

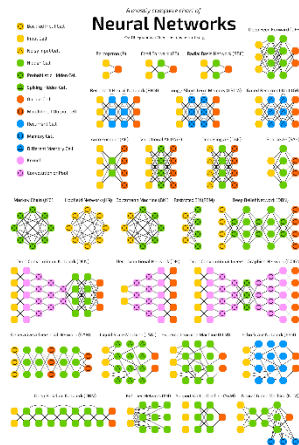
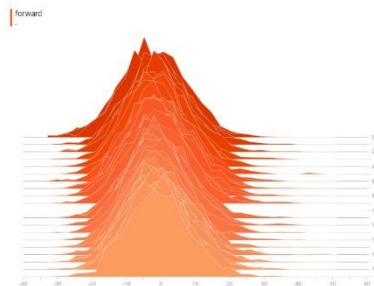
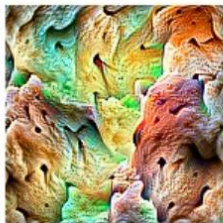
深度学习应用开发

基于TensorFlow的实践

基于TensorFlow的实践

计算机与计算科学学院

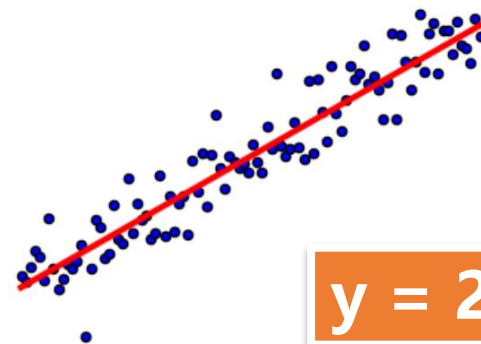
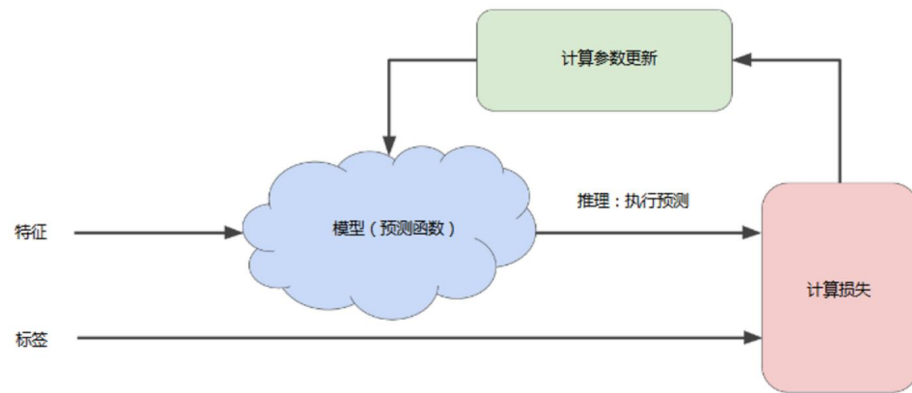
Dept. of Computer Science
Zhejiang University City College



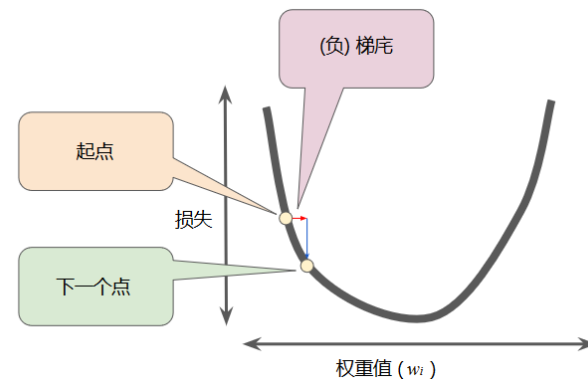
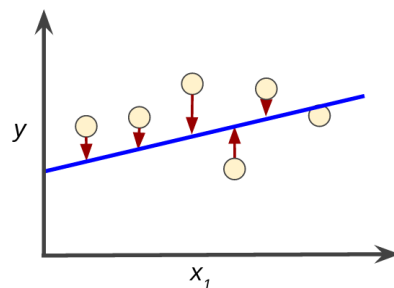
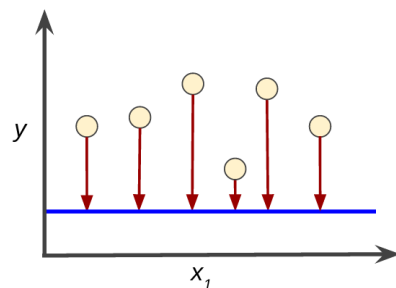


