

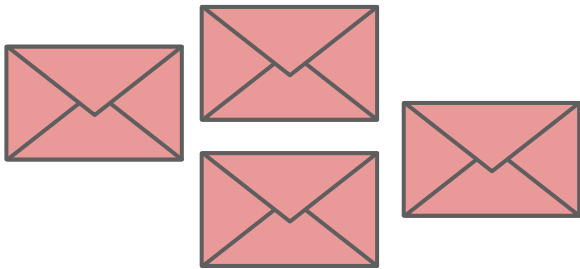
데이터 과학

L13: Naive Bayes

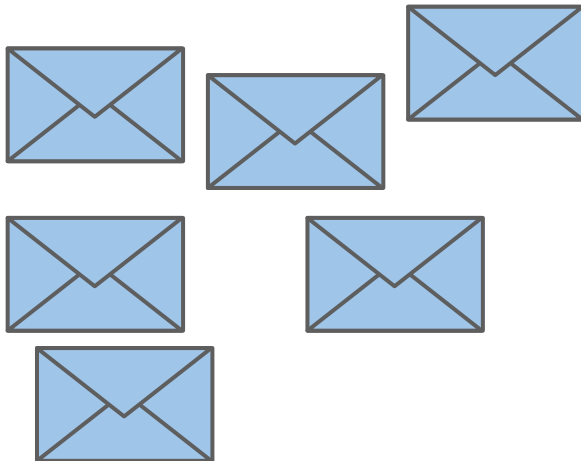
Kookmin University

스팸 필터

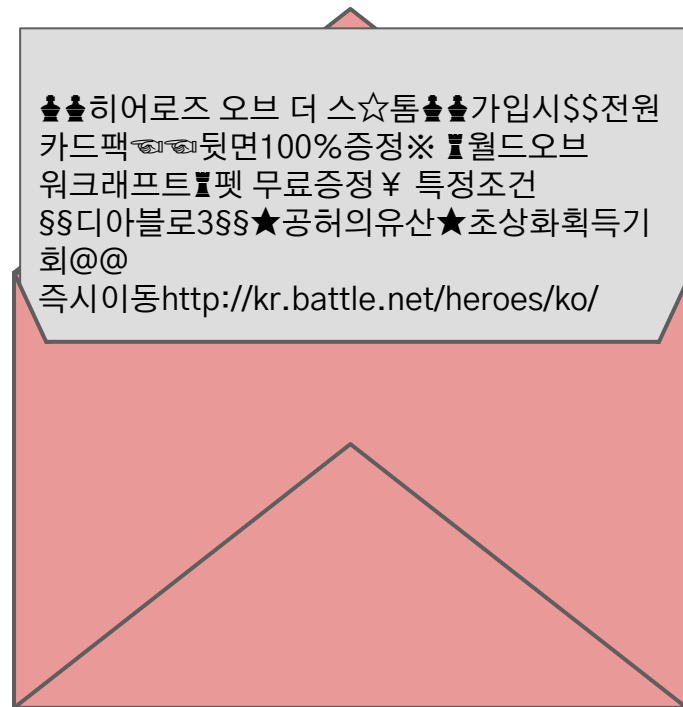
- 스팸 메일은 어떻게 분류 할 수 있을까?



스팸메일



정상메일



Naïve Bayes

- 통계적 분류기
- 각 분류별 확률 값을 계산
- 베이즈 정리에 따라 확률 계산
- 확률 계산의 단서들이 서로 조건부독립임을 가정
 - → 확률 계산이 단순해짐

$$\rightarrow P(A, B | c) = P(A | c) P(B | c)$$

Naïve Bayes 스팸 필터

새로운 메일이 왔다. 스팸인지 알아보려면?

- 스팸 메일과 일반 메일의 비율을 보고 판단
 - 예) 일반 메일 80%, 스팸 메일 20% →
아무런 단서가 없을 경우 일반 메일로 판단
- 메일에 포함된 단어들이 **스팸 메일**에 자주 나오는 단어인지, **일반 메일**에 자주 나오는 단어인지를 살펴보고 스팸 여부 판단

Naïve Bayes 스팸 필터

스팸을 확률

스팸 메일과 일반 메일들을 수집했다.

