Navie Bayes Classification
(나이브 베이즈 분류)

예제로 시작합시다.

남자일 확률, P(A) = 0.6

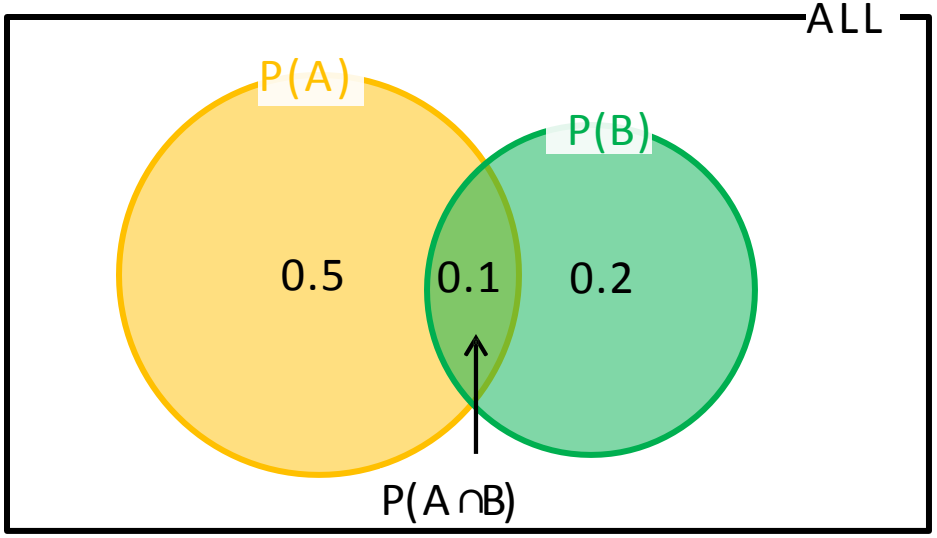
키가 180 이상일 확률, P(B) = 0.3

남자이고 키가 180 cm 이상일 확률, $P(A \cap B) = 0.1$

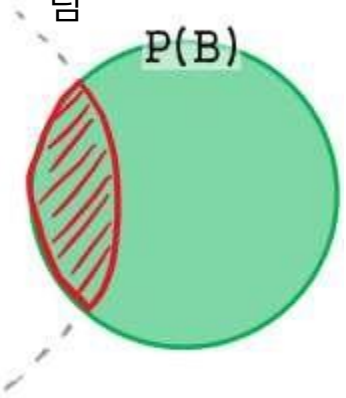
그렇다면,

키가 180 cm 일 때, 남자일 확률은?

(Probability of A given B?)



힌트 : 아래 그림



정답은 다음 페이지에,

확률, probability

확률, probability

P(A), A 가 일어날 확률 (probability of A)

