Lecture 3 Logistic Regression

$\underline{h_\theta(x)} \cancel{=} \theta^T x \notin [0, 1]$

$g(\theta^T x)$

$\underline{\theta}$

$$P(y|x;\theta) = \begin{cases} h_\theta(x) & y=1 \\ 1-h_\theta(x) & y=0 \end{cases}$$

$$g(z) = \frac{1}{1+e^{-z}} \qquad h_\theta(x) = g(\theta^T x)$$

$$g'(z) = g(z)(1-g(z))x$$

$$L(\theta) = y \log h_\theta(x) + (1-y)[\log(1-h_\theta(x))]$$

$$\frac{\partial L(\theta)}{\partial \theta_j} = y \cdot \frac{1}{g(\theta^T x)} \cdot \frac{\partial}{\partial \theta_j} g(\theta^T x) + (1-y)\frac{1}{1-g(\theta^T x)}\left(-\frac{\partial g(\theta^T x)}{\partial \theta_j}\right)$$

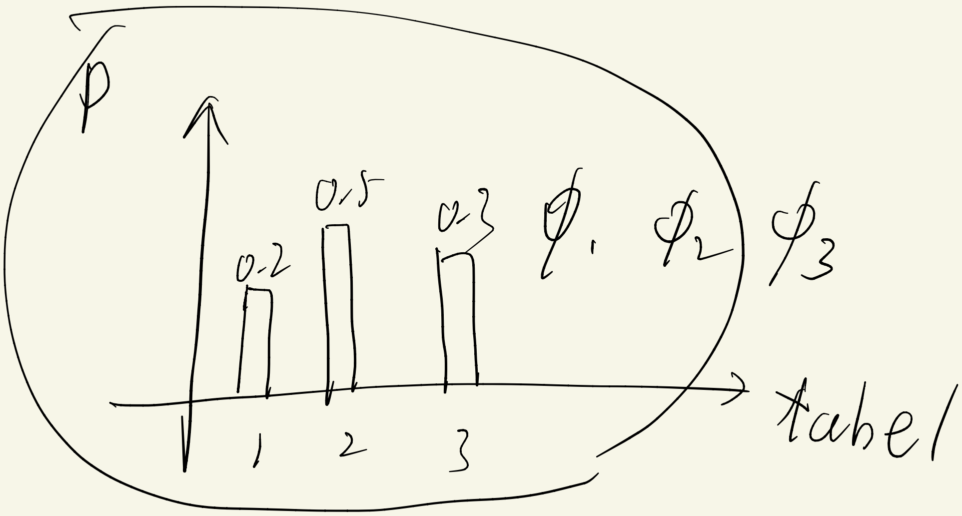
$$= \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)}\right)\frac{\partial}{\partial \theta_j} g(\theta^T x)$$

$$= \left(y\frac{1}{g(\theta^T x)} - (1-y)\frac{1}{1-g(\theta^T x)}\right) g(\theta^T x)(1-g(\theta^T x)) \rightarrow x_j$$

$$= (y - h_\theta(x))\, x_j$$

chain rule

$$\frac{\partial g(\theta^T x)}{\partial \theta} = \frac{\partial g(\theta^T x)}{\partial (\theta^T x)} \cdot \frac{\partial (\theta^T x)}{\partial \theta} \qquad x$$



$0.2 + 0.5 + 0.3 = 1$

$\phi_i = \theta_i^T x$ $\quad [0, 1] \quad \frac{1}{1+e^{-z}}$

$$\sum_{i=1}^{K} \frac{\exp(t_i)}{\sum_{j=1}^{K} \exp(t_j)} = 1$$

$$\frac{\exp(t_i)}{}$$ label i

t_i larger $\exp(\ )$

$\exp(t_i)$ larger

t_i logit

$$P(y=i \mid t_1, t_2 \cdots t_K) \propto \exp(t_i)$$

$$\frac{P(y=i)}{P(y=j)} = \frac{\exp(t_i)}{\exp(t_j)}$$

$$\begin{bmatrix} \frac{\exp(t_1)}{\Sigma \cdots} \\ \frac{\exp(t_2)}{\cdots} \\ \frac{\exp(t_k)}{\cdots} \end{bmatrix}$$

