Logistic Regression

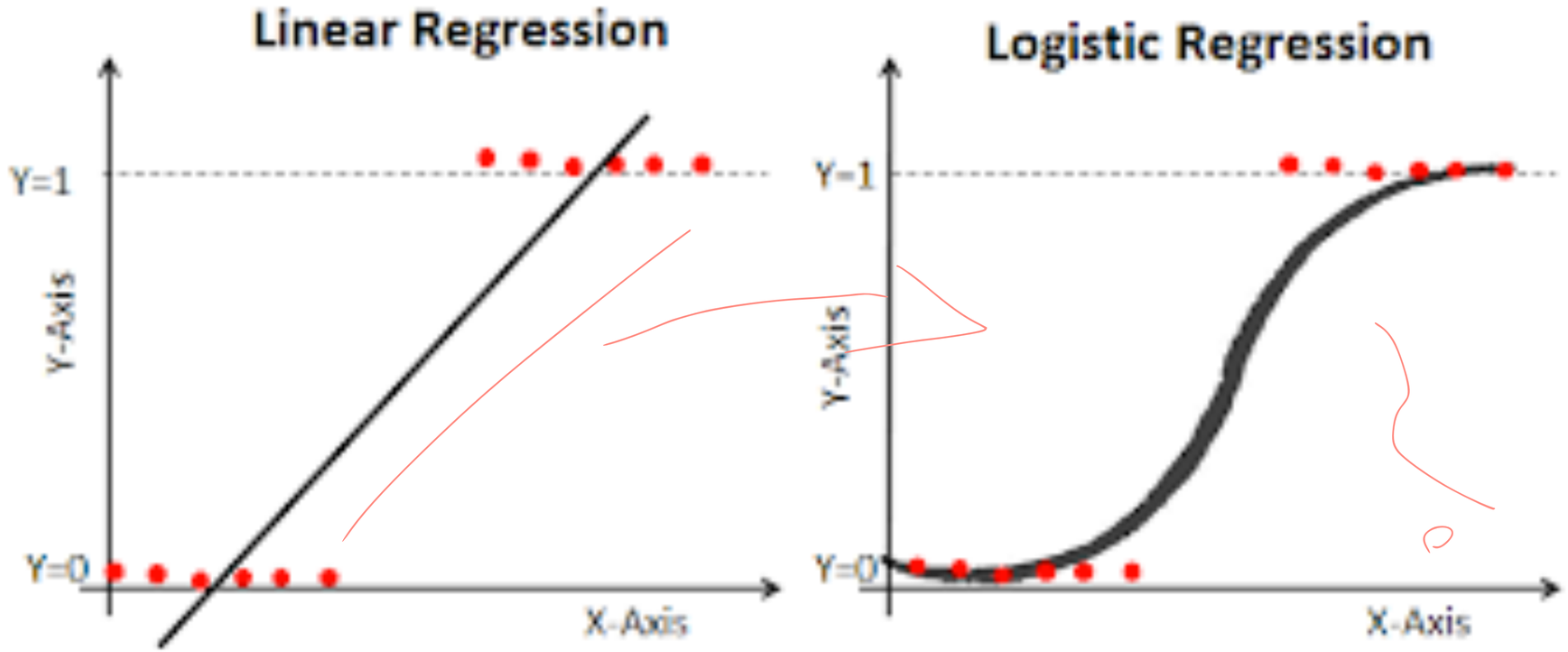
데이터 분석 모델링반 (ML1)

로지스틱회귀

선형회귀분석과 유사하지만 종속변수가 양적척도가 아닌 질적척도

특정 수치를 예측하는 것이 아니라 어떤 카테고리에 들어갈지 분류하는 모델

기본 모형은 종속변수가 0과 1이라는 이항으로 이루어짐 (구매/미구매, 성공/실패)



- 선형회귀의 사상은 그대로 유지하되 종속변수가 1이될 확률로 변환, 그 확률에 따라 0과 1의 여부를 예측한다.
- 이를 위해서는 오즈(Odds)와 로짓변환을 이용 (S자 커브로 변하는 것 오즈, 로짓변환 표현)
- 확률을 표현하기 위해서는 선형회귀가 아닌 S자 모형으로 변환,