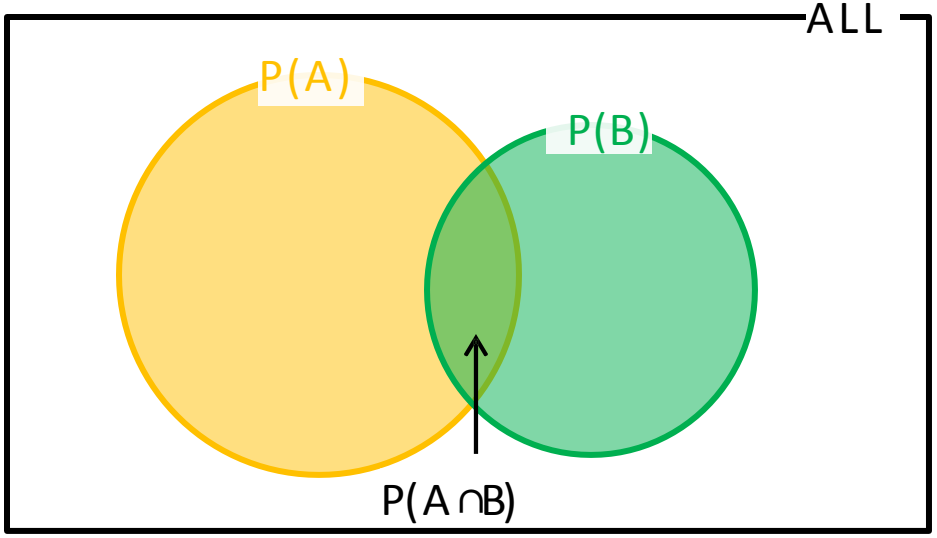
조건부 확률, conditional probability

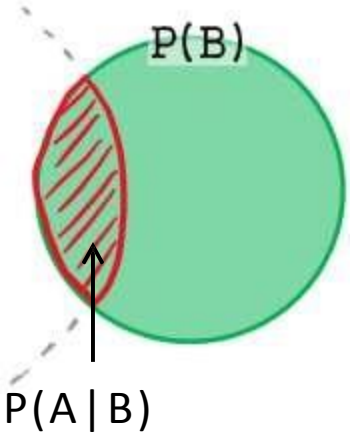
P(A|B), B가 주어졌을 때, A가 일어날 확률 (probability of A given B)

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

P(B)가 주어졌을 때, (P(B)의 확률이 1이라 보고)

붉은 부분의 확률이 P(A|B) 이다





따라서 키가 180cm일 때 남자일 확률은, P(A|B)

0.1 / 0.3 = 0.33, 즉 33.3%

Bayes Rule

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

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}, \text{ Bayes Rule}$$

Thomas Bayes(1701-1761)

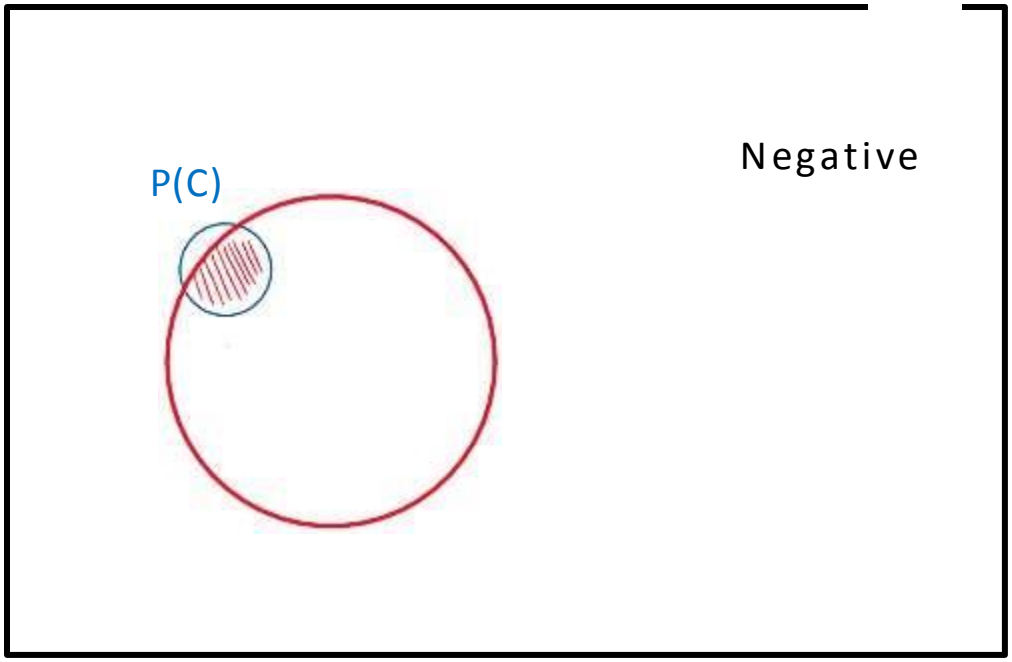
Bayes Rule, $P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$

예제 하나 더 갑니다.

암에 걸릴 확률 P(C) = 0.01, 1%

암에 걸렸을 때, 검사결과가 양성(Positive)일 확률 90% 암에 안 걸렸을 때, 검사결과가 음성(Negative)일 확률 90%

문제, 결과가 양성(Positive) 일 때, 암일 확률은? P(C|Pos)?



ALL

1. 1%
2. 9%
3. 90%
4. 99%

가장 가까운 답은 뭘까요…???