$$\sum_i \phi_i = 1$$

$\underline{K}$ $\underline{k-1}$

$$\frac{\text{degree of freedom}}{k-1}$$

$$\begin{bmatrix} 0.1 \\ 0.2 \\ 0.7 \end{bmatrix}$$

$1 - 0.3 = 0.7$

$\theta_i \in R^d$ $\qquad$ $x \in R^d$

$\theta_1 \quad \theta_2 \cdots \theta_k$ $\qquad$ $\underline{\underline{k \times d}}$ parameters

input

$x \xrightarrow{\theta} \theta^T x \longrightarrow \text{softmax} \longrightarrow p(y|x)$

MLP $\qquad$ nonlinear activation

$\theta^T x$ + softmax $\qquad$ Relu()

$x \to \square \to \square \to \square \longrightarrow \text{softmax}$

$\to p(y|x)$

$$l_{ce} \quad \underbrace{\mathbb{R}^k}_{\substack{t \\ \theta^T x}} \times \underbrace{\{1, \dots, k\}}_{y} \longrightarrow \underline{\underline{\mathbb{R}_{\geq 0}}}$$

$$l_{ce}([t_1 \cdots t_k], y) = -\log\left[\frac{\exp(t_y)}{\sum_{j=1}^{k} \exp(t_j)}\right]$$

$$l_{ce} = -\log\left(\frac{\exp(t_y)}{\sum_{j=1}^{k}\exp(t_j)}\right) \qquad t = \theta^T x$$

$$\frac{\exp(t_i)}{\sum_{j=1}^{k}\exp(t_j)} = \phi_i$$

$$l_{ce} = -\log\phi_i$$

$$\frac{\partial l_{ce}}{\partial t_i} = -\frac{1}{\phi_y}\cdot\frac{\partial}{\partial t_i}\frac{\exp(t_y)}{\sum_{j=1}^{k}\exp(t_j)}$$

$$= -\frac{1}{\phi_y}\cdot\left[\frac{\partial}{\partial t_i}(\exp(t_y))\cdot\frac{1}{\sum_{j=1}^{k}\exp(t_j)} + \exp(t_y)\cdot\frac{\partial}{\partial t_i}\left(\frac{1}{\sum\cdots}\right)\right]$$

$$= -\frac{1}{\phi_y}\left[\exp(t_y)\cdot\frac{-\exp(t_i)}{\left(\sum_{j=1}^{k}\exp(t_j)\right)^2} + \begin{cases}\frac{\exp(t_i)}{\sum_{j=1}^{k}\cdots} & i=y\\ 0 & i\neq y\end{cases}\right]$$

$$\left(-\phi_y\cdot\phi_i + \begin{cases}\phi_i & i=y\\ 0 & i\neq y\end{cases}\right)$$

