

선수 지식 - 통계

베이즈 정리

베이즈 정리 | 딥러닝의 기초가 되는 확률 개념 알아보기

강사 나동빈

선수 지식 - 통계

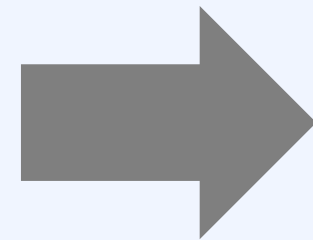
베이지 정리

문제 상황 - 스팸 분류 모델

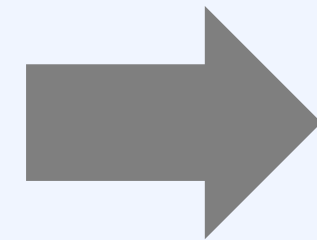
- 하나의 소프트웨어를 만들고 싶다고 가정하자.
- 입력: 하나의 텍스트(text)
- 출력: 텍스트가 특정 클래스(스팸 혹은 햄)에 속할 확률
- **목표)** 하나의 텍스트(text) x 가 스팸(spam) y 일 확률 계산



텍스트(text)



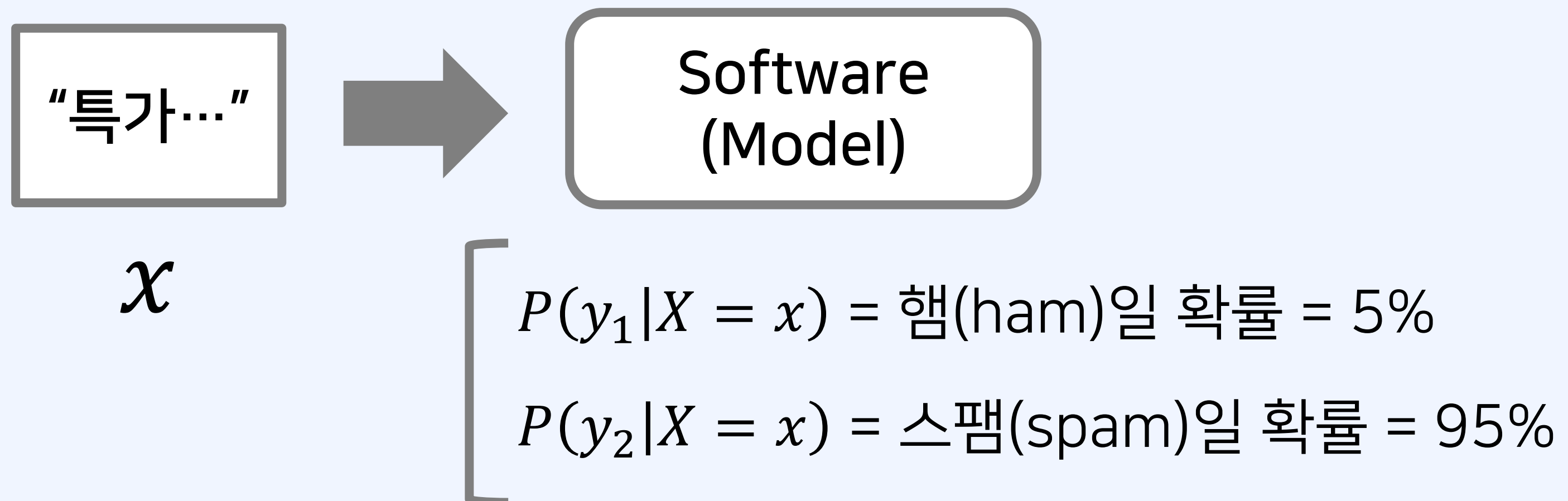
분류 모델



스팸(spam) or 햄(ham)

문제 상황 - 스팸 분류 모델

- **목표)** 하나의 텍스트(text) x 가 스팸(spam)일 확률 계산
- 이때 텍스트의 확률 변수를 X , 클래스의 확률 변수를 Y 라고 하자.
- 클래스는 오직 두 개만 존재한다고 가정하자. ($y_1 = \text{햄}$, $y_2 = \text{스팸}$)