线性回归 TensorFlow实战

第六讲 线性回归问题？用一个神经元搞定！



$$y = 2x + 1$$





浙江大学城市学院
ZHEJIANG UNIVERSITY CITY COLLEGE

监督式机器学习



监督式机器学习



浙江大學城市學院
ZHEJIANG UNIVERSITY CITY COLLEGE



机器学习系统：

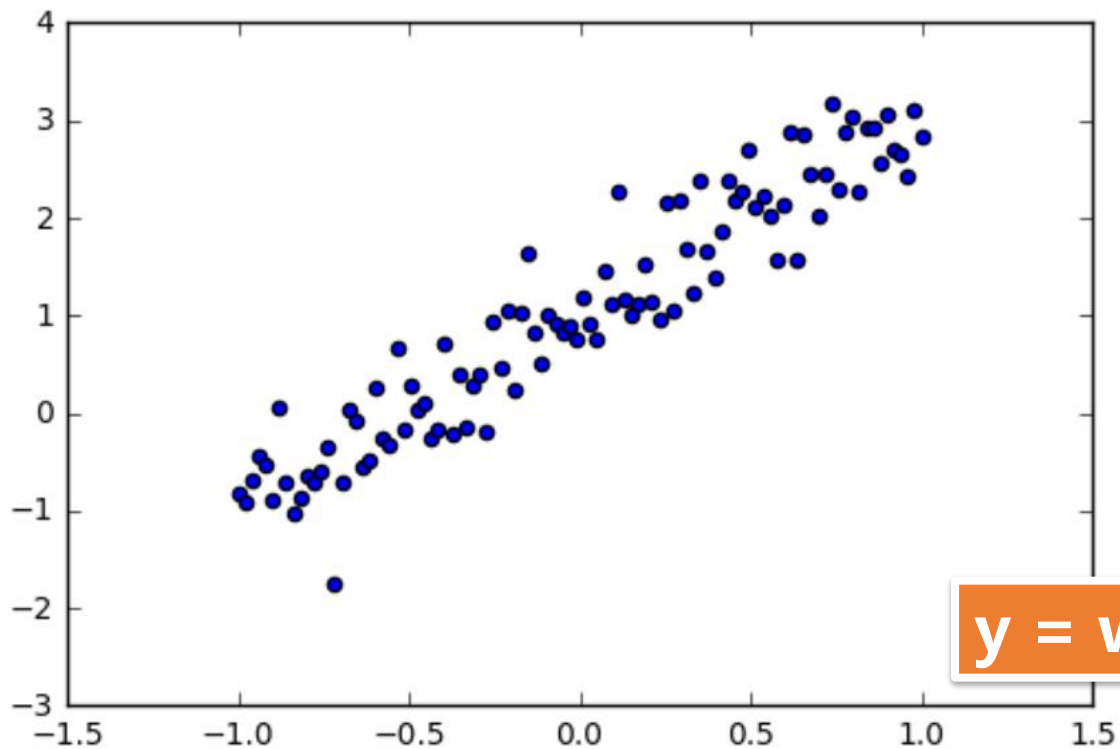
通过学习如何组合输入信息
来对未见过的数据
做出有用的预测

本讲课程部分内容基于“[机器学习速成课程](https://developers.google.cn/machine-learning/crash-course/)”