$$\phi_i\cdot\mathbb{1}[y=i]^{\text{true}}$$

$$\frac{d}{dx} f(x) \cdot g(x)$$

$$= g(x) \cdot \frac{d}{dx} f(x) + f(x) \cdot \frac{d}{dx} g(x)$$

$0 \le \phi_i \le 1$ scalar

$$\theta_i^{new} = \theta_i^{old} - [\phi_i - 1(y=i)]\, x$$

feature associated with label i

negative $y=i$ $1(y=i)=1$

positive $y \neq i$ $1(y=i)=0$

$t_i = \theta_i^T x$ logit label i softmax$(t) = p$

$$t_i^{new} = [\theta_i^{old} - [\phi_i^{old} - 1(y=i)]\, x]^T x$$

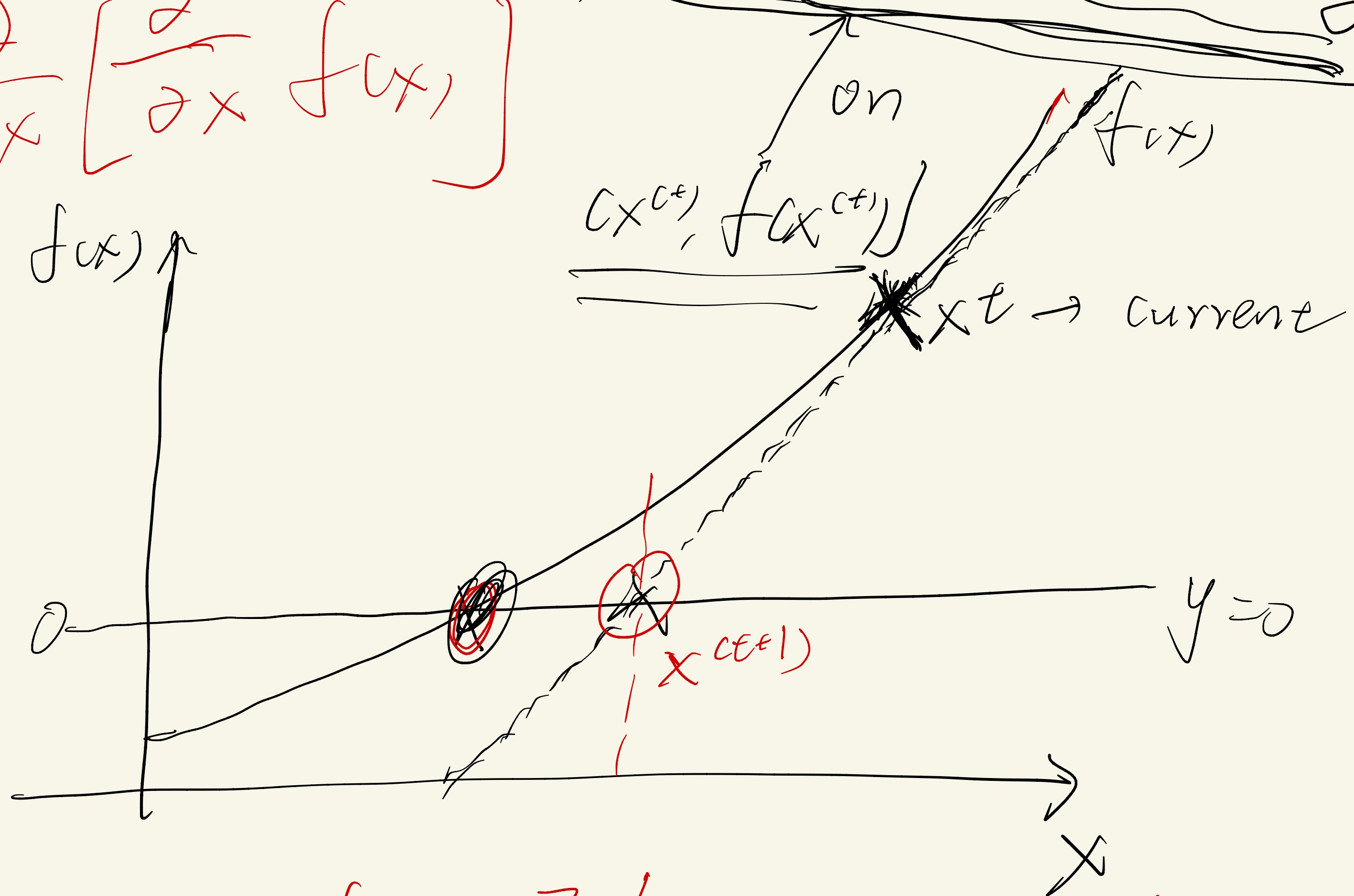
$$= \theta_i^{T\,old} x - [\phi_i^{old} - 1(y=i)]\, x^T x$$

