上海交通大學
SHANGHAI JIAO TONG UNIVERSITY
学生创新中心
Student Innovation Center

深度学习入门——回归与分类
Regression & Classification

学生创新中心：肖雄子彦

SJTU

目录 / CONTENTS

线性回归
Linear Regression

学习目标：

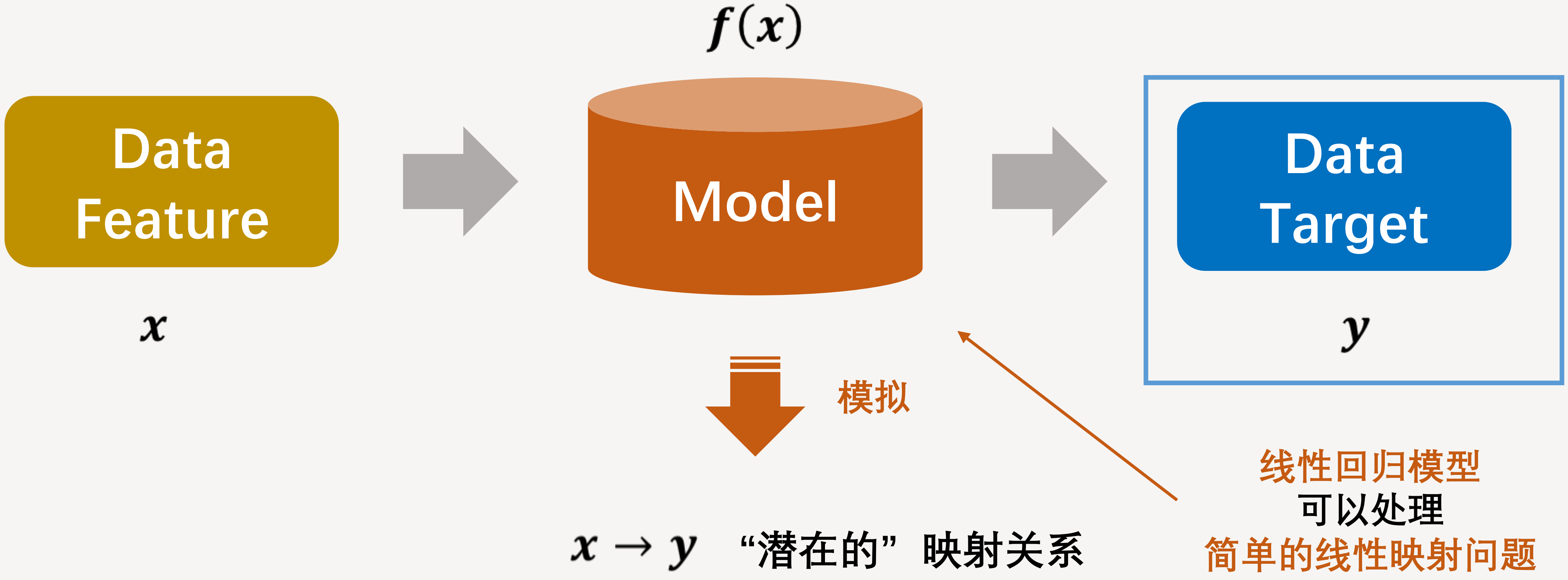
- **掌握**线性回归相关**概念、建模方法**
- **识别**生活中**可用线性回归**解决的问题
- **运用线性回归**对案例进行**训练**和**预测**

What’s Linear Regression

Regression —— foundation of Deep Learning

机器学习/深度学习本质上都是寻找一组最优的函数映射关系

Supervised Learning(监督学习)



What’s Linear Regression

线性回归：从以某种概率分布的样本中找到数据特征与目标值之间的函数关系

回归任务	input 输入		output 输出
房价预测	$f($ 房屋面积、地理位置 $)$	$=$	房价
期末考评	$f($ 课堂表现、打榜分数 $)$	$=$	课程成绩
质量评估	$f($ 红酒的年代、成色 $)$	$=$	质量分数

