

机器学习优化目标

目录

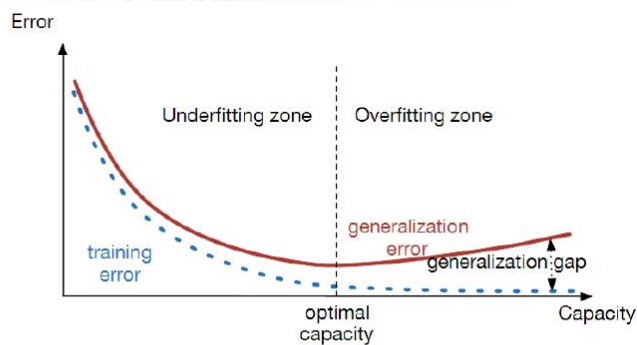
- ◆ 模型优化概述

- ◆ 常见优化目标

模型优化概述

优化目标

- ◆ 机器学习用有限训练集上的期望损失作为优化目标(即代理损失函数 **loss function**)，损失代表预测值 $f(x)$ 与真实值 Y 的不一致程度。



一般损失函数越小，模型的性能就越好，观察训练集和测试集的误差就能知道模型的收敛情况，估计模型的性能

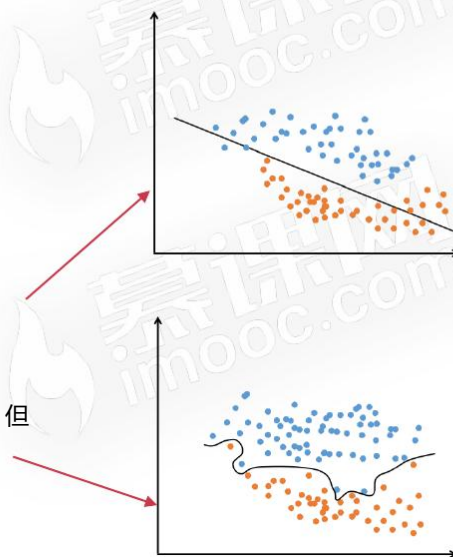
过拟合与欠拟合状态

◆ 模型在训练集和测试集上的不同表现

训练集表现	测试集表现	结果
不好	不好	欠拟合
好	不好	过拟合
好	好	适当拟合

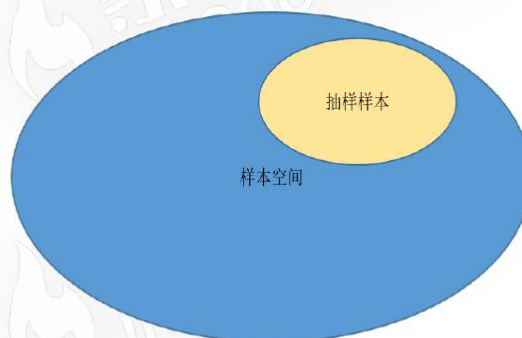
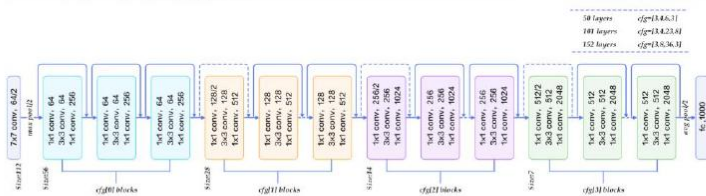
欠拟合 (under-fitting) 也称为欠学习, 指模型在训练集上精度差

过拟合 (over-fitting) 也称为过学习, 指模型在训练集上精度高, 但在测试集上精度低, 泛化性能差



过拟合的原因

◆ 模型过大, 数据太少



模型本身过于复杂, 拟合了训练样本集中的噪声

训练样本太少或者缺乏代表性

两类常见的优化目标

◆ 分类任务优化目标与回归任务优化目标

分类任务



0-1损失
交叉熵损失
KL散度
Softmax损失
.....

回归任务



L1距离
L2距离
Smooth L1距离
.....

常见优化目标

0-1损失

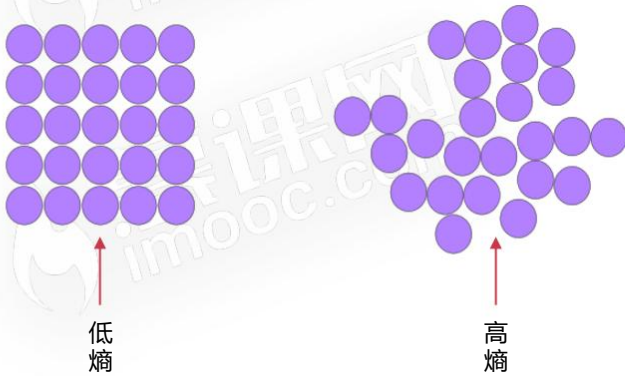
- ◆ 只看分类的对与错，当标签与预测类别相等时，损失为0，否则为1

(())

- ◆ 真实的优化目标，但是无法求导和优化，只有理论意义

熵与交叉熵(cross entropy)

- ◆ 熵表示热力学系统的无序程度，在信息学中用于表示信息多少，不确定性越大，概率越低，则信息越多，熵越高



熵函数是概率的单调递减的函数

交叉熵损失(cross entropy loss)

◆ 衡量两个概率分布的相似性

令 y_i 表示第*i*个样本属于分类*j*的标签, 分类的概率

\hat{y}_i 表示的是样本*i*预测为*j*

()

Softmax loss

◆ 交叉熵损失(cross entropy loss)与softmax loss

()

当 () ——

() 是

, 为交叉熵的特例!

L1 Loss

- ◆ L1损失即Mean absolute loss(MAE loss)，以绝对误差作为距离。

—

主要问题：梯度在零点不平滑

L1/L2-loss

- ◆ L2损失即Mean Squared Loss(MSE loss)，也被称为欧氏距离，以误差的平方和作为距离。

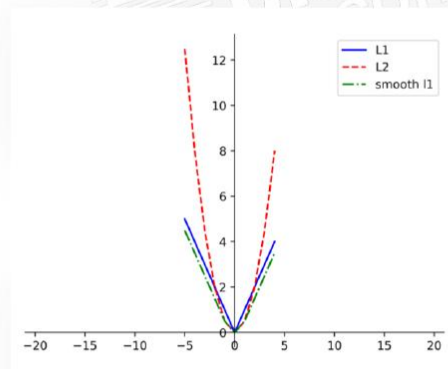
—

当预测值与目标值相差很大(异常值)时, 梯度容易爆炸, 因为梯度里有两者差值。

Smooth L1 loss

- ◆ 解决L1 loss梯度不平滑，L2 loss梯度爆炸的问题

()



在x比较小时，等价于L2 loss，保持平滑。
在x比较大时，等价于L1 loss，可以限制数值的大小。

下次预告：机器学习案例实战