AI → ML → Supervised

ML → Unsupervised

ML ↓

حل مسائل با بینش غریزه

$x \xrightarrow{f} y$ → regression ✓linear reg } tree

→ prediction ✓logisti reg

x → Clustering K means

→ Association

→ Dimension reduct PCA

h_θ $\underbrace{ax + b}$ + noise

آمار کلاسیک

Least square

↓

Analytic حل دقیق

جبر و اثبات خط

best fit

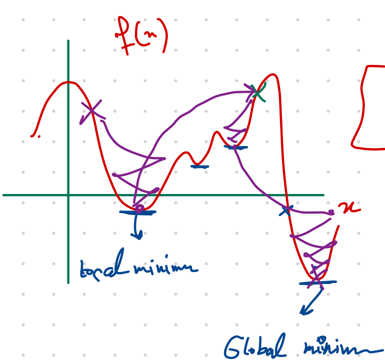
ML

1. رندوم
2. اصلاح
3. loss ↓

a b ≈ a^* b^*

↓

پیش‌بینی !



① $h_\theta = \boxed{\theta_0 x_0 + \theta_1 x_1}$

↓

$\theta^T X$

رندوم

② Loss function

sum of square distance

③ optimize

Stochastic Gradient descent

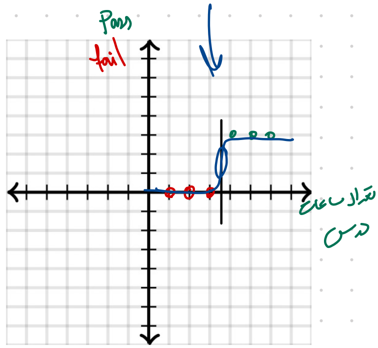
converge

$\theta \approx \theta$

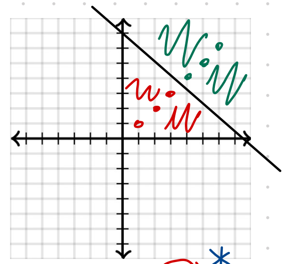
Learning Algorithm

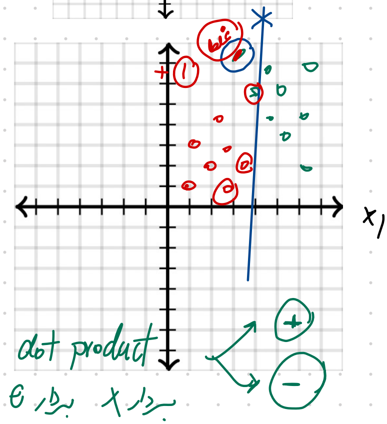
Logistic regression - supervised - prediction $y \in \{0,1\}$

x_0	x_1	x_2	y
1	1	1	0
1	2	2	0
1	3	3	0
1	4	4	1
1	5	5	1
1	6	6	1



decision boundary





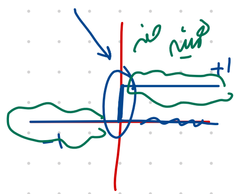
① $h_\theta \sim h_\theta(x) = \text{sign}\{\theta^T x\}$

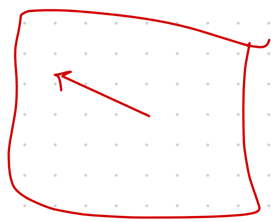
+1 عدد > 0

−1 عدد < 0

عدد = 0

Loss ↗





سیگموئید $g(z) = \dfrac{1}{1+e^{-z}}$

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

① $h_\theta : \dfrac{1}{1+e^{-(\theta^T x)}}$

$$\begin{cases} \mathbb{P}(y=1) = \underbrace{g(\theta^T x)} \\ \mathbb{P}(y=0) = 1 - h_\theta(x) \end{cases} \quad \begin{matrix} [0,1] \\ \{0,1\} \end{matrix}$$

$$\mathbb{P}(y=1 \mid x;\theta) = \left(h_\theta(x)\right)^{y} \cdot \left(1 - h_\theta(x)\right)^{1-y}$$

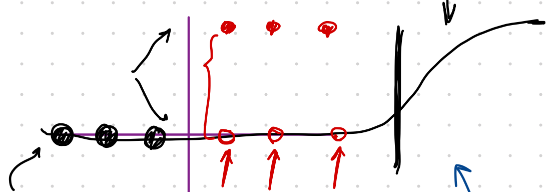
$$\begin{cases} P(y=1) \geq .5 & 1 \\ P(y=1) < .5 & 0 \end{cases}$$

