Probability

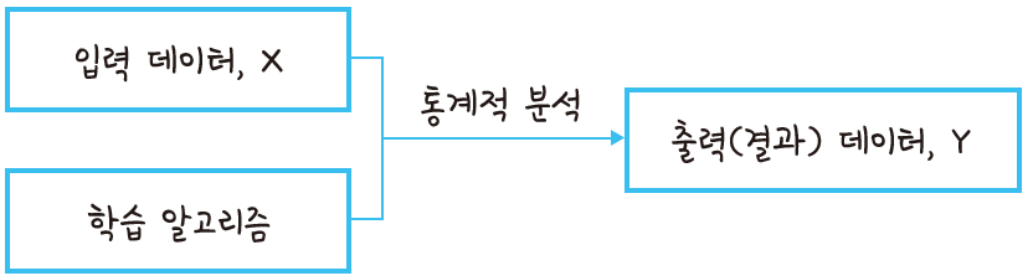
What is probability

❖ Relationship between probability and AI

- The purpose of artificial intelligence is to perform predictions using big data
- For example, you can predict the probability of a particular illness based on patient medical records or predict abnormal operation of factory equipment based on data collected from sensors
- A recommendation system that analyzes the music patterns that the user listens to and recommends the best music should also consider probabilities
- Probability and statistics are the foundation of data analysis

❖ Relationship between probability and AI

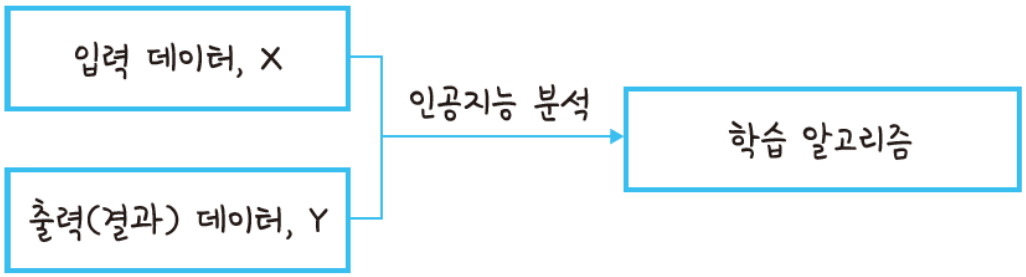
- Let's find out the difference between statistical analysis and data analysis using artificial intelligence
  - In statistical analysis, when the independent variable (X) and the mathematical model are informed as inputs, the dependent variable (Y) is output



What is probability

❖ **Relationship between probability and AI**

- In the analysis using artificial intelligence, when the independent variable (X) and the dependent variable (Y, label) are informed, the computer creates a learning model by itself



What is probability

❖ **Relationship between probability and AI**

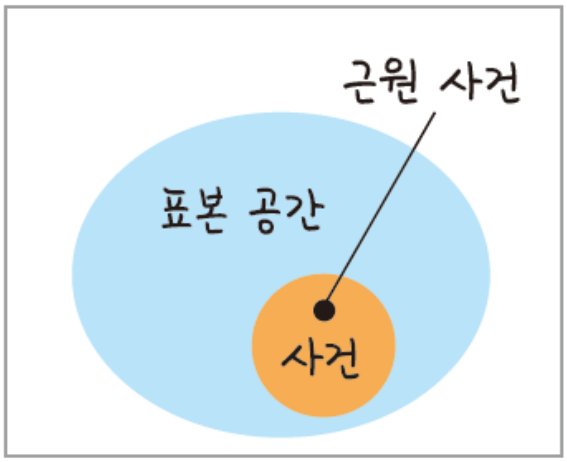
- Statistical analysis extracts output (result) data by applying an appropriate algorithm each time the input data is changed (every time the data is changed)
- On the other hand, data analysis using artificial intelligence automatically generates learning algorithms with only input and output (result) data, so algorithms can be recycled for data analysis for similar purposes
- What should not be misunderstood here is that *probability/statistics* ***have not been replaced by*** *artificial intelligence*, but that ***probability/statistics concepts have been added to artificial intelligence***
- AI adds a probability/statistics concept called weight when processing input data

