \max_{\boldsymbol{\theta}} \Pr(X|Y, \boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta}) \\ \Pr(\boldsymbol{\theta}_c|X_c) &= \frac{\Pr(X_c|\boldsymbol{\theta}_c) \times \Pr(\boldsymbol{\theta}_c)}{\Pr(X_c)} = \frac{\Pr(X_c|\boldsymbol{\theta}_c) \times \Pr(\boldsymbol{\theta}_c)}{\Pr(X_c)}\end{aligned}$$

- Bernoulli 분포의 pdf는  $\theta_{k|c}^x (1 - \theta_{k|c})^{1-x}$ 으로

$$\Pr(\boldsymbol{\theta}_c|X_c) \propto \Pr(X_c|\boldsymbol{\theta}_c) \Pr(\boldsymbol{\theta}_c) = \prod_{k=1}^K \left( \Pr(\theta_{k|c}) \prod_{i=1}^n \left( (1 - \theta_{k|c})^{(1-x_{ik})} \times \theta_{k|c}^{x_{ik}} \right) \right)$$

## Contents ⚙

1	Discrete Naïve Bayes
1.1	모형의 구분
1.2	주의
2	추정방법
2.1	Applicaiton of Bayes' rule
2.2	Posteriori probability
2.3	조건부 확률
2.4	결합확률분포
2.5	Bayes' rule
3	Naïve Bayes' Rule
3.1	조건부 독립
3.2	Bayes' Rule
3.2.1	성인 남자
3.2.2	여자 청소년
3.3	모형의 평가
4	독립과 조건부 독립
4.1	조건부 독립
4.2	(unconditional) independence
4.3	조건부 독립
5	Discrete naïve EM algorithm
5.1	Scikit-learn Implementation
5.2	모형의 특징
6	Bernoulli Naïve Bayes
6.1	Posterior, likelihood
6.2	Bernoulli Naïve Bayes
6.2.1	Bernoulli Naïve Bayes
6.2.2	일계조건과 추정량
6.3	Bernoulli Naïve Bayes
6.4	MLE, MAP, Bayesian
6.4.1	Frequentist approach
6.4.2	Bayesian approach
7	예제: Spam Filter
7.1	전처리
7.2	OHE vs BoW
7.3	Sklearn Preprocessor
7.4	모형의 평가
8	Multinomial Naïve Bayes
8.1	Multinomial Naïve Bayes
8.2	조건부 확률
8.3	조건부 확률
8.3.1	multinomial distribution
8.3.2	Dirichlet distribution
8.3.3	Loss Function
8.3.4	일계조건과 추정량
8.3.5	Sklearn implementation
8.4	Multinomial Naïve Bayes
8.5	Updating prior
8.5.1	Sklearn implementation
8.5.2	update function
9	예제: Spam Filter
10	Categorical Naïve Bayes

### 6.2.1 Bernoulli distribution과 Beta prior

- Prior  $\Pr(\theta_{k|c})$ 로 Bernoulli distribution의 conjugate prior distribution인  $\text{Beta}(\theta_{k|c}; \alpha_{k|c}, \beta_{k|c}) = \text{Beta}(\theta_{k|c}; \alpha_c, \beta_c)$ 를 가정하자.

$$\Pr(\boldsymbol{\theta}_c | X_c) = \prod_{k=1}^K \left[ \text{Beta}(\theta_{k|c}; \alpha_c, \beta_c) \prod_{i=1}^{n_c} \theta_{k|c}^{x_{ik|c}} (1 - \theta_{k|c})^{(1-x_{ik|c})} \right]$$

- $\alpha, \beta > 1$ 이고  $0 < \theta < 1$ 이라면 beta distribution은 다음과 같다.

$$\text{Beta}(\theta_{k|c}; \alpha_c, \beta_c) = \frac{(\alpha_c + \beta_c - 1)!}{(\alpha_c - 1)!(\beta_c - 1)!} \theta_{k|c}^{\alpha_c - 1} (1 - \theta_{k|c})^{\beta_c - 1}$$

- $n_c$ 를  $y_i = c$  표본의 수라고 하면 posterior는 다음과 같이 정리할 수 있다.

$$\begin{aligned} \ln \Pr(\boldsymbol{\theta}_c | X_c) &= \text{const} + \sum_{k=1}^K \left[ (\alpha_c - 1) \ln \theta_{k|c} + (\beta_c - 1) \ln (1 - \theta_{k|c}) + \sum_{i \in \text{class } c} \{x_{ik|c} \ln \theta_{k|c} + (1 - x_{ik|c}) \ln (1 - \theta_{k|c})\} \right] \\ &= \text{const} + \sum_{k=1}^K \left[ \left( \sum_{i=1}^{n_c} x_{ik|c} + \alpha_c - 1 \right) \ln \theta_{k|c} + \left( n_c - \sum_{i=1}^{n_c} x_{ik|c} + \beta_c - 1 \right) \ln (1 - \theta_{k|c}) \right] \end{aligned}$$

### 6.2.2 일계조건과 추정량

$$\begin{aligned} \frac{\sum_{i=1}^{n_c} x_{ik|c} + \alpha_c - 1}{\hat{\theta}_{k|c}} &= \frac{n_c - \sum_{i=1}^{n_c} x_{ik|c} + \beta_c - 1}{1 - \hat{\theta}_{k|c}} \\ \left( n_c - \sum_{i=1}^{n_c} x_{ik|c} + \beta_c - 1 + \sum_{i=1}^{n_c} x_{ik|c} + \alpha_c - 1 \right) \hat{\theta}_{k|c} &= \sum_{i=1}^{n_c} x_{ik|c} + \alpha_c - 1 \end{aligned}$$

- 따라서 추정량은 다음과 같이 계산할 수 있다.

$$\hat{\theta}_{k|c} = \frac{\sum_{i=1}^{n_c} x_{ik|c} + (\alpha_c - 1)}{n_c + (\alpha_c + \beta_c - 2)} = \frac{\sum_{i=1}^n 1(y_i = c)x_{ik} + (\alpha_c - 1)}{\sum_{k=1}^K \sum_{i=1}^n 1(y_i = c)x_{ik} + (\alpha_c + \beta_c - 2)}$$

- 여기서  $\sum_{i=1}^n x_{ik|c}$ 은 class  $c$  중에서  $x_{ik} = 1$ 인 관찰값의 수이므로 표본에서 필요한 정보는 각 class  $c$ 와 feature  $k$ 에 해당하는  $x_{ik|c} = 1$ 인 표본의 수이다. 총  $C \times K$  개의 모수를 계산해야 한다.
- Closed form solution으로 연산속도와 update에 장점이 있다.
- Sklearn의 naïve Bayes 모형에선 prior의 분포와 관계없이 adaptive smoothing을 사용한다.

$$\frac{\sum_{i=1}^{n_c} x_{ik|c} + \alpha}{n_c + K\alpha_c}$$

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) 독립
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, threshold
6.4.1 Frequentist
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial distribution
8.3.2 Dirichlet distribution
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating prior
8.5.1 Sklearn
8.5.2 updating prior
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 6.3 Bernoulli Naïve Bayes classifier

- Bernoulli Naïve Bayes classifier는 모든 feature들이 조건부 Bernoulli 분포를 따른다고 가정한다.
- Bernoulli 분포의 pdf는  $\theta_{k|c}^x (1 - \theta_{k|c})^{1-x}$  이므로

$$\Pr(Y|X, \boldsymbol{\theta}) \propto \Pr(X|Y, \boldsymbol{\theta}) \Pr(Y) = \pi_c \times \prod_{k=1}^K \theta_{k|c}^{x_k} (1 - \theta_{k|c})^{(1-x_k)}$$

classifier는 다음과 같다.

$$y^{\text{pred}}(x^{\text{new}}) = \arg \max_c \left( \ln \pi_c + \sum_{k=1}^K \left[ x_k^{\text{new}} \ln \hat{\theta}_{k|c} + (1 - x_k^{\text{new}}) \ln (1 - \hat{\theta}_{k|c}) \right] \right)$$

## 6.4 MLE, MAP, MLE, threshold

### 6.4.1 Frequentist

- Frequentist의 방법에선  $X$ 와  $\theta$ 는 확률변수로 취급하지 않는다 (혹은 분산이 0인 확률변수).
- Bayes formula에서 왼쪽항은 우리가 알 수 없는 영역이므로 특별한 의미를 갖지 않는다

$$\Pr(\theta|Y, X) = \frac{\Pr(Y|X, \theta) \Pr(\theta)}{\Pr(Y)}$$

- Likelihood 인  $\Pr(Y|X, \theta) (= L(\theta; X, Y))$ 을 극대화하는  $\theta$ 를 추정하고, 이를 예측에도 사용한다.

$$\Pr(Y = c|X, \theta) = \sum_{i=1}^n \sum_{k=1}^K \left[ x_{ik|c} \ln \theta_{k|c} + (1 - x_{ik|c}) \ln (1 - \theta_{k|c}) \right]$$
$$y^{\text{pred}} = \Pr(y = c|x, \hat{\theta}) \left( = \arg \max_c L(\hat{\theta}_c; x^{\text{new}}) \right)$$

- 따라서 최적 예측은  $(x, \hat{\theta})$ 에 대해 발생확률이 가장 큰 class이므로 이항분류문제에선 0.5를 threshold로 사용한다.