오즈값이란? (Odds Ratio)

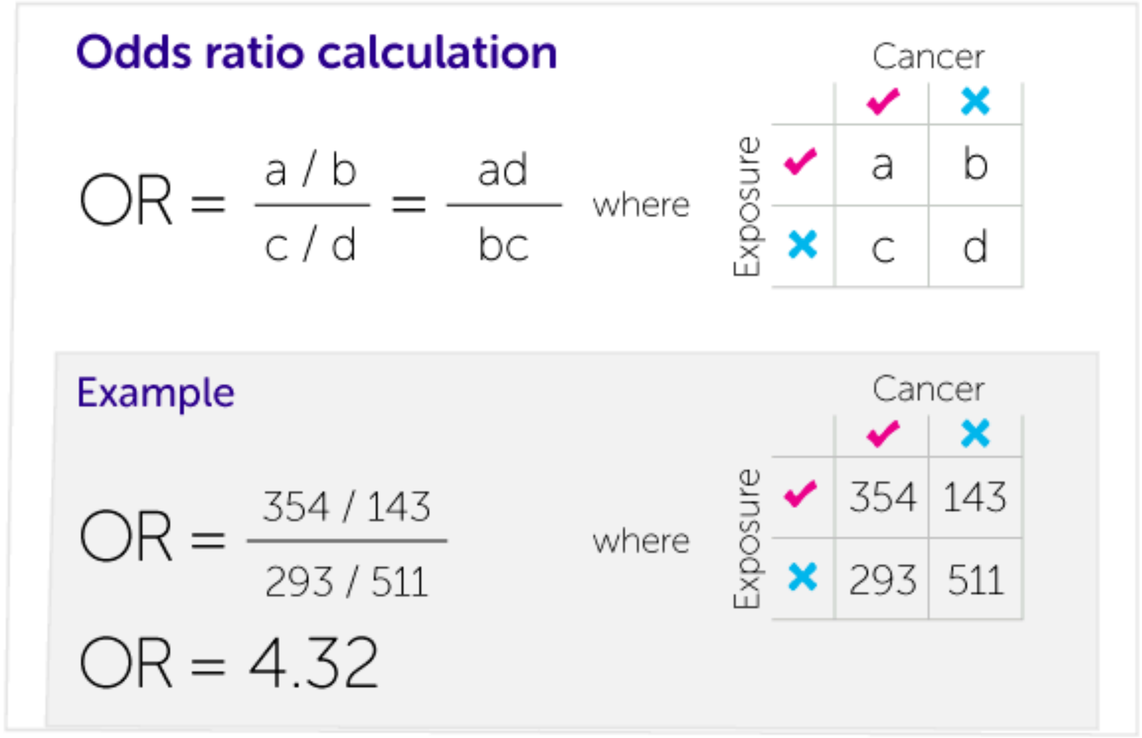
: 오즈(Odds)는 어떤 사건이 일어날 가능성으로 P/(1-P)으로 표현됨