具体数值

……

<u>线性回归特点：输出的数值连续而非离散</u>

Linear Regression

Example

Self-motivation自驱力！

$f($  $) = ($ Final Result $)$

x Feature

y Target

$f\ ()\ \ ?$

$x_1,\ x_2,\ x_3\ \ldots$

活跃度
课堂答题
小组pre得分
…

$$y = w_1 x_1 + w_2 x_2 + \cdots + b$$
$$= \sum w_i x_i + b$$

权重 weight

偏置量 bias

Parameters $\boldsymbol{\theta}$

Linear Regression

假如现在只考虑这一个影响因素

Training data

$(x^1, \hat{y}^1)$

$(x^2, \hat{y}^2)$

⋮

$(x^{10}, \hat{y}^{10})$

⋮

$\hat{y}^i$

x^i

希望每一组数据的
Feature X

$$y = w \cdot x + b$$

得到的值尽可能
接近Target Y

Linear Regression

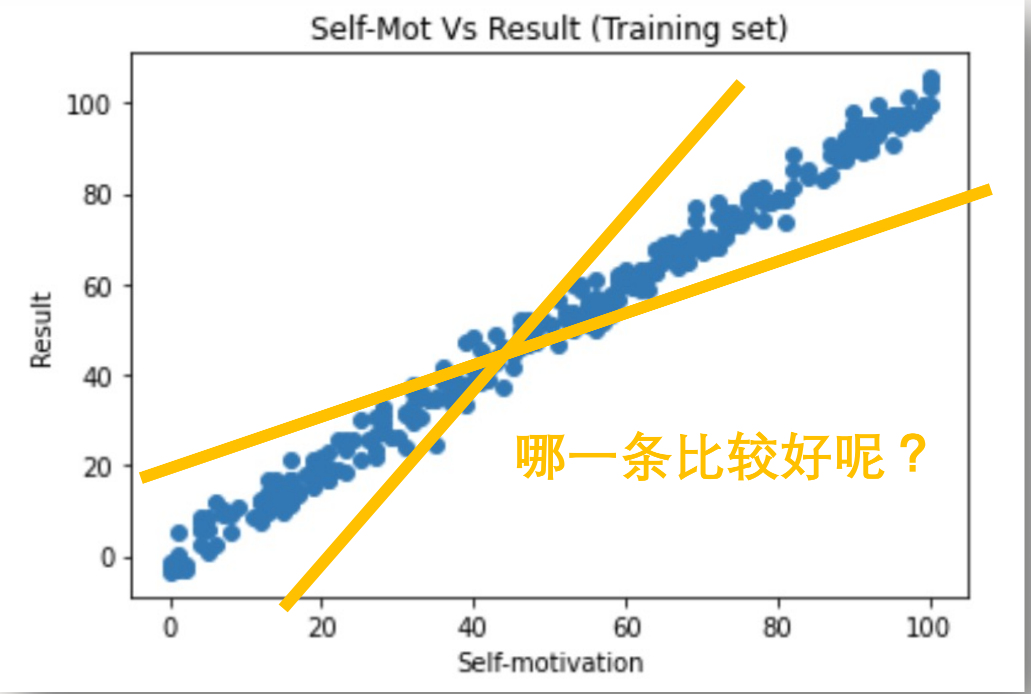
怎么找到最合适的一组（w，b）？

- Loss function损失函数：来衡量这组参数得到的模型性能。
  （计算网络输出的值与实际Target间的差距）

Square Error

$$L(f) = \frac{1}{2}\sum_{i=1}^{n}(f(x^i) - \hat{y}^i)^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}((wx^i + b) - \hat{y}^i)^2$$