풀이

암에 걸릴 확률 1% ----------> P(C) = 0. 01

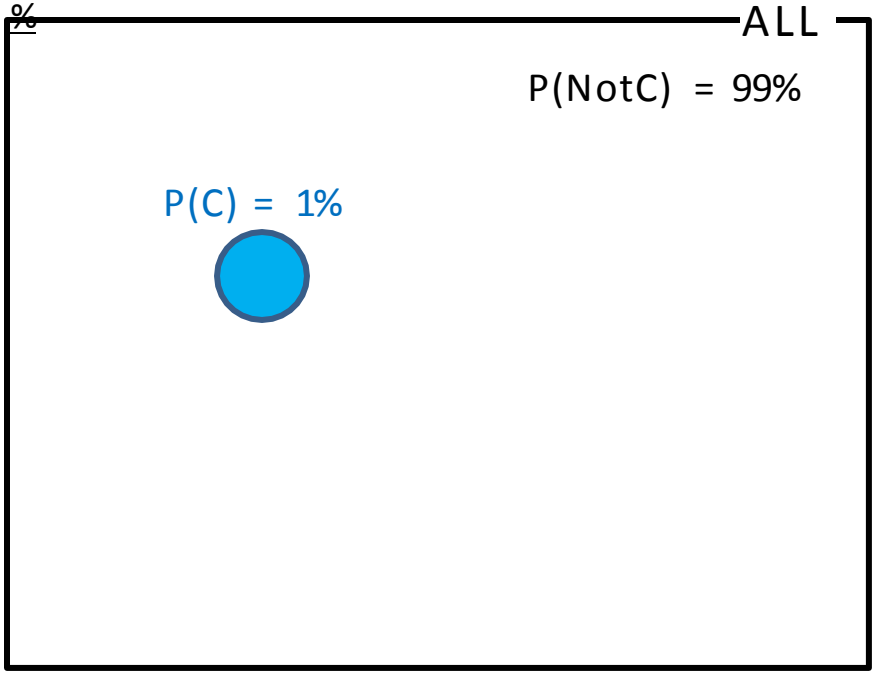
암에 걸리지 않을 확률 99% ----------> P(NotC) = 0. 99

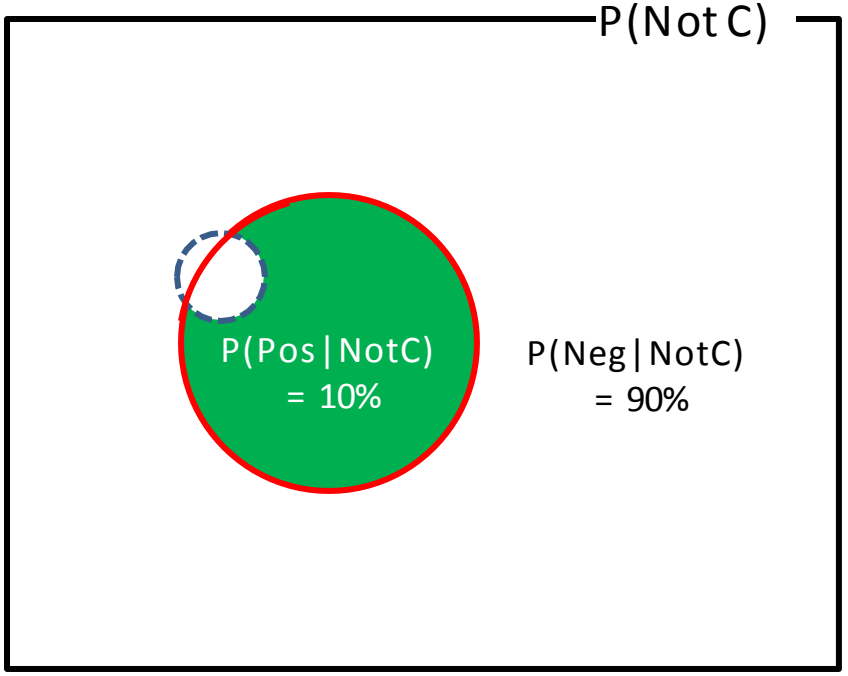
암에 걸렸을 때, 검사결과가 양성(Positive)일 확률 90% ----------> P(Pos|C) = 0.9

