

나이브 베이즈

Naïve Bayes

나이트베이즈 개념

- 베이즈 정리
- 가능도 Likelihood
- 가우시안 나이트베이즈

베이즈 정리

- 나이브 베이즈 알고리즘의 기본이 되는 개념
- 두 확률 변수의 사전 확률과 사후 확률 사이의 관계를 나타내는 정리
- 사건 A, B가 있을 때, 사건 B가 일어난 것을 전제로 한 사건 A의 조건부 확률을 구하고자 한다
- 하지만 현재 가지고 있는 정보는 사건 A가 일어난 것을 전제로 한 사건 B의 조건부 확률과 A의 확률, B의 확률 뿐이다.
- 이 때, 원래 구하고자 했던 것을 다음과 같이 구할 수 있다는 것이 베이즈 정리이다.

베이즈 정리

$$P(A_k|B)$$
$$= \frac{P(A_k \cap B)}{P(B)}$$

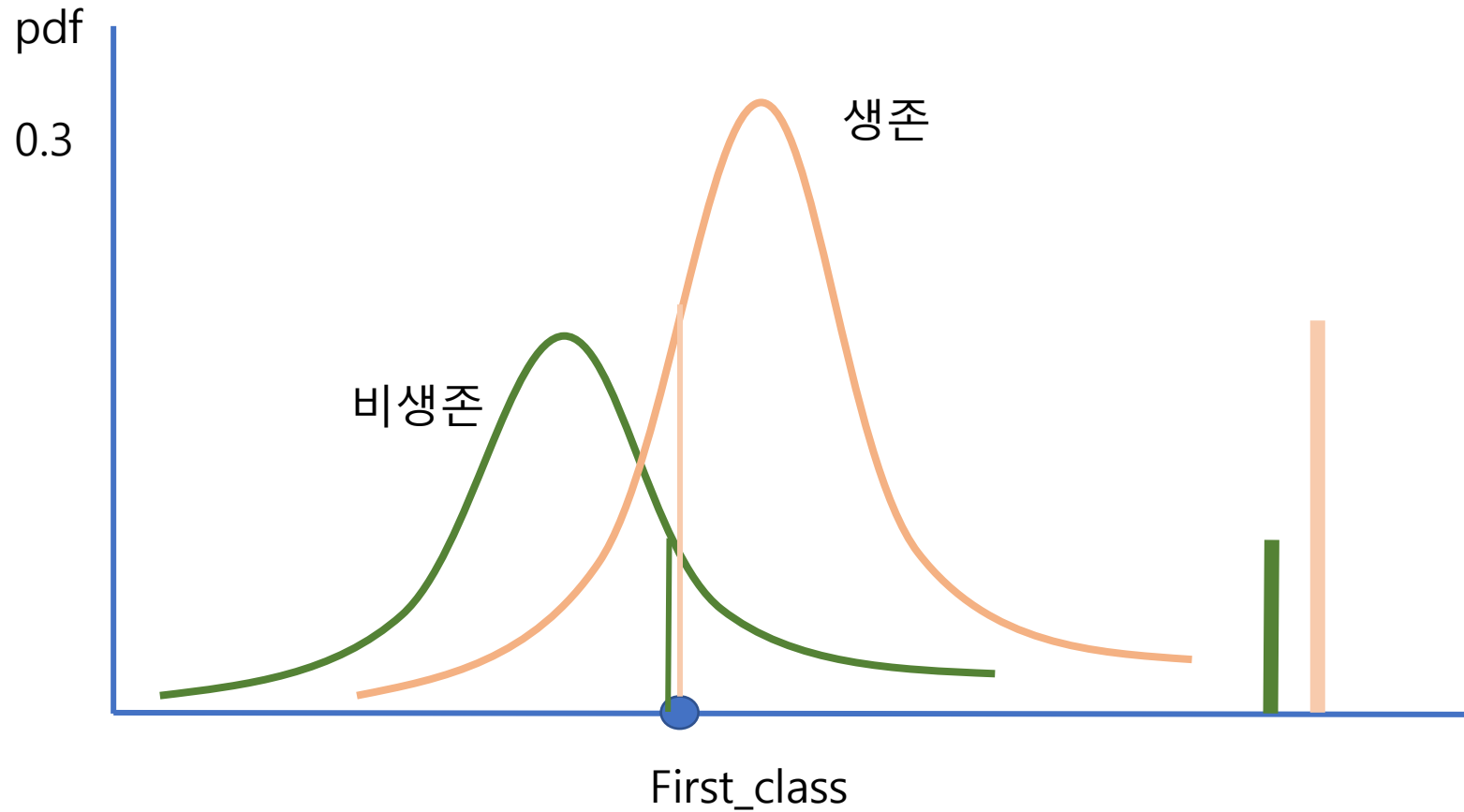
- $P(A|B)$: 사건 B가 발생했을 때 사건 A가 발생할 확률 = 사후확률(posterior)
- $P(B|A)$: 사건 A가 발생했을 때 사건 B가 발생할 확률 = 우도(likelihood)
- $P(A \cap B)$: 사건 A와 B가 동시에 발생할 확률
- $P(A)$: 사건 A가 발생할 확률 = 사전확률 (prior)
- $P(B)$: 사건 B가 발생할 확률 = 관찰 값(evidence)

$$= \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)}$$

$$= \frac{P(B|A_k)P(A_k)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}$$

(단 k 가 $k = 1, 2, 3, \cdots, n$)

가능도 (likelihood, 우도)



•나이브베이지즈

$$\begin{aligned} p(C_k|x_1, \dots, x_n) &\propto p(C_k, x_1, \dots, x_n) \\ &\propto p(C_k) p(x_1|C_k) p(x_2|C_k) p(x_3|C_k) \cdots \\ &\propto p(C_k) \prod_{i=1}^n p(x_i|C_k) . \end{aligned}$$

- 나이브 베이지즈는 예측에 사용되는 독립변수들이 상호 독립이라는 가정하에 확률 계산을 단순화함
- 모든 특성들이 분류하는데 동등한 역할을 한다는 것
- X가 클래스 전체의 확률 분포에 대비하여 특정 클래스에 속할 확률을 베이지즈 정리를 기반으로 계산

나이브 베이즈 분류 종류

1. GaussianNB: 정규분포 나이브베이즈

- 독립변수가 연속형일 때

2. BernoulliNB: 베르누이분포 나이브베이즈

- 독립변수가 이산형일 때

3. MultinomialNB: 다항분포 나이브베이즈

- 독립변수가 이산형이 아닌, 범주형일 때

•가우시안 나이브베이지즈

- 분류 문제에서는 주로 가우시안 나이브베이지즈를 사용
- 가우시안은 가우스 분포 즉, 정규분포상에서 발생확률을 계산함
(일상생활의 대부분은 정규분포를 이룬다는 개념)
- 정규분포 상에서의 확률 (likelihood)를 구해서 계산함