What is probability

❖ **Relationship between probability and AI**

- For example, suppose you create an algorithm that categorizes music genres into categories such as 'hip-hop, jazz, and pop'
- The first thing is an algorithm for classification based on the frequency of use of certain words (composer names, music titles, etc.)
- At this time, "specific words" become variables such as X and Y, and "frequency of use" becomes a weight
- Algorithm accuracy varies depending on which weight is used, so if the programmer inputs historical data into the computer and uses the extracted frequency value (probability/statistics) as a weight, it can be closer to the accuracy of the desired result
- Data analysis using AI introduces and uses the concept of probability/statistics in AI data analysis rather than replacing probability/statistics

❖ **Probability basic terminology**



What is probability

용어	설명	표현
실험(trial)	동일한 조건에서 여러 번 반복할 수 있고 그 결과가 우연으로 결정되는 관찰이나 실험	= epoke
표본 공간 (sample space)	한 실험에서 나올 수 있는 모든 가능한 결과의 집합	Ω
근원사건 (elementary outcome)	표본 공간을 이루는 개개의 결과	$\omega_1, \omega_2, \cdots$
사건(event)	근원사건의 집합, 표본 공간의 부분 집합	
합사건	두 사건 A와 B의 합집합으로 표현할 수 있는 사건	$A \cup B$
곱사건	두 사건 A와 B의 교집합으로 표현할 수 있는 사건	$A \cap B$
여사건	사건 A가 일어나지 않는 사건	A^C
배반사건	사건 A와 B가 동시에 일어나지 않는 사건	$A \cap B = 0$

❖ Probability basic terminology

- For example, if you throw a dice once, let's find out the sample space and the elementary event
  - Sample space (Ω): Ω = {1, 2, 3, 4, 5, 6}
  - Elementary event (ω_1, ω_2, …): 1, 2, 3, 4, 5, 6
  - Even-numbered incident (A, B, …): A = {2, 4, 6}

❖ Probability basic terminology

The probability of an event

- In an experiment, the number of possibilities of event A is called the probability of event A occurring, and it is expressed as follows

$$P(A)$$

- The probability of an event is the percentage of times the event is expected to occur when the experiment is repeated under the same conditions
- If the number of elements in the sample space is n(S) and the number of elements in event A is n(A), the probability of event A occurring can be expressed as follows

$$P(A) = \frac{A\text{가 일어나리라 예상되는 횟수}}{\text{전체 실험의 횟수}}$$

❖ Probability basic terminology

- This can be expressed more mathematically as follows, which is called mathematical probability

$$P(A) = \frac{n(A)}{n(S)}$$

- Probability has three properties
  - (1) For any event A, the probability P(A) is $0 \leq P(A) \leq 1$
  - (2) Probability P(S) = 1 for event S that must occur
  - (3) P(ø) = 0 for an event ø that never happens
- The minimum probability is 0 and the maximum is 1

❖ Independent event and dependent event

- In two events A and B, A and B are called independent events when the outcome of one event does not affect the other
- Events A and B are to be independent of each other when whether or not event A occurs does not affect the probability of event B occurring

$$P(B \mid A) = P(B \mid A^c) = P(B)$$

- The necessary and sufficient conditions for the two events to be independent of each other are

$$P(A \cap B) = P(A)P(B \mid A) = P(A)P(B) \text{ (단 } P(A) > 0,\ P(B) > 0)$$

❖ **Independent event and dependent event**

- In two cases A and B, A and B are referred to as dependent events when the outcome of one event affects another
- Events A and B are said to be dependent when whether or not event A occurs affects the probability of event B occurring, and are expressed as follows

$$P(B \mid A) \neq P(B \mid A^c) \neq P(B)$$

❖ Conditional probability

- Conditional probability is the ***probability that another event B will occur under the condition that event A has occurred***

$$P(B \mid A)$$

- When event A occurs, the conditional probability law for event B is

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

  - P(B|A): conditional probability of B occurring under condition A
  - P(A, B) = P(AB) = P(A ∩ B): joint probability that occurs together/at the same time
  - P(A): the marginal probability of focusing only on a particular event A