## Contents ⚙

1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayesian rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (unconditional) independence
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter - BernoulliNB
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocessor
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부확률
8.3 조건부확률
8.3.1 multinomial NB
8.3.2 Dirichlet prior
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn
8.5.2 updating probabilities
9 예제: Spam Filter - Categorical Naïve Bayes

### 6.4.2 Bayesian

- Bayesian 방법은 자료와 모수 전체의 결합확률분포를 추정하므로 추정과 예측은 다음 posterior의 극대화 문제의 해로 구한다.

$$\Pr(\theta|X, Y) = \frac{\Pr(X|Y, \theta) \Pr(Y) \Pr(\theta)}{\Pr(X, Y)}$$

$$\Pr(Y|X, \theta) = \frac{\Pr(X|Y, \theta) \Pr(Y)}{\Pr(X)}$$

- Bernoulli 분포의 pdf는  $\theta_{k|c}^x (1 - \theta_{k|c})^{1-x}$  이므로

$$\Pr(Y = c|X, \theta_c) \propto \Pr(X|\theta_c) \Pr(Y = c) = \pi_c \times \prod_{i=1}^n \prod_{k=1}^K \theta_{ik|c}^{x_{ik|c}} (1 - \theta_{ik|c})^{(1-x_{ik|c})}$$

- 따라서 예측은 다음과 같다.

$$y^{\text{pred}} = \arg \max_c \left( \ln \pi_c + \sum_{i=1}^n \sum_{k=1}^K \left[ x_{ik|c}^{\text{new}} \ln \hat{\theta}_{ik|c} + (1 - x_{ik|c}^{\text{new}}) \ln (1 - \hat{\theta}_{ik|c}) \right] \right)$$

- 이 값은 posterior  $\Pr(\theta|X, Y)$ 의 likelihood에 해당하는 값이고, 앞서 logistic regression에서 설명한 바와 같이 예측오차를 극소화하는 값이 된다. 하지만 예측에선  $\ln \pi_c$ 가 상대적인 크기에 영향을 미치므로 제에서 threshold는 0.5와는 다르다.
- 예측에 대한 posterior에 대응하는 likelihood  $\Pr(Y = c|X, \theta_c)$ 를 극대화하도록  $\theta_c$ 를 결정하면 목적함수는 logistic regression과 다르지만 동일한 해를 갖는다.
- 예측을 위해 필요한 총 모수의 수는 모수추정에  $(C \times K)$ 개, 그리고 각 class의 빈도수  $C$ 개가 필요하므로 총  $(C \times K + C)$ 개의 값을 계산해야 한다.

## 7 예제: Spam Filter - BernoulliNB

- 이 자료는 4,825의 ham과 747개의 spam 등 총 5,572개의 SMS message로 구성
- Grumbletext Web site와 NUS SMS Corpus 등에서 모은 자료로 상당히 unbalanced 되어 있다.
- Kaggle (<https://www.kaggle.com/uciml/sms-spam-collection-dataset>), UCI repository (<https://archive.ics.uci.edu/ml/datasets/sms+spam+collection>)에서 내려받는다.
- Spam 문자는 ham에 비해 상대적으로 spam에서 자주 나타나는 단어들의 빈도로 식별가능
- 간단한 전처리 후 분석에 사용한 '단어'의 수는 총 4,659개로 각 단어들이 개별 feature가 된다.
- Feature 생성에는 일반적으로 빈도수를 고려한 bag of words를 많이 사용하지만 BernoulliNB는 단어들의 출현 여부만을 고려하므로 OneHotEncoding을 적용한 것과 동일하다.로 분석
- 일반적인 분석 단계 (<https://www.kdnuggets.com/2017/03/email-spam-filtering-an-implementation-with-python-and-scikit-learn.html>)
  - 전처리
    - 맞춤법, 띄어쓰기, 단어추출/분리, 문장부호 및 기호 제거, 원형으로 변환, 불용어 제거
  - Feature extraction
    - OneHotEncoe, Bag of Words, TF-IDF(Term Frequency-Inverse Document Frequency), Word2Vec (unknown 처리가능)
  - 분류기 선택
    - Logistic regression, discrete Naïve Bayes, SVM, decision tree

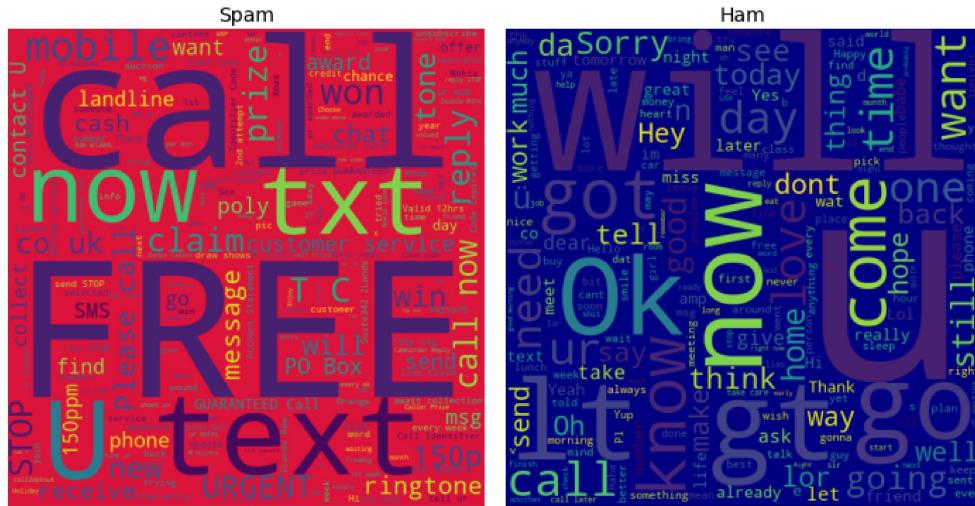
Contents ☰

- |     |                |      |      |
|-----|----------------|------|------|
| ▼ 1 | Discrete Naïve | 5567 | spam |
| 1.1 | 모형의 구분         | 5568 | ham  |
| 1.2 | 주의             | 5569 | ham  |
| ▼ 2 | 주정방법           | 5570 | ham  |
|     |                | 5571 | ham  |

label w  
5567 spam  
5568 ham  
5569 ham  
5570 ham  
5571 ham

message  
5567 This is the 2nd time we have tried 2 contact u. U have won the £750 Pound prize. 2 claim is easy, call 087187272008 NOW! Only 10p per minute. BT-national-rate.  
5568 Will u b going to esplanade fr home?  
5569 Pity, \* was in mood for that. So...any other suggestions?  
5570 The guy did some bitching but I acted like i'd be interested in buying something else next week and he gave it to us for free  
5571 Rofl. Its true to its name

Figure 5. WordCloud



## Contents ⚙

1 Discrete Naïve Bayes	1.1 모형의 구분
1.2 주의	
2 추정방법	2.1 Application of Bayes' Rule
2.2 Posterior probability	2.3 조건부 확률
2.4 결합확률분포	2.5 Bayes' rule
3 Naïve Bayes' Rule	3.1 조건부 독립
3.2 Bayes' Rule	3.2.1 성인 남자
3.3 모형의 평가	3.2.2 여자 청소년
4 독립과 조건부 독립	4.1 조건부 독립
4.2 (unconditional)	4.3 조건부 독립
5 Discrete naïve Bayes	5.1 Scikit-learn Implementation
5.2 모형의 특징	
6 Bernoulli Naïve Bayes	6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes	6.2.1 Bernoulli Rule
6.3 Bernoulli Naïve Bayes	6.2.2 일계조합
6.4 MLE, MAP, Bayesian	6.4.1 Frequency
6.4.2 Bayesian	
7 예제: Spam Filter	7.1 전처리
7.2 OHE vs BoW	7.2 OHE vs BoW
7.3 Sklearn Preprocessing	7.4 모형의 평가
8 Multinomial Naïve Bayes	8.1 Multinomial Naïve Bayes
8.2 조건부 확률	8.3 조건부 확률
8.3.1 multinomial	8.3.2 Dirichlet
8.3.3 Loss Function	8.3.4 일계조합
8.3.5 Sklearn	8.4 Multinomial NB
8.5 Updating probabilities	8.5.1 Sklearn
8.5.2 updating	
9 예제: Spam Filter	
10 Categorical Naïve Bayes	

## 7.1 전처리

- 어떤 programming language를 사용하던 같은 것과 다른 것의 구분은 확실하다.
- 분류모형은 다른 것들 중 같은 것을 찾거나 같은 것들 중 다른 것을 식별한다.
- 자연어처리의 시작은 포함시킬 단어 종류를 결정하고 오탈자를 수정이다.
- 문자의 경우 줄임말과 이모티콘 역시 처리할 필요가 있다.
- 전화번호, 통화, 숫자 등은 제거하거나 feature로 이용할 수 있다.
- Feature로 사용할 때는 내용이 중요한 것이 아니므로 Tel로 변환하여 사용
  - 전화번호는 [tel], 숫자는 [num], 통화는 [money]
- NLTK의 stopwords는 처음 사용할 때 따로 내려받아야 한다.  
Sklearn의 stopwords list는 318개, NLTK는 179개의 불용어를 포함하고 있다.

```
import nltk  
nltk.download('stopwords')
```

the number of features: 8403

## 7.2 OHE vs BoW

- Bag of words는 text 분석에서 가장 흔히 사용하는 방법으로 단어의 순서나 배열을 무시하고 단순히 빈도수를 이용한다.
- Bag of words 모형을 적용하기 위해서는 text를 단어로 분리한 후, 각 단어의 빈도수를 dict(sparse data)나 혹은 행렬로 정리한다. 이를 다음과 같은 형태로 만들어 분석에 사용한다.