<https://developers.google.cn/machine-learning/crash-course/>



简单的线性回归案例



$$y = w * x + b$$



术语：标签和特征

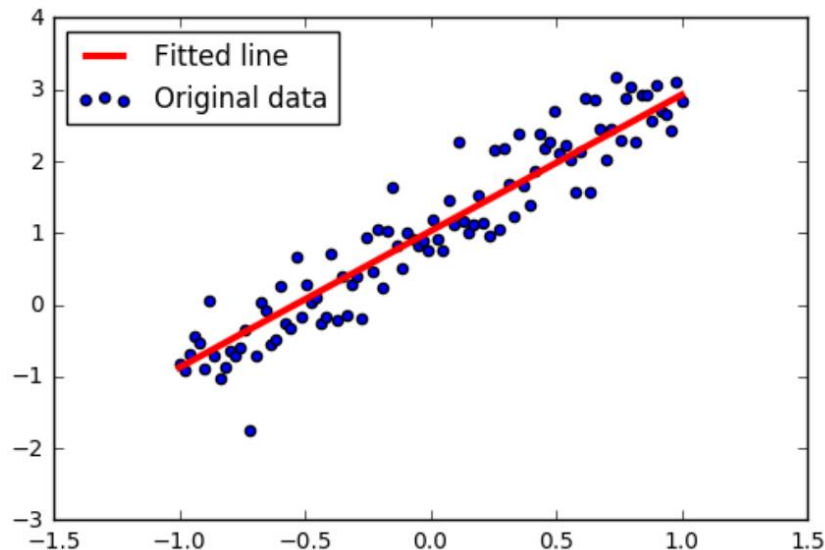


标签是我们要预测的真实事物： y

线性回归中的 y 变量

特征是指用于描述数据的输入变量： x_i

线性回归中的 $\{x_1, x_2, \dots, x_n\}$ 变量



术语：样本和模型

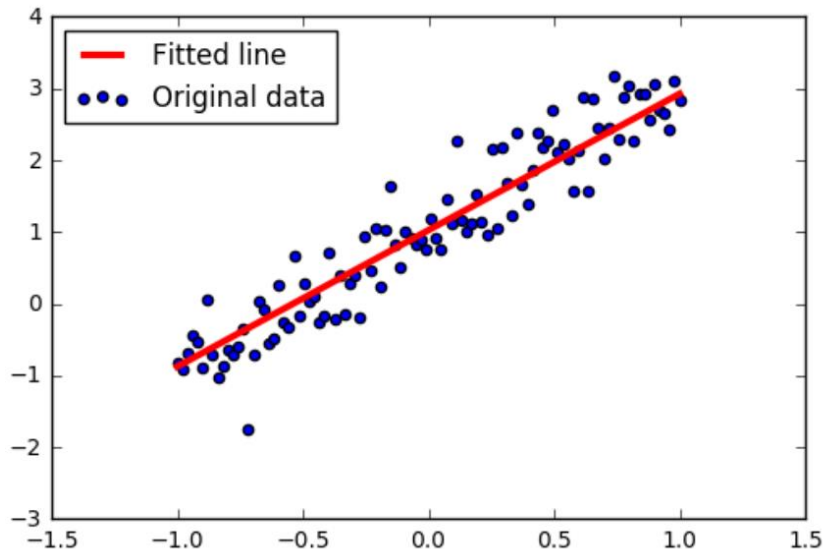
样本是指数据的特定实例： x

有标签样本具有{特征, 标签}： $\{x, y\}$

用于训练模型

无标签样本具有{特征, ? }： $\{x, ?\}$

用于对新数据做出预测





术语：样本和模型

样本是指数据的特定实例： x

有标签样本具有{特征, 标签}： $\{x, y\}$

用于训练模型

无标签样本具有{特征, ? }： $\{x, ?\}$

用于对新数据做出预测

模型可将样本映射到预测标签： y'

由模型的内部参数定义，这些内部参数值是通过学习得到的



术语：训练



训练模型表示通过**有标签样本**来学习（确定）所有**权重**和**偏差**的理想值

在监督式学习中，机器学习算法通过以下方式构建模型：

检查多个样本并尝试找出可最大限度地**减少损失**的模型

这一过程称为**经验风险最小化**



术语：损失



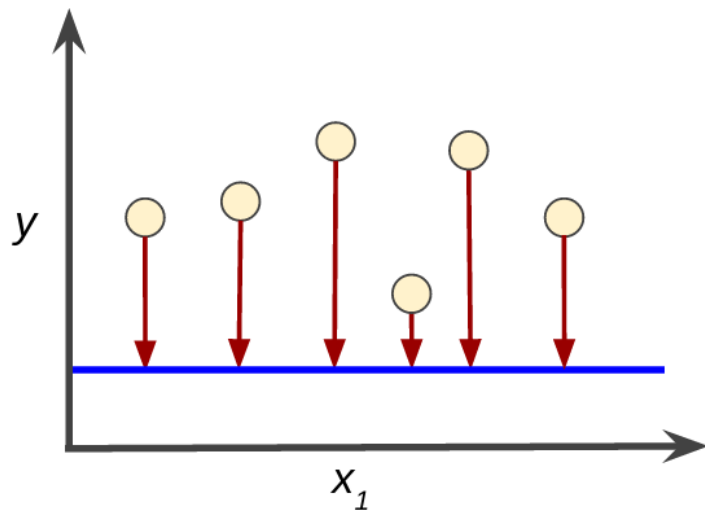
损失是对糟糕预测的惩罚：**损失**是一个数值，表示对于**单个样本**而言模型预测的准确程度

