

Task02

2021年8月18日 16:08

P3 回归

- 应用可以有：PM2.5预测，股市预测，无人驾驶，推荐；
- Step1 Model: (Linear model: 其中b表示偏置，w表示权重，x表示属性特征)

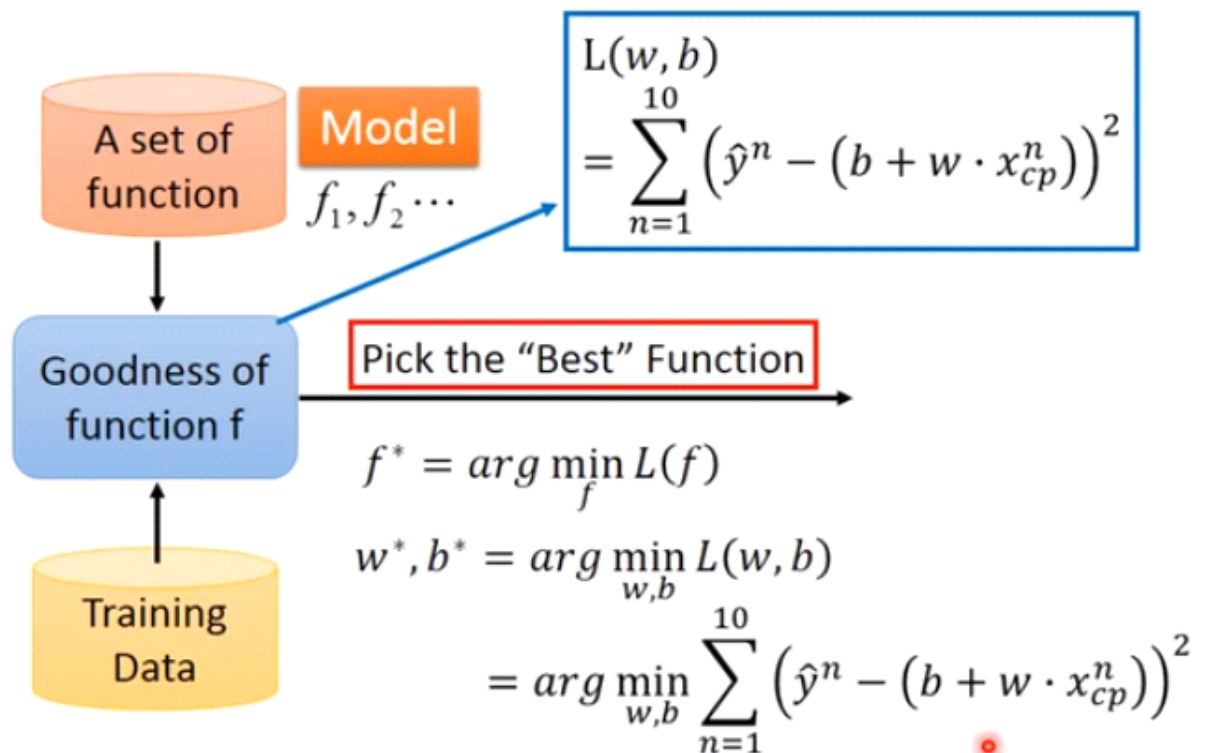
$$y = b + \sum w_i x_i$$

- Step2 Goodness of Function: 判别Function好坏的方式，定义损失函数 loss function，输入是函数function，即参数是w和b，如，定义loss function L为

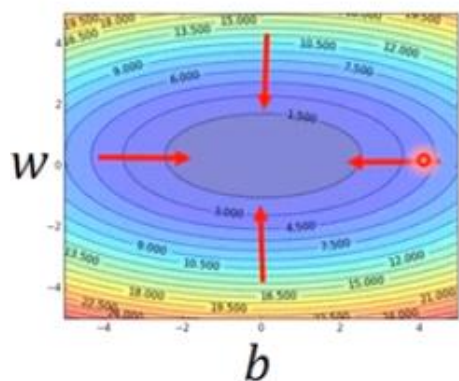
$$L(w, b) = \sum_{n=1}^{10} (\hat{y}^n - (b + w \cdot x_{cp}^n))^2$$