BoW	said	fox	little	prince	But	And	shall	It	rose	like	tame	one
Frequency	30	29	25	20	14	11	11	11	9	9	9	8

- One-hot encoding 방법은 해당단어의 유무에 따라 dummy variable과 같이 0과 1의 값을 사용한다. Dummy와 이름만 다를 뿐 실질적인 차이는 없다.

OHE	said	fox	little	prince	But	And	shall	It	rose	like	tame	one
Frequency	1	1	1	1	1	1	1	1	1	1	1	1

- BernoulliNB에는 binarize keyword가 있어 default로 BoW를 OHE로 변환하여 적용한다. binarize의 default값은 0.0으로 필요에 따라 조정한다.

## Contents ⚙

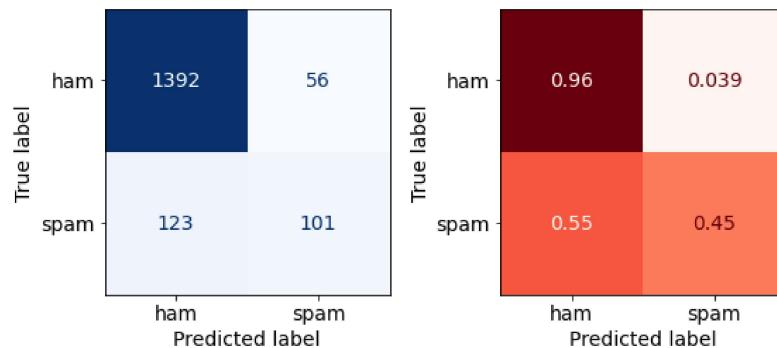
1 Discrete Naïve Bayes
1.1 모형의 구분
1.2 주의
2 추정방법
2.1 Application of Bayes' Rule
2.2 Posterior probability
2.3 조건부 확률
2.4 결합확률분포
2.5 Bayes' rule
3 Naïve Bayes' Rule
3.1 조건부 독립
3.2 Bayes' Rule
3.2.1 성인 남자
3.2.2 여자 청소년
3.3 모형의 평가
4 독립과 조건부 독립
4.1 조건부 독립
4.2 (un)conditional independence
4.3 조건부 독립
5 Discrete naïve Bayes
5.1 Scikit-learn Implementation
5.2 모형의 특징
6 Bernoulli Naïve Bayes
6.1 Posterior, likelihood
6.2 Bernoulli Naïve Bayes
6.2.1 Bernoulli Naïve Bayes
6.2.2 일계조합
6.3 Bernoulli Naïve Bayes
6.4 MLE, MAP, Bayesian
6.4.1 Frequency
6.4.2 Bayesian
7 예제: Spam Filter
7.1 전처리
7.2 OHE vs BoW
7.3 Sklearn Preprocess
7.4 모형의 평가
8 Multinomial Naïve Bayes
8.1 Multinomial NB
8.2 조건부 확률
8.3 조건부 확률
8.3.1 multinomial NB
8.3.2 Dirichlet prior
8.3.3 Loss Function
8.3.4 일계조합
8.3.5 Sklearn Implementation
8.4 Multinomial NB
8.5 Updating probabilities
8.5.1 Sklearn Implementation
8.5.2 updating probabilities
9 예제: Spam Filter
10 Categorical Naïve Bayes

## 7.3 Sklearn Preprocess

- CountVectorizer() 는 text 분석에서 각 문자열의 단어 수를 sparse data 형식으로 한다. 선택할 수 있는 option이 다양하며, 속도가 상당히 빠르다.
  - CountVectorizer는 형태가 다른 모든 token을 하나의 feature로 구분한다. 한번만 나타나거나 모든 문장에 공통적으로 나타나는 token을 제거할 때는 max\_df과 min\_df를 사용한다. 모형의 성과에 상당한 영향을 미친다.  

```
sklearn.feature_extraction.text.CountVectorizer(*, input='content', encoding='utf-8', decode_error='strict', strip_accents=None, lowercase=True, preprocessor=None, tokenizer=None, token_pattern='(?u)\\b\\w+\\w+\\b', ngram_range=(1, 1), analyzer='word', max_df=1.0, min_df=1, max_features=None, vocabulary=None, binary=False, dtype=<class 'numpy.int64'>)
```
  - input: {'filename', 'file', 'content'}, default='content'
  - strip\_accents: {'ascii', 'unicode'}, default=None
  - preprocessor: callable, default=None - Override the preprocessing (strip\_accents and lowercase) stage
  - tokenizer: callable, default=None
  - stop\_words: {'english'}, list, default=None
  - token\_pattern: str, default=r"(?u)\\b\\w+\\w+\\b"
  - ngram\_range: tuple (min\_n, max\_n), default=(1, 1)
  - max\_df: float in range [0.0, 1.0] or int, default=1.0
  - min\_df: float in range [0.0, 1.0] or int, default=1
  - max\_features: int, default=None
- ```
cv = CountVectorizer(keywords)
cv.fit(contents)
cv.transform(contents)
```

Figure 6. Confusion Matrix: Test dataset



## 7.4 모형의 평가

- Spam의 비율이 13.4%인 것과 단어의 출현여부만을 고려한 것을 감안해도 accuracy 90%는 썩 만족스러운 결과는 아니다. 하지만 spam 식별률은 45%로 추정관련 비용을 고려하면 '상당히' 효과적인 모형이다.
- 모형이 단순해서인지 overfitting 문제는 '전혀' 없다.

## Contents ☰

|                                 |
|---------------------------------|
| 1 Discrete Naïve Bayes          |
| 1.1 모형의 구분                      |
| 1.2 주의                          |
| 2 추정방법                          |
| 2.1 Application of Bayes' rule  |
| 2.2 Posterior probability       |
| 2.3 조건부 확률                      |
| 2.4 결합확률분포                      |
| 2.5 Bayes' rule                 |
| 3 Naïve Bayes' Rule             |
| 3.1 조건부 독립                      |
| 3.2 Bayes' Rule                 |
| 3.2.1 성인 남자                     |
| 3.2.2 여자 청소년                    |
| 3.3 모형의 평가                      |
| 4 독립과 조건부 독립                    |
| 4.1 조건부 독립                      |
| 4.2 (unconditional) 독립          |
| 4.3 조건부 독립                      |
| 5 Discrete naïve EM algorithm   |
| 5.1 Scikit-learn implementation |
| 5.2 모형의 특징                      |
| 6 Bernoulli Naïve Bayes         |
| 6.1 Posterior, likelihood       |
| 6.2 Bernoulli Naïve Bayes       |
| 6.2.1 Bernoulli Naïve Bayes     |
| 6.2.2 일계조합                      |
| 6.3 Bernoulli Naïve Bayes       |
| 6.4 MLE, MAP, Bayesian          |
| 6.4.1 Frequency                 |
| 6.4.2 Bayesian                  |
| 7 예제: Spam Filter               |
| 7.1 전처리                         |
| 7.2 OHE vs BoW                  |
| 7.3 Sklearn Preprocessor        |
| 7.4 모형의 평가                      |
| 8 Multinomial Naïve Bayes       |
| 8.1 Multinomial Naïve Bayes     |
| 8.2 조건부 확률                      |
| 8.3 조건부 확률                      |
| 8.3.1 multinomial distribution  |
| 8.3.2 Dirichlet distribution    |
| 8.3.3 Loss Function             |
| 8.3.4 일계조합                      |
| 8.3.5 Sklearn Preprocessor      |
| 8.4 Multinomial Naïve Bayes     |
| 8.5 Updating probabilities      |
| 8.5.1 Sklearn Preprocessor      |
| 8.5.2 updating probabilities    |
| 9 예제: Spam Filter               |
| 10 Categorical Naïve Bayes      |

| Training Data |      | precision    | recall | f1-score | support |
|---------------|------|--------------|--------|----------|---------|
| ham           | 0.91 | 0.96         | 0.94   | 3377     |         |
| spam          | 0.60 | 0.42         | 0.50   | 523      |         |
|               |      | accuracy     |        | 0.89     | 3900    |
|               |      | macro avg    | 0.76   | 0.69     | 3900    |
|               |      | weighted avg | 0.87   | 0.89     | 3900    |

| Test Data |      | precision    | recall | f1-score | support |
|-----------|------|--------------|--------|----------|---------|
| ham       | 0.92 | 0.96         | 0.94   | 1448     |         |
| spam      | 0.64 | 0.45         | 0.53   | 224      |         |
|           |      | accuracy     |        | 0.89     | 1672    |
|           |      | macro avg    | 0.78   | 0.71     | 1672    |
|           |      | weighted avg | 0.88   | 0.89     | 1672    |

