DATE. NO.

~~선형회귀모델.~~

◦ 로지스틱 회귀모델.

preview) (선형회귀 ,,

핵심; Y값이 연속형,

(ex) Y: 혈압(mmHg), 나이(years)

혈압 = 81.54 + 1.22 (나이),,

but. 로지스틱 회귀의 필요성.

X	Y (범주형)

범주형 반응변수 ; 이진변수 (반응변수 값 ∈ 0, 1)

멀티변수 (반응변수 값 ∈ 1 or 2 or 3 이상)

" 선형회귀 모델과는 다른 방식으로 접근해야 될 필요성.

반면 (ex) 선형회귀의 경우 $\varepsilon \sim N(0, \sigma^2)$

Error가 정규분포를 따를 것.

평균이 0이고, 분산이 σ^2 (Constant),,

변하지 않고 항상 일정한 수를 가리키는 것.

를 따를 것.

$f(x) = ax^2 + bx + c$.

; x는 변수 임으로 상수 X

; a.b.c는 변하지 않는 상수이며,

; c는 상수 항.

로지스틱 회귀는 분류예제를 푸는 것 이라 할 수 있다.

로지스틱 회귀모델 사용.

; 새로운 관측치가 왔을 때 이를 기존 범주 중 하나로 예측. (범주예측)

로지스틱 회귀모델 이론배경.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad Y_i = 0 \text{ or } 1$$

가정. Y의 Error가 0인 것을 가정하면, Y_i의 기댓값은 아래와 같다.,, *

Assume $E(\varepsilon_i) = 0$ / $E(Y_i) = \beta_0 + \beta_1 X_i$

Consider. Y_i to be a Bernouli random Variable

$P(Y_i = 1) = \pi_i$

$P(Y_i = 0) = 1 - \pi_i$

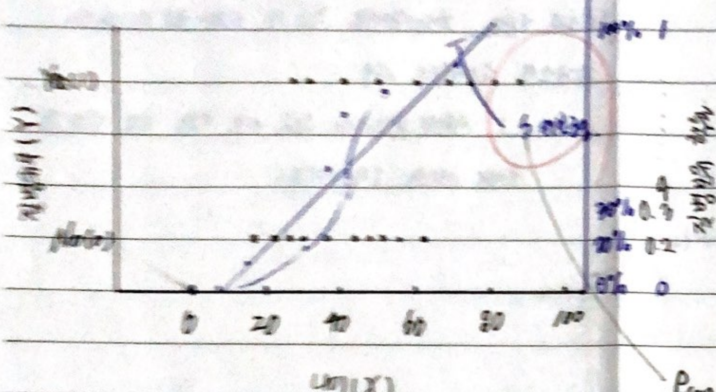
결과 $E(Y_i) = 1 \cdot \pi_i + 0 \cdot (1 - \pi_i) = \pi_i$

해석:

어떤 X 값이 들어왔을 때, Y가 특정 범주를 가질 확률

1 이 값을 가질 ,,

예: 나이 대비 관상질환 (Yes / No)



$$f(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

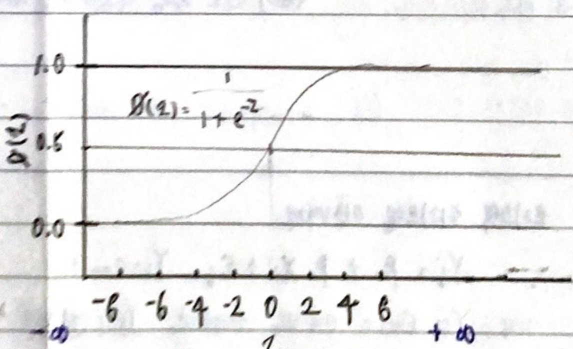
Logistic 함수

Sigmoid 함수, Squashing function.

* X값 $-\infty \sim +\infty$ 즉, 모든 실수값에서 들어올 수 있는데 나오는 0과 1 사이인 것이 특징

질병

나이층	그룹내 수	관상질환자	%
20-29	5	0	0
30-39	6	1	17
40-49	7	2	29
50-59	7	4	57
60-69	5	4	80
70-79	2	2	100
80-89	1	1	100



- * Large input > Small output.
- Output의 범위 : 0-1
- 미분값이 쉽게 산출된다.(혹은 산출된다) 함수.
- 이를 기반으로 output의 계산도 용이함.

(Gradient learning method에서 유용하게 사용)

$$\frac{d\phi(z)}{dz} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \phi(z)\cdot(1-\phi(z))$$

DATE. NO.

로지스틱 함수.. ; 기본정의 ; 연속형 데이터를 Input으로 하여 이산형 output 측정.

* $E(y) = \pi(X=x) = P(Y=1 \mid X=x) = 1 - P(Y=0 \mid X=x)$

부수설명.

장점 이유 ; 확률적인 표현이 가능하다는 것.

식 $y = \frac{e^x}{1+e^x}$,, x가 $-\infty$; 0 , $+\infty$; 1

x가 0이면 ; $\frac{1}{2}$

Sigmoid function (None linear)

Linear Regression 과 비교,, $Loss = \sum_i (Y_i - f(x_i, w))^2$.

-

Weight에 대해서 Non-linear.

즉, 위를 Loss라 정의할 수 없고,

Gradient를 써도 0값이 여러개이기에 적용 불가능.

즉, 확률적 접근의 필요성,,

Sigmoid function에 맞는 (적합한) likelihood

를 통해 lost function을 세움.

k(a),,

√ Conditional PDF (Probability Density Function) 과 유사.

or PMF (〃 Mass 〃) 와 유사.

간단, 요약 $\int_a^b f(x)dx$ = 상대도수. > 확률밀도함수 그래프의 넓이는 $dx \times \frac{상대도수}{dx}$

√ 이산에 대한 값은 = 0 , $P(a < x < b) = P(a \le x \le b)$

√ $f(x) \ge 0$; 특정 값에 대한 확률값 X , 구간에 대한 적분만이 확률.

상대도수 / 계급의 크기 이산에 대한 값은 0

√ $P(\alpha \le x \le \beta) = 1$.

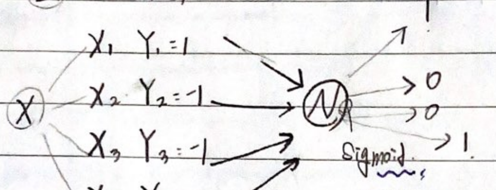
결국 상대도수들의 합이므로

* 단, 차이 $P(X|Y)$, Y given X 中 X는 고정 Y값이 변함. > Next page.

예시.

원래 Model.

In label



(출력해서 내고자 하는 값이라고 판단할 확률. 0 ~ 1 사이).

Single layer NN 이라고 사용한다. 가정.

$$P(Y_i \mid X) = \text{Sig}(f(x, w)) \quad Y_i = 1$$
$$\times\ 1 - \text{Sig}(f(x, w)) \quad Y_i = -1$$
$$\times\ 1 - \text{Sig} \quad '' \quad -1$$
$$\times\ \text{Sig} \quad '' \quad 1$$

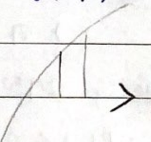
N개

즉, ~~곱~~

곱. MAX 지향

$$= \text{Loss} \stackrel{\Delta}{=} -\prod_{T=1}^{N} P(Y_i \mid X_i)$$

Step 2. log 가. 양수에서만 정의 되므로.



~~log~~ 확률(양수) MAXimize 하기 위해 log.

Result

$$= -\sum \log P(Y_i \mid X_i)$$

< ~~Binary~~

Likelihood. Cost function.