

# 자연어 처리 – AI 전문가 양성(기본) Day 1

Naïve Bayes Classification

# 실습 소개

## ■ 실습 목표

- Naïve Bayes 알고리즘을 직접 구현하여 Spam 메일과 그렇지 않은 메일을 분류

## ■ 실습 내용

- 데이터 확인
  - 입력 예시)
    - "LookAtMe!: Thanks for your purchase of a video clip from LookAtMe!, you've been charged 35p. ..."
  - 출력 예시)
    - "spam" / "ham"
- Naïve Bayes 알고리즘을 위한 확률 계산
- Naïve Bayes 알고리즘 구현
- Test set에 대한 분류 성능 평가

# 데이터 소개

## ■ Data Analysis

- Columns: label(str), mail(str)
- train.csv: 5,014 rows (4334, 680)
- test.csv: 558 rows (491, 67)

label	mail
spam	LookAtMe!: Thanks for your purchase of a video clip from LookAtMe!, you've been charged 35p. Think you can do better? Why not send a video in a MMSto 32323.
ham	Aight, I'll hit you up when I get some cash
ham	Don no da:)whats you plan?
ham	Going to take your babe out ?

# Naïve Bayes Classification

- Bayes' theorem

- $P(A|B)$ 를 직접 구하지 못하거나,  $P(B|A)$ 를 쉽게 구할 수 있는 상황일 때.

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

- How to Classify?

- $P(\text{ham} | \text{sentence})$  = 입력 문장이 있을 때 정상 메일일 확률
- $P(\text{spam} | \text{sentence})$  = 입력 문장이 있을 때 **스팸** 메일일 확률

$$P(\text{ham}|\text{sentence}) = \frac{P(\text{sentence}|\text{ham}) * P(\text{ham})}{P(\text{sentence})}$$

# 실습 소개

## ▪ Naïve Bayes Classification

- Bayes 정리를 적용한 확률 값을 비교
- 확률 값이 높은 클래스를 해당 문장의 label이라고 예측

$$P(ham|sentence) = \frac{P(sentence|ham) * P(ham)}{P(sentence)}$$

$$P(spam|sentence) = \frac{P(sentence|spam) * P(spam)}{P(sentence)}$$

크기만 비교하면 된다.

따라서 공통된 분모는 제거할 수 있음!



$$P(ham|sentence) = P(sentence|ham) * P(ham)$$

$$P(spam|sentence) = P(sentence|spam) * P(spam)$$


# 실습 소개

## ▪ Naïve Bayes Classification

- 모든 단어의 등장은 독립적이라고 가정
- 이 때, 단어 순서를 무시한다

$$P(ham|sentence) = P(sentence|ham) * P(ham)$$

$$P(spam|sentence) = P(sentence|spam) * P(spam)$$

각 문장의 단어들의 등장 확률을 독립적으로 가정하고, 곱해줌 (순서 무시) 

$$P(ham|sentence) = P(word_1|ham) * P(word_2|ham) * P(word_3|ham) * ... * P(ham)$$

$$P(spam|sentence) = P(word_1|spam) * P(word_2|spam) * P(word_3|spam) * ... * P(spam)$$

# 실습 소개

- Code 작성 가이드

1. Prior 확률 계산:  $P(\text{ham}) / P(\text{spam})$
2. Likelihood 계산 - word count:  $P(\text{word}/\text{ham}) / P(\text{word}/\text{spam})$
3. Test 데이터 spam 메일 예측
4. 결과 리포팅