## 8 Multinomial Naïve Bayes classifier

- A model of  $C \geq 2$  categories and discrete features

### 8.1 Multinomial distribution

- 주사위 던지기 실험에서 가능한 결과의 수는 6개이다. 주사위의 각 면이 나올 확률을  $\theta_1, \dots, \theta_6$ 라고 하면 주사위를  $n$ 번 던지는 실험의 결과는 다항분포 multinomial distribution으로 정리할 수 있다.

$$\Pr(n_1, n_2, \dots, n_6; n, \theta_1, \dots, \theta_6) = \binom{n}{n_1, n_2, \dots, n_6} \theta_1^{n_1} \theta_2^{n_2} \cdots \theta_6^{n_6}$$

- Class의 수가  $C$ 일 때 일반적인 다항분포의 확률밀도함수는 다음과 같다.

$$\Pr(n_1, \dots, n_C; n, \theta_1, \dots, \theta_C) = \binom{n}{n_1, \dots, n_C} \theta_1^{n_1} \cdots \theta_C^{n_C}$$

- 한편 특정 결과  $(n_1, \dots, n_C)$ 가 발생할 확률인 likelihood는 다음과 같다.

$$\Pr(n_1, \dots, n_C | \theta_1, \dots, \theta_C) = \theta_1^{n_1} \theta_2^{n_2} \cdots \theta_C^{n_C}$$

## Contents ⚙

|                                 |
|---------------------------------|
| 1 Discrete Naïve Bayes          |
| 1.1 모형의 구분                      |
| 1.2 주의                          |
| 2 추정방법                          |
| 2.1 Application of Bayes' Rule  |
| 2.2 Posterior probability       |
| 2.3 조건부 확률                      |
| 2.4 결합확률분포                      |
| 2.5 Bayes' rule                 |
| 3 Naïve Bayes' Rule             |
| 3.1 조건부 독립                      |
| 3.2 Bayes' Rule                 |
| 3.2.1 성인 남자                     |
| 3.2.2 여자 청소년                    |
| 3.3 모형의 평가                      |
| 4 독립과 조건부 독립                    |
| 4.1 조건부 독립                      |
| 4.2 (unconditional) 독립          |
| 4.3 조건부 독립                      |
| 5 Discrete naïve Ensemble       |
| 5.1 Scikit-learn Implementation |
| 5.2 모형의 특징                      |
| 6 Bernoulli Naïve Bayes         |
| 6.1 Posterior, likelihood       |
| 6.2 Bernoulli Naïve Bayes       |
| 6.2.1 Bernoulli                 |
| 6.2.2 일계조합                      |
| 6.3 Bernoulli Naïve Bayes       |
| 6.4 MLE, MAP, Bayesian          |
| 6.4.1 Frequency                 |
| 6.4.2 Bayesian                  |
| 7 예제: Spam Filter               |
| 7.1 전처리                         |
| 7.2 OHE vs BoW                  |
| 7.3 Sklearn Preprocessor        |
| 7.4 모형의 평가                      |
| 8 Multinomial Naïve Bayes       |
| 8.1 Multinomial                 |
| 8.2 조건부 확률                      |
| 8.3 조건부 확률                      |
| 8.3.1 multinomial               |
| 8.3.2 Dirichlet                 |
| 8.3.3 Loss Function             |
| 8.3.4 일계조합                      |
| 8.3.5 Sklearn                   |
| 8.4 Multinomial                 |
| 8.5 Updating prior              |
| 8.5.1 Sklearn                   |
| 8.5.2 updating prior            |
| 9 예제: Spam Filter               |
| 10 Categorical Naïve Bayes      |

## 8.2 조건부 확률의 추정량

- MLE with multinomial distribution

$$\hat{\pi}_c = \frac{n_c}{n}, \quad \hat{\theta}_{k|c} = \frac{\sum_{i=1}^n x_{ik|c}}{\sum_n \sum_k x_{ik}}$$

- MAP with Multinomial distribution and Dirichlet prior

$$\hat{\pi}_c = \frac{n_c + a_c}{n + \sum_c a_c}, \quad \hat{\theta}_{k|c} = \frac{\sum_{i=1}^n x_{ik|c} + b_c}{\sum_n \sum_k x_{ik} + \sum_c b_c} = \frac{n_{k|c} + b_c}{\sum_n \sum_k x_{ik} + \sum_c b_c}$$

- 새로운 표본이 추가되면 분자와 분모에 해당 표본의 수를 더해서 update 할 수 있다.

## 8.3 조건부 확률의 추정

- Posterior recap

$$\Pr(\boldsymbol{\theta}|X, Y) = \frac{\Pr(X, Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)} = \frac{\Pr(X|\boldsymbol{\theta}) \times \Pr(Y|\boldsymbol{\theta}) \times \Pr(\boldsymbol{\theta})}{\Pr(X, Y)}$$

- 다음 유도과정에선  $Y = c$ 에 대한 조건부분포를 사용한다.

$$\Pr(\boldsymbol{\theta}_c|X_c) = \frac{\Pr(X_c|\boldsymbol{\theta}_c) \times \Pr(\boldsymbol{\theta}_c)}{\Pr(X_c)} = \frac{\Pr(X_c|\boldsymbol{\theta}_c) \times \Pr(\boldsymbol{\theta}_c)}{\Pr(X_c)}$$

- Multinomial Naïve Bayes에서는 likelihood로는 multinomial distribution, prior는 multinomial distribution의 conjugate prior인 Dirichlet distribution을 사용한다.

$$X \sim \text{Multimomial}(n, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$$

### 8.3.1 multinomial distribution

$$\text{multinomial}(x_1, \dots, x_K; \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K) = \frac{m!}{x_1! \dots x_K!} \theta_1^{x_1} \dots \theta_K^{x_K} = \frac{\Gamma\left(\sum_{k=1}^K x_k + 1\right)}{\prod_{k=1}^K \Gamma(x_k + 1)} \prod_{k=1}^K \theta_k^{x_k}$$
$$x_k \in \mathbb{N}, \quad \sum_{k=1}^K x_k = m$$

### 8.3.2 Dirichlet distribution

$$\text{Dir}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K; \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^K \theta_k^{\alpha_{k-1}}$$
$$\theta_k \in [0, 1], \quad \sum_{k=1}^K \theta_k = 1, \quad B(\boldsymbol{\alpha}) = \frac{\prod_{c=1}^C \Gamma(\alpha_c)}{\Gamma\left(\sum_{c=1}^C \alpha_c\right)}$$

## Contents ⚙

|       |                             |
|-------|-----------------------------|
| 1     | Discrete Naïve Bayes        |
| 1.1   | 모형의 구분                      |
| 1.2   | 주의                          |
| 2     | 추정방법                        |
| 2.1   | Applicaiton of Bayes' rule  |
| 2.2   | Posterior probability       |
| 2.3   | 조건부 확률                      |
| 2.4   | 결합확률분포                      |
| 2.5   | Bayes' rule                 |
| 3     | Naïve Bayes' Rule           |
| 3.1   | 조건부 독립                      |
| 3.2   | Bayes' Rule                 |
| 3.2.1 | 성인 남자                       |
| 3.2.2 | 여자 청소년                      |
| 3.3   | 모형의 평가                      |
| 4     | 독립과 조건부 독립                  |
| 4.1   | 조건부 독립                      |
| 4.2   | (unconditional) 독립          |
| 4.3   | 조건부 독립                      |
| 5     | Discrete naïve E            |
| 5.1   | Scikit-learn Implementation |
| 5.2   | 모형의 특징                      |
| 6     | Bernoulli Naïve Bayes       |
| 6.1   | Posterior, likelihood       |
| 6.2   | Bernoulli Naïve Bayes       |
| 6.2.1 | Bernoulli                   |
| 6.2.2 | 일계조건부 확률                    |
| 6.3   | Bernoulli Naïve Bayes       |
| 6.4   | MLE, MAP, Bayesian          |
| 6.4.1 | Frequentist                 |
| 6.4.2 | Bayesian                    |
| 7     | 예제: Spam Filter             |
| 7.1   | 전처리                         |
| 7.2   | OHE vs BoW                  |
| 7.3   | Sklearn Preprocessor        |
| 7.4   | 모형의 평가                      |
| 8     | Multinomial Naïve Bayes     |
| 8.1   | Multinomial                 |
| 8.2   | 조건부 확률                      |
| 8.3   | 조건부 확률                      |
| 8.3.1 | multinomial                 |
| 8.3.2 | Dirichlet                   |
| 8.3.3 | Loss Function               |
| 8.3.4 | 일계조건부 확률                    |
| 8.3.5 | Sklearn                     |
| 8.4   | Multinomial                 |
| 8.5   | Updating probabilities      |
| 8.5.1 | Sklearn                     |
| 8.5.2 | updating probabilities      |
| 9     | 예제: Spam Filter             |
| 10    | Categorical Naïve Bayes     |