암에 안 걸렸을 때, 검사결과가 음성(Negative)일 확률 90 % ----------> P(Neg|NotC) = 0.9

----------> P(Pos|NotC) = 0.1

안에 안 걸렸을 때, 검사결과가 양성(Positive)일 확률 10 %



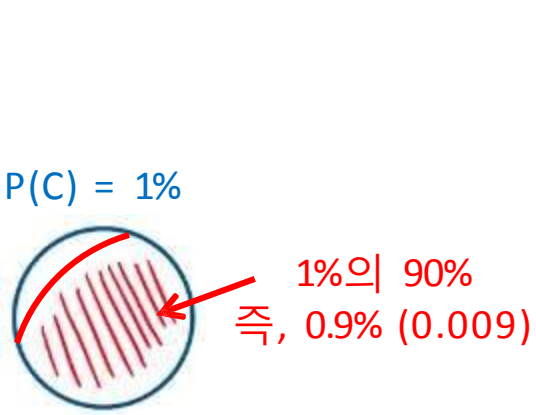


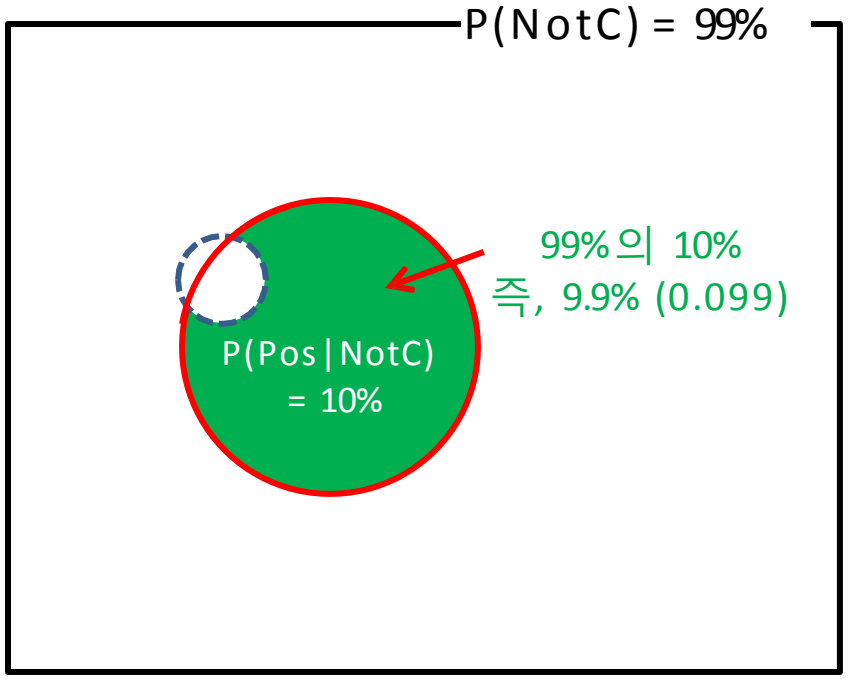
암에 걸리지 않을 확률 99%
암에 걸리지 않았을 때, 양성일 확률 10%