如果模型的预测完全准确，则损失为零，否则损失会较大

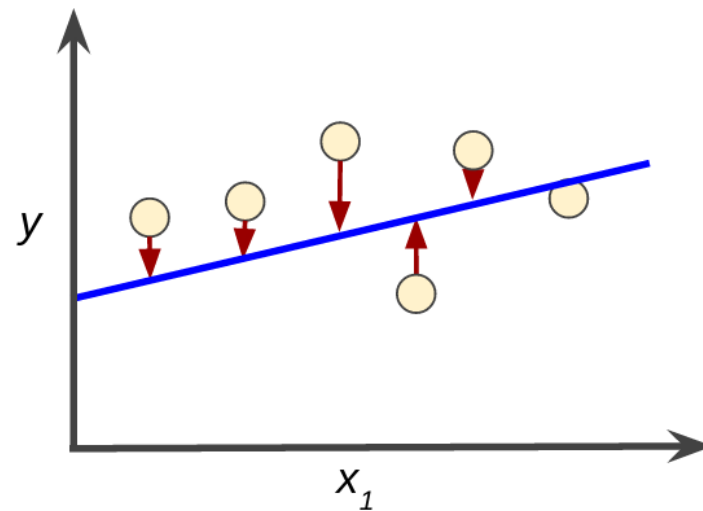
训练模型的目标是从所有样本中找到一组**平均损失“较小”**的权重和偏差



损失



左侧模型的损失较大



右侧模型的损失较小



定义损失函数

L_1 损失：基于模型预测的值与标签的实际值之差的绝对值

平方损失：一种常见的损失函数，又**称为 L_2 损失**

均方误差 (MSE) 指的是每个样本的平均平方损失

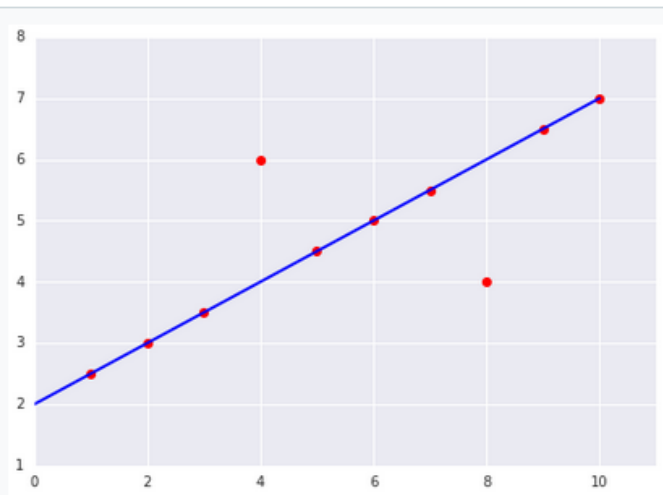
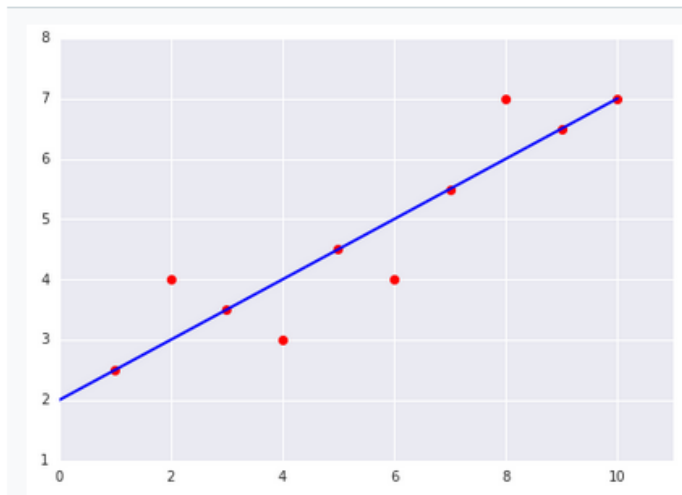
$$MSE = \frac{1}{N} \sum_{(x,y) \in D} (y - prediction(x))^2$$