❖ Conditional probability

- Conditional probability P(B|A) has the following meanings
  - Probability of event B if event A occurs
  - How the accuracy (credibility) of the fact that this sample belongs to event B changes when you learn a new fact that the sample belongs to event A

❖ Conditional probability

- For example, 65 percent of all students at a high school called ABC wore glasses, and 45 percent of them wore glasses in boy
- Let's find the probability that if one of the students at this school was a student wearing glasses, that student was a boy
- Among the total students, the student wearing glasses (P(A)) was 0.65 and the student wearing glasses (P(A ∩ B)) was 0.45
- The probability of being a student with glasses and a male student is as follows

$$P(B \mid A) = \frac{P(A \bigcap B)}{P(A)} = \frac{0.45}{0.65} = \frac{9}{13}$$

❖ Conditional probability

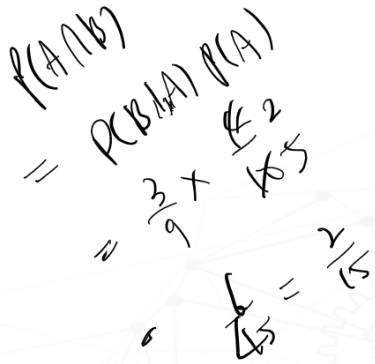
연습 문제

(1) 주머니 속에 흰색 공 네 개와 붉은색 공 여섯 개가 있습니다. 공을 한 개씩 두 번 꺼낼 때 다음 각 경우에서 두 개가 모두 흰색 공일 확률을 구하세요.

① 처음 꺼낸 공을 다시 넣지 않은 경우

② 처음 꺼낸 공을 다시 넣는 경우

0.16



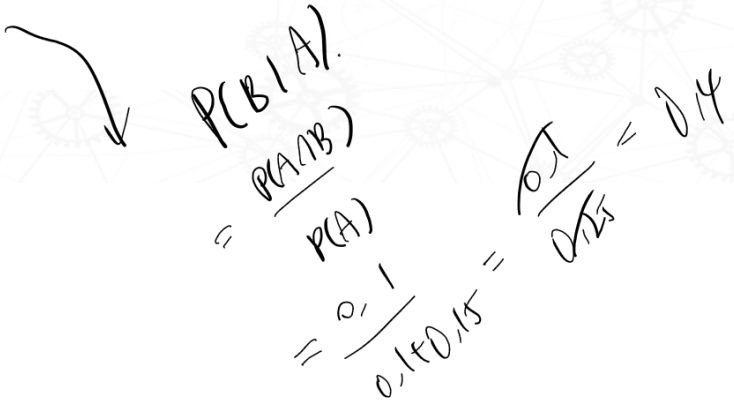
(2) 다음 표가 주어졌을 때 물음에 답하세요.

구분	비만	정상	저체중	합계
고혈압	0.10	0.07	0.03	0.20
정상혈압	0.15	0.55	0.10	0.80
합계	0.25	0.62	0.13	1.00

① 비만일 경우 고혈압일 확률

② 비만이고 고혈압일 확률

0.1



❖ Conditional probability

문제 풀이

(1) ① 처음 꺼낸 공을 다시 넣지 않은 경우

- 사건 A: 첫 번째에 흰색 공이 나오는 사건
- 사건 B: 두 번째에 흰색 공이 나오는 사건

구하고자 하는 확률은 $P(A \cap B)$입니다.

$$P(A \cap B) = P(B \mid A)P(A) = \frac{3}{9} \times \frac{4}{10} = \frac{2}{15}$$