풀이

암이면서, 양성일 확률 (P(Pos)∩P(C))
P(C) * P(Pos|C) = 0.01 * 0.9 = 0.009

암이 아니면서, 양성일 확률(P(Pos)∩P(NotC))
P(NotC) * P(Pos|NotC) = 0.99 * 0.1 = 0.099

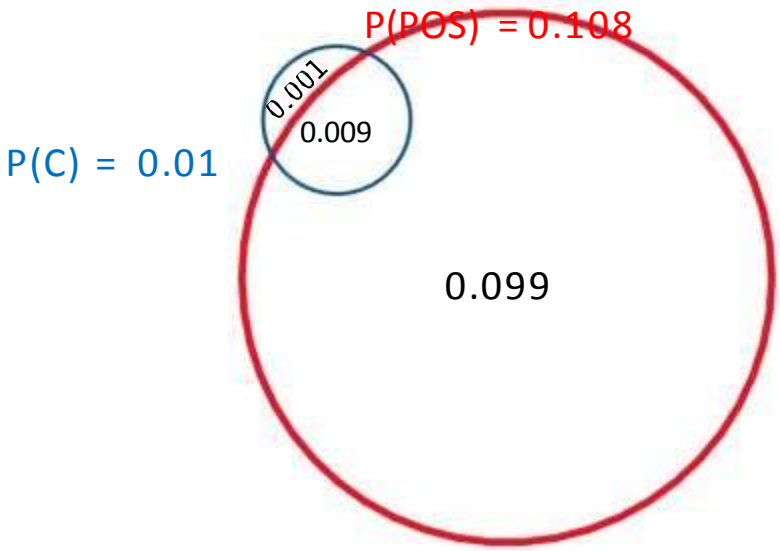




양성일 확률, P(pos)
0.09 + 0.099 = 0.108

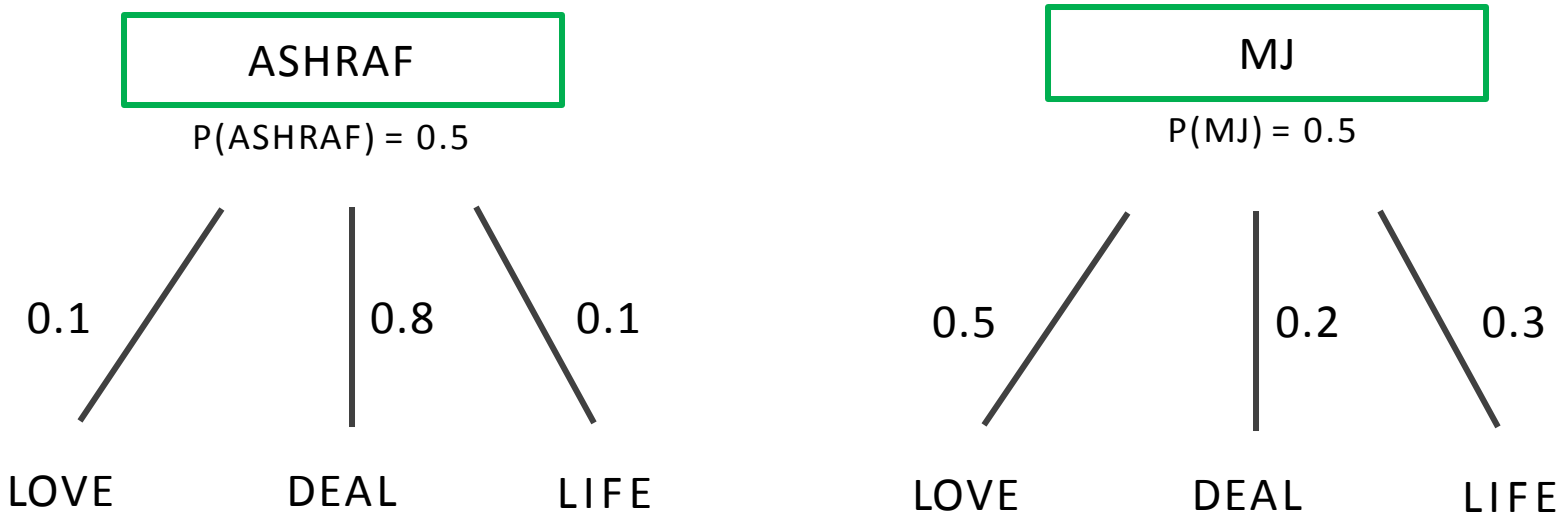
풀이

따라서 양성일 때, 암일 확률은
P(C|Pos)
= 0.009 / 0.108
= 0.083 (8.3%)



정답 : 8.3 %
(양성일 때 암이 아닐 확률은 91.7% 다)