问题



以下曲线图中显示的两个数据集，哪个数据集的均方误差（MSE）较高？

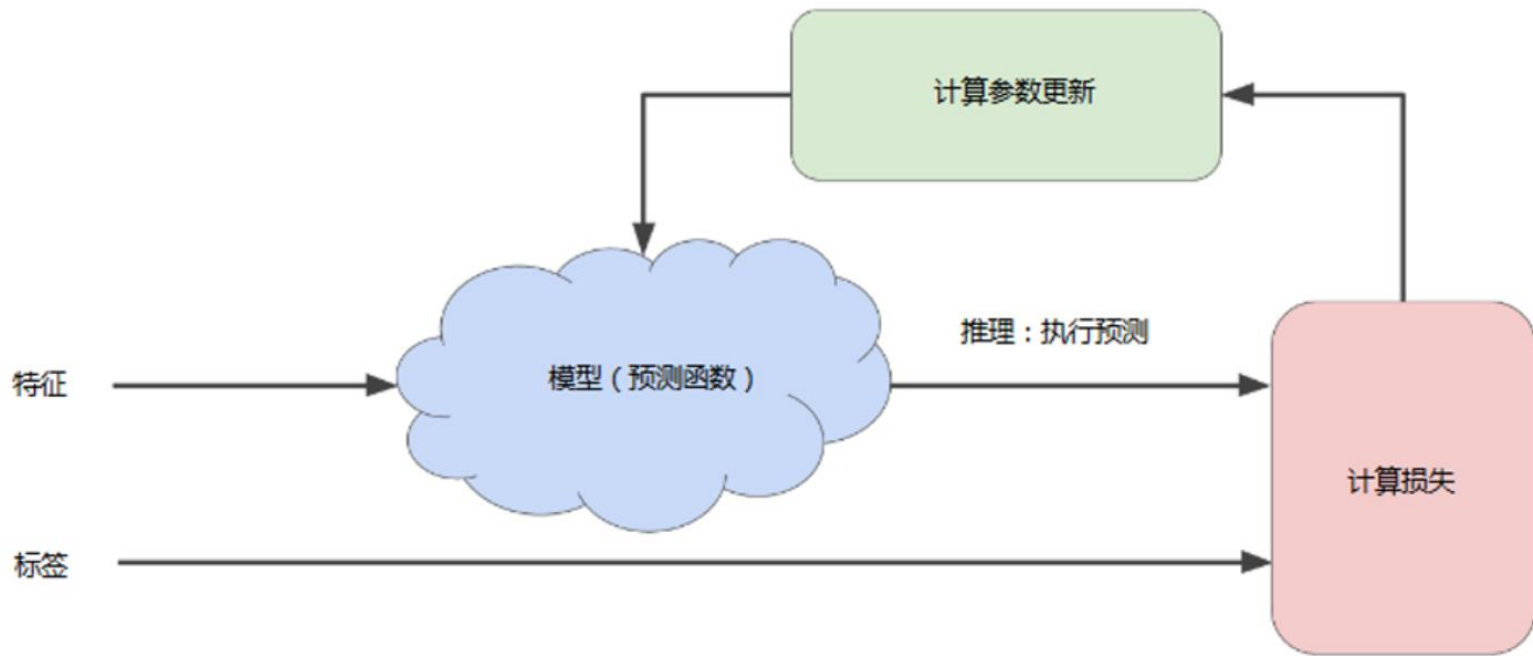




模型训练与降低损失



训练模型的迭代方法





要点

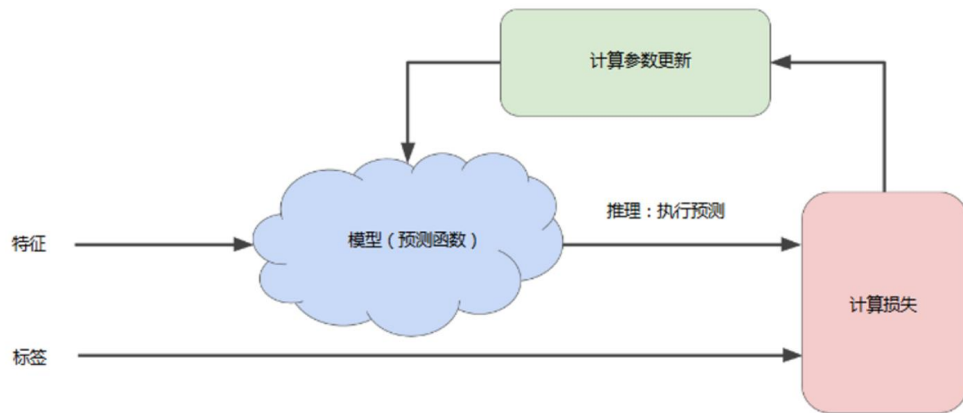
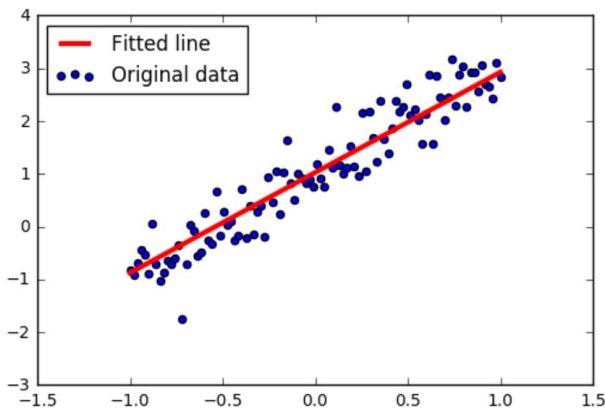


模型训练要点

首先对权重 w 和偏差 b 进行初始猜测

然后反复调整这些猜测

直到获得损失可能最低的权重和偏差为止



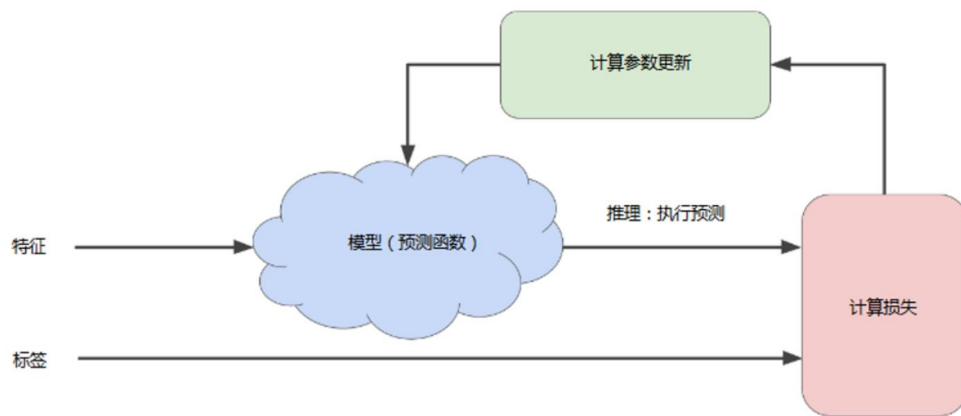
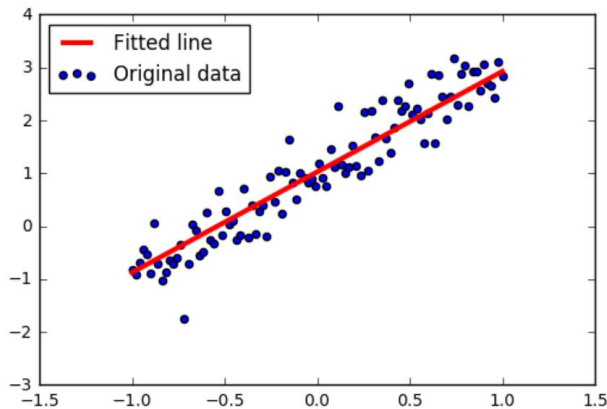