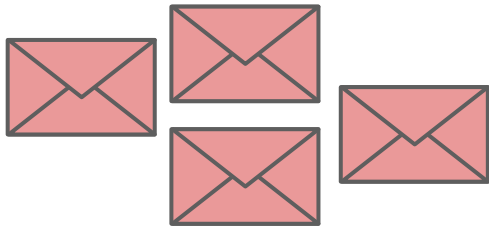
- $P(S)$: 스팸 메일과 일반 메일의 비율을 계산
- $P(w_i | S=\text{True})$: 각 단어가 스팸 메일에서 얼마나 자주 등장하는지 계산
- $P(w_i | S=\text{False})$: 각 단어가 일반 메일에서 얼마나 자주 등장하는지 계산

새로운 메일(M)이 왔다. 스팸인지 알아보려면?

- 스팸? $P(S = T) \prod_{w_i \in M} P(w_i | S = T)$
- 일반? $P(S = F) \prod_{w_i \in M} P(w_i | S = F)$

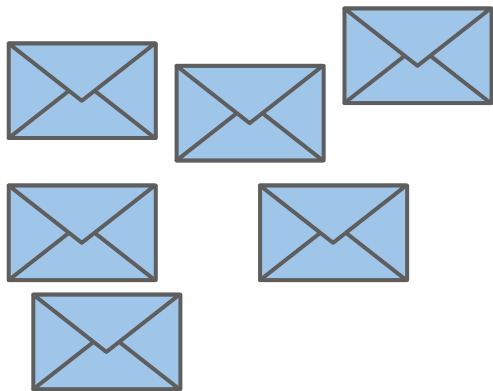
사전확률 Prior probability

- $P(S)$: 스팸 메일과 일반 메일의 비율을 계산



스팸메일

$$P(S=T) = 0.4$$

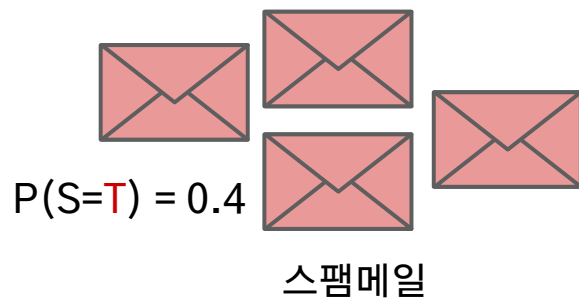


정상메일

$$P(S=F) = 0.6$$

가능도 Likelihood

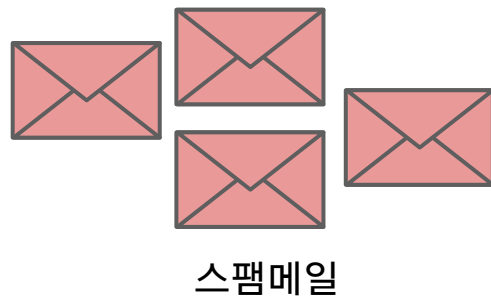
- $P(w_i | S=\text{True})$: 각 단어가 **스팸 메일**에서 얼마나 자주 등장하는지 계산



단어	빈도
룰	1
시공	5
조아	3
ㄱㄱ	0
폭풍	4
접속	2

가능도 Likelihood

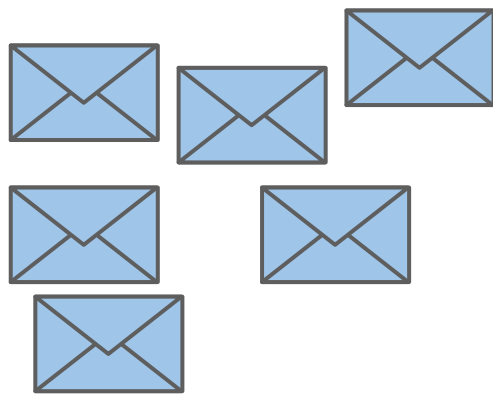
- $P(w_i | S=\text{True})$: 각 단어가 **스팸 메일**에서 얼마나 자주 등장하는지 계산



단어	빈도	가능도
롤	■	$P(\text{롤} S=\text{T}) = 1/15$
시공	■ ■ ■ ■ ■	$P(\text{시공} S=\text{T}) = 5/15$
조아	■ ■ ■	$P(\text{조아} S=\text{T}) = 3/15$
ㄱ ㄱ		$P(\text{옴치} S=\text{T}) = 0/15$
폭풍	■ ■ ■ ■	$P(\text{폭풍} S=\text{T}) = 4/15$
접속	■ ■	$P(\text{접속} S=\text{T}) = 2/15$

가능도 Likelihood

- $P(w_i | S=\text{False})$: 각 단어가 일반 메일에서 얼마나 자주 등장하는지 계산



정상메일

단어	빈도	가능도
롤	4	$P(\text{롤} S=\text{F}) = 4/18$
시공	1	$P(\text{시공} S=\text{F}) = 1/18$
조아	2	$P(\text{조아} S=\text{F}) = 2/18$
ㅋㅋ	6	$P(\text{옹치} S=\text{F}) = 6/18$
폭풍	1	$P(\text{폭풍} S=\text{F}) = 1/18$
접속	4	$P(\text{접속} S=\text{F}) = 4/18$

스팸 분류하기

- 새로운 메일이 왔다. 스팸인지 알아보려면?

