No.

Date

Binary Classification

(x, y) $x \in \mathbb{R}^{n_x}$, $y \in \{0, 1\}$

$\downarrow$ $\downarrow$

数据 标签

$\{x^{(1)}, y^{(1)}\}, \{x^{(2)}, y^{(2)}\}, \ldots \{x^{(m)}, y^{(m)}\}$

training set = m / m_{train} :

test set = m_{test}

矩阵形式.

$$X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{bmatrix} \updownarrow n_x$$

$\longleftarrow m \longrightarrow$

m个 n_x维的向量.

$X \in \mathbb{R}^{n_x \times m}$.

X.shape $= (n_x, m)$.

$Y = [y^{(1)}, y^{(2)}, \ldots y^{(m)}]$ $Y \in \mathbb{R}^{1 \times m}$

Y.shape $= (1, m)$

No.

Date

Logistic Regression.

X 图像 → 由像素值组成的特征向量. $X \in R^{n_x}$

$\hat{y}$ 条件概率 $P(y=1|x)$.

在x的条件下, $y=1$ 的概率

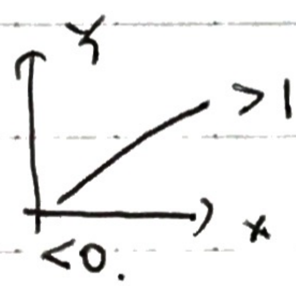
总结, $0 \le \hat{y} \le 1$.

参数 $\begin{cases} \text{权值 } w \in R^{n_x} \\ \text{偏置值 } b \in R. \end{cases}$

为什么线性回归模型不能解决?

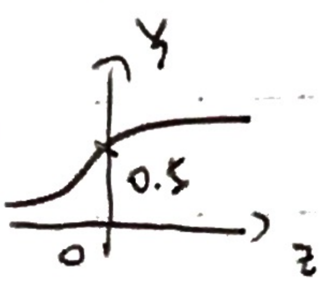
$\hat{y} = w^T x + b$.

$\because \hat{y} \in [0, 1]$, 但



$\therefore$ 激活函数 σ sigmoid.

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad z \in R.$$



$$\lim_{z \to +\infty} \sigma(z) = \frac{1}{1+0} = 1$$

$$\lim_{z \to -\infty} \sigma(z) = \frac{1}{+\infty} = 0.$$ 符合 $\hat{y} \in [0, 1]$

No.

Date

LR Cost func

$y^{(i)} = \sigma(w^T x^{(i)} + b)$，其中 $\sigma(z) = \frac{1}{1+e^{-z}}$

有一组训练集 $\{(x^1, y^1), \dots (x^m, y^m)\}$,

期望 $\hat{y}^{(i)} \to y^{(i)}$ 预测值接近真实值.

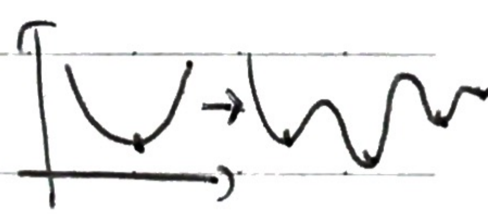
定义 Loss func.

没理解这个 Loss 的选取理由.

$$L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

在LR中不这样定义 Loss func的原因是?

因为在LR中优化问题是非凸的. 这样一来梯度下降只能得到多个局部最优解, 而非全局最优.



在LR中定义 Loss func:

$$L(\hat{y}, y) = -(y\log\hat{y} + (1-y)\log(1-\hat{y}))$$

$y=1$, $L = -\log\hat{y} \to L\downarrow$, $\log\hat{y}\uparrow$, $\hat{y}\uparrow \to \hat{y}\uparrow$ 与 $y=1$ 接近

$y=0$, $L = -(1-y)\log(1-\hat{y}) \to L\downarrow$, $\hat{y}\downarrow \to \hat{y}\downarrow$ 与 $y=0$ 接近

这是针对一个 (x^i, y^i) 组合而言的.

No.

Date

当我们对于样本训练集 $\{(x^1,y^1),\dots,(x^m,y^m)\}$.

Cost func $= J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^i, y^i)$.

参数.

即 $-\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log \hat{y}^i + (1-y^i)\log(1-\hat{y}^i)\right]$.