❖ Conditional probability

② 처음 꺼낸 공을 다시 넣는 경우

이 경우 A와 B는 서로 독립사건입니다.

$$P(A \cap B) = P(B \mid A)P(A) = P(B)P(A) = \frac{4}{10} \times \frac{4}{10} = \frac{4}{25}$$

(2) A: 고혈압, B: 비만일 경우

① 비만일 경우 고혈압일 확률

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.10}{(0.10 + 0.15)} = \frac{0.1}{0.25} = 0.4$$

② 비만이고 고혈압일 확률

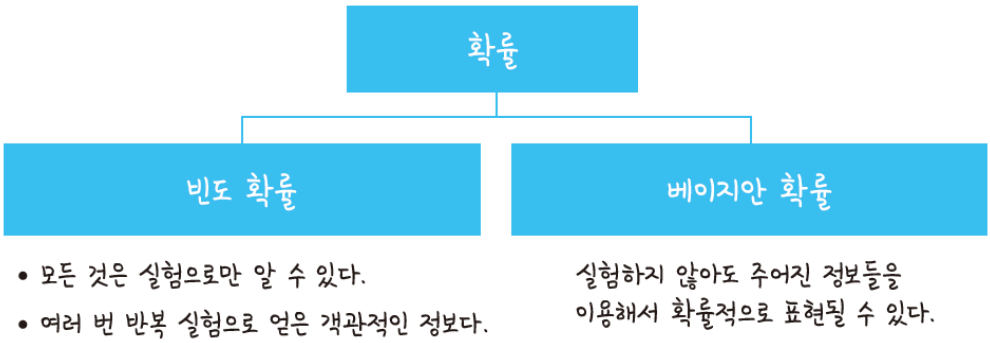
$$P(A \cap B) = P(A \mid B)P(B) = 0.4 \times 0.25 = 0.1$$

❖ Bayesian theory

Frequentist probability vs. Bayesian probability

- Statistically, studying AI requires understanding two probability:

  frequentist probability and Bayesian probability



❖ Bayesian theory

- The frequentist probability determines the frequency of the event being repeated
- The model is ***verified by observing how frequently a particular event occurs repeatedly and hypothesizing based on it***
- For example, a probability value of 0.5 for an "event with a front surface" by tossing a coin repeatedly means that the event occurred by multiplying the total number of times the coin was thrown by the probability value
- Bayesian probabilities are

  (1) *predisposed to subjective hypotheses as probabilities for non-occurring or uncertain events,*
  (일어나지 않았거나 불확실한 사건에 대한 확률로 주관적인 가설의 사전 확률을 정하고)

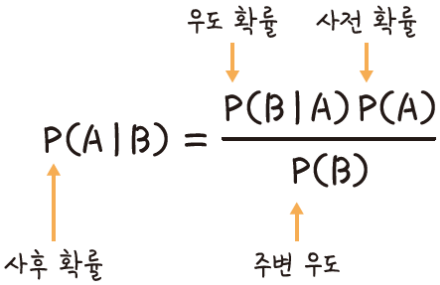
  (2) *calculating probabilities based on observed data* and
  (관찰된 데이터를 기반으로 가능도를 계산)

  (3) *correcting subjective probabilities initially established*
  (처음 설정한 주관적 확률을 보정함)

❖ Bayesian theory

Bayesian probability

- Theorems that show the relationship between the prior and posterior probabilities of two random variables, and if we know the prior probability P(A) and the likelihood probability P(B|A), we know the posterior probability P(A|B)
- Bayesian probability is expressed as a conditional probability as follows, and the posterior probability P(A|B) is obtained while updating the information



❖ Bayesian theory

- P(A), prior probability: the probability of cause A( ) determined before the result appears
  (결과가 나타나기 전에 결정된 원인A의 확률)
- P(B|A), likelihood probability: the probability that the outcome (B) will occur under the assumption that the cause (A) has occurred
  (원인(A)이 발생했다는 가정하에 결과(B)가 발생할 확률)
- P (A|B), Posterior probability: The probability that the cause (A) occurred under the assumption that the result (B) occurred
  (결과(B)가 발생했다는 가정하에 원인(A)이 발생했을 확률)
- P(B), peripheral likelihood (marginal probability): Expression probability of event (B)
  (사건(B)의 발현 확률)