收敛

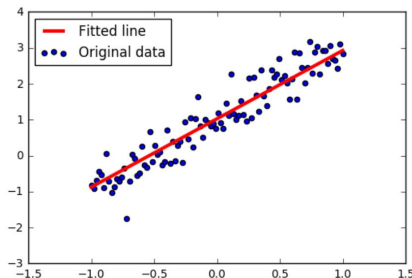


在学习优化过程中，机器学习系统将根据所有标签去重新评估所有特征，为损失函数生成一个新值，而该值又产生新的参数值。

通常，您可以不断迭代，直到总体损失不再变化或至少变化极其缓慢为止。这时候，我们可以说该模型已**收敛**

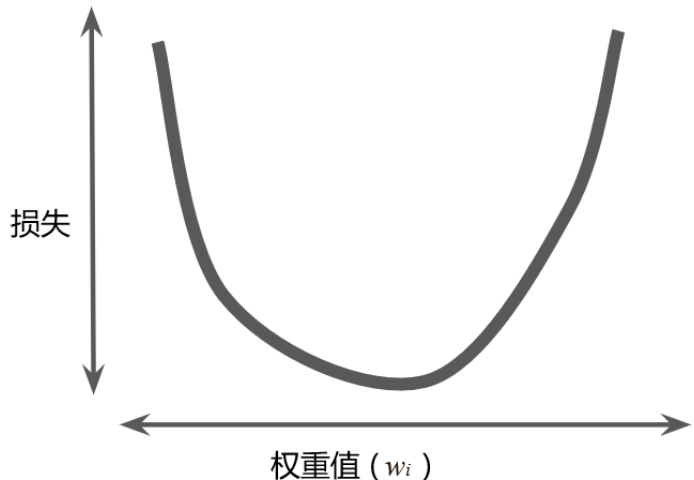


计算损失例子



$$MSE = \frac{1}{N} \sum_{(x,y) \in D} (y - prediction(x))^2$$

$$\text{pred} = w * x + b$$

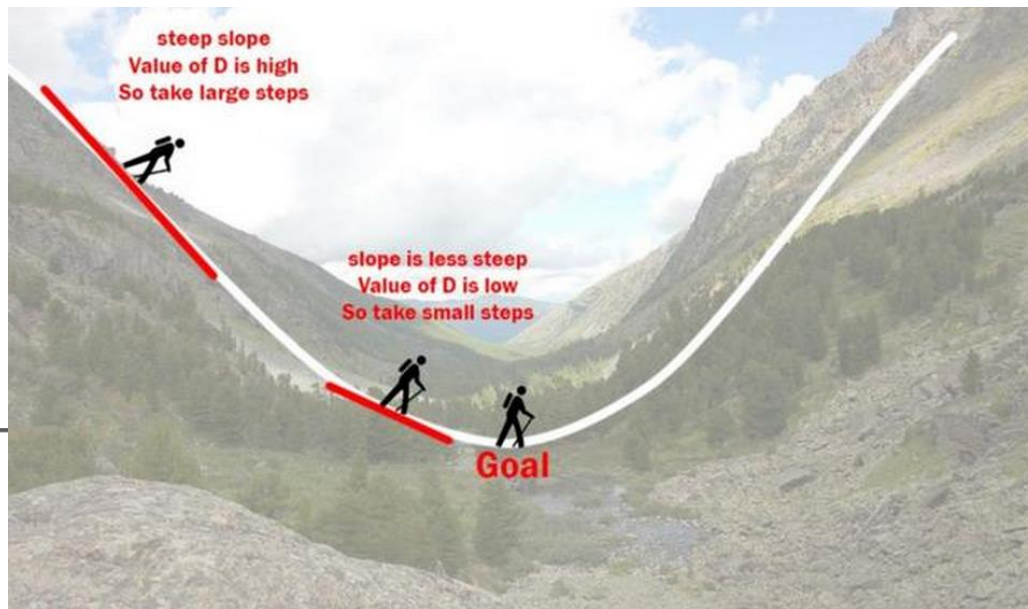
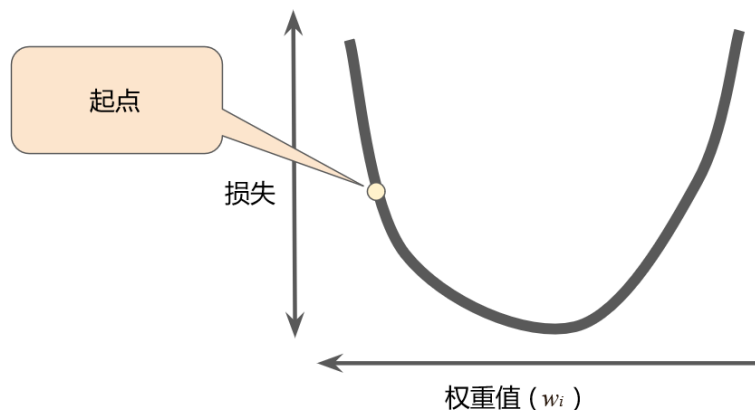