- Step3 Best function

Step 3: Best Function



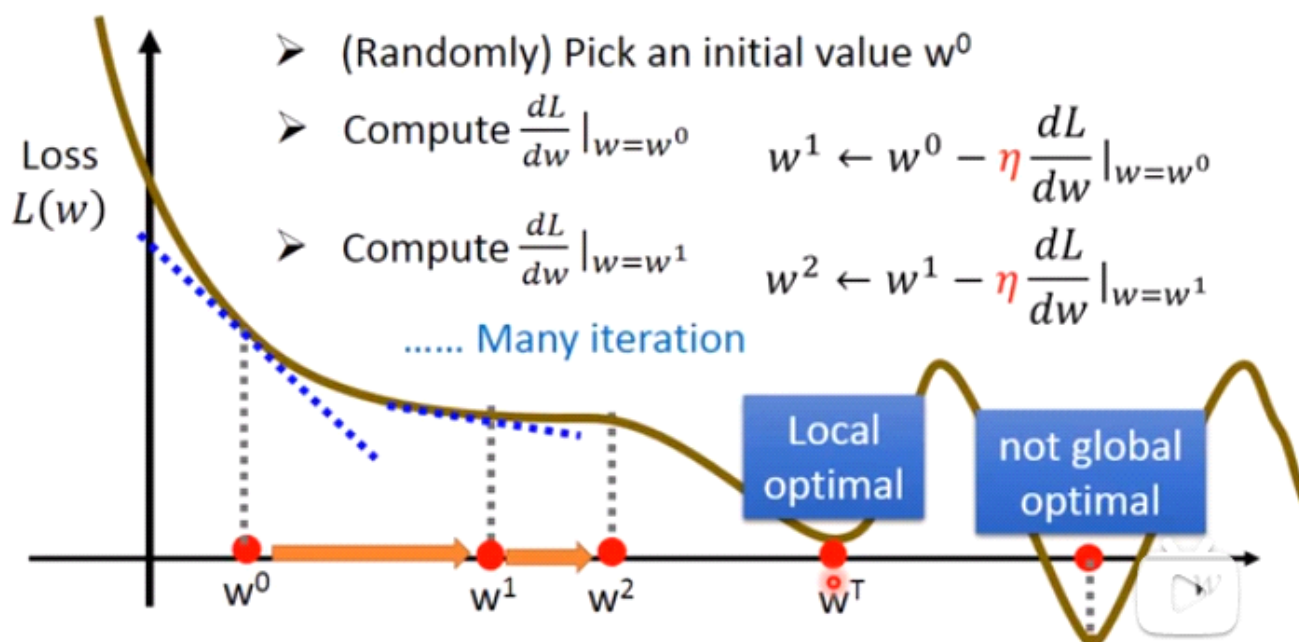
- 以上方程的求解方法，Gradient descent 梯度下降，主要思想，L函数的参数是w和b，要寻找使得L最小的w和b，初始化参数，求偏导，如果导数小于0，那增大参数值，导数大于0，则减小参数值，增大或减小的程度要视当前所求偏导的大小决定，如果导数大，即变化较大，则选择一个较大的变化步长能够取得更好的优化效果，如果导数较小，则选择一个较小的步长。逐步优化后直到步长足够小则停止。
- 问题：只能找到局部最优，不能找到全局最优，因此比较适合linear regression



Step 3: Gradient Descent

$$w^* = \arg \min_w L(w)$$

- Consider loss function $L(w)$ with one parameter w :



- 训练数据时如果表现不好，则需另外选择model，引入二次式更好的拟合

Selecting another Model

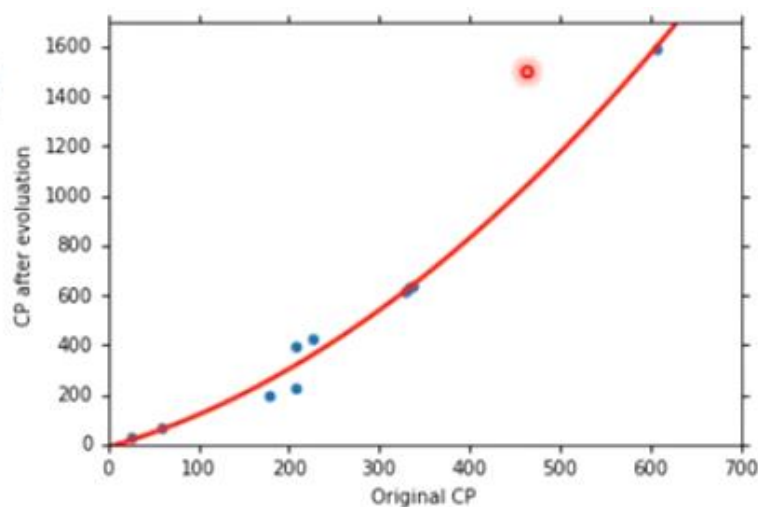
$$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2$$