메일내용: 시공 조아 폭풍 조아

이 메일이 **스팸메일**일 확률

$$= P(S=T \mid \text{시공, 조아, 폭풍, 조아})$$

베이즈 정리에 의해:

$$= P(\text{시공, 조아, 폭풍, 조아} \mid S=T)P(S=T) / P(\text{시공, 조아, 폭풍, 조아})$$

나이브 베이즈의 조건부독립 가정에 의해:

$$= P(S=T)P(\text{시공} \mid S=T)P(\text{조아} \mid S=T)P(\text{폭풍} \mid S=T)P(\text{조아} \mid S=T) / P(\text{시공, 조아, 폭풍, 조아})$$

불구당

이 메일이 **일반메일**일 확률

$$= P(S=F)P(\text{시공} \mid S=F)P(\text{조아} \mid S=F)P(\text{폭풍} \mid S=F)P(\text{조아} \mid S=F) / P(\text{시공, 조아, 폭풍, 조아})$$

참고: 베이즈 정리 Bayes' theorem

- 두 확률 변수의 사전확률과 사후확률 사이의 관계를 나타내는 정리

The diagram illustrates the components of Bayes' theorem. At the top, '가능도' (Likelihood) and '사전확률' (Prior) are labeled. Arrows point from '가능도' to the numerator's first term $P(B|A)$ and from '사전확률' to the numerator's second term $P(A)$. Another arrow points from '사전확률' to the denominator $P(B)$. Below the equation, '사후확률' (Posterior) is labeled with an arrow pointing to $P(A|B)$.

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

스팸 분류하기

- 새로운 메일이 왔다. 스팸인지 알아보려면?

메일내용: 시공 조아 폭풍 조아

스팸메일 가능성: $P(S=T)P(\text{시공}|S=T)P(\text{조아}|S=T)P(\text{폭풍}|S=T)P(\text{조아}|S=T)$

vs

일반메일 가능성: $P(S=F)P(\text{시공}|S=F)P(\text{조아}|S=F)P(\text{폭풍}|S=F)P(\text{조아}|S=F)$

$P(S=T) = 0.4$

단어	빈도	가능도
롤	■	$P(\text{롤} S=T) = 1/15$
시공	■■■■■	$P(\text{시공} S=T) = 5/15$
조아	■■■	$P(\text{조아} S=T) = 3/15$
ㅋㅋ		$P(\text{옴치} S=T) = 0/15$
폭풍	■■■■	$P(\text{폭풍} S=T) = 4/15$
접속	■■	$P(\text{접속} S=T) = 2/15$

$P(S=F) = 0.6$

단어	빈도	가능도
롤	■■■■	$P(\text{롤} S=F) = 4/18$
시공	■	$P(\text{시공} S=F) = 1/18$
조아	■■	$P(\text{조아} S=F) = 2/18$
ㅋㅋ	■■■■■	$P(\text{옴치} S=F) = 6/18$
폭풍	■	$P(\text{폭풍} S=F) = 1/18$
접속	■■■■	$P(\text{접속} S=F) = 4/18$

스팸 분류하기

- 새로운 메일이 왔다. 스팸인지 알아보려면?

메일내용: 시공 조아 폭풍 조아

스팸메일 가능성: $0.4 \times (5/15) \times (3/15) \times (4/15) \times (3/15) = 0.00142222$

VS

일반메일 가능성: $0.6 \times (1/18) \times (2/18) \times (1/18) \times (2/18) = 0.00002286$

P(S=T) = 0.4			P(S=F) = 0.6		
단어	빈도	가능도	단어	빈도	가능도
롤	■	$P(\text{롤} S=T) = 1/15$	롤	■■■■	$P(\text{롤} S=F) = 4/18$
시공	■■■■■	$P(\text{시공} S=T) = 5/15$	시공	■	$P(\text{시공} S=F) = 1/18$
조아	■■■	$P(\text{조아} S=T) = 3/15$	조아	■■	$P(\text{조아} S=F) = 2/18$
ㅋㅋ		$P(\text{옴치} S=T) = 0/15$	ㅋㅋ	■■■■■	$P(\text{옴치} S=F) = 6/18$
폭풍	■■■■	$P(\text{폭풍} S=T) = 4/15$	폭풍	■	$P(\text{폭풍} S=F) = 1/18$
접속	■■	$P(\text{접속} S=T) = 2/15$	접속	■■■■	$P(\text{접속} S=F) = 4/18$

스팸 분류하기

- 또다른 메일이 왔다. 스팸인지 알아보려면?