该线性回归问题产生的损失与权重图为**凸形**
凸形问题只有**一个最低点**；即只存在一个**斜率正好为 0** 的位置
这个最小值就是损失函数收敛之处



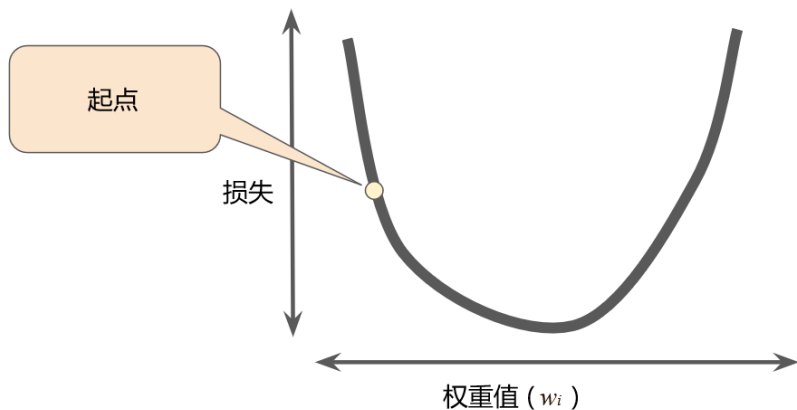
梯度下降法

梯度下降法

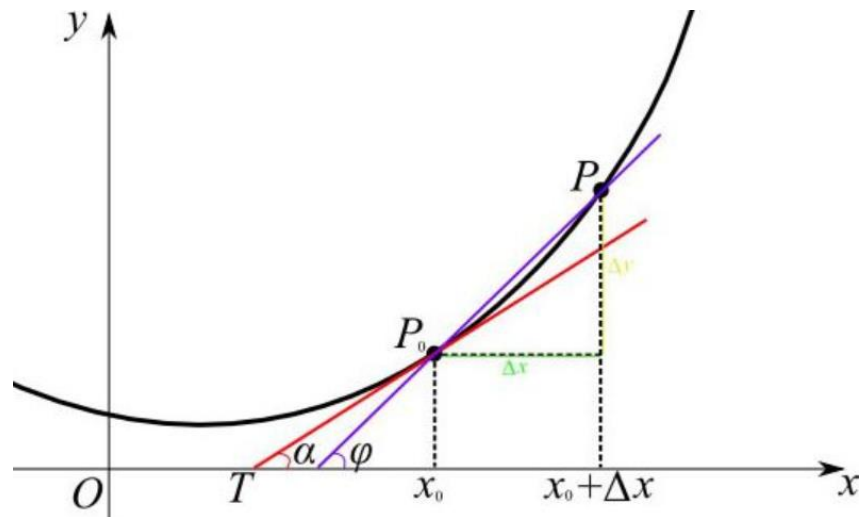


梯度：一个向量（矢量），表示某一函数在该点处的方向**导数**沿着该方向取得**最大值**，即函数在该点处沿着该方向（此梯度的方向）变化最快，变化率最大

梯度下降法



$$f'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$



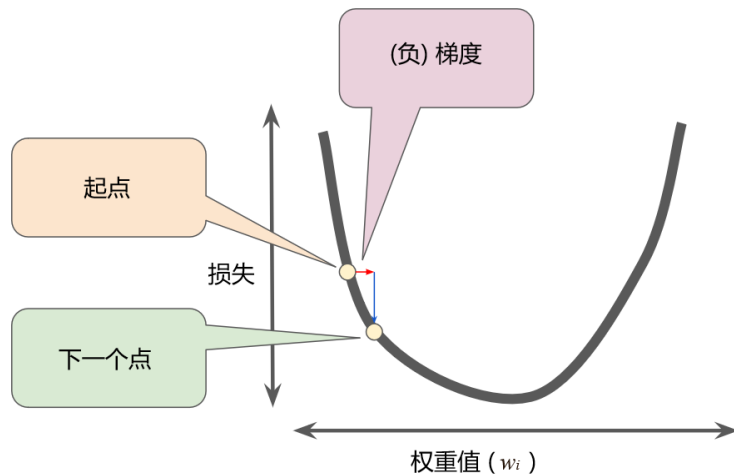
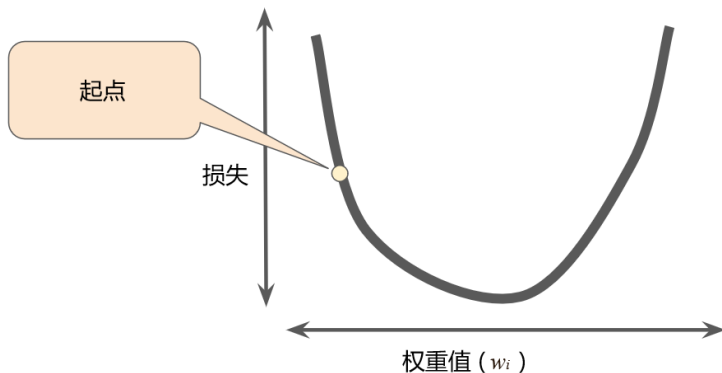
梯度：一个向量（矢量），表示某一函数在该点处的方向**导数**沿着该方向取得**最大值**，即函数在该点处沿着该方向（此梯度的方向）变化最快，变化率最大