*오즈(Odds) = 성공확률/실패확률

: 위험인자에 노출된 사람 중에서 암환자인 오즈값 = a/b = **Odds1**

: 위험인자에 노출되지 않은 사람 중에서 암환자인 오즈값 = c/d = **Odds2**

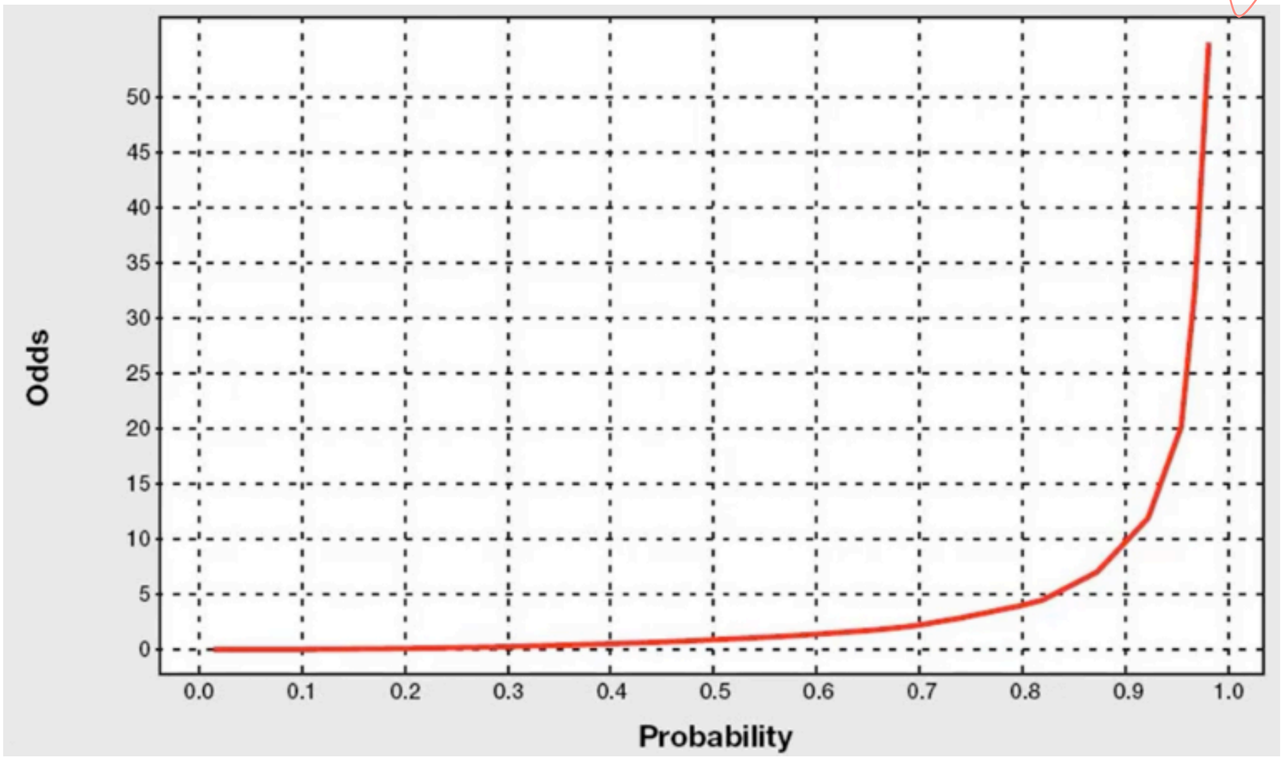
▶**Odds Ratio(오즈비;교차비;승산비)** = **Odds1**/**Odds2** = 'a/b' / 'c/d'



→ 위험인자에 노출된 사람은 노출되지 않은 사람에 비해 4.32배 정도로 더 암에 걸리는 경향을 보임

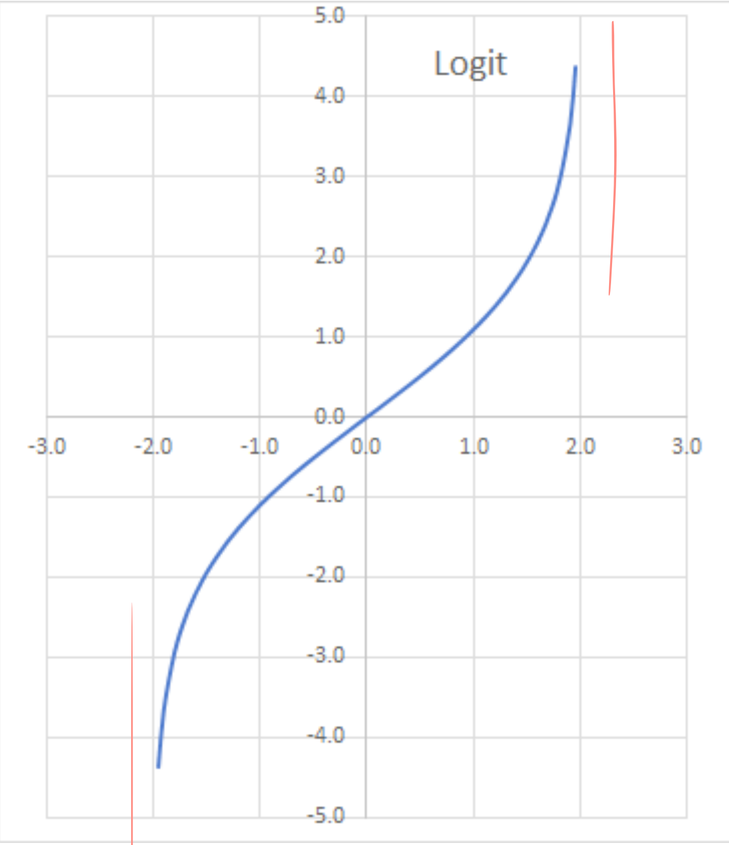
- 사건이 발생할 가능성이 사건이 발생하지 않을 가능성보다 어느정도 큰지 나타내는 값
- 분모는 사건이 발생하지 않을 확률 분자는 사건이 발생할 확률
- Odds = p(사건이 발생할 확률)/ 1-p(사건이 발생하지 않을 확률)
- 만약 발생 활률이 60% 발생하지 않을 확률 40% 오즈비는 1.5