❖ Bayesian theory

- The formula P(B) for Bayesian probability can be obtained as the intersection between A and A complements and P(B) as follows

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

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

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

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

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

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A')(P(A'))}$$

❖ Bayesian theory

- Let's give an example of Bayesian probability
  - 100 students in a class
  - 3 percent of female students are foreigners
  - 8 percent of the 70 percent of boys are foreigners
  - Let's find the probability that this student is a female student if the one randomly chosen from this class is a foreigner

❖ Bayesian theory

- Let's solve the problem using $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

$$
\begin{aligned}
P(\text{여학생} \mid \text{외국인}) &= \frac{P(\text{여학생} \cap \text{외국인})}{P(\text{외국인})} \\
&= \frac{P(\text{여학생} \cap \text{외국인})}{P(\text{여학생} \cap \text{외국인}) + P(\text{남학생} \cap \text{외국인})} \\
&= \frac{P(\text{여학생})\ P(\text{외국인} \mid \text{여학생})}{P(\text{여학생})\,P(\text{외국인} \mid \text{여학생}) + P(\text{남학생})\,P(\text{외국인} \mid \text{남학생})} \\
&= \frac{0.3 \times 0.03}{0.3 \times 0.03 + 0.7 \times 0.08} \\
&= \frac{0.009}{0.009 + 0.056} = \frac{0.009}{0.065} \\
&= 0.14
\end{aligned}
$$

❖ **Bayesian theory**

- If one randomly chosen person is a foreigner, the probability that this student is a female student is 0.14 (rounded from three decimal places)
- The most representative use of Bayesian probability is spam-mail filters
- Let's look at an example of using the text of mail to distinguish whether a mail is spam or not
- First of all, the presence or absence of spam in the mail can be checked

$$P(\text{정상메일} \mid \text{메일본문}) = \text{메일본문이 정상일 확률}$$

$$P(\text{스팸메일} \mid \text{메일본문}) = \text{메일본문이 스팸일 확률}$$

❖ Bayesian theory

- Using Naive Bayes to organize

$$P(\text{정상메일} \mid \text{메일본문}) = (P(\text{메일본문} \mid \text{정상메일}) \times P(\text{정상메일})) \,/\, P(\text{메일본문})$$

$$P(\text{스팸메일} \mid \text{메일본문}) = (P(\text{메일본문} \mid \text{스팸메일}) \times P(\text{스팸메일})) \,/\, P(\text{메일본문})$$

- If P (normal mail | mail text) is larger than P (spam mail | mail text) when input text is given, it is likely to be a normal mail, otherwise it is likely to be a spam mail

❖ **Bayesian theory**

- The preceding two expressions have P (mail text) in the denominator for both probabilities, so if you remove both, you can simplify as follows

$$P(\text{정상메일} \mid \text{메일본문}) = P(\text{메일본문} \mid \text{정상메일}) \times P(\text{정상메일})$$

$$P(\text{스팸메일} \mid \text{메일본문}) = P(\text{메일본문} \mid \text{스팸메일}) \times P(\text{스팸메일})$$

- For example, assuming that there are two words in the text, if the words are expressed as w1 and w2, the classification of normal mail and spam mail using Naive Bayes is as follows

$$P(\text{정상메일} \mid \text{메일본문}) = P(w1 \mid \text{정상메일}) \times P(w2 \mid \text{정상메일}) \times P(\text{정상메일})$$

$$P(\text{스팸메일} \mid \text{메일본문}) = P(w1 \mid \text{스팸메일}) \times P(w2 \mid \text{스팸메일}) \times P(\text{스팸메일})$$