- Best function 损失函数最小

$$f^* = \underset{w,b}{\operatorname{argmin}} L(f)$$

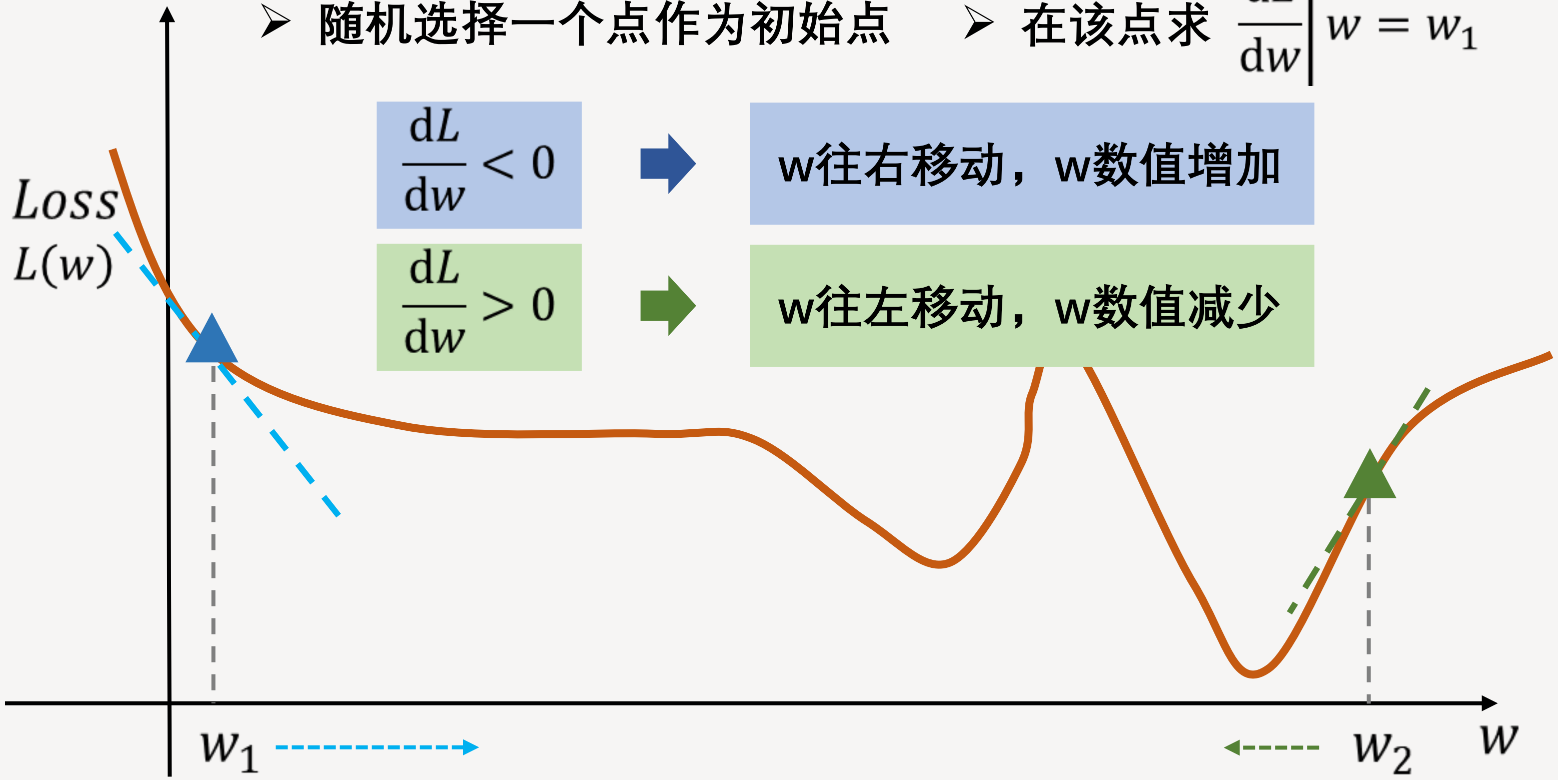
Method : Gradient Descent

怎么找到Loss的最小值？（以一个参数w为例）

Target: 找到使Loss值最小的w

- 随机选择一个点作为初始点
- 在该点求 $\left.\frac{\mathrm{d}L}{\mathrm{d}w}\right|_{w=w_1}$

$\frac{\mathrm{d}L}{\mathrm{d}w} < 0$	➡	w往右移动，w数值增加
$\frac{\mathrm{d}L}{\mathrm{d}w} > 0$	➡	w往左移动，w数值减少