梯度下降法



梯度是矢量: 具有**方向**和**大小**



沿着**负梯度方向**进行下一步探索



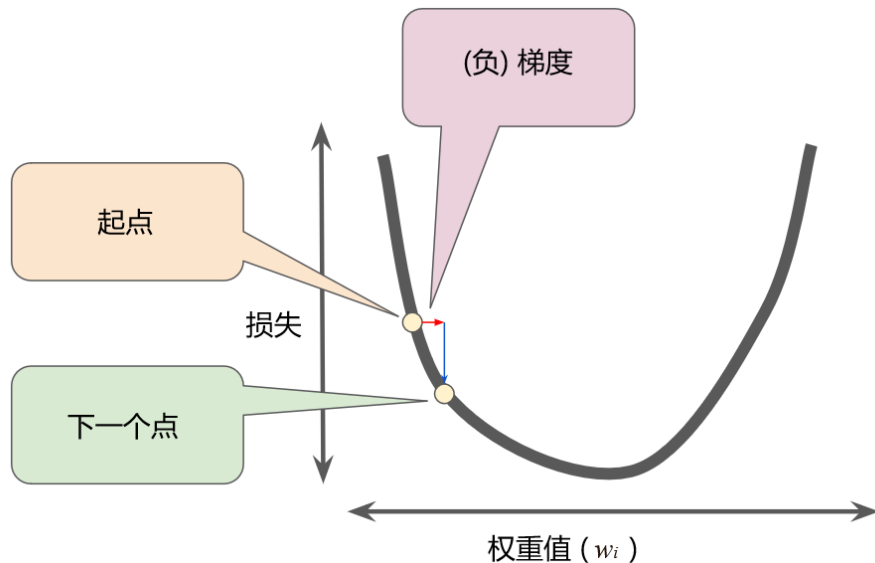
学习率



沿着负梯度方向进行下一步探索，**前进多少合适？**

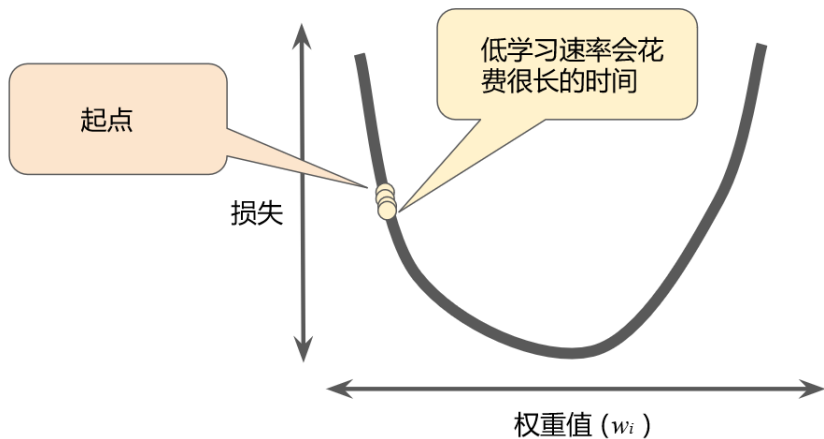
用梯度乘以一个称为**学习速率**
(有时也称为**步长**)的标量，
以确定下一个点的位置

例如：如果梯度大小为 2.5，学习速率为 0.01，则梯度下降法算法会选择距离前一个点 0.025 的位置作为下一个点

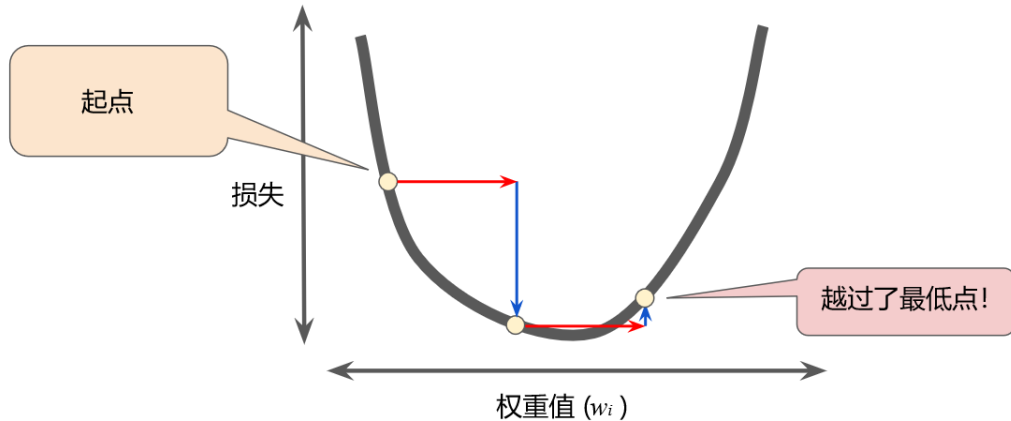




学习率



梯度学习速率过小



梯度学习速率过大



学习率

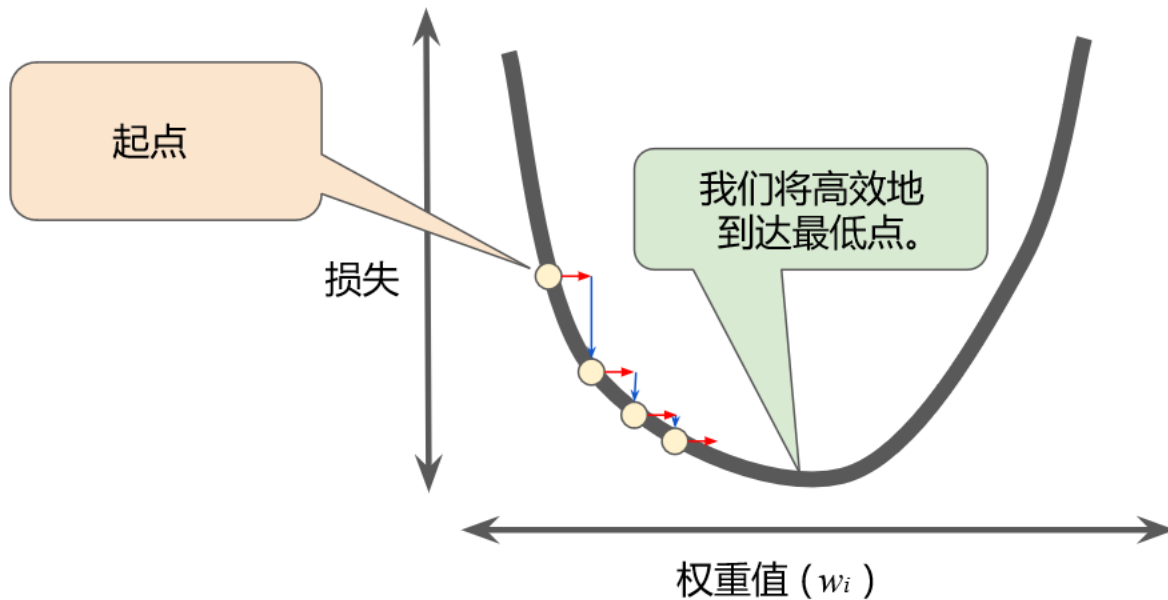


浙江大學城市學院
ZHEJIANG UNIVERSITY CITY COLLEGE





学习率





超参数



在机器学习中，**超参数**是在开始学习过程**之前**设置值的参数，而不是通过训练得到的参数数据

通常情况下，需要对超参数进行优化，选择一组好的超参数，可以提高学习的性能和效果

超参数是编程人员在机器学习算法中用于调整的旋钮

典型超参数：学习率、神经网络的隐含层数量……