❖ Bayesian theory

- For reference, all words in the mail body are changed to vectors for the computer to understand and used as inputs in the Naive Bayes classifier
- Ignore word order and consider frequency only

순서	메일 본문 단어들	분류
1	your free lottery	스팸
2	free lottery free you	스팸
3	your free apple	정상
4	free to contact me	정상
5	I won award	정상
6	my lottery ticket	스팸

❖ **Bayesian theory**

- Given the training data, let's find the probability of normal and spam mails in the input text my free lottery
- The probability of normal and spam mails in the input text can be obtained by applying the following formula

$$P(\text{정상메일} \mid \text{메일본문}) = P(\text{my} \mid \text{정상메일}) \times P(\text{free} \mid \text{정상메일}) \times P(\text{lottery} \mid \text{정상메일}) \times P(\text{정상메일})$$

$$P(\text{스팸메일} \mid \text{메일본문}) = P(\text{my} \mid \text{스팸메일}) \times P(\text{free} \mid \text{스팸메일}) \times P(\text{lottery} \mid \text{스팸메일}) \times P(\text{스팸메일})$$

❖ Bayesian theory

- The probability can be omitted because the number of normal and spam mails is the same

$$P(\text{정상메일}) = P(\text{스팸메일}) = \text{총 메일 여섯 개 중 세 개} = 0.5$$

- Since P(normal mail) and P(spam mail) have the same value, probability can be omitted for both expressions
- Probability can be obtained by applying the following formula

$$P(\text{정상메일} \mid \text{메일본문}) = P(\text{my} \mid \text{정상메일}) \times P(\text{free} \mid \text{정상메일}) \times P(\text{lottery} \mid \text{정상메일})$$

$$P(\text{스팸메일} \mid \text{메일본문}) = P(\text{my} \mid \text{스팸메일}) \times P(\text{free} \mid \text{스팸메일}) \times P(\text{lottery} \mid \text{스팸메일})$$

❖ Bayesian theory

- Here's how to get P(my | normal mail)

$$\frac{\text{정상메일에서 } my\text{가 등장한 총 빈도수}}{\text{정상메일에 등장한 모든 단어의 빈도수 총합}}$$

- In this case, $\frac{0}{10} = 0$
- Developing the equation with this principle is as follows

$$P(\text{정상메일} \mid \text{메일본문}) = \frac{0}{10} \times \frac{2}{10} \times \frac{0}{10} = 0$$

$$P(\text{스팸메일} \mid \text{메일본문}) = \frac{1}{10} \times \frac{3}{10} \times \frac{3}{10} = 0.009$$

- As a result, since P (normal mail | mail text) < P (spam mail | mail text), the input test "my free lottery" is classified as spam mail

Conditional probability

num	label	text	label_num
1	spam	your free lottery	1
2	spam	free lottery free you	1
3	ham	your free apple	0
4	ham	free to contact me	0
5	ham	I won award	0
6	spam	my lottery ticket	1

mean 평균

median 중앙값

mode 최빈값

❖ Bayesian theory

- Before practicing Bayesian probability examples, first install the following libraries from the Anaconda prompt

```
〉pip install pandas          또는  conda install pandas
〉pip install scikit-learn    또는  conda install scikit-learn
〉pip install nltk            또는  conda install nltk
〉import nltk                 또는  conda install nltk
〉nltk.download()
```

❖ Bayesian theory

- Implemented as follows

In [3]:

```
# 라이브러리를 호출합니다
import numpy as np
import pandas as pd
from nltk.corpus import stopwords
import string
import nltk

# Pandas 라이브러리를 호출하여 표 14-3의 csv 파일 데이터를
# 데이터프레임에 넣어 줍니다
df = pd.read_csv("./spam.csv")
df
```

❖ **Bayesian theory**

Out [3]:

	num	label	text	label_num
0	1	spam	your free lottery	1
1	2	spam	free lottery free you	1
2	3	ham	your free apple	0
3	4	ham	free to contact me	0
4	5	ham	I won award	0
5	6	spam	my lottery ticket	1

