Keywords 关键词 | Notes 笔记

## P3. 回归 regression

- Stock market forecast
- Self driving car
- recommendation

example: $f(x_{cp}, x_s, x_{hp}, x_h, x_w) = y$

**step 1: model: a set of function**

$y = b + w \cdot x_{cp}$ ⇒ Linear model

$$y = \underbrace{b}_{\text{bias}} + \sum \underbrace{w_i}_{\text{weight}} \underbrace{x_i}_{\text{feature}}$$

**step 2: goodness of function**

training data: $(x^1, \hat{y}^1) \cdots (x^{10}, \hat{y}^{10})$

↓

loss function L: $L(f) = L(w, b)$

$$= \sum_{i=1}^{n} \left(\hat{y}^i - (b + w \cdot x^i)\right)^2$$

(w,b作直角坐标系图,函数为点)

**step 3: Best function** $f^* = \arg\min_f L(f)$

gradient descent，处理任意L(f)可微

| Keywords 关键词 | Notes 笔记 |
|---|---|

## Gradient descent

1. 随机选取初始点 $w^0$

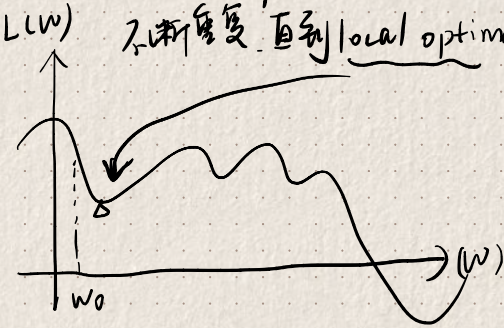
2. 算微分 $\left.\frac{dL}{dw}\right|_{w=w^0}$

$\begin{cases} \text{negative} \rightarrow \text{increase } w \\ \text{positive} \rightarrow \text{decrease } w \end{cases}$ $\Downarrow$ L变小

学习率

$$w^1 = w^0 - \eta \left.\frac{dL}{dw}\right|_{w=w^0}$$

不断重复 直到 local optimal (局部)




(对于 w, b, 需计算两个偏微分)

$$\nabla L = \begin{bmatrix} \frac{\partial L}{\partial w} \\ \frac{\partial L}{\partial b} \end{bmatrix} \text{gradient}$$

此时 Loss L(w, b) 是椭圆等高线?

等高线法线方向

线性模型中 Loss 是没有局部最优值的 convex

$$\frac{\partial L}{\partial w} = \sum 2(y^n - (b + w x^n))(-x^n)$$

$$\frac{\partial L}{\partial b} = \sum 2(y^n - (b + w x^n))(-1)$$

Keywords 关键词 | Notes 笔记

generalization 外推 test data

loss ( test ) > loss ( train )

selecting another model (step 1)

$$y = b + w_1 x_{cp} + w_2 (x_{cp})^2$$

$$y = b + w_1 x_{cp} + w_2 (x_{cp})^2 + w_3 (x_{cp})^3$$

... ...

⇒ 模式复杂后，测试集误差可能会非常大
overfitting

⇒ 引入其他特征，优选模型

$\boxed{\delta(x)}$ 真值 int(bool) [冲激函数?]

regularization 正则

$$L = \sum (y - \hat{y}) + \boxed{\lambda \sum (w_i)^2}$$

参数值接近0的函数更平滑（斜率小）

（太平滑变成水平线，Loss大）