오즈값의 문제



- 오즈값은 발생확률이 1에 가까워질수록 기하급수적으로 커지고, 최솟값은 0이 된다.
- 따라서 균형을 잡지 못하는 형태

로짓logit = log(오즈비)

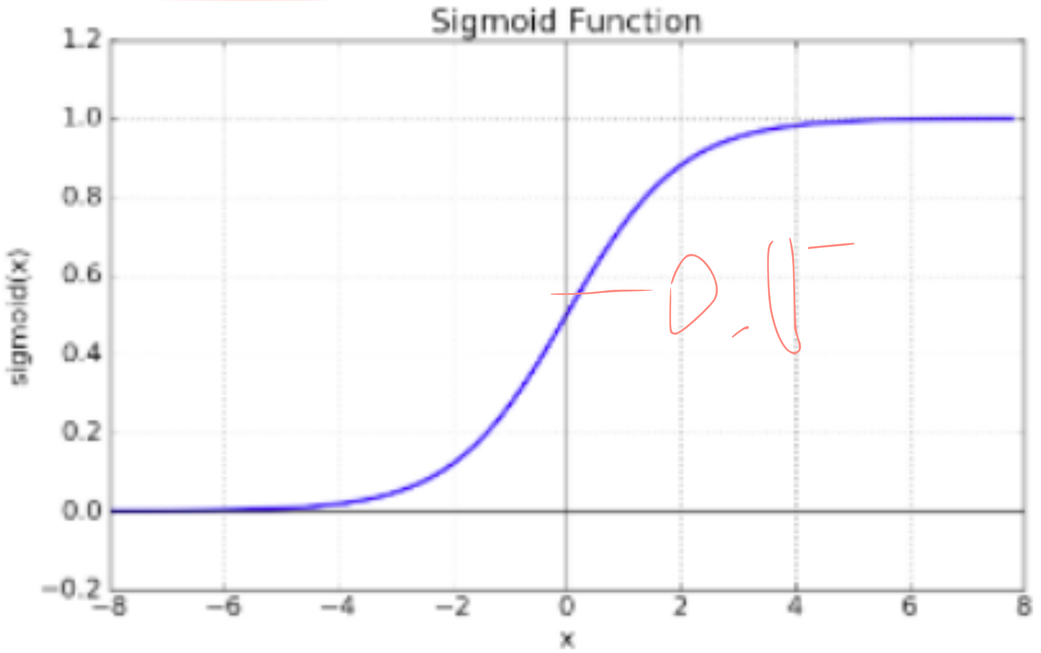


- 오즈에 log를 취하여 0<p<1 & 0<1-p<1
- p가 0에 가까울수록 오즈비는 0
- p가 1에 가까울수록 오즈비는 무한대
- 결국 -무한대 < 로짓 < + 무한대
- 하지만 여전히 0과 1 사이에서 범위 나타내지 못하므로
- 이 것을 치환해 주는 변환식