### 8.3.3 Loss Function

$$\Pr(\theta|X) = \frac{\Pr(X|\theta) \times \Pr(\theta)}{\Pr(X)}$$

$$\Pr(\theta|X) \propto \frac{\Gamma\left(\sum_{k=1}^K x_{k|c} + 1\right)}{\prod_{k=1}^K \Gamma(x_{k|c} + 1)} \prod_{k=1}^K \theta_{k|c}^{x_{k|c}} \times \frac{1}{B(\alpha)} \prod_{k=1}^K \theta_{k|c}^{\alpha_k - 1}$$

$$\boldsymbol{\theta}_{\text{MAP}|c} = \arg \max_{\boldsymbol{\theta} \in \Theta} \left( \prod_{i=1}^n \prod_{k=1}^K \theta_{k|c}^{x_{ik}} \times \prod_{k=1}^K \theta_{k|c}^{\alpha_k - 1} \right) = \arg \max_{\boldsymbol{\theta} \in \Theta} \sum_{k=1}^K \left( \sum_{i=1}^n x_{ik} + \alpha_k - 1 \right) \ln \theta_{k|c}$$

- $\sum_{k=1}^K \theta_{k|c} = 1$  이므로

$$\boldsymbol{\theta}_{\text{MAP}|c} = \arg \max_{\boldsymbol{\theta} > 0} \left( \sum_{k=1}^{K-1} \left( \sum_{i=1}^n x_{ik} + \alpha_k - 1 \right) \ln \theta_{k|c} + \left( \sum_{i=1}^n x_{iK} + \alpha_K - 1 \right) \ln \left( 1 - \sum_{k=1}^{K-1} \theta_{k|c} \right) \right)$$

### 8.3.4 일계조건

- 모든  $k$ 에 대한  $\theta_k$ 의 1계조건은

$$\frac{\sum_{i=1}^n x_{ik|c} + \alpha_k - 1}{\hat{\theta}_{k|c}} - \frac{\sum_{i=1}^n x_{ik|c} + \alpha_K - 1}{1 - \sum_{k=1}^{K-1} \hat{\theta}_{k|c}} = \frac{\sum_{i=1}^n x_{ik|c} + \alpha_k - 1}{\hat{\theta}_{k|c}} - \frac{\sum_{i=1}^n x_{ik|c} + \alpha_K - 1}{\hat{\theta}_{K|c}} = 0$$

- Multinomial regression과 같이 모수들은 비율로만 식별이 가능하다.
- 모든  $k$ 에 대해  $(\sum_{i=1}^n x_{ik|c} + \alpha_k - 1)$ 와  $\hat{\theta}_k$ 의 비율이 같아야 한다.

$$\frac{\sum_{i=1}^n x_{ik|c} + \alpha_k - 1}{\hat{\theta}_{k|c}} = A, \quad \text{for all } k$$

- 따라서 제약  $\sum_{k=1}^K \theta_k = 1$ 을 이용하여 다음과 같이 정리할 수 있다.

$$\hat{\theta}_{k|c} = \frac{\sum_{i=1}^n x_{ik|c} + \alpha_k - 1}{\sum_{k=1}^K (\sum_{i=1}^n x_{ik|c} + \alpha_k - 1)} = \frac{\sum_{i=1}^n 1(y_i = c)x_{ik|c} + (\alpha_k - 1)}{\sum_{k=1}^K \sum_{i=1}^n 1(y_i = c)x_{ik|c} + \sum_{k=1}^K (\alpha_k - 1)}$$

- $\alpha_k = 1$ 이면 MLE가 된다.
- 최종적으로 각 class마다  $K$ 개의 모수를 추정하며 posterior 추정에는 총  $C \times K$ 개의 모수가 필요하다.

$$(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_c, \dots, \boldsymbol{\theta}_C), \quad \boldsymbol{\theta}_c \in [0, 1]^K$$

## Contents ⚙

|                                 |
|---------------------------------|
| 1 Discrete Naïve Bayes          |
| 1.1 모형의 구분                      |
| 1.2 주의                          |
| 2 추정방법                          |
| 2.1 Application of MLE          |
| 2.2 Posterior probability       |
| 2.3 조건부 확률                      |
| 2.4 결합확률분포                      |
| 2.5 Bayes' rule                 |
| 3 Naïve Bayes' Rule             |
| 3.1 조건부 독립                      |
| 3.2 Bayes' Rule                 |
| 3.2.1 성인 남자                     |
| 3.2.2 여자 청소년                    |
| 3.3 모형의 평가                      |
| 4 독립과 조건부 독립                    |
| 4.1 조건부 독립                      |
| 4.2 (unconditional) 독립          |
| 4.3 조건부 독립                      |
| 5 Discrete naïve Bayes          |
| 5.1 Scikit-learn Implementation |
| 5.2 모형의 특징                      |
| 6 Bernoulli Naïve Bayes         |
| 6.1 Posterior, likelihood       |
| 6.2 Bernoulli Naïve Bayes       |
| 6.2.1 Bernoulli                 |
| 6.2.2 일계조합                      |
| 6.3 Bernoulli Naïve Bayes       |
| 6.4 MLE, MAP, Bayesian          |
| 6.4.1 Frequency                 |
| 6.4.2 Bayesian                  |
| 7 예제: Spam Filter               |
| 7.1 전처리                         |
| 7.2 OHE vs BoW                  |
| 7.3 Sklearn Preprocessor        |
| 7.4 모형의 평가                      |
| 8 Multinomial Naïve Bayes       |
| 8.1 Multinomial NB              |
| 8.2 조건부 확률                      |
| 8.3 조건부 확률                      |
| 8.3.1 multinomial               |
| 8.3.2 Dirichlet                 |
| 8.3.3 Loss Function             |
| 8.3.4 일계조합                      |
| 8.3.5 Sklearn                   |
| 8.4 Multinomial NB              |
| 8.5 Updating prior              |
| 8.5.1 Sklearn                   |
| 8.5.2 updating prior            |
| 9 예제: Spam Filter               |
| 10 Categorical Naïve Bayes      |

### 8.3.5 Sklearn implementation

- Sklearn의 MultinomialNB에서는
  - MLE를 추정하려면  $\alpha = 0$ 을 사용한다.
  - $\alpha = 1$ 은 Laplace smoothing,  $0 < \alpha < 1$ 은 Lidstone smoothing이라고 한다. Training data에 나타나지 않은 모집단의 특성을 반영하고 zero probability로 인한 문제를 방지한다.

$$\hat{\theta}_{k|c} = \frac{\sum_{i=1}^n x_{ik|c} + \alpha}{\sum_{k=1}^K \sum_{i=1}^n x_{ik|c} + K\alpha}$$

## 8.4 Multinomial Naïve Bayes classifier

- Multinomial Naïve Bayes classifier는 모든 feature들이 조건부 multinomial 분포를 따른다고 가정하므로,

$$\Pr(Y = c | X, \boldsymbol{\theta}) \propto \Pr(X|Y = c, \boldsymbol{\theta}) \Pr(Y) = \pi_c \times \theta_{1|c}^{x_1} \times \cdots \times \theta_{K|c}^{x_K}$$

- 따라서 classifier는 다음과 같다.

$$y^{\text{pred}}(x^{\text{new}}) = \arg \max_c \left( \ln \pi_c + \sum_{k=1}^K x_k^{\text{new}} \ln \hat{\theta}_{k|c} \right)$$

## 8.5 Updating priors

### 8.5.1 Sklearn의 prior, $\Pr(Y)$

- Posterior에는  $\boldsymbol{\theta}$ 와  $Y$ 에 대한 두 개의 prior가 포함되어 있다.
- Sklearn에선  $\boldsymbol{\theta}$ 의 분포에 대한 prior나  $Y$  분포의 prior를 update하지 않는다.
- Keyword 중 `fit_prior=True` 은 표본분의 MLE로 prior를 추정하여 사용하며 `class_prior` 는 prior distribution을 지정할 때 사용한다.