ضرب پایه بدون پیش‌بینی

②

$\mathbb{P}(\text{data} \mid \theta)$

$$\underbrace{\text{Likelihood}(\theta)} = \mathbb{P}(\text{data} \mid \theta)$$

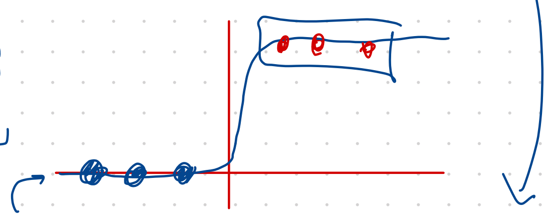


$\approx 1 \times 1 \times 1 \times \approx 0 \times \approx 0 \approx 0$

is more likely

$$\underbrace{\mathbb{P}(\text{data} \mid \theta)}$$

درست‌نمایی



$\approx 1 \quad 1 \quad 1 \quad \approx 1 \times 1 \times 1$

درست‌تر به نظر میرسه

likelihood

$$\mathcal{L}(\theta) = \mathbb{P}(\text{data} \mid \theta) = \prod_{i=1}^{n} h_\theta(x^{(i)})^{y^{(i)}} \cdot \left(1 - h_\theta(x^{(i)})\right)^{(1-y^{(i)})}$$

log likelihood

$$\ell(\theta) = \log \mathcal{L}(\theta) = \log \prod_{i=1}^{n} h_\theta(x^{(i)})^{y^{(i)}} \cdot \left(1 - h_\theta(x^{(i)})\right)^{(1-y^{(i)})}$$

$$= \sum \log \left[h_\theta(x^{(i)})^{y^{(i)}} \cdot \left(1 - h_\theta(x^{(i)})\right)^{(1-y^{(i)})} \right]$$

$$= \sum \log \left[h_\theta(x) \right]^{y} + \log \left[1 - h_\theta(x) \right]^{(1-y)}$$

$$= \sum_{i=1}^{n} y^{(i)} \log \left[h_\theta(x^{(i)}) \right] + (1 - y^{(i)}) \log \left[1 - h_\theta(x^{(i)}) \right]$$

$\ell(\theta)$ بیشینه

gradient ascent

- θ رندم شروع
- $\theta_{t+1} := \theta_t + \alpha \underbrace{\frac{\partial}{\partial \theta} \ell(\theta)}_{\text{مشتق؟}}$
- repeat

$$\frac{\partial}{\partial\theta} l(\theta) = \frac{\partial}{\partial\theta}\, y \log\left[h_\theta(x)\right] + (1-y)\log\left[1-h_\theta(x)\right]$$

$$\frac{\partial}{\partial\theta}\; \boxed{y \log\left(g(\theta^T x)\right)} + \boxed{(1-y)\log\left[1-g(\theta^T x)\right]}$$

θ

چهارم ضرب

سوم لگاریتم

دوم سیگموید

ورودی

اول ضرب

$$\frac{\partial}{\partial\theta}\, \underline{y\left(\log g(\theta^T x)\right)} = \frac{\partial}{\boxed{\partial\theta}} \underbrace{\theta^T x}_{} = x$$

$$\times\; \frac{\partial}{\partial \theta^T x}\, \underbrace{g(\theta^T x)}_{} = g(\theta^T x)\left(1-g(\theta^T x)\right)$$

$$\frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right] = \frac{1+e^{-x}\times(1)' - (1)(1+e^{-x})'}{(1+e^{-x})^2},$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} \times \frac{e^{-x}}{(1+e^{-x})}$$

$$[g(x)]'$$

$$= g(x)\left(1-g(x)\right)$$

$$\times\; \frac{\partial}{\partial g(\theta^T x)}\, \underbrace{\log\left(g(\theta^T x)\right)}_{} = \frac{1}{g(\theta^T x)}$$

$$\times\; \frac{\partial}{\partial \log\left(g(\theta^T x)\right)}\, y \boxed{\log\left(g(\theta^T x)\right)} = y$$

$$\left[y \times \frac{1}{g(\theta^T x)} \times \boxed{g(\theta^T x)(1-g(\theta^T x)) \times X} \right] +$$

$$\left[-(1-y) \frac{1}{(1-g(\theta^T x))} \times \boxed{g(\theta^T x)(1-g(\theta^T x)) \times X} \right]$$

$$= \left[\frac{y}{g(\theta^T x)} - \frac{(1-y)}{1-g(\theta^T x)} \right] g(\theta^T x)(1-g(\theta^T x)) X$$

$$= \frac{y(1-g(\theta^T x)) - (1-y)(g(\theta^T x))}{\cancel{g(\theta^T x)\ 1-g(\theta^T x)}} \cancel{g(\theta^T x)(1-g(\theta^T x))} \boxed{X}$$

$$\underbrace{y - \cancel{yg(\theta^T x)} - 1(g(\theta^T x))}\ \cancel{+ yg(\theta^T x)}$$

$$\frac{\partial}{\partial \theta_j} = \underbrace{(y - \hat{y})}\, x_i$$