参数会：
沿着导数的反方向更新

$$w_{new} = w_{old} - \alpha \cdot \frac{\mathrm{d}L}{\mathrm{d}w}$$

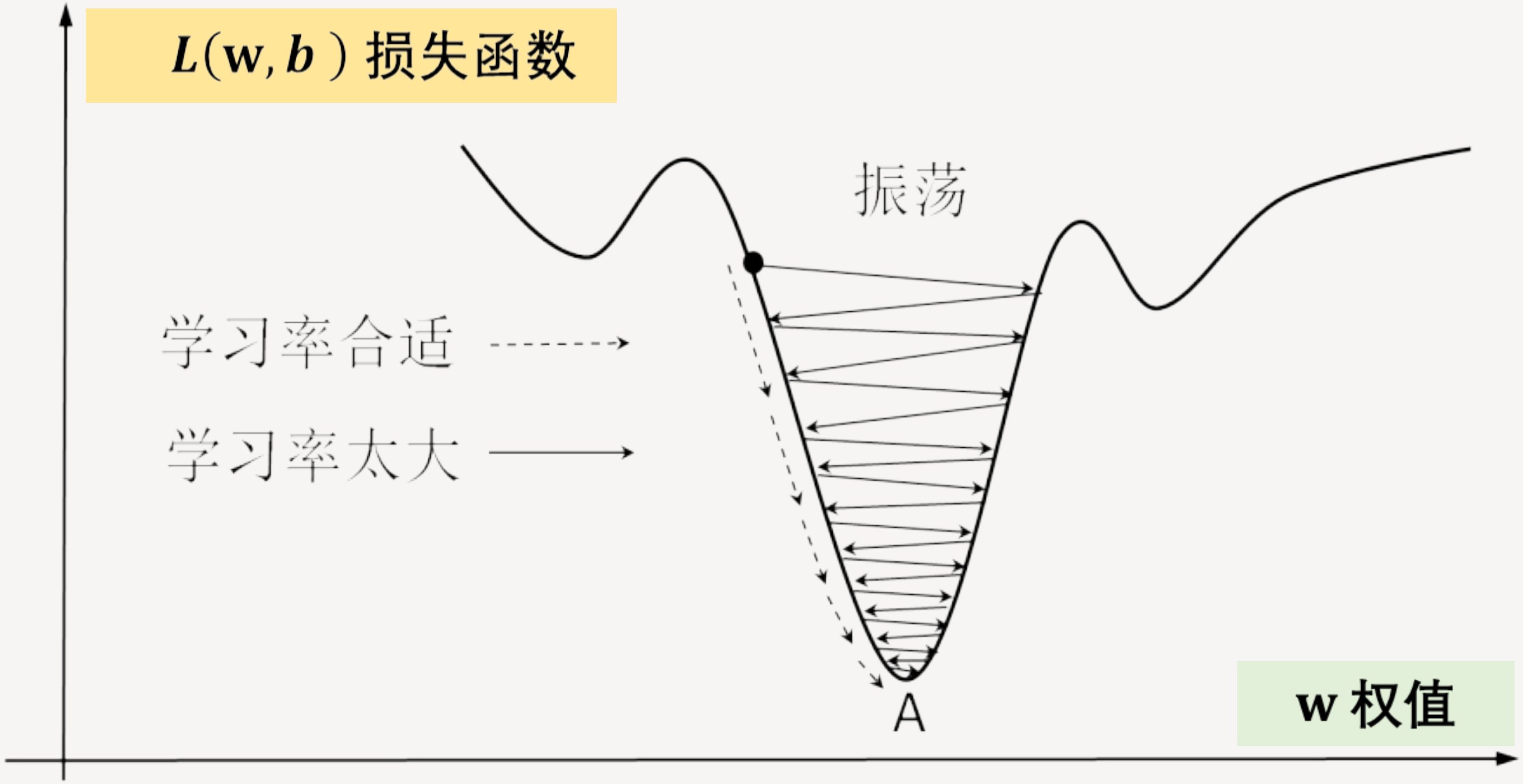
这里 α为学习率
即下降的速度和步长

学习率（Learning Rate）

学习率（Learning Rate）表示参数学习的效率，即参数更新的速度。

$$w_1 = w_0 - \alpha \cdot \frac{\partial L}{\partial w_0}$$

$$b_1 = b_0 - \alpha \cdot \frac{\partial L}{\partial b_0}$$



那学习率到底怎么设置才合适呢？

$$\alpha^t = \frac{\alpha}{\sqrt{t+1}}$$

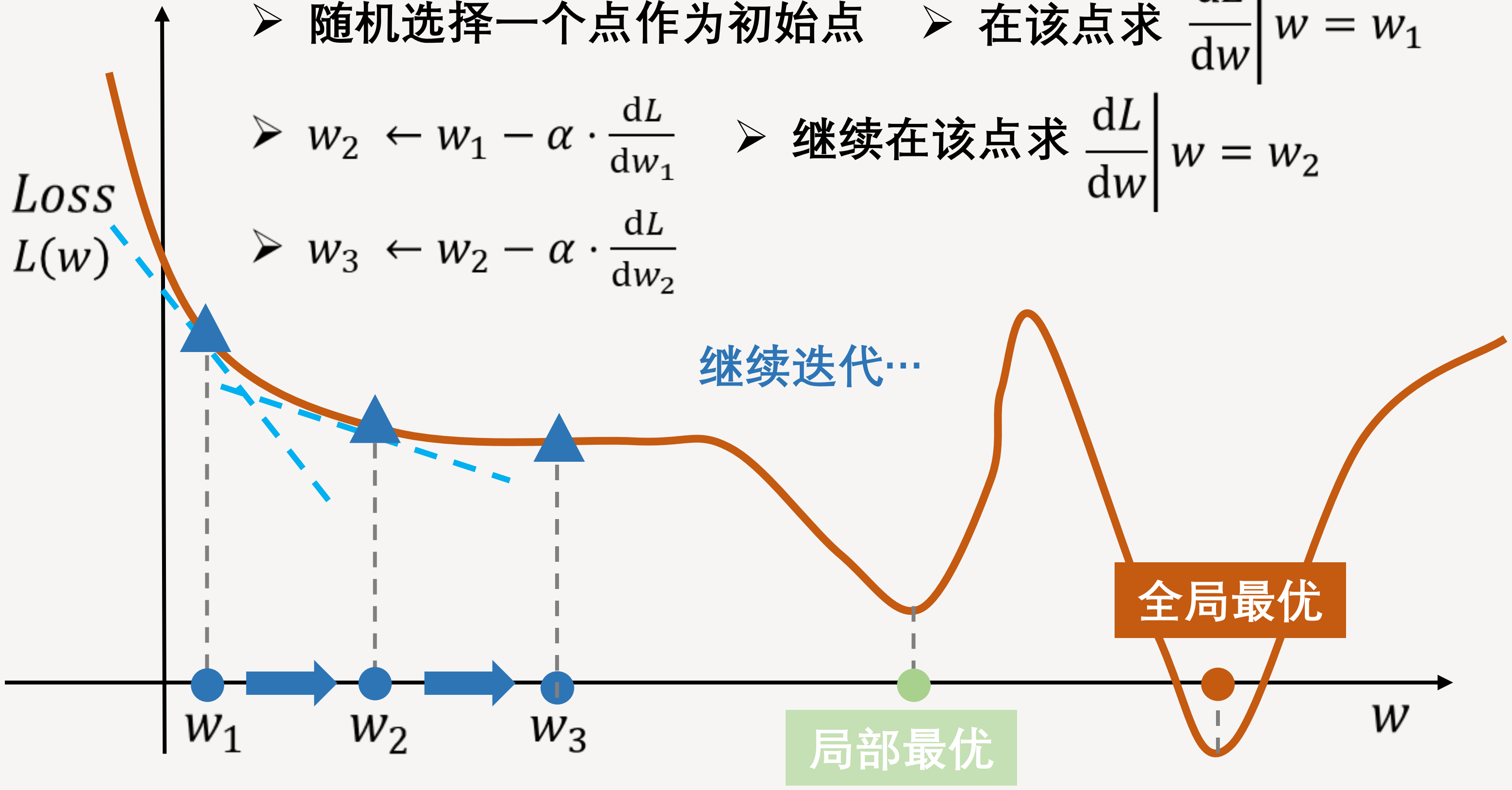
策略之一：可以随着更新次数增多选择更小的Learning Rate…

Method : Gradient Descent

设置好learning rate之后，接下来开始参数的更新

Target: 找到使Loss值最小的w

- 随机选择一个点作为初始点
- 在该点求 $\left.\frac{\mathrm{d}L}{\mathrm{d}w}\right|w = w_1$
- $w_2 \leftarrow w_1 - \alpha \cdot \frac{\mathrm{d}L}{\mathrm{d}w_1}$
- 继续在该点求 $\left.\frac{\mathrm{d}L}{\mathrm{d}w}\right|w = w_2$
- $w_3 \leftarrow w_2 - \alpha \cdot \frac{\mathrm{d}L}{\mathrm{d}w_2}$



