2019150445 신백록

1. $(x_1, x_2, \cdots, x_n)$가 주어졌을 때 모수가 나올 가능성을 모수에 대한 함수로 나타낸 것을 우도함수라고 한다.

우도함수가 maximize 된다는 것은 범주형일때는 각각의 Class를 맞추는 확률이 가장 클 때이고, 회귀일때는 y_i값에 가장 가깝게 $\hat{y}_i$을 추정하는 것이다.

이것은 손실함수가 minimize 되는 것과 같고, 그것을 만족하는 모수 θ를 찾는 것이 딥러닝의 목표이다.

2. 결합분포는 모수가 주어졌을 때의 표본에 대한 함수이다.

반대로 우도함수는 표본이 주어졌을 때의 모수에 대한 함수이다.

우도함수는 θ에 대한 함수이기에 θ에 대해 integration을 하면 1이 나온다.

$f(x_1, \cdots, x_n | \theta) = L(\theta | x_1, \cdots, x_n)$을 만족한다.

3. 만약 n개의 표본 $(x_1, \cdots, x_n)$이 서로 독립적으로 추출됐다면,

① $x_1, \cdots, x_n \sim Ber(p_i)$

$$\log L(p) = \log \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{(1-x_i)} = \sum_{i=1}^{n}\left(x_i \log p_i + (1-x_i)\log(1-p_i)\right)$$

$Ber(p_i)$

S

성공횟수

② $x_1, \cdots, x_n \sim Bin(r, p_i)$

$$\log L(r, p_i) = \log \prod_{i=1}^{n} \underbrace{\cancel{\binom{n}{r}}}_{Constant} p_i^{r}(1-p_i)^{n-r} = \sum_{i=1}^{n}\left(r \log p_i + (n-r)\log(1-p_i)\right)$$

m개 Class 각각이 나올 확률 $\left(p_m = 1 - \sum_{j=1}^{m-1} p_j\right)$ Class 1이 나온 횟수. $\left(n_m = n - \sum_{j=1}^{m-1} n_j\right)$

$Multinoulli(p_1, p_2, \cdots, p_{m-1})$

S

③ $x_1, \cdots, x_n \sim Multinomial(n_1, n_2, \cdots, n_{m-1}, p_{1i}, p_{2i}, p_{3i}, \cdots p_{(m-1)i})$

$$\log L(p) = \sum_{i=1}^{n} \log\left(p_{1i}^{n_1} p_{2i}^{n_2} \cdots p_{(m-1)i}^{n_{m-1}} p_{mi}^{n_m}\right) = \sum_{i=1}^{n}\left(n_1 \log p_{1i} + n_2 \log p_{2i} + \cdots + n_m \log p_{mi}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(n_j \log p_{ji}\right)$$

(순서 = Constant)

④ $x_1, \cdots, x_n \sim N(\mu_i, \sigma_i^2)$

$$\log L(\mu, \sigma^2) = \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) = \sum_{i=1}^{n}\left(\log \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)$$

4. 이항분포에서의 확률은 p로 observation마다의 p가 i에 의존하지 않고 identical하다.

이항분류에서의 확률은 p_i로 observation이 주어졌을 때의 p_i를 예측하고 싶어한다. (identical ×)

하지만 이렇게 되면 p_i가 Sample에 의존하게 되어 p_i의 추정이 불가능해진다.

따라서 특성변수 x_i를 정의하고, p_i를 모수 θ에 대한 함수인 $p_i = \theta x_i + b$로 만들면 된다.

5. Bernoulli 분포의 로그우도함수는 p가 identical해서 $\log p \cdot \sum_{i=1}^{n} y_i + (n - \sum_{i=1}^{n} y_i)\log(1-p)$가 되고,

binary Crossentropy는 p가 identical하지 않기에 p_i로 표시하고 각각의 x_i가 주어졌을 때 모수 β를 추정하여 p_i를 추정하는 과정을 거친다.

다항분포는 범주가 여러 개일 때이고, Bernoulli와 마찬가지로 설명가능하다.

정규분포의 로그우도함수도 μ가 identical할 때이고,

MSE에서의 $\mu_i = \beta * x_i$로 identical하지 않고, 모수 β를 추정함으로써 μ_i를 추정하게 된다.