로지스틱

로지스틱을 이해하기 위해서는 사전 지식이 필요함.

사전 지식

1. 우도

확률≠우도

일단 이걸 먼저 생각 해야함.

확률이란 사건 A가 발생할 고정된 확률이지만, 우도란 어떤 표본이 주어 질 때(모집단 모름), 모집단에 속할 확률을 우도라고 한다.

2. 최대우도추정

샘플 데이터를 이용해서 우도를 최대화 하는 방향으로 모집단을 추정하는 방법.

그러니까 우도를 최대화 한다 → **관측된 샘플 데이터가 나올 확률을 최대**로 하는 모수를 추정하는 방법임(정리하자면 MLE는 특정 샘플이 속할 수 있는 가장 높은 우도의 모수를 추정)

이걸 반복적으로 최적화를 돌리는데, 손실함수가 최소가 되는 방향으로 최적화를 진행한다.

3. 베르누이 분포

0과 1, 이런 식으로 이진으로 데이터를 나누어지는 데이터 분포를 베르누이 분포라고 하고, 이 확률질량함수를 수식으로 시각화를 해보면,

$$P(y_i \mid \theta) = \sigma(z_i)^{y_i} \left[1 - \sigma(z_i)\right]^{1-y_i}$$

이런 확률질량함수를 가지게 된다.

이 때, 이 확률질량함수는 입력값에 대해서, 입력값이 확률질량함수에 속할 확률이 나옴

로지스틱이란

베르누이 분포(0과1)를 가정하고 회귀식을 이용해서 데이터를 분류하는 모델.

로지스틱 함수 수식

$$sigma(z) = \frac{1}{1+e^{-z}}$$

해당 수식은 로지스틱 회귀는 아니고, 로지스틱 함수임.

그리고 여기 z에 회귀식을 추가한다면

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

$$\sigma(z) = \frac{1}{1+e^{-a\,(x_1+x_2)}}$$

이런 로지스틱 회귀식이 된다.

학습 방법

OLS와 다르게 파라미터를 MLE 라는 최대 우도 추정으로 갱신한다.

1. 초기 weight랑 bias를 구한다.
2. 손실함수(로지스틱 회귀에서는 음의 로그 우도)가 최소가 되는 파라미터를 최적화로 구한다

$$L(\theta) = \prod_{i=1}^{n} \sigma(z_i)^{y_i}\ \left[1-\sigma(z_i)\right]^{1-y_i}$$

이게 로지스틱에 있는 우도함수를 그린거임.

이 우도함수를 기반으로 손실함수에 대한 미분을 진행 할 예정

의문점

이 우도 함수는 어떤 뜻인가?

각 관측치가
독립적으로 발생한다는 가정이므로, 연쇄 곱(전체 row 갯수 만큼)을 한다는 거고, y=1인 로짓 확률과 아닌 확률을 row 갯수 만큼 돌리는 거임

왜 하필 y=1인 시그모이드 출력, y=0인 시그모이드 출력의 곱으로 보는 걸까?

베르누이 분포의 확률 질량 함수가

$$p^y(1-p)^(1-y)$$

로 정의되기 때문에, 로지스틱 회귀에서는 p 대신 σ(z)를 대입하여

$$\sigma(z)^y[1-\sigma(z)]^(1-y)$$

로 표현합니다. 이 식은 각 데이터 포인트가 주어진 모수 하에서 관측될 우도를 나타내며,

로지스틱 회귀에서는 최대 우도 추정(MLE)을 통해, 관측된 데이터의 우도를 최대화하는 방향으로 파라미터를 갱신합니다.

근데 이게 그럼 손실함수로 표현하면 왜 자연로그로 변할까?

연쇄곱에 대한 미분을 편하게 하기 위해서임

$$L(\theta) = \prod_{i=1}^{n} \sigma(z_i)^{y_i} \, [1-\sigma(z_i)]^{1-y_i}$$

자 이 우도 함수를 미분 할려면 엄청 복잡해진다.

자연로그는 상용로그의 성질을 모두 가지므로,

로그의 진수 곱은 로그의 합으로 분해할 수 있으니까 다음과 같은 형태로 변형시킨다.

$$
\begin{aligned}
\ln\left(\prod_{i=1}^{n}\sigma(z_i)^{y_i}\,[1-\sigma(z_i)]^{1-y_i}\right) &= \sum_{i=1}^{n}\ln\Big(\sigma(z_i)^{y_i}\,[1-\sigma(z_i)]^{1-y_i}\Big)\\
&= \sum_{i=1}^{n}\Big[\ln\Big(\sigma(z_i)^{y_i}\Big)+\ln\Big([1-\sigma(z_i)]^{1-y_i}\Big)\Big]\\
&= \sum_{i=1}^{n}\Big[y_i\ln\sigma(z_i)+(1-y_i)\ln(1-\sigma(z_i))\Big].
\end{aligned}
$$

클래스가 불균형한 데이터에 대해서 회귀 계수 및 해당 오즈비로 정확한 판단을 할 수 없는 이유

최대우도 추정(MLE) 과정에서 전체 우도 함수에 각 관측치가 동일하게 기여하지 않고, 다수 클래스의 관측치가 우도에 더 큰 영향을 미치게 되는 점 때문에, 편향적이고 과장된 결과를 뱉어냄.