

나이브 베이즈 항상은

1- (1)

$$P(\text{comedy} | \text{문서}) = P(\text{comedy}) \cdot P(\text{fun} | \text{comedy}) \cdot P(\text{fast} | \text{comedy}) \cdot P(\text{furious} | \text{comedy}) = 0$$

$$P(\text{action} | \text{문서}) = P(\text{action}) \cdot P(\text{fun} | \text{action}) \cdot P(\text{fast} | \text{action}) \cdot P(\text{furious} | \text{action}) = 0.0018$$

$\therefore$  문서는 0.0018의 확률로 action으로 분류된다.

1- (2)

등장하지 않아 0이되는 feature들을 보정해주는 라플라스 스무딩을 이용하자. (+1을 해주어 확률이 0이 되지 못하게 한다.)

ex) comedy 문서에 furious 라는 feature은 0번 나옴.

조건부 확률이라는 가정 자체가 비현실적이다.

## 과제

1. 다음의 나이브베이즈 문제를 풀어보세요.

문서번호	주요단어	문서분류
1	fun, couple, love, love	comedy
2	fast, furious, shoot	action
3	couple, fly, fast, fun, fun	comedy
4	furious, shoot, shoot, fun	action
5	fly, fast, shoot, love	action

1.1 입력문서가 {fast, furious, fun}만을 주요단어로 가질 때, 이 문서는 얼마의 확률로 어떤 문서로 분류되는가?

1.2 어떠한 문제점이 있고, 이를 해결하기 위해 어떻게 할 것인가? (방법론만 제시)

\*식과 함께 답을 같이 제출해주세요~

\*TIP: 나이브 베이즈 함수가 어떻게 생겼죠??