### 8.5.2 updating MAP estimator

- 예측을 하면서 지속적인 feedback이 가능하다면 prior update가 필요하다.
- MLE는 MAP에서  $\alpha_c = 1$ 인 경우이므로 MAP만 유도해 본다.
- 다항분포의 conjugate prior인 Dirichlet 분포의 확률밀도함수이다.

$$\text{Dirichlet}(\boldsymbol{\pi}; \boldsymbol{\alpha}) = \frac{\prod_{c=1}^C \pi_c^{\alpha_c - 1}}{B(\alpha_1, \dots, \alpha_C)}$$

- Likelihood에  $n = 1$ 인 다항분포, 혹은 categorical distribution을 사용하고 prior로 Dirichlet 분포를 사용하면 posterior는 다음과 같은 Dirichlet 분포를 따른다. 유도과정은 앞과 동일하다.

$$\Pr(\boldsymbol{\pi}|Y) = \frac{\Pr(Y|\boldsymbol{\pi}) \Pr(\boldsymbol{\pi})}{\Pr(Y)} = \text{Dirichlet}(n_1 + \alpha_1, \dots, n_C + \alpha_C)$$

- MAP estimate

$$\hat{\pi}_c = \frac{n_c + \alpha_c - 1}{n + \sum_{c=1}^C (\alpha_c - 1)}$$

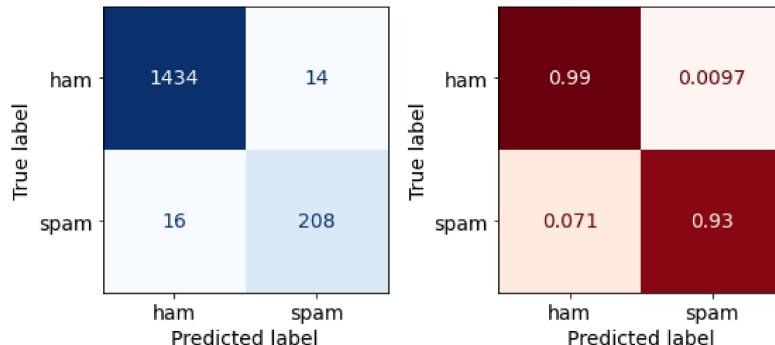
## Contents ⚙

|                                |
|--------------------------------|
| 1 Discrete Naïve Bayes         |
| 1.1 모형의 구분                     |
| 1.2 주의                         |
| 2 추정방법                         |
| 2.1 Application of Bayes' Rule |
| 2.2 Posterior probability      |
| 2.3 조건부 확률                     |
| 2.4 결합확률분포                     |
| 2.5 Bayes' rule                |
| 3 Naïve Bayes' Rule            |
| 3.1 조건부 독립                     |
| 3.2 Bayes' Rule                |
| 3.2.1 성인 남자                    |
| 3.2.2 여자 청소년                   |
| 3.3 모형의 평가                     |
| 4 독립과 조건부 독립                   |
| 4.1 조건부 독립                     |
| 4.2 (unconditional) 독립         |
| 4.3 조건부 독립                     |
| 5 Discrete naïve Bayes         |
| 5.1 Scikit-learn 활용            |
| 5.2 모형의 특징                     |
| 6 Bernoulli Naïve Bayes        |
| 6.1 Posterior, likelihood      |
| 6.2 Bernoulli Naïve Bayes      |
| 6.2.1 Bernoulli Naïve Bayes    |
| 6.2.2 일계조합                     |
| 6.3 Bernoulli Naïve Bayes      |
| 6.4 MLE, MAP, Bayesian         |
| 7 예제: Spam Filter              |
| 7.1 전처리                        |
| 7.2 OHE vs BoW                 |
| 7.3 Sklearn Preprocessor       |
| 7.4 모형의 평가                     |
| 8 Multinomial Naïve Bayes      |
| 8.1 Multinomial NB             |
| 8.2 조건부 확률                     |
| 8.3 조건부 확률                     |
| 8.3.1 multinomial distribution |
| 8.3.2 Dirichlet prior          |
| 8.3.3 Loss Function            |
| 8.3.4 일계조합                     |
| 8.3.5 Sklearn 활용               |
| 8.4 Multinomial NB             |
| 8.5 Updating prior             |
| 8.5.1 Sklearn 활용               |
| 8.5.2 updating prior           |
| 9 예제: Categorical Naïve Bayes  |
| 10 Categorical Naïve Bayes     |

## 9 예제: Spam Filter - MultinomialNB

- 자료를 Bag of Words로 정리하고 MultinomialNB로 만든 spam filter의 performance를 BernoulliNB 결과와 비교
- 다음과 같은 연습을 해볼 수 있다.
  - GridSearchCV를 이용하여 '최적' min\_df 와 max\_df 를 구해 분류기를 만든다.
  - 단어의 frequency를 feature로 하는 logistic regression 모형을 이용해 spam filter를 만들고 예측결과를 평가
  - L1-regularized logistic regression 모형을 이용하여 만든 spam filter의 예측결과를 평가
    - 컴퓨터가 감당하지 못하면 CountVectorizer 의 max\_features option을 이용하여 feature의 수를 줄여서 추정

Figure 7. Confusion Matrix: Training dataset



## 10 Categorical Naïve Bayes

- Categorical distribution은 Titanic dataset에서 sex, passager class나 embarked city 와 같이  $J_k$  개의 값 중 하나의 값을 갖을 수 있는 확률변수를 모형화할 때 사용한다.
- Categorical distribution의 밀도함수는 다음과 같다.

$$\Pr(x_k = j; \theta_1, \dots, \theta_{J_k}) = \prod_{j=1}^{J_k} \theta_k^{1(x_k=j)}, \quad x_k \in \{1, \dots, J_k\}, k = 1, \dots, K$$

- Categorical distribution의 conjugate prior는 Dirichlet 분포이며 MAP estimator는 다음과 같다.

$$\Pr(x_{ik} = j | Y = c) = \hat{\theta}_{kj|c} = \frac{n_{kj|c} + \alpha}{n_c + \alpha J_k} = \frac{\sum_{i=1}^n 1(y_i = c) \cdot 1(x_{ik} = j) + \alpha}{\sum_{i=1}^n 1(y_i = c) + \alpha J_k}$$

여기서  $n_{kj|c}$ 는 class  $c$ 에서  $x_k$ 의 값이  $j$ 인 표본의 수이다.

- Categorical naïve Bayes 모형과 multinomial naïve Bayes 모형의 추정량은 동일하다. 전체 시행횟수가 1 이므로 분모의 값을 비교해보면 MultinomialNB는 해당 class에 속한 모든 feature들의 값이고 CategoricalNB는 각각의 표본의 수가 된다.
- 하지만 categorical distribution은 범주형, multinomial distribution은 count 변수에 사용하므로 쓰임새가 다르다.

## Contents ⚙

|                                 |
|---------------------------------|
| 1 Discrete Naïve Bayes          |
| 1.1 모형의 구분                      |
| 1.2 주의                          |
| 2 추정방법                          |
| 2.1 Application of Bayes' Rule  |
| 2.2 Posterior probability       |
| 2.3 조건부 확률                      |
| 2.4 결합확률분석                      |
| 2.5 Bayes' rule                 |
| 3 Naïve Bayes' Rule             |
| 3.1 조건부 독립                      |
| 3.2 Bayes' Rule                 |
| 3.2.1 성인 남성                     |
| 3.2.2 여자 청소년                    |
| 3.3 모형의 평가                      |
| 4 독립과 조건부 독립                    |
| 4.1 조건부 독립                      |
| 4.2 (unconditional) 독립          |
| 4.3 조건부 독립                      |
| 5 Discrete naïve Bayes          |
| 5.1 Scikit-learn Implementation |
| 5.2 모형의 특징                      |
| 6 Bernoulli Naïve Bayes         |
| 6.1 Posterior, likelihood       |
| 6.2 Bernoulli Naïve Bayes       |
| 6.2.1 Bernoulli                 |
| 6.2.2 일계조합                      |
| 6.3 Bernoulli Naïve Bayes       |
| 6.4 MLE, MAP, Bayesian          |
| 6.4.1 Frequency                 |
| 6.4.2 Bayesian                  |
| 7 예제: Spam Filter               |
| 7.1 전처리                         |
| 7.2 OHE vs BoW                  |
| 7.3 Sklearn Preprocessor        |
| 7.4 모형의 평가                      |
| 8 Multinomial Naïve Bayes       |
| 8.1 Multinomial NB              |
| 8.2 조건부 확률                      |
| 8.3 조건부 확률                      |
| 8.3.1 multinomial               |
| 8.3.2 Dirichlet                 |
| 8.3.3 Loss Function             |
| 8.3.4 일계조합                      |
| 8.3.5 Sklearn                   |
| 8.4 Multinomial NB              |
| 8.5 Updating prior              |
| 8.5.1 Sklearn                   |
| 8.5.2 updating prior            |
| 9 예제: Spam Filter               |
| 10 Categorical Naïve Bayes      |

# 11 Discrete Naïve Bayesian의 특징

## 11.1 Linear decision boundary

- 조건부 독립성 가정이 성립하면 naïve Bayes 모형들은 logistic regression 모형과 같다.