메일내용: 시공 시공 시공 ㄱㄱ

스팸메일 가능성: $P(S=T)P(\text{시공}|S=T)P(\text{시공}|S=T)P(\text{시공}|S=T)P(\text{ㄱㄱ}|S=T)$

VS

일반메일 가능성: $P(S=F)P(\text{시공}|S=F)P(\text{시공}|S=F)P(\text{시공}|S=F)P(\text{ㄱㄱ}|S=F)$

$P(S=T) = 0.4$			$P(S=F) = 0.6$		
단어	빈도	가능도	단어	빈도	가능도
롤	■	$P(\text{롤} S=T) = 1/15$	롤	■■■■	$P(\text{롤} S=F) = 4/18$
시공	■■■■■	$P(\text{시공} S=T) = 5/15$	시공	■	$P(\text{시공} S=F) = 1/18$
조아	■■■	$P(\text{조아} S=T) = 3/15$	조아	■■	$P(\text{조아} S=F) = 2/18$
ㄱㄱ		$P(\text{옴치} S=T) = 0/15$	ㄱㄱ	■■■■■	$P(\text{옴치} S=F) = 6/18$
폭풍	■■■■	$P(\text{폭풍} S=T) = 4/15$	폭풍	■	$P(\text{폭풍} S=F) = 1/18$
접속	■■	$P(\text{접속} S=T) = 2/15$	접속	■■■■	$P(\text{접속} S=F) = 4/18$

스팸 분류하기

- 또다른 메일이 왔다. 스팸인지 알아보려면?

메일내용: 시공 시공 시공 ㄱㄱ

스팸메일 가능성: $0.4 \times (5/15) \times (5/15) \times (5/15) \times (0/15) = 0$

VS

일반메일 가능성: $0.6 \times (1/18) \times (1/18) \times (1/18) \times (6/18) = 0.00003429$

P(S=T) = 0.4			P(S=F) = 0.6		
단어	빈도	가능도	단어	빈도	가능도
롤	■	$P(\text{롤} S=T) = 1/15$	롤	■■■■	$P(\text{롤} S=F) = 4/18$
시공	■■■■■	$P(\text{시공} S=T) = 5/15$	시공	■	$P(\text{시공} S=F) = 1/18$
조아	■■■	$P(\text{조아} S=T) = 3/15$	조아	■■	$P(\text{조아} S=F) = 2/18$
ㄱㄱ		$P(\text{옴치} S=T) = 0/15$	ㄱㄱ	■■■■■	$P(\text{옴치} S=F) = 6/18$
폭풍	■■■■	$P(\text{폭풍} S=T) = 4/15$	폭풍	■	$P(\text{폭풍} S=F) = 1/18$
접속	■■	$P(\text{접속} S=T) = 2/15$	접속	■■■■	$P(\text{접속} S=F) = 4/18$

라플라스 스무딩 (Smoothing)

- 기존 스팸메일이나 일반메일에서 한 번도 등장하지 않은 단어가 나올 경우 계산 결과가 이상해짐
- 스무딩: 모든 단어가 일반/스팸메일에 한 번씩은 등장했다고 가정 (한 번씩 $\rightarrow \alpha$ 번씩)

P(S=T) = 0.4			P(S=F) = 0.6		
단어	빈도	가능도	단어	빈도	가능도
롤	■ ■	$P(\text{롤} S=T) = 2/21$	롤	■ ■ ■ ■ ■	$P(\text{롤} S=F) = 5/24$
시공	■ ■ ■ ■ ■ ■	$P(\text{시공} S=T) = 6/21$	시공	■ ■	$P(\text{시공} S=F) = 2/24$
조아	■ ■ ■ ■	$P(\text{조아} S=T) = 4/21$	조아	■ ■ ■	$P(\text{조아} S=F) = 3/24$
ㄱㄱ	■	$P(\text{옹치} S=T) = 1/21$	ㄱㄱ	■ ■ ■ ■ ■ ■ ■	$P(\text{옹치} S=F) = 7/24$
폭풍	■ ■ ■ ■ ■ ■	$P(\text{폭풍} S=T) = 5/21$	폭풍	■ ■	$P(\text{폭풍} S=F) = 2/24$
접속	■ ■ ■	$P(\text{접속} S=T) = 3/21$	접속	■ ■ ■ ■ ■	$P(\text{접속} S=F) = 5/24$

스팸 분류하기

- 스무딩 후 다시 메일을 분류해보면...

