Naïve bayes : 선형모델과 유사함. Linear SVC 선형 분류기 보다 훈련 속도 빠름. 일반화 성능이 뒤쳐진다.

각 특성을 개별로 취급해 파라미터를 학습하고 , 각 특성에서 클래스별 통계를 단순히 취합한다.

#함수 : scikit-learn에 구현된 함수는 총 3가지, GaussianNB, BernoulliNB, MultinormalNB

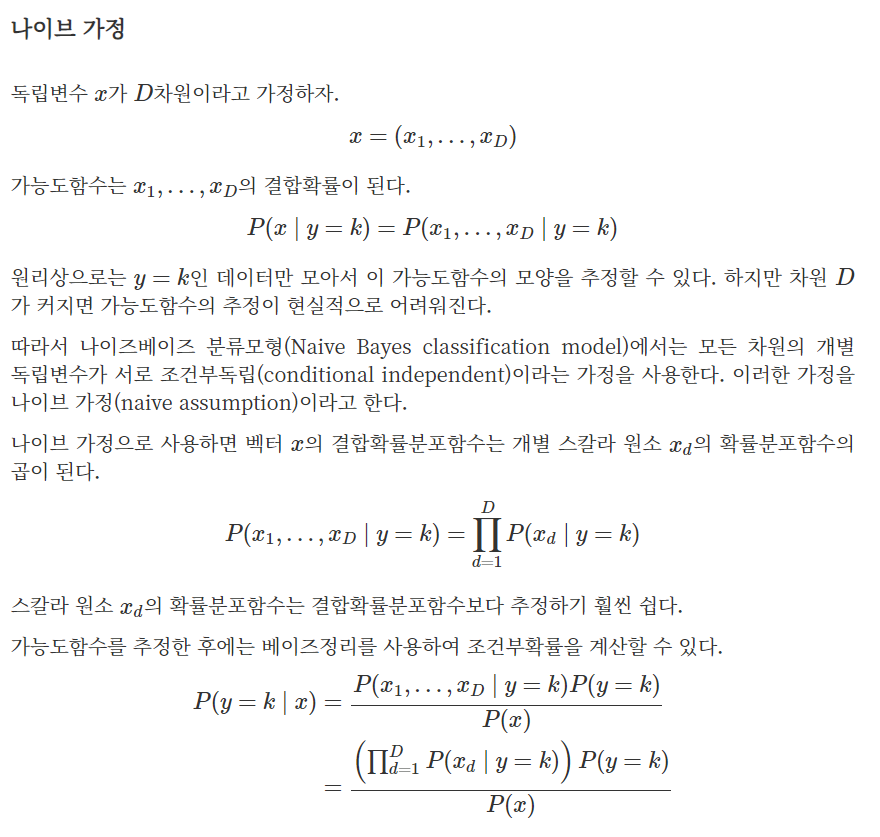
GaussianNB : 어떠한 연속적인 데이터에도 사용 가능

클래스별로 각 특성의 평균 / 표준편차를 계산

BernoulliNB : 이진 데이터에 사용

MultinormalNB : 카운트 데이터(정수 카운트, 문장에 나타난 단어의 횟수 같은 거)

클래스별로 각 특성의 평균을 계산



**오류 : 확률 =0, 구분 불가(happy word / word happy)**