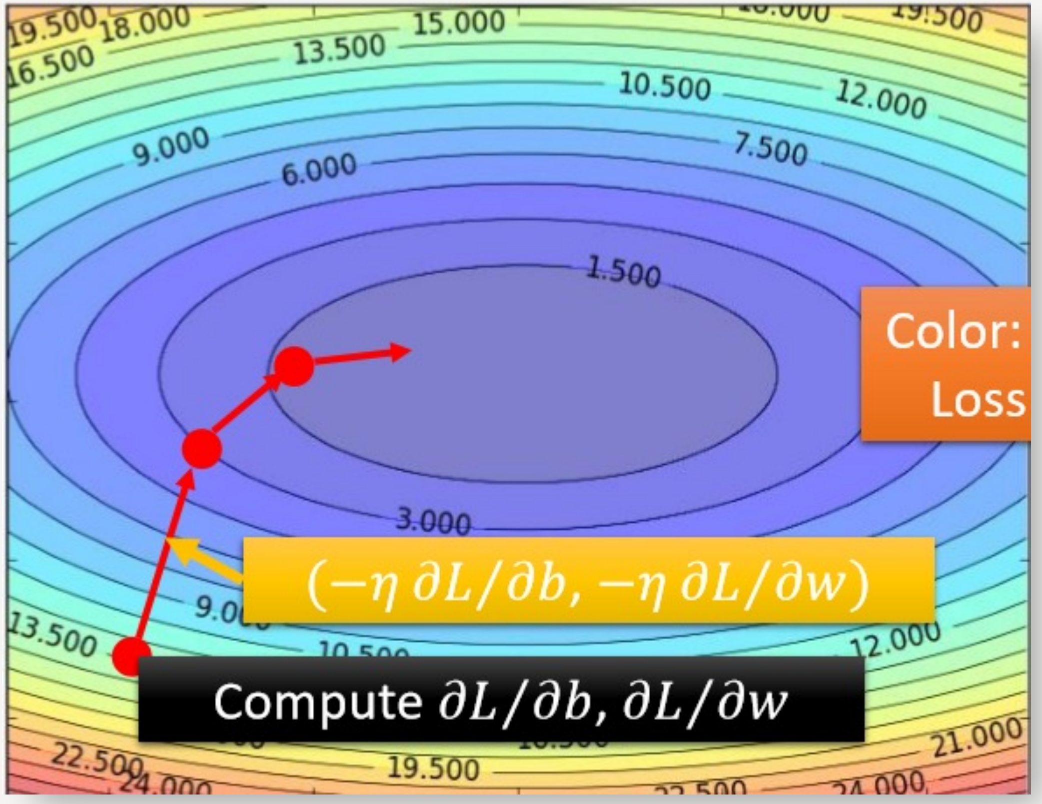
一个参数情况下
参数会
沿着导数的反方向更新

$$w_{new} = w_{old} - \alpha \cdot \frac{\mathrm{d}L}{\mathrm{d}w}$$

这里 α为学习率
即下降的速度和步长

Linear Regression

线性回归都是凸集的，不用担心他陷入局部最优



- 没有局部最优
- 从任意一个地方都能到达整体最优

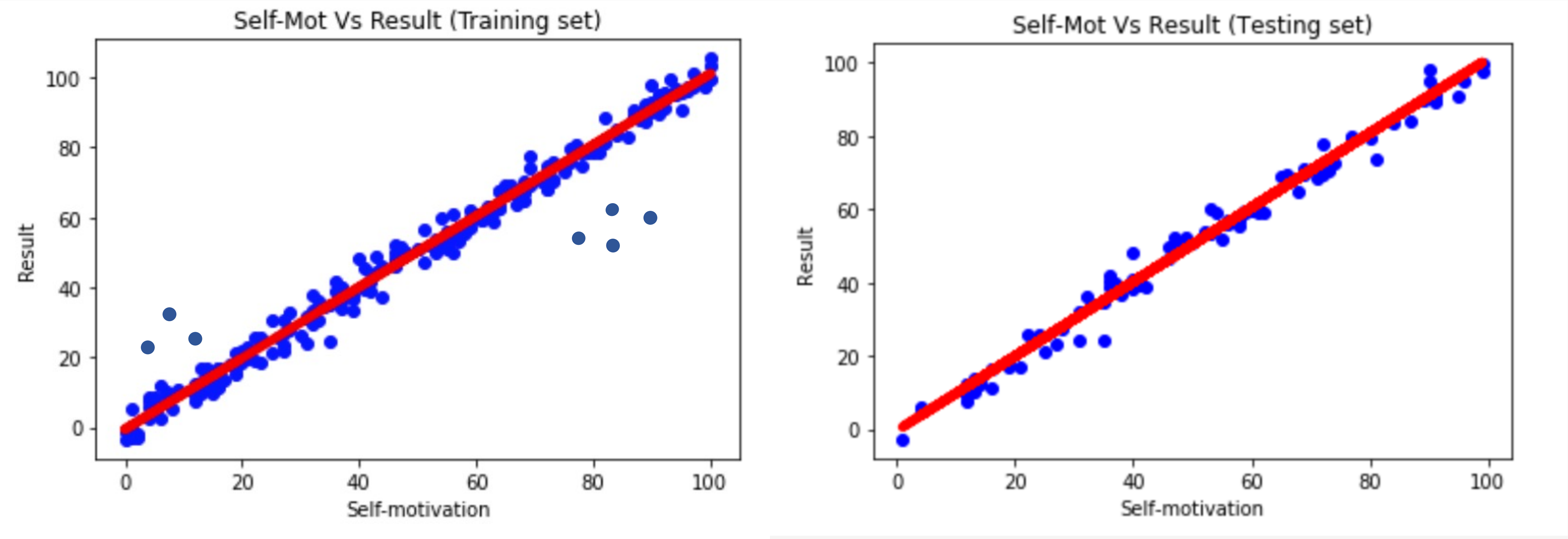
$$\frac{\partial L}{\partial w} = \sum_{i=1}^{n} [f(x^i) - \hat{y}^i]\, x^i$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} [f(x^i) - \hat{y}^i]$$