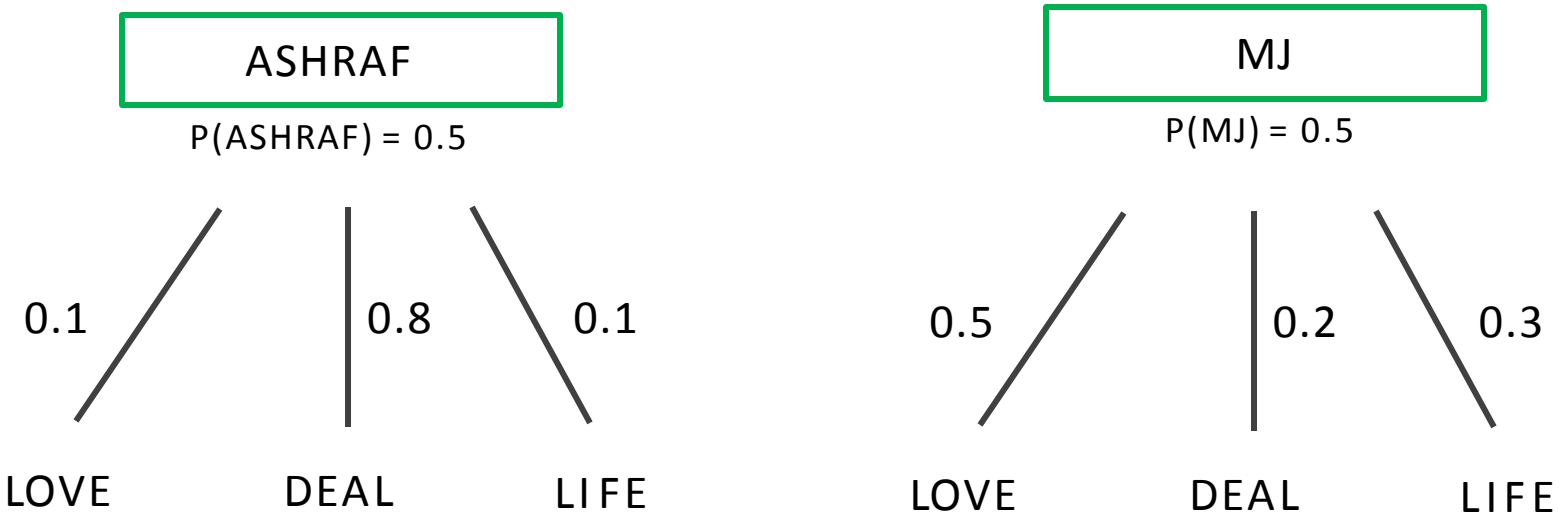
누가 Deal Life 라는 단어가 많이 나올까요?



ASHRAF = 0.8 * 0.1 * 0.5 = 0.04

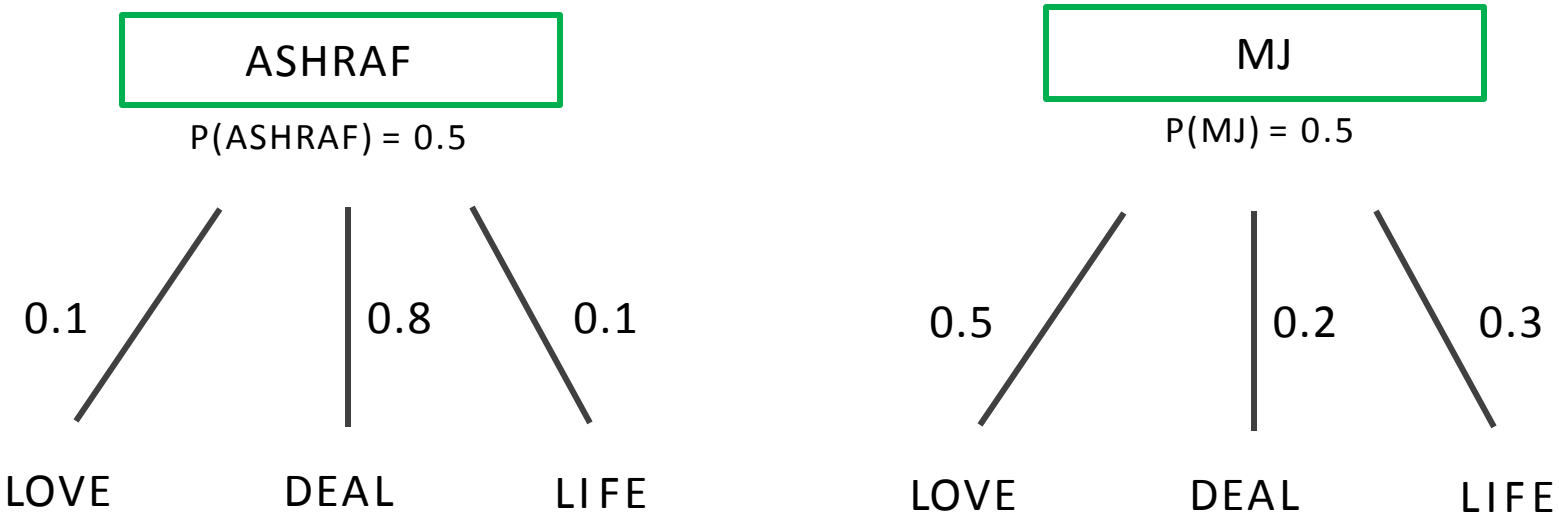
MJ = 0.2 * 0.3 * 0.5 = 0.03

P (ASHRAF|"LIFE DEAL") = ?
P (MJ|"LIFE DEAL") = ?



