Bayes Theorem과

Sigmoid와 Softmax 사이의 관계

Bayes Theorem

Likelihood / Class-conditional Data Distribution

$$\underbrace{P(Y|X)}_{\text{Posterior (사후 확률)}} = \frac{P(X|Y)P(Y)}{\underbrace{P(X)}_{\text{Evidence}}}$$

Prior (사전 확률)

Posterior (사후 확률)

Evidence 확률의 marginal 특징에 의해

$$P(X) = \sum_j P(X|Y_j)P(Y_j)$$

X : data
Y : Classification 문제의 Class
Prior : 위의 class 들의 분포가 어떻게 되어 있는지에 대한 사전 지식
P(X|Y) : 각 class에 속해 있는 data의 확률 분포
P(Y|X) : 새로운 data가 들어왔을 경우 Y의 분포 → 어떤 class Y에 속할 지 정할 수 있음

Logistic Function : 2-Class Classification

- 분류를 하고자 할 경우 → 새로운 data를 분류할 때 ⇒ Posterior(사후 확률)을 기준으로 삼음!

$$\left.\begin{array}{l} X : Y_1 \ if \ P(Y_1|X) > P(Y_2|X) \\ X : Y_2 \ if \ P(Y_1|X) < P(Y_2|X) \end{array}\right\} \text{ 1과 2 두개의 class}$$

⇒ 판별 기준 Posterior

$$P(Y_1|X) = \frac{P(X|Y_1)P(Y_1)}{P(X)}$$

$$P(Y_2|X) = \frac{P(X|Y_2)P(Y_2)}{P(X)}$$

$$P(X) = P(X|Y_1) + P(X|Y_2)P(Y_2)$$

⇒ Posterior : likelihood와 Prior의 곱이 결정!

⇒ 각 class 별로 likelihood와 Prior의 곱을 정의하는 것은 의미가 있음!

Logistic Function : 2-Class Classification

- Likelihood와 Prior의 곱 정의 $a_k = \ln(P(X|Y_k)P(Y_k))$

  Log 붙이는 이유 -> log likelihood
  - Log : 단조 증가 함수 → 극점 변화 X
  - 각 시행이 독립일 경우 → 곱에 덧셈으로 바뀜 ⇒ 계산 용이
  - 확률 분포가 Gaussian등의 Exponential Family의 경우 → 계산 용이

- Class 1 Posterior

$$P(Y_1|X) = \frac{P(X|Y_1)P(Y_1)}{P(X|Y_1)P(Y_1) + P(X|Y_2)P(Y_2)} = \frac{e^{a_1}}{e^{a_1} + e^{a_2}}$$

  분모와 분자를 모두 분자로 나누면

$$P(Y_1|X) = \frac{1}{1 + e^{a_2 - a_1}}$$

$$a_2 - a_1 = \ln\left(\frac{P(X|Y_2)P(Y_2)}{P(X|Y_1)P(Y_1)}\right)$$

$$a = -(a_2 - a_1) = \ln\left(\frac{P(X|Y_1)P(Y_1)}{P(X|Y_2)P(Y_2)}\right)$$ ← 이를 Log Odds라 부름

Logistic Function : 2-Class Classification

- Posterior 새롭게 정의 $P(Y_1|X) = \dfrac{1}{1+e^{-a}}$

  위 식을 a에 대한 함수로 그려 보면
  → sigmoid function(정식 명칭 : logistic function)

- Naïve Bayes 예로 → 보통 likelihood를 Gaussian으로 정의
  ⇒ 한 class 내에서 확률 분포 : 평균과 분산을 가지는 Gaussian 적합

  Prior의 경우 : 두 가지 중 선택의 문제 → Bernoulli 적합

- Log odds : 해당하는 부분에 위의 분포 대입 Naïve Bayes 분류기 ⇒ Generative Model

- Log odds : likelihood와 Prior의 확률 분포 구분X → $P(Y_1|X) = \frac{1}{1+e^{-\theta x}}$ fitting problem

  ⇒ Logistic Regression : 대표적인 discriminative problem

soua

Softmax Function : K-Class Classification

- 2-Class → K-Class 확대

$$P(Y_1|X) = \frac{P(X|Y_1)P(Y_1)}{P(X|Y_1)P(Y_1) + P(X|Y_2)P(Y_2) + \cdots + P(X|Y_k)P(Y_k)}$$

앞에서 정의한 것 대입

$$P(Y_1|X) = \frac{e^{a_1}}{e^{a_1} + e^2 + \cdots + e^{a_k}} = \frac{e^{a_1}}{\sum_j e^j}$$

⇒ Softmax Function

- Normalization Term(분모) → 한 class의 확률이 다른 class에 영향

Generative Model(Naïve Bayes / Gaussian Discriminant Analysis 등) →Prior P(Y)에도 영향!

- Prior 고려 없이 바로 P(Y|X) fitting ⇒ Softmax Regression(Discriminant Model)

$$P(Y|X) = \frac{e^{\theta_1 x}}{\sum_j e^{\theta_j x_j}}$$