Linear Regression

用训练集训练出来的model，还需要在测试集上测试

在training set 和testing set上loss都小才√

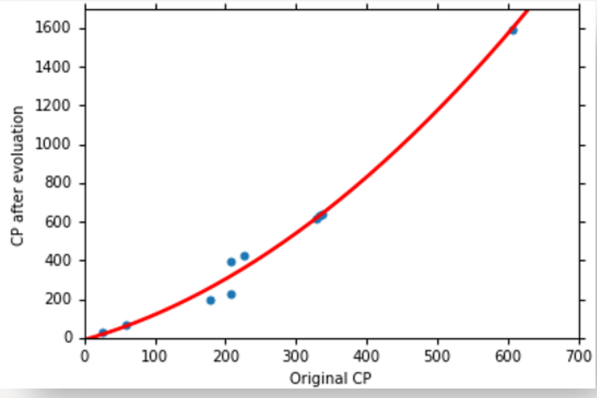


Training data → Validation data → Testing data

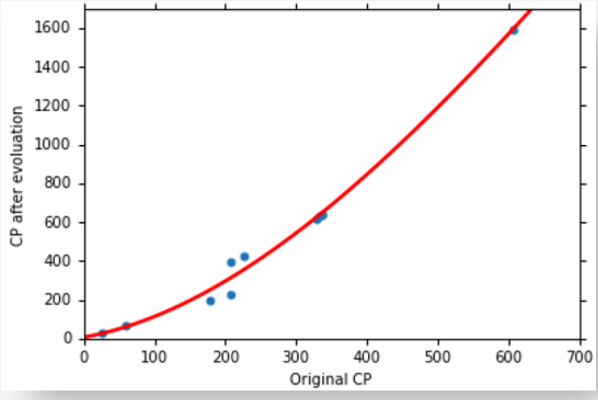
但真实的数据，往往不这么好训练，可能分布并不会这么“集中”。

Linear Regression

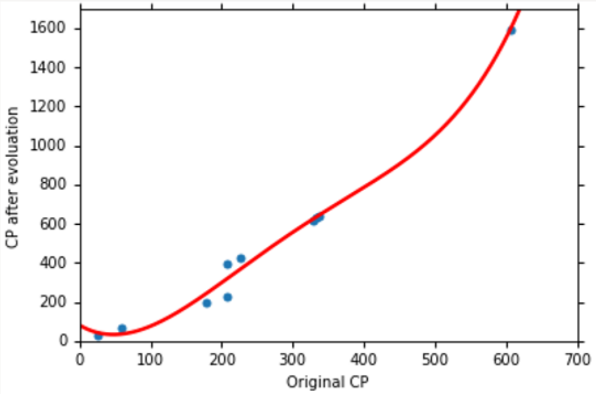
如果想更好去拟合训练集，可以设计更复杂的模型



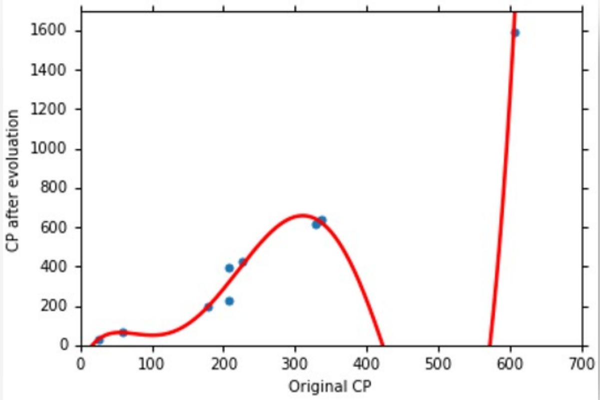
$$y = w_1 x_1 + w_2 x_1^2 + b$$



$$y = w_1 x_1 + w_2 x_1^2 + w_3 x_1^3 + b$$



$$y = w_1 x_1 + w_2 x_1^2 + w_3 x_1^3 + w_4 x_1^4 + b$$



$$y = w_1 x_1 + w_2 x_1^2 + w_3 x_1^3 + w_4 x_1^4 + w_5 x_1^5 + b$$

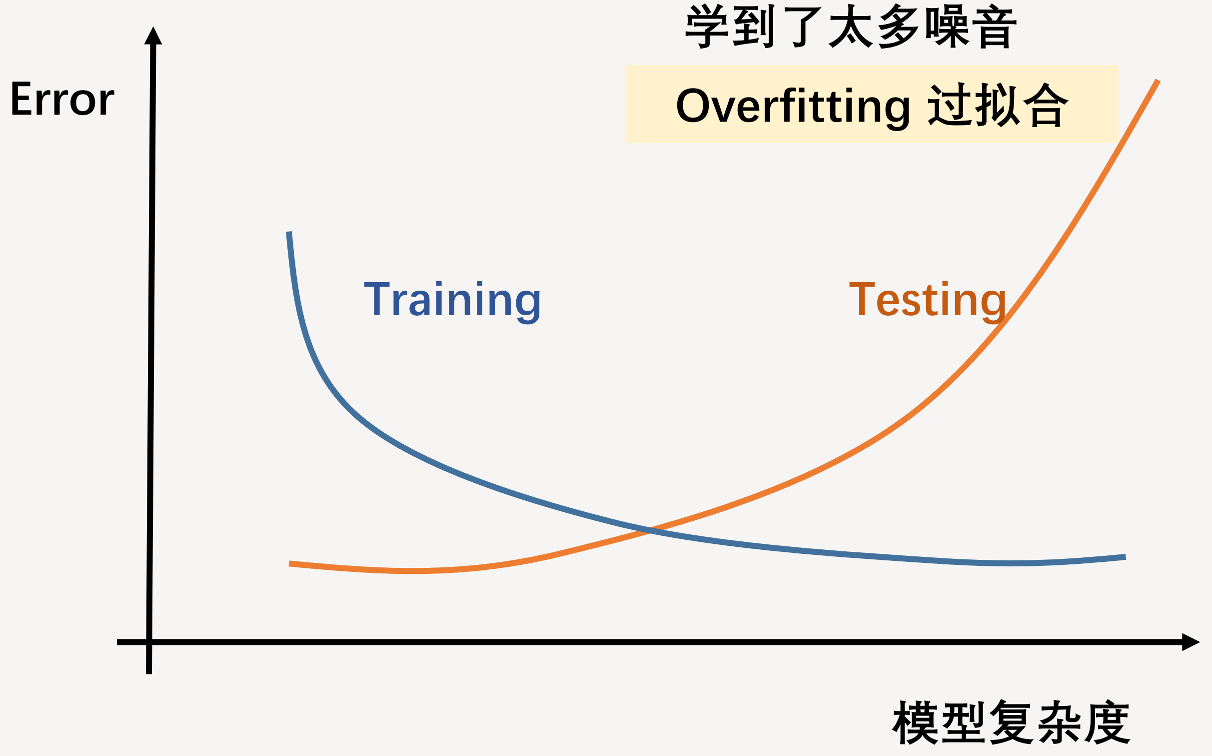
	Training	Testing
1	31.9	35.0
2	15.4	18.4
3	15.3	18.1
4	14.9	28.2
5	12.8	232.1

