Classification

$y \in C$ 인경우, $f(x) \in C$ 인 f를 찾는것이 목적이다.

$f(x)$가 0, 1로 classification이 가능하지만 but 어떠한 경우도 존재.

그럴경우 $f(x)$가 속할 C의 확률을 최대한 ↑ 하자.

방법1) Just Linear model에 넣어서 class를 {0,1}로 설정하면

되는가? NO, Linear model의 result는 $-\infty \sim \infty$ 사이의 값이라 $\hat{y}$ 설정X

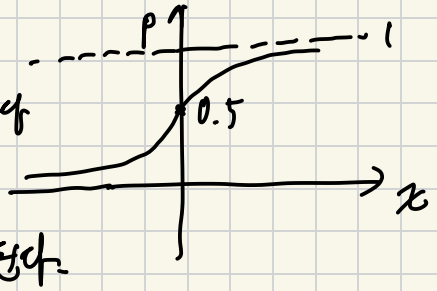
& class가 3개 이상인 경우 {0,1,2}일 시 1과 2의 2배 차이가 생긴다.

< Logistic Regression > Linear Regression + sigmoid.

(이진분류는 베르누이를 따른다는 가정)

$x \in R^d$ 이고 $y \in \{-1, +1\}$ 인 Binary classification 가정

이때 $f(x)$가 1일 확률이 높다면 1에 가깝게
" +1 or -1 정말 어떠한 경우 0.5 "
" -1 이 될 확률이 높다면 0에 "
} 하도록 Mapping 하고싶다.



그래서 쓰는 function이 $\text{sigmoid} = \frac{1}{1+e^{-x}}$ 이다

Linear Regression 결과 ∞ 라면 1
" 0 라면 0.5 로 mapping 된다.
$-\infty$ 라면 -1

확률의 정의 : 모든 outcome들이 0~1 사이에 존재)
Σ 사건 = 1 를 만족해야한다.

그러면 $p(y=1|x) + p(y=-1|x) = 1$ 을 만족해야한다.

$\frac{1}{1+e^{-\beta^T x}}$ $\frac{1}{1+e^{\beta^T x}}$ → $\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$ 이다.

그리고 $P(y=1|x) = \frac{1}{1+e^{-\beta x}}$ 을 식을 정리해보면

$\log\left(\frac{P(x;\beta)}{1-P(x;\beta)}\right) = \beta_0 + \beta_1 x_1$ 이다.

→ $y=1$ 일 확률

→ $y=0$ 일 확률.

$P(x;\beta)$가 ↑일수록 값이↑ 이고 $P(x;\beta)$가 ↓일수록 값이↓ 된다.

이렇게 $\log\left(\frac{P(x;\beta)}{1-P(x;\beta)}\right)$ 를 "log odds" 라고 부르고

Logistic Regression은 log odds가 "선형적" 이라는 가정을 하고있다.

odds는 승산을 의미한다. y가 일어날확률/일어나지않을확률 이기에

$-\infty < \log \text{odds} < \infty$ 의 범위를 갖는다.

그래서 $\log \text{odds} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$ 처럼 선형적 해석가능.

Linear Regression의 경우 데이터들의 관계 자체가 선형적이라는 가정을 한 모델이라면, Logistic Regression의 경우 log odds가 선형적이라는 가정을 한 모델이다.

〈MLE of Logistic Regression〉

$$L(\beta) = \prod_{i,y=1} p(y_i|x_i) \cdot \prod_{j,y=-1} (1-p(y_i|x_i))$$

$$\log L(\beta) = \sum \log\left(\frac{1}{1+e^{-\beta^T x_i}}\right) + \sum \log\left(\frac{1}{1+e^{\beta^T x_i}}\right)$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{1+e^{-y_i \beta^T x_i}}\right) = -\sum_{i=1}^{n} \log\left(1+e^{-y_i \beta^T x_i}\right)$$

$$\Rightarrow \sum_{i=1}^{n} \frac{x_i y_i}{1+e^{y_i \beta^T x_i}} = 0$$

not closed solution.

β를 구하는 공식이 존재하지 않는다.