시그모이드 함수(Sigmoid)함수

종속변수가 여러 개인 경우는
하나를 잡고 나머지 다른 범주랑 비교해서 계산하는 것
모든 확률의 합은 1이니깐

선형회귀 계산 방식 - 최소제곱법 계산해서 진행
로지스틱 계산 방식 - 최대우도법 MLE



- 로그를 취한 오즈에 시그모이드 함수를 적용한 최종의 로지스틱 회귀식

- $-\infty < x < \infty$
- $0 \leq y \leq 1$
- $sigmoid(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$
- $\frac{dsigmoid(x)}{dx} = sigmoid(x)(1 - sigmoid(x))$

비용함수(Cost Function)

$$\text{if } y = 1 \rightarrow \text{cost}\,(H(x), y) = -\log(H(x))$$

$$\text{if } y = 0 \rightarrow \text{cost}\,(H(x), y) = -\log(1 - H(x))$$

$$J(w) = -\frac{1}{n}\sum_{i=1}^{n}[y^{(i)} log H(x^{(i)}) + (1 - y^{(i)}) log(1 - H(x^{(i)}))]$$

최대우도법 , 최적화 진행,

- 로지스틱회귀는 평균제곱오차를 사용하지 않음 (로컬미니멈 이슈)
- 시그모이드 함수는 0과 1사이의 값
- 0일 때 y값이 1에 가까워지면 오차가 커지며 실제값이 1일 때 y값이 0에 가까워지면 오차카 커짐
- 크로스 엔트로피 함수라고도 함 (Cross Entropy)함수, 가중치를 찾기 위해 크로스 엔트로피 함수의 평균을 취한 함수 사용

로지스틱회귀 모델 MLE

- 선형 회귀 분석과는 다르게 로지스틱 회귀분석은 직선으로 회귀계수 추정이 어려운 상황 따라서 로지스틱 회귀분석은 Maximum Likelihood Estimation 회귀 계수 구하기!
- MLE 주어진 데이터를 이용해 가능도를 최대화하는 파라미터를 찾는 통계기법, 가중치와 절편을 추정

$$\arg\max_{\mathbf{w}}\left(\prod_{x_i \text{ is RED}} P(\text{Red}\mid \mathbf{x}_i,\mathbf{w})\right)\left(\prod_{x_i \text{ is BLUE}} P(\text{Blue}\mid \mathbf{x}_i,\mathbf{w})\right)$$

$$P(y_i \mid \mathbf{x}_i,\mathbf{w}) = \begin{cases} h(\mathbf{x}_i) & \text{if } y_i = 1 \text{ (Red)} \\ 1-h(\mathbf{x}_i) & \text{if } y_i = 0 \text{ (Blue)} \end{cases}$$

$$= \arg\max_{\mathbf{w}} \log\left(\prod_{i\in\{i\mid y_i=1\}} h(\mathbf{x}_i)\right)\left(\prod_{i\in\{i\mid y_i=0\}} 1-h(\mathbf{x}_i)\right)$$

$$= \arg\max_{\mathbf{w}}\left(\sum_{i\in\{i\mid y_i=1\}} \log h(\mathbf{x}_i) + \sum_{i\in\{i\mid y_i=0\}} \log\left(1-h(\mathbf{x}_i)\right)\right)$$

$$= \arg\max_{\mathbf{w}}\left(\sum_{i\in\{i\mid y_i=1\}} \log h(\mathbf{x}_i) + \sum_{i\in\{i\mid y_i=0\}} \log\left(1-h(\mathbf{x}_i)\right)\right)$$

$$= \arg\max_{\mathbf{w}}\left(\sum_{i\in\{i\mid y_i=1\}} \left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right) + \sum_{i\in\{i\mid y_i=0\}} \left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)\right)$$

$$= \arg\max_{\mathbf{w}}\left(\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)\right)$$

Find w which maximize this!!