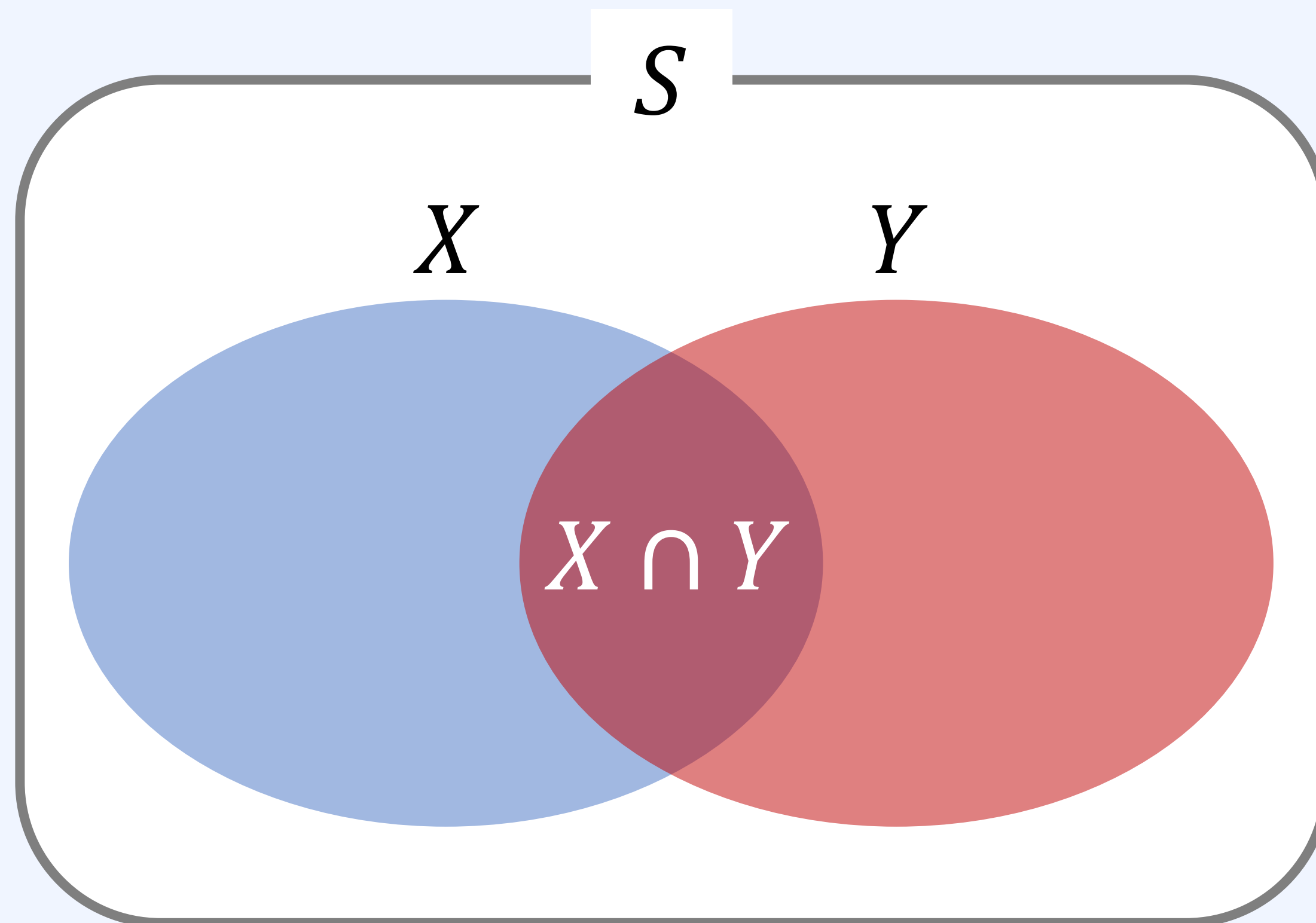
문제 상황 - 스팸 분류 모델

- 지금까지 받은 메일을 확인해 보았더니, 70%는 스팸 메일, 30%는 정상 메일이었다.
- 스팸 메일 중에 90%는 "대출"이라는 단어가 포함되어 있었다.
- 정상 메일 중에 3%는 "대출"이라는 단어가 포함되어 있었다.

Q. 이때, "대출"이라는 단어가 들어있는 메일이 스팸 메일일 확률을 구해보자.

[배경지식] 조건부 확률(Conditional Probability)

- 사건 X 가 발생했을 때, 사건 Y 가 발생할 확률을 의미한다.
- $P(Y|X) = n(X \cap Y)/n(X)$



베이지 정리(Bayes' Theorem)

- $P(Y|X)$ 를 계산할 수 있으면, 우리가 원하는 프로그램을 만들 수 있다.
- 하지만 $P(Y|X)$ 를 직접적으로 계산하는 것이 어려울 때, 베이지 정리에 기반한 방법을 이용한다.