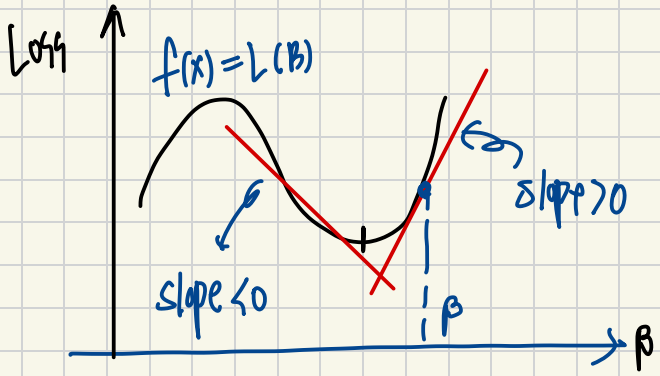
Closed-form 해와 convex 문제는 완전히 별개의 개념입니다. 우선 closed-form 해의 존재 여부는 MLE나 최소제곱으로 세운 손실 함수 $L(\beta)$에 대해 $\nabla_\beta L = 0$을 풀었을 때, 이 식이 $-A\beta + b = 0$처럼 β에 대해 2차식 또는 선형 방정식 형태로 정리되어 $\beta = A^{-1}b$와 같은 해로 바로 구해질 수 있는지를 의미합니다. 반면 목적 함수가 convex라는 것은, 그 함수가 볼록 형태여서 모든 국소 최소값이 곧 전역 최소값이 된다는 뜻이며, 이 경우 경사 하강법 같은 수치 최적화 기법을 사용해 전역 최적해를 안정적으로 찾을 수 있다는 보장을 제공합니다. 따라서 "손실이 2차식이면 closed-form 해가 가능"하고 "손실이 convex이면 경사법으로 전역 최적해 보장"이라는 두 판단 기준은 서로 다른 개념임을 명확히 구분해야 합니다.

⇒ MLE의 결과로 나온 방정식이 선형식이라면 closed-form이 존재하지 않는다.

〈gradient Descent〉

optimization : 특정 set에서 조건을 만족하는 최선의 element를 구해보자.

그래서 "최적화" 하려는 함수를 구하기 위한 β에 대해서 미분을 계속해줌.



기울기와 반대방향.

$$\beta_{new} = \beta_{old} - \gamma \cdot \frac{\partial}{\partial \beta_{old}} L(\beta)$$

미분값이

추가 업데이트 이동방향

different1) Not Convex problem인 경우 ⇒ Local optim에 빠질수 있다.

different2) 미분이 가능만 cost 함만 사용 가능

different3) 수렴이 느려진다 (거의 수렴안하는 경우) ⇒ SGD로 개선.

〈Stochastic Gradient Descent〉

data를 특정 개수의 Batch 만큼만 활용해서 최적화 진행 (mini-batch)

diminishing returns 효과 (Batch크기 ↑ 일수록 얻는 정확 ↓)

지금까지는 1개의 변수와 이진분류문제를 풀었다.
But 변수가 N개가 된다면 아래와 같이 표현이 가능해진다.

$$\log\left(\frac{p(x;\beta)}{1-p(x;\beta)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \text{ 이다 } (\beta\text{이 선형적이기에})$$

$$p(x;\beta) = \frac{e^{\beta_0+\beta_1+\cdots+\beta_n x_n}}{1+e^{-\beta_0+\beta_1 x_1+\cdots+\beta_n x_n}} \text{ 이 된다.}$$

Class가 3개 이상인 경우

이진분류에서 $p(y=1|x) = \frac{1}{1+e^{-\beta^T x}}$ & $p(y=-1|x) = \frac{1}{1+e^{\beta^T x}}$ 였다

이들의 분모를 통일하게 해주면

$$p(y=1|x) = \frac{e^{\beta^T x}}{e^{-\beta^T x}+e^{\beta^T x}},\quad p(y=-1|x) = \frac{e^{-\beta^T x}}{e^{-\beta^T x}+e^{\beta^T x}} \text{ 가정}$$

이를 확장해서 N개의 class인 경우

$$p(y=k|x) = \frac{e^{x_k}}{e^{x_1}+e^{x_2}+\cdots+e^{x_n}} \text{ 로 설정이 가능}$$

↳ 모든 class에 대한 확률값 sum.

이게 Softmax function이다.