11.1.1 Binomial Naïve Bayes classifier (<https://svivek.com/teaching/lectures/slides/naive-bayes/naive-bayes-linear.pdf>)

$$\begin{aligned}\ln \frac{\Pr(y=1|x)}{\Pr(y=0|x)} &= \ln \frac{\pi}{1-\pi} + \sum_{i=1}^k \ln \frac{(1-a_i)}{(1-b_i)} + \sum_{k=1}^K x_k \ln \frac{a_k(1-b_k)}{b_k(1-a_k)} \\ &= \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K\end{aligned}$$

## 유도과정

$$\begin{aligned}\frac{\Pr(Y=1|X_1, X_2, \dots, X_K)}{\Pr(Y=0|X_1, X_2, \dots, X_K)} &= \frac{\Pr(Y=1) \prod_{k=1}^K \Pr(X_k|Y=1)}{\Pr(Y=0) \prod_{k=1}^K \Pr(X_k|Y=0)} \\ \ln \frac{\Pr(Y=1|X_1, X_2, \dots, X_K)}{\Pr(Y=0|X_1, X_2, \dots, X_K)} &= \ln \frac{\Pr(Y=1)}{\Pr(Y=0)} + \sum_{k=1}^K \ln \frac{\Pr(X_k|Y=1)}{\Pr(X_k|Y=0)}\end{aligned}$$

- $\pi_c = \Pr(y=c)$ ,  $a_k = \Pr(x_k = 1|y=1)$ ,  $b_k = \Pr(x_k = 1|y=0)$  라고 하자.

$$\Pr(x_k|y=1) = a_k^{x_k} (1-a_k)^{1-x_k}, \quad \Pr(x_k|y=0) = b_k^{x_k} (1-b_k)^{1-x_k}$$

이므로

$$\begin{aligned}\ln \frac{\Pr(y=1|x)}{\Pr(y=0|x)} &= \ln \frac{\pi_1}{\pi_0} + \sum_{k=1}^K \ln \frac{a_k^{x_k} (1-a_k)^{1-x_k}}{b_k^{x_k} (1-b_k)^{1-x_k}} \\ &= \ln \frac{\pi_1}{\pi_0} + \sum_{k=1}^K \ln \frac{(1-a_k)}{(1-b_k)} + \sum_{k=1}^K \ln \frac{a_k^{x_k} (1-b_k)^{x_k}}{b_k^{x_k} (1-a_k)^{x_k}} \\ &= \ln \frac{\pi_1}{\pi_0} + \sum_{k=1}^K \ln \frac{(1-a_k)}{(1-b_k)} + \sum_{k=1}^K x_k \ln \frac{a_k(1-b_k)}{b_k(1-a_k)}\end{aligned}$$

## Contents ⚙

|                                  |
|----------------------------------|
| 1 Discrete Naïve Bayes           |
| 1.1 모형의 구분                       |
| 1.2 주의                           |
| 2 추정방법                           |
| 2.1 Application of Bayes' Rule   |
| 2.2 Posterior probability        |
| 2.3 조건부 확률                       |
| 2.4 결합확률분석                       |
| 2.5 Bayes' rule                  |
| 3 Naïve Bayes' Rule              |
| 3.1 조건부 독립                       |
| 3.2 Bayes' Rule                  |
| 3.2.1 성인 남자                      |
| 3.2.2 여자 청소년                     |
| 3.3 모형의 평가                       |
| 4 독립과 조건부 독립                     |
| 4.1 조건부 독립                       |
| 4.2 (un)conditional independence |
| 4.3 조건부 독립                       |
| 5 Discrete naïve Bayes           |
| 5.1 Scikit-learn Implementation  |
| 5.2 모형의 특징                       |
| 6 Bernoulli Naïve Bayes          |
| 6.1 Posterior, likelihood        |
| 6.2 Bernoulli Naïve Bayes        |
| 6.2.1 Bernoulli Naïve Bayes      |
| 6.2.2 일계조합                       |
| 6.3 Bernoulli Naïve Bayes        |
| 6.4 MLE, MAP, Prior              |
| 6.4.1 Frequency                  |
| 6.4.2 Bayesian prior             |
| 7 예제: Spam Filter                |
| 7.1 전처리                          |
| 7.2 OHE vs BoW                   |
| 7.3 Sklearn Preprocessor         |
| 7.4 모형의 평가                       |
| 8 Multinomial Naïve Bayes        |
| 8.1 Multinomial Naïve Bayes      |
| 8.2 조건부 확률                       |
| 8.3 조건부 확률                       |
| 8.3.1 multinomial distribution   |
| 8.3.2 Dirichlet distribution     |
| 8.3.3 Loss Function              |
| 8.3.4 일계조합                       |
| 8.3.5 Sklearn Implementation     |
| 8.4 Multinomial Naïve Bayes      |
| 8.5 Updating probabilities       |
| 8.5.1 Sklearn Implementation     |
| 8.5.2 updating probabilities     |
| 9 예제: Spam Filter                |
| 10 Categorical Naïve Bayes       |

### 11.1.2 Multinomial Naïve Bayes classifier

- Multinomial regression 모형과 마찬가지로 OVO에 근거한 예측이다.
- 두 개의 class의 확률을 비교하여 확률이 더 큰 class를 선택하므로 decision boundary는 두 확률이 같아지는 surface이다.

$$\Pr(Y = a|x) \geq \Pr(Y = b|x) \iff \ln \pi_a + \sum_{k=1}^K x_k \ln \hat{\theta}_{k|a} \geq \ln \pi_b + \sum_{k=1}^K x_k \ln \hat{\theta}_{k|b}$$
$$\iff (\ln \pi_a - \ln \pi_b) + \sum_{k=1}^K x_k (\ln \hat{\theta}_{k|a} - \ln \hat{\theta}_{k|b})$$

### 11.2 zero-frequency problem

- 훈련에 사용한 표본에 범주  $X_{k|c} = x_{k|j|c}$  중 하나가 빠져 있으면 분류기는 그런 조건에 대한 예측을 어떻게 하는 것이 바람직할까? 예를 들어 표본에 사망한 여자아이가 없다면 이런 조건에 대한 예측은 생략된다.
- Titanic 자료에서 가족의 수가 9명과 10명인 관찰값이 없다. 이를 극대값이 11개의 값을 갖는 범주형 변수로 취급하여 CategoricalNB로 예측을 구해보면 class의 label에 따라 생존확률이 0이 되거나 1이 될 확률이 있다.
- Laplace smoothing, additive smoothing, Lidstone smoothing

표본에 해당 관찰값이 포함되어 있지 않은 경우 예측 확률이 0이 되는 것을 방지하는 역할이다.

$$\hat{\theta}_{k|j|c} = \frac{n_{k|j|c} + \alpha}{n_c + \alpha C}$$

- Sklearn에서 default는 1이며, 값을 지정하여 사용할 때는  $\alpha \geq 1.0e-10$ 로 설정한다.

### 11.3 conditional independence

- Feature set이 조건부 독립성을 만족하지 않는 경우에도 naïve Bayes 모형의 zero-one loss function에 대한 performance는 상당히 만족스러운 것으로 알려져 있다.
- [Domingos, Pedro and Pazzani, Michael \(1996\), Simple Bayesian Classifiers Do Not Assume Independence \(\[http://engr.case.edu/ray\\\_soumya/mlrg/optimality\\\_of\\\_nb.pdf\]\(http://engr.case.edu/ray\_soumya/mlrg/optimality\_of\_nb.pdf\)\)](http://engr.case.edu/ray_soumya/mlrg/optimality_of_nb.pdf)
- Transformed Weight-normalized Complement Naive Bayes (TWCNB)는 관련 문제를 보완한 모형이다.
- [Jason D. M. Rennie Lawrence Shih Jaime Teevan David R. Karger \(2003\), Tackling the Poor Assumptions of Naive Bayes Text Classifiers \(<http://people.csail.mit.edu/jrennie/papers/icml03-nb.pdf>\)](http://people.csail.mit.edu/jrennie/papers/icml03-nb.pdf)