베이지 정리 공식

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

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

- $P(X)$: 특정 텍스트가 나올 확률
- $P(Y)$: 특정 클래스가 나올 확률
- $P(X|Y)$: 특정 클래스에서 특정 텍스트가 나올 확률
- $P(Y|X)$: 특정 텍스트가 특정 클래스에서 나올 확률

베이지 정리(Bayes' Theorem)

- 베이지 정리란, 조건부확률을 구하는 공식이다.
- 베이지 정리를 이용하면, 다양한 확률 문제를 해결할 수 있다.
- 베이지 정리 공식: $P(A|B) = P(B|A)P(A)/P(B)$

[베이지 정리 유도]

- 조건부 확률 정의: $P(A|B) = P(A, B)/P(B) \rightarrow P(A, B) = P(A|B)P(B)$
- 마찬가지로, $P(B|A) = P(A, B)/P(A) \rightarrow P(A, B) = P(B|A)P(A)$
- 따라서 $P(A, B) = \mathbf{P(A|B)P(B) = P(B|A)P(A)}$
- 그러므로, $P(A|B) = P(B|A)P(A)/P(B)$

문제 해결 - 스팸 분류 모델

- 가지고 있는 정보를 활용한다.
- 지금까지 받은 메일을 확인해 보았더니, 70%는 스팸 메일, 30%는 정상 메일이었다.
- 스팸 메일 중에 90%는 "대출"이라는 단어가 포함되어 있었다.
- 정상 메일 중에 3%는 "대출"이라는 단어가 포함되어 있었다.

Q. 이때, "대출"이라는 단어가 들어있는 메일이 스팸 메일일 확률을 구해보자.

문제 해결 - 스팸 메일 분류 문제

- 주어진 정보는 다음과 같다.
- $P(\text{스팸}) = 0.7$
- $P(\text{정상}) = 0.3$
- $P(\text{대출}|\text{스팸}) = 0.9$
- $P(\text{대출}|\text{정상}) = 0.03$
- [알고자 하는 것] $P(\text{스팸}|\text{대출})$ 이다.
- 베이즈 정리에 따르면, $P(\text{스팸}|\text{대출}) = P(\text{대출}|\text{스팸})P(\text{스팸})/P(\text{대출})$
- 따라서 먼저 $P(\text{대출})$ 을 구해야 한다.

베이지 정리 - 스팸 메일 분류 문제

- 가장 먼저 $P(\text{대출})$ 을 구해야 한다.
- $$\begin{aligned} P(\text{대출}) &= P(\text{대출} \cap \text{스팸}) + P(\text{대출} \cap \text{정상}) \\ &= P(\text{대출}|\text{스팸})P(\text{스팸}) + P(\text{대출}|\text{정상})P(\text{정상}) \\ &= 0.9 * 0.7 + 0.03 * 0.3 = 0.639 \end{aligned}$$

[정답] $P(\text{스팸}|\text{대출}) = P(\text{대출}|\text{스팸})P(\text{스팸})/P(\text{대출}) = 0.9 * 0.7 / 0.639$

확률 모델(Probabilistic Models)