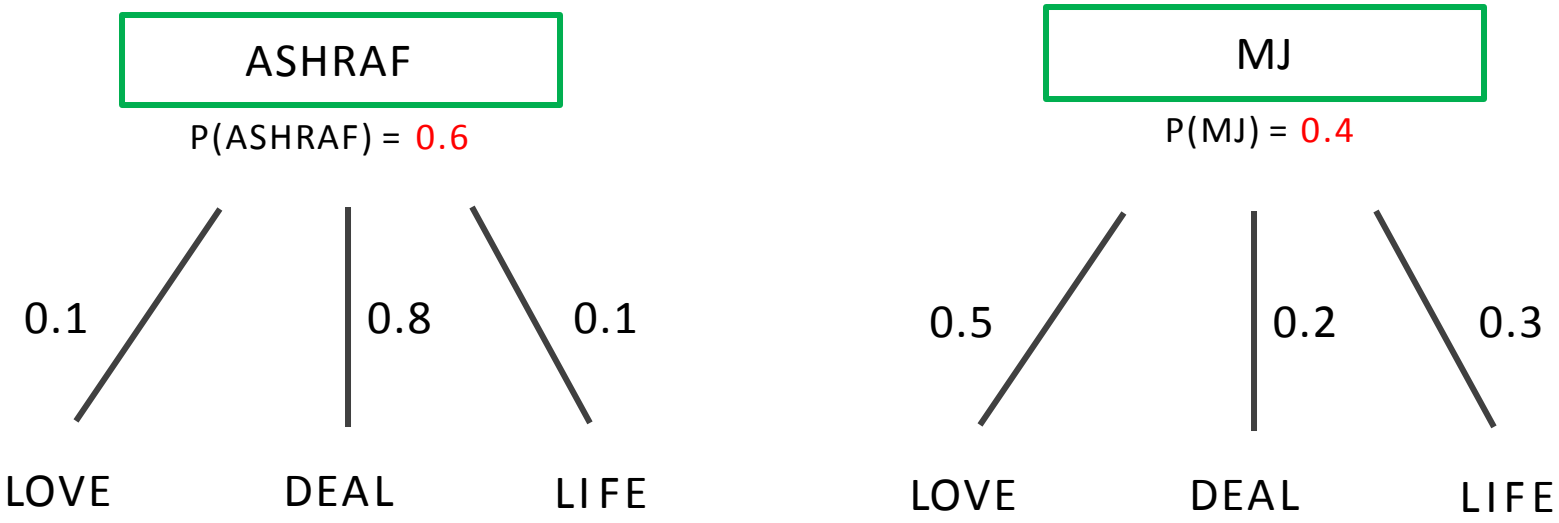
P (ASHRAF|"LIFE DEAL") = 0.04 / (0.04+0.03) = 0.57
P (MJ|"LIFE DEAL") = 0.03 / (0.04+0.03) = 0.43

P (ASHRAF|"LOVE DEAL") = ?
P (MJ|"LOVE DEAL") = ?



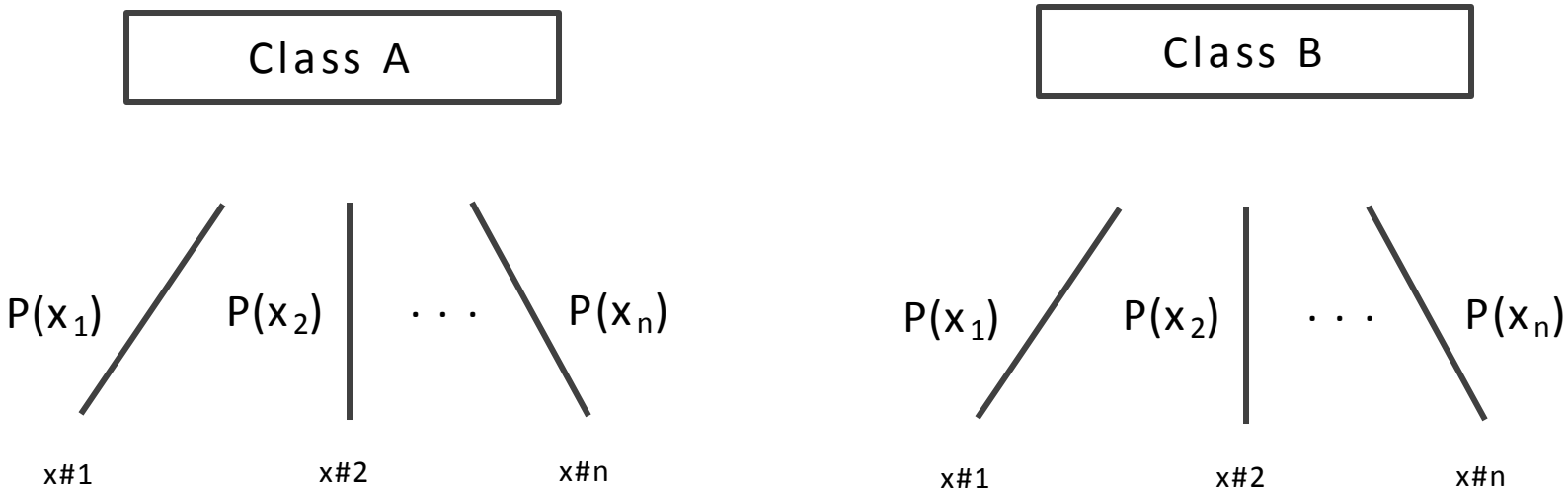
P (ASHRAF|"LOVE DEAL") = 0.44
P (MJ|"LOVE DEAL") = 0.55