数据来源
李宏毅教授

SJTU

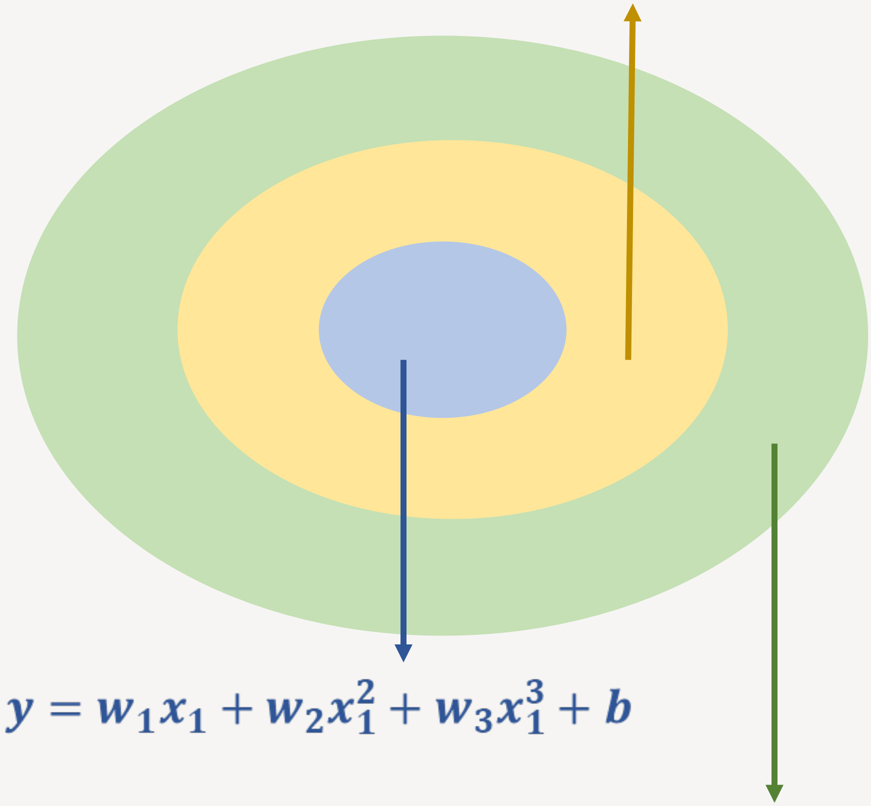
Linear Regression

不同模型在训练集和测试集上的表现



复杂的模型可表达的内容更丰富，学到的东西更多。

$$y = w_1x_1 + w_2x_1^2 + w_3x_1^3 + w_4x_1^4 + b$$



$$y = w_1x_1 + w_2x_1^2 + w_3x_1^3 + b$$

$$y = w_1x_1 + w_2x_1^2 + w_3x_1^3 + w_4x_1^4 + w_5x_1^5 + b$$

多选题 5分

设置

以下对线性回归表述正确的是？

A 线性回归可以确定自变量X对目标值Y的影响程度

B 线性回归的目标值Y必须是定量的，如“得分”、“收入”

C 线性回归的自变量X不一定是定量的，可以是分类变量，如“性别”、“民族”等。

提交

Question
&
Demo