❖ Bayesian theory

- Select from text given only meaningful word tokens in the data
- Stopwords, such as surveys or suffixes, such as I, my, me, over, and on, should be removed and used, but the dataset in Example is omitted because there is not much data

In [4]:

```
def process_text(text):

    # text에서 구두점을 삭제합니다
    nopunc = [char for char in text if char not in string.punctuation]
    nopunc = ''.join(nopunc)

    # text에서 무의미한 단어(접미사, 조사 등)는 삭제합니다
    cleaned_words = [word for word in nopunc.split()
                     if word.lower() not in stopwords.words('english')]
    return cleaned_words
```

❖ **Bayesian theory**

```
# process_text 함수를 적용하여 데이터 세트의 텍스트 데이터를 토큰화합니다
df['text'].head().apply(process_text)

# text를 토큰 수의 행렬로 변환합니다
from sklearn.feature_extraction.text import CountVectorizer
messages_bow = CountVectorizer(analyzer=process_text).fit_transform
(df['text'])
```

❖ **Bayesian theory**

```
# 데이터를 80%의 training과 20%의 testing 데이터셋으로 분리합니다
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(messages_bow,
df['label_num'], test_size = 0.20, random_state = 0)

# 다항식 나이브베이즈 모델을 만들고 훈련시킵니다
from sklearn.naive_bayes import MultinomialNB
classifier = MultinomialNB()
classifier.fit(X_train, y_train)
```

❖ **Bayesian theory**

```
# 데이터셋 분류에 대한 예측 및 실제 관측 값을 보여 줍니다
print(classifier.predict(X_train)) # 예측 값 출력
print(y_train.values)              # 실제 관측 값 출력
```

```
[1 0 1 0]
[1 0 1 0]
```

❖ Bayesian theory

In [5]:

```
# 학습 데이터셋에서 모델의 정확도를 표현합니다
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix, accuracy_score

pred = classifier.predict(X_train) # 예측 값 출력

# 사이킷런 패키지의 metrics 패키지에서는 정밀도, 재현율, F1 점수를 구합니다
print(classification_report(y_train,pred))

# 혼동행렬로 표현합니다
print('Confusion Matrix: \n', confusion_matrix(y_train,pred))
print()
print('Accuracy: ', accuracy_score(y_train,pred)) # 정확도 점수로 표현
```

❖ **Bayesian theory**

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2
1	1.00	1.00	1.00	2
accuracy			1.00	4
macro avg	1.00	1.00	1.00	4
weighted avg	1.00	1.00	1.00	4

Continuous Variable

↳ word embedding → CoN ... → description

NLP (자연어 처리)

self-attention CV NC SP) transformer

❖ **Bayesian theory**

```
Confusion Matrix:
 [[2 0]
  [0 2]]

Accuracy:  1.0
```

```
In [6]:
# 테스트 데이터셋(X_test & y_test)에서 모델의 정확도를 테스트합니다
print('Predicted value: ', classifier.predict(X_test))
```

```
Predicted value: [1 1]
```


RAG

Retrival - Agment ..? — Generate.

❖ **Bayesian theory**

In [7]:

```
# 실제 관측 값 출력
print('Actual value: ', y_test.values)
```

```
Actual value: [1 0]
```

In [8]:

```
# 테스트 데이터셋에서 모델을 평가합니다
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix, accuracy_score
pred = classifier.predict(X_test)
print(classification_report(y_test,pred))
print('Confusion Matrix: \n', confusion_matrix(y_test,pred))
```

❖ **Bayesian theory**

```
print()
print('Accuracy: ', accuracy_score(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         1
           1       0.50      1.00      0.67         1

    accuracy                           0.50         2
   macro avg       0.25      0.50      0.33         2
weighted avg       0.25      0.50      0.33         2
```

❖ **Bayesian theory**

```
Confusion Matrix:
 [[0 1]
  [0 1]]

Accuracy:  0.5
```