

PRML 책에서 언급한 Naive Bayes Classifier에 대해 간단히 정리하였습니다.

Optimal Classifier

$$\hat{y} = \operatorname{argmax}_{Y=y} P(Y = y/X = x) \approx \operatorname{argmax}_{Y=y} P(X = x/Y = y)P(Y = y)$$

input data 주어졌을 때 y 의 확률을 가장 높게하는 class로 분류한다. (PRML chapter1 에서 알게 되었다) 그리고 Bayes's rule에 따라 우변과 같이 근사할 수 있다. $P(Y = y)$ 는 쉽게 구할 수 있다.

- 예를 들어, 100개의 data중에서 남자가 60, 여자가 40이면 $P(Y = male) = 0.6$ 으로 구할 수 있다.

하지만 $P(X = x/Y = y)$ 의 경우는 다르다. input X 의 차원이 1이면 쉽게 likelihood를 이용하여 MLE를 구할 수 있겠지만 차원이 높아지면 높아질수록 구하기가 어려워진다.

- 예를 들어, Y =남자 인 경우의 해당 식을 구하기 위해서는 남자인 데이터들만 뽑은 뒤에 $P(X_1, X_2, \dots, X_d)$ 같은 joint distribution을 구해야 한다. 얼마나 어려운가!

Naive Bayes Classifier는 이를 해결하기 위해 conditional independent라는 가정을 사용한다.

- conditional independent
 - 의미 : $p(x_1/x_2, y) = p(x_1/y)$
 - 응용 : $p(x_1, x_2/y) = p(x_1/x_2, y)p(x_2/y) = p(x_1/y)p(x_2/y)$
 - NB : $p(x_1, x_2, \dots, x_d/y)p(y) = p(x_1/y)p(x_2/y)\dots p(x_d/y)p(y)$

Naive Bayes Classifier

- Given:
 - class prior $P(Y)$
 - conditionally independent X given the class Y 라는 가정
 - likelihood of $P(X_i/Y)$
 - input data의 차원은 d
- Naive Bayes Classifier Function

$$\hat{y} = \operatorname{argmax}_{Y=y} P(Y = y) \prod_{1 \leq i \leq d} P(X_i = x/Y = y)$$

분포의 parameter는 MLE, MAP를 추정하여 구한다.

문제점

- 가정 자체가 한계가 존재한다. X 의 각 변수끼리 conditionally independent라는 가정은 현실에서 쉽게 보이지는 않는다.
- MLE, MAP가 갖고 있는 단점을 갖고 있다.

나이브 베이즈 분류기는 주로 텍스트마이닝에서 긍정부정 분류로 많이들 쓰였던 것 같다. bag of words를 이용하여 구하는 실습 자료를 쉽게 찾아볼 수 있다.

참고 사이트

- https://www.youtube.com/watch?v=ndvZKwZw5tQ&list=PLbhbGI_ppZISMV4tAWHlytBqNq1-lb8bz&index=11
- [https://ratsgo.github.io/machine learning/2017/05/18/naive/](https://ratsgo.github.io/machine%20learning/2017/05/18/naive/)
- <https://gomguard.tistory.com/69>