P (ASHRAF|"LOVE DEAL") = ?
P (MJ|"LOVE DEAL") = ?



P(ASHRAF) * P (ASHRAF|"LOVE DEAL") = 0.6 * 0.44 = 0.264
P(MJ) = 0.4 * P (MJ|"LOVE DEAL") = 0.4 * 0.55 = 0.22

NAÏVE BAYES



클래스 확률 예) P(ASHRAF), P(MJ)

$$\hat{y} = \underbrace{\operatorname*{argmax}_{k \in \{1,\ldots,K\}}}_{\text{클래스 별로 계산한 확률 중 최고 값}} \underbrace{p(C_k)}_{\text{클래스 확률 예) P(ASHRAF), P(MJ)}} \underbrace{\prod_{i=1}^{n} p(x_i|C_k).}_{\text{주어진 클래스에서 } x_i \text{ 데이터일 확률 예) Love일 확률, deal일 확률}}$$

순서는 고려하지 않음 (각 확률이 서로 독립적이라 가정함)
그래서 NAÏVE BAYES 임

따라 해 봅시다

Part 1: train, test data 정의

```
from textblob.classifiers import NaiveBayesClassifier
# from textblob import TextBlob


train = [
    ('I love this sandwich.', 'pos'),  ('
    This is an amazing place!', 'pos'),
    ('I feel very good about these beers.', 'pos'),
    ('This is my best work.', 'pos'),
    ("What an awesome view", 'pos'),
    ('I do not like this restaurant', 'neg'),
    ('I am tired of this stuff.', 'neg'),
    ("I can't deal with this", 'neg'),
    ('He is my sworn enemy!', 'neg'),
    ('My boss is horrible.', 'neg')
]
test = [
    ('The beer was good.', 'pos'),  (
    'I do not enjoy my job', 'neg'),
    ("I ain't feeling dandy today.", 'neg'),
    ("I feel amazing!", 'pos'),
    ('Gary is a friend of mine.', 'pos'),
    ("I can't believe I'm doing this.", 'neg')
]
```

Part 2 : 학습 및 학습결과 확인

```
# make a classifier
clf = NaiveBayesClassifier(train)  print("test 결과:
", clf.accuracy(test))


 # Classify some text
 print(clf.classify("Their burgers are amazing."))  # "pos"
 print(clf.classify("I don't like their pizza."))  # "neg"
 print(clf.classify("This is amazing library!"))  #  "Pos"
 print(clf.classify("But the hangover is horrible"))  #"Neg
 "
new_data = [('She is my best friend.', 'pos'),
            ("I'm happy to have a new friend.", 'pos'),
            ("Stay thirsty, my friend.", 'pos'),
            ("He ain't from around here.", 'neg')]

# update new data  clf.up
date(new_data)  print(clf
.accuracy(test))
```

준비하기

```
> pip install textblob
> pip install -U textblob nltk
> python -m textblob.download_corpora
```

요약: ML 라이브러리를 활용하여
간단한 Naïve Bayes 모델을 만들 수 있었습니다.