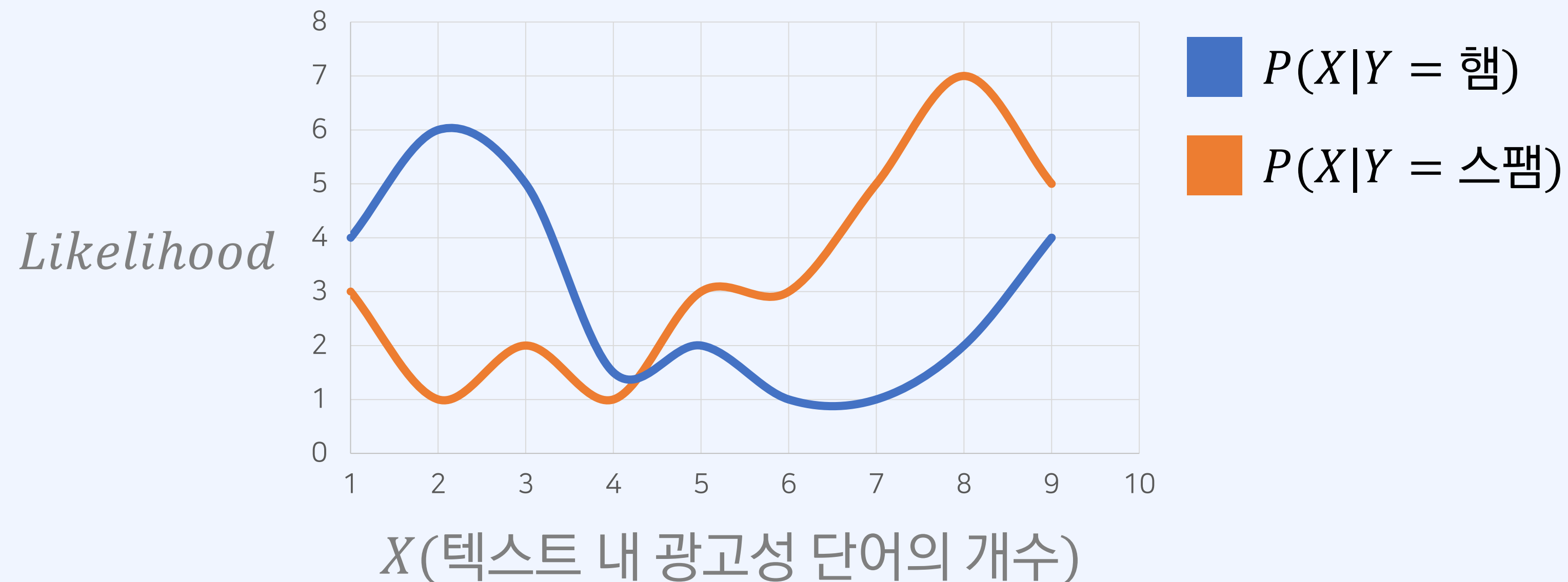
- 일반적인 분류 모델 $P(y|x)$ 은 다음과 같은 공식으로 예측 결과 \hat{y} 를 계산한다.
- $\hat{y} = \operatorname{argmax}_y P(y|x)$

황금 법칙(Golden Rule)

- 텍스트 x 가 주어졌을 때, 이 텍스트는 어떤 클래스(y)로 분류될까?
- $\operatorname{argmax}_y P(y|x) = \operatorname{argmax}_y P(x|y)P(y)/P(x) = \operatorname{argmax}_y P(x|y)P(y)$
- 따라서 분모에 해당하는 $P(x)$ 를 고려하지 않아도 된다.

최대 우도 추정(Maximum Likelihood Estimation)

- 가능도(likelihood)가 가장 높은 클래스를 선택하는 방법이다.
- X 는 특징(feature) 혹은 데이터를 말한다.
- X = 광고성 단어의 개수("특가" 등)라고 해보자.
- 현재 예시에선 $X \geq 5$ 라면 스팸으로 분류하면 된다.



최대 우도 추정 시 유의할 점

- 우리는 사후 확률(posterior)을 직접적으로 계산하기 어렵기 때문에 가능도를 이용한다.
- 하지만 가능도만으로 사후확률을 완전히 근사할 수 없다.

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

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

- 만약 *prior*가 *uniform distribution*을 따르지 않는다면?
 - 전체 메일 중에서 스팸 메일의 수 자체가 적다고 하자.
 - $P(Y = \text{스팸}) = 1/3$, $P(Y = \text{햄}) = 2/3$
 - 분포는 어떻게 바뀔까?

최대 우도 추정(Maximum Likelihood Estimation)

- 전체 메일 중에서 스팸 메일의 수 자체가 적다고 해보자.
- $P(Y = \text{스팸}) = 1/3$, $P(Y = \text{햄}) = 2/3$
- 이를 반영하면, 더욱 정교한 분류가 가능하다.
- 이제는 $X \geq 6$ 이라면 스팸으로 분류하면 된다.

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$



■ $P(X|Y = \text{햄})P(Y = \text{햄})$

■ $P(X|Y = \text{스팸})P(Y = \text{스팸})$

베이지 정리(Bayes' Theorem)

- 정리하면, *prior*를 고려할 때 *posterior*를 더욱 잘 계산할 수 있다.
- 베이지 정리는 머신러닝 분야에서 끊임없이 등장하는 개념이므로 중요하다.
- 나이브 베이지 분류기 (Naïve Bayes Classifier)에서 사용된다.
- 최신 인공지능 기술보다 성능은 많이 뒤떨어지지만 기본적인 모델로 자주 언급된다.