## Contents ⚙

|                             |                                  |
|-----------------------------|----------------------------------|
| 1 Discrete Naïve Bayes      | 1.1 모형의 구분                       |
|                             | 1.2 주의                           |
| 2 추정방법                      | 2.1 Application of Bayes' Rule   |
|                             | 2.2 Posterior probability        |
|                             | 2.3 조건부 확률                       |
|                             | 2.4 결합확률분포                       |
|                             | 2.5 Bayes' rule                  |
| 3 Naïve Bayes' Rule         | 3.1 조건부 독립                       |
|                             | 3.2 Bayes' Rule                  |
|                             | 3.2.1 성인 남자                      |
|                             | 3.2.2 여자 청소년                     |
|                             | 3.3 모형의 평가                       |
| 4 독립과 조건부 독립                | 4.1 조건부 독립                       |
|                             | 4.2 (unconditional) independence |
|                             | 4.3 조건부 독립                       |
| 5 Discrete naïve Estimation | 5.1 Scikit-learn Implementation  |
|                             | 5.2 모형의 특징                       |
| 6 Bernoulli Naïve Bayes     | 6.1 Posterior, likelihood        |
|                             | 6.2.1 Bernoulli Naïve Bayes      |
|                             | 6.2.2 일계조합                       |
|                             | 6.3 Bernoulli Naïve Bayes        |
|                             | 6.4 MLE, MAP, Prior              |
|                             | 6.4.1 Frequency                  |
|                             | 6.4.2 Bayesian                   |
| 7 예제: Spam Filter           | 7.1 전처리                          |
|                             | 7.2 OHE vs BoW                   |
|                             | 7.3 Sklearn Preprocessor         |
|                             | 7.4 모형의 평가                       |
| 8 Multinomial Naïve Bayes   | 8.1 Multinomial Naïve Bayes      |
|                             | 8.2 조건부 확률                       |
|                             | 8.3 조건부 확률                       |
|                             | 8.3.1 multinomial distribution   |
|                             | 8.3.2 Dirichlet distribution     |
|                             | 8.3.3 Loss Function              |
|                             | 8.3.4 일계조합                       |
|                             | 8.3.5 Sklearn                    |
|                             | 8.4 Multinomial Naïve Bayes      |
|                             | 8.5 Updating prior               |
|                             | 8.5.1 Sklearn                    |
|                             | 8.5.2 updating prior             |
| 9 예제: Spam Filter           |                                  |
| 10 Categorical Naïve Bayes  |                                  |

## multicollinearity

- (Uncondition) independence 와 conditional independence의 관계에서는 하나가 다른 하나를 의미하지 않는다.
- 마찬가지로 conditional independence와 multicollinearity 사이에도 특별한 관계가 없다.  
(multicollinearity는 independence와 더 관련이 있다.)

다음 예에서는 두 확률변수가 조건부독립이지만 표본이 balanced 되어 있다면 둘 변수는 거의 완전한 선형관계에 있다.

$$\mu_{y=0} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad \mu_{y=1} = \begin{bmatrix} -10 \\ -10 \end{bmatrix}, \quad \Sigma_{y=0} = \Sigma_{y=1} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$$

- 어느 정도의 조건부독립성은 예측에 큰 영향을 미치지 않지만 다중공선성에 대한 처치는 필요하다.
- 아래 예는  $X_1 = X_2$ 와 같은 극단적인 다중공선성이 상당한 편의를 발생시킬 수 있음을 보여준다.

### 11.4 예제: perfectly collinear features

- 다음과 같은 확률변수  $X_1$ 과  $Y$ 는  $\Pr(X_1 \cap Y) = \Pr(X_1) \times \Pr(Y)$ 이므로 독립적이다.
- 따라서  $X_1$ 은  $Y$ 의 값을 예측하는데 전혀 도움이 되지 않는다.

$$\Pr(X_1 = 0|Y = 0) = \Pr(X_1 = 0|Y = 1) = 0.3$$

|  |  | $Y = 0$   | $Y = 1$ |
|--|--|-----------|---------|
|  |  | $X_1 = 0$ | $0.12$  |
|  |  | $X_1 = 1$ | $0.28$  |
|  |  |           | $0.4$   |
|  |  |           | $0.6$   |

- 위 모형에 확률변수  $X_1$ 과 완전히 '동일한'  $X_2$ 를 추가해 보자.  $\text{corr}(X_1, X_2) = 1$ ,  $\text{corr}(X_1, X_2|y) = 0$

|  |  | $Y = 0$   | $Y = 1$   |
|--|--|-----------|-----------|
|  |  | $X_1 = 0$ | $X_2 = 0$ |
|  |  | $X_1 = 1$ | $X_2 = 1$ |
|  |  |           | $0.12$    |
|  |  |           | $0.18$    |
|  |  |           | $0.0$     |
|  |  |           | $0.0$     |
|  |  |           | $0.28$    |
|  |  |           | $0.42$    |

- 둘은 완벽한 선형관계에 있으므로 상관관계는 1이다.
- 독립성을 만족하지 않으면 예측확률에는 큰 오차가 발생한다.

$$\Pr(X_1 = 0 \cap X_2 = 0 \cap Y = 0) = 0.12,$$
$$\Pr(X_1 = 0) \times \Pr(X_2 = 0) \times \Pr(Y = 0) = 0.3 \times 0.3 \times 0.4 = 0.036$$

## Contents ⚙

|                                 |
|---------------------------------|
| 1 Discrete Naïve Bayes          |
| 1.1 모형의 구분                      |
| 1.2 주의                          |
| 2 추정방법                          |
| 2.1 Application of Bayes' Rule  |
| 2.2 Posterior probability       |
| 2.3 조건부 확률                      |
| 2.4 결합확률분포                      |
| 2.5 Bayes' rule                 |
| 3 Naïve Bayes' Rule             |
| 3.1 조건부 독립                      |
| 3.2 Bayes' Rule                 |
| 3.2.1 성인 남성                     |
| 3.2.2 여자 청소년                    |
| 3.3 모형의 평가                      |
| 4 독립과 조건부 독립                    |
| 4.1 조건부 독립                      |
| 4.2 (unconditional) 독립          |
| 4.3 조건부 독립                      |
| 5 Discrete naïve Bayes          |
| 5.1 Scikit-learn Implementation |
| 5.2 모형의 특징                      |
| 6 Bernoulli Naïve Bayes         |
| 6.1 Posterior, likelihood       |
| 6.2 Bernoulli Naïve Bayes       |
| 6.2.1 Bernoulli Rule            |
| 6.2.2 일계조합                      |
| 6.3 Bernoulli Naïve Bayes       |
| 6.4 MLE, MAP, Bayesian          |
| 6.4.1 Frequency                 |
| 6.4.2 Bayesian                  |
| 7 예제: Spam Filter               |
| 7.1 전처리                         |
| 7.2 OHE vs BoW                  |
| 7.3 Sklearn Preprocessor        |
| 7.4 모형의 평가                      |
| 8 Multinomial Naïve Bayes       |
| 8.1 Multinomial Rule            |
| 8.2 조건부 확률                      |
| 8.3 조건부 확률                      |
| 8.3.1 multinomial rule          |
| 8.3.2 Dirichlet prior           |
| 8.3.3 Loss Function             |
| 8.3.4 일계조합                      |
| 8.3.5 Sklearn                   |
| 8.4 Multinomial Rule            |
| 8.5 Updating probabilities      |
| 8.5.1 Sklearn                   |
| 8.5.2 updating probabilities    |
| 9 예제: Spam Filter               |
| 10 Categorical Naïve Bayes      |

## 11.5 Pros and Cons

- Pros
  1. 표본의 수나 class 수가 많아도 계산이 간단하고 빠르다.
  2. 조건부 독립성이 만족하면 naïve Bayes classifier는 logistic regression보다 상대적으로 나은 결과를 보여주며 작은 표본으로도 학습이 가능하다.
- Cons
  1. 범주형 자료의 경우 "Zero Frequency" 문제가 발생할 수 있다. Laplace smoothing technique으로 해결이 가능하다.
  2. 조건부 독립성은 상당히 만족되기 어려운 조건이다.
  3. 상관관계가 높은 feature들은 추정확률에 편의를 발생시킨다.
  4. 가정이 만족하지 않을 때는 Naïve Bayes로 추정한 계수값들의 신뢰도는 낮다고 알려져 있으므로 predict\_proba에는 별다른 의미를 부여하지 않는 것이 좋다.

## 12 분류모형의 특징

|                                 | Linear Probability model                    | Logistic Regression model      | Naïve Bayes                    |
|---------------------------------|---------------------------------------------|--------------------------------|--------------------------------|
| Decision boundary의 형태           | linear                                      | linear                         | linear                         |
| cost of computation, 모형의 유지와 보수 | low                                         | high                           | extremely low                  |
| class의 비율이 추정 결과에 미치는 영향        | sensitive                                   | less                           | less                           |
| 추정 결과의 해석의 용이성                  | easiest                                     | easy                           | moderate                       |
| 추정에 필요한 표본의 수                   | balanced                                    | relatively smaller sample size | relatively smaller sample size |
| 특징                              | linear approximation of Logistic regression | log odds                       | conditional independence       |
| 유형                              | probabilistic discriminant                  | probabilistic discriminant     | generative                     |

<http://www2.aueb.gr/users/ion/data/enron-spam/> (<http://www2.aueb.gr/users/ion/data/enron-spam/>)