Best Function

$$b = -10.3$$

$$w_1 = 1.0, w_2 = 2.7 \times 10^{-3}$$

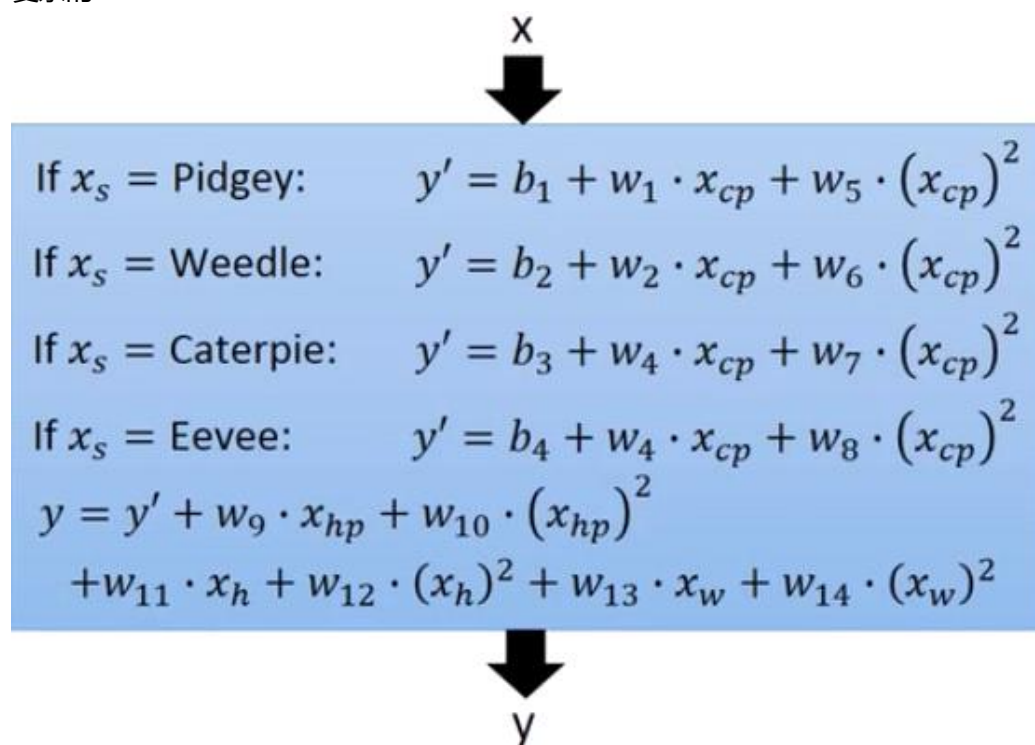
$$\text{Average Error} = 15.4$$



- 并不是训练数据时越贴切越好，一定程度后会过拟合（Overfitting），即在训练数据时，

loss极小, 但在testing data上, 误差会很大

- 复杂的function



- 更好的 loss function, 想要更小的w, 因为w更小, 浮动更小, 函数会更平滑。其中 lambda越大, 函数越平滑; 但此时训练数据的误差考虑更少,

Back to step 2: Regularization

$$y = b + \sum w_i x_i$$

The functions with
smaller w_i are better

$$L = \sum_n \left(\hat{y}^n - \left(b + \sum w_i x_i \right) \right)^2 + \lambda \sum (w_i)^2$$

Regularization



λ	Training	Testing
0	1.9	102.3
1	2.3	68.7
10	3.5	25.7
100	4.1	11.1
1000	5.6	12.8
10000	6.3	18.7
100000	8.5	26.8

➤ Training error: larger λ , considering the training error less

- 因此，合适的lambda能取得足够好的function
- regularization时，不需要考虑bias，因为b不会影响函数的平滑程度