특정 모수 집합 하에 관측될 확률

$$f_i(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}, i = 1,2,\cdots,n \qquad \begin{matrix} P(y_i=1)=\pi_i \\ P(y_i=0)=1-\pi_i \end{matrix}$$

$$L = \prod_i f_i(y_i) = \prod_i \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$$

$$\arg\max_{\mathbf{w}}\left(\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)\right)$$

$$= \arg\min_{\mathbf{w}}\left(-\sum_{i=1}^{n} \left(y_i \log h(\mathbf{x}_i) - (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)\right)$$

$$\text{where } h(\mathbf{x}) = \frac{1}{1+\exp\left(-\mathbf{w}\cdot\mathbf{x}\right)}$$

$$E(\mathbf{w}) = -\sum_{i=1}^{n}\left(y_i \log h(\mathbf{x}_i) + (1-y_i)\log\left(1-h(\mathbf{x}_i)\right)\right)$$

$$\text{where } h(\mathbf{x}) = g(f(\mathbf{x})) = \frac{1}{1+\exp\left(-f(\mathbf{x})\right)} = \frac{1}{1+\exp\left(-(w_0x_0+w_1x_1+\cdots+w_dx_d)\right)}$$

$$\frac{\partial}{\partial w_j}E(\mathbf{w}) = \sum_{i=1}^{n}\left(h(\mathbf{x}_i)-y_i\right)\mathbf{x}_{ij}$$

$$f(\mathbf{x}) = -\mathbf{w}^T\mathbf{x} = -(w_0x_0+w_1x_1+\cdots+w_dx_d)$$

$$\frac{df(\mathbf{x})}{dw_j} = -x_j$$

$$h(x) = \frac{1}{1+\exp(-x)}$$

$$\frac{d}{dx}h(x) = h(x)(1-h(x))$$

Repeat forever for all j $\quad w_j^{t+1} = w_j^t - \eta\left.\frac{\partial E(\mathbf{w})}{\partial w_j}\right|_{\mathbf{w}=\mathbf{w}^t}$

$$w_j^{t+1} = w_j^t - \eta\sum_{i=1}^{n}\left(h(\mathbf{x}_i)-y_i\right)\mathbf{x}_{ij}$$

$$\text{where } h(\mathbf{x}_i) = \frac{1}{1+\exp\left(-f(\mathbf{x}_i)\right)}, \text{ and } f(\mathbf{x}_i) = \sum_{j=0}^{d} w_j^t x_{ij}$$

$$\log h_\theta(x^i) = \log \frac{1}{1+e^{-\theta x^i}} = -\log(1+e^{-\theta x^i}),$$

$$\log(1-h_\theta(x^i)) = \log(1-\frac{1}{1+e^{-\theta x^i}}) = \log(e^{-\theta x^i}) - \log(1+e^{-\theta x^i}) = -\theta x^i - \log(1+e^{-\theta x^i}),$$

[this used: $1 = \frac{(1+e^{-\theta x^i})}{(1+e^{-\theta x^i})}$, the 1's in numerator cancel, then we used: $\log(x/y) = \log(x) - \log(y)$]

Since our original cost function is the form of:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m} y^i \log(h_\theta(x^i)) + (1-y^i)\log(1-h_\theta(x^i))$$

Plugging in the two simplified expressions above, we obtain

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[-y^i(\log(1+e^{-\theta x^i})) + (1-y^i)(-\theta x^i - \log(1+e^{-\theta x^i}))\right]$$

, which can be simplified to:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \theta x^i - \log(1+e^{-\theta x^i})\right] = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i\theta x^i - \log(1+e^{\theta x^i})\right], \quad (*)$$

where the second equality follows from

$$-\theta x^i - \log(1+e^{-\theta x^i}) = -\left[\log e^{\theta x^i} + \log(1+e^{-\theta x^i})\right] = -\log(1+e^{\theta x^i}).$$

[we used $\log(x) + \log(y) = log(xy)$]

All you need now is to compute the partial derivatives of $(*)$ w.r.t. θ_j. As

$$\frac{\partial}{\partial\theta_j} y_i\theta x^i = y_i x_j^i,$$

$$\frac{\partial}{\partial\theta_j}\log(1+e^{\theta x^i}) = \frac{x_j^i e^{\theta x^i}}{1+e^{\theta x^i}} = x_j^i h_\theta(x^i),$$

the thesis follows.

경사하강법, 미니 배치, 뉴턴방법
sklearn 제공하는 패키지에서는 계산하는 방식이나, 메서드에 대한 선택을 다할 수 있다.

실제 직접 구현하는 경우는 이런 부분을 생각해야 한다.
코드로 비용함수, 예측하는 것들 한 번 코드로 구현해서 어떤식으로 되는지

https://youtu.be/het9HFqo1TQ