메일내용: 시공 시공 시공 ㄱㄱ

스팸메일 가능성: $0.4 \times (6/21) \times (6/21) \times (6/21) \times (1/21) = 0.0004442$

VS

일반메일 가능성: $0.6 \times (2/24) \times (2/24) \times (2/24) \times (7/24) = 0.0001012$

P(S=T) = 0.4			P(S=F) = 0.6		
단어	빈도	가능도	단어	빈도	가능도
롤	■ ■	$P(\text{롤} S=T) = 2/21$	롤	■ ■ ■ ■ ■ ■	$P(\text{롤} S=F) = 5/24$
시공	■ ■ ■ ■ ■ ■ ■	$P(\text{시공} S=T) = 6/21$	시공	■ ■	$P(\text{시공} S=F) = 2/24$
조아	■ ■ ■ ■	$P(\text{조아} S=T) = 4/21$	조아	■ ■ ■	$P(\text{조아} S=F) = 3/24$
ㄱㄱ	■	$P(\text{옴치} S=T) = 1/21$	ㄱㄱ	■ ■ ■ ■ ■ ■ ■ ■	$P(\text{옴치} S=F) = 7/24$
폭풍	■ ■ ■ ■ ■ ■	$P(\text{폭풍} S=T) = 5/21$	폭풍	■ ■	$P(\text{폭풍} S=F) = 2/24$
접속	■ ■ ■	$P(\text{접속} S=T) = 3/21$	접속	■ ■ ■ ■ ■ ■	$P(\text{접속} S=F) = 5/24$

언더플로우

- 메일에 단어가 많을 경우 가능성이 0으로 수렴
- 너무 0에 가까워지면 컴퓨터 연산의 특성상 정확도가 떨어짐 → 언더플로우
- Log를 활용하여 개선가능

$$\text{Log}(A*B*C) = \text{Log}(A) + \text{Log}(B) + \text{Log}(C)$$













언더플로우

- 로그를 사용하여 다시 계산해보면?

메일내용: 롤 접속 ㄱㄱ

$$\begin{aligned}\text{스팸: } & \log(P(S=\textcolor{red}{T})) + \log(P(\text{롤}|S=\textcolor{red}{T})) + \log(P(\text{접속}|S=\textcolor{red}{T})) + \log(P(\text{ㄱㄱ}|S=\textcolor{red}{T})) \\ & = (-0.92) + (-2.35) + (-1.95) + (-3.04) = -8.26\end{aligned}$$

$$\begin{aligned}\text{일반: } & \log(P(S=\textcolor{blue}{F})) + \log(P(\text{롤}|S=\textcolor{blue}{F})) + \log(P(\text{접속}|S=\textcolor{blue}{F})) + \log(P(\text{ㄱㄱ}|S=\textcolor{blue}{F})) \\ & = (-0.51) + (-1.57) + (-1.57) + (-1.23) = -4.88\end{aligned}$$

$P(S=\textcolor{red}{T}) = 0.4$			$P(S=\textcolor{blue}{F}) = 0.6$		
단어	빈도	가능도	단어	빈도	가능도
롤		$P(\text{롤} S=\textcolor{red}{T}) = 2/21$	롤		$P(\text{롤} S=\textcolor{blue}{F}) = 5/24$
시공		$P(\text{시공} S=\textcolor{red}{T}) = 6/21$	시공		$P(\text{시공} S=\textcolor{blue}{F}) = 2/24$
조아		$P(\text{조아} S=\textcolor{red}{T}) = 4/21$	조아		$P(\text{조아} S=\textcolor{blue}{F}) = 3/24$
ㄱㄱ		$P(\text{옴치} S=\textcolor{red}{T}) = 1/21$	ㄱㄱ		$P(\text{옴치} S=\textcolor{blue}{F}) = 7/24$
폭풍		$P(\text{폭풍} S=\textcolor{red}{T}) = 5/21$	폭풍		$P(\text{폭풍} S=\textcolor{blue}{F}) = 2/24$
접속		$P(\text{접속} S=\textcolor{red}{T}) = 3/21$	접속		$P(\text{접속} S=\textcolor{blue}{F}) = 5/24$

Questions?