

第八课 机器学习简单介绍