old prob

$x^T x = \|x\|_2^2 \geq 0$

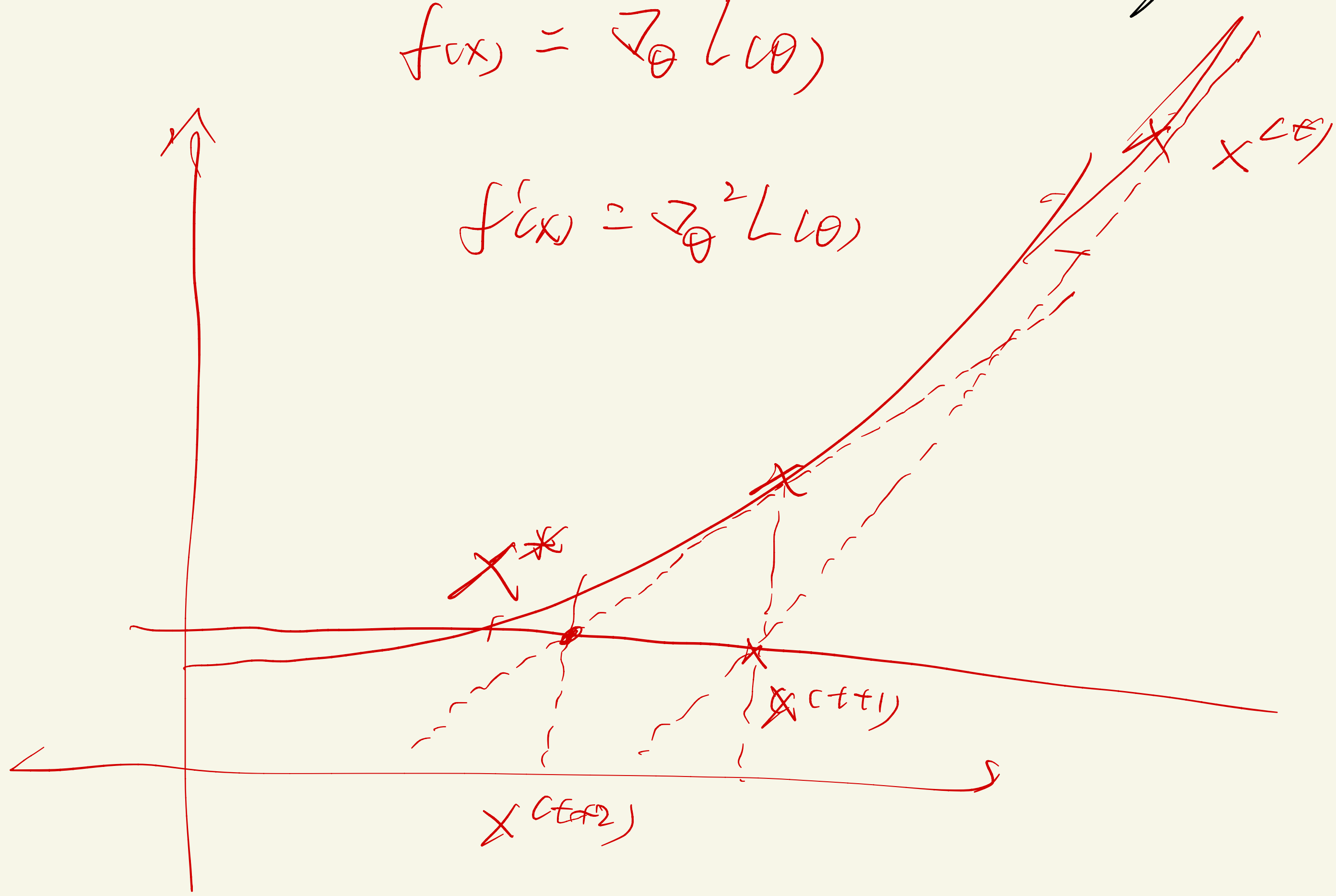
$$f'(x^{(t)})x^{(t+1)} + f(x^{(t)}) - x^{(t)} f'(x^{(t)}) = y$$

$$\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} f(x)\right]$$



$$f(x) = \nabla_\theta L(\theta)$$

$$f'(x) = \nabla_\theta^2 L(\theta)$$



$$\phi_i = h_\theta(x) = g(\theta^T x)$$

$$\phi_i = \underline{\theta^T x}$$

$$\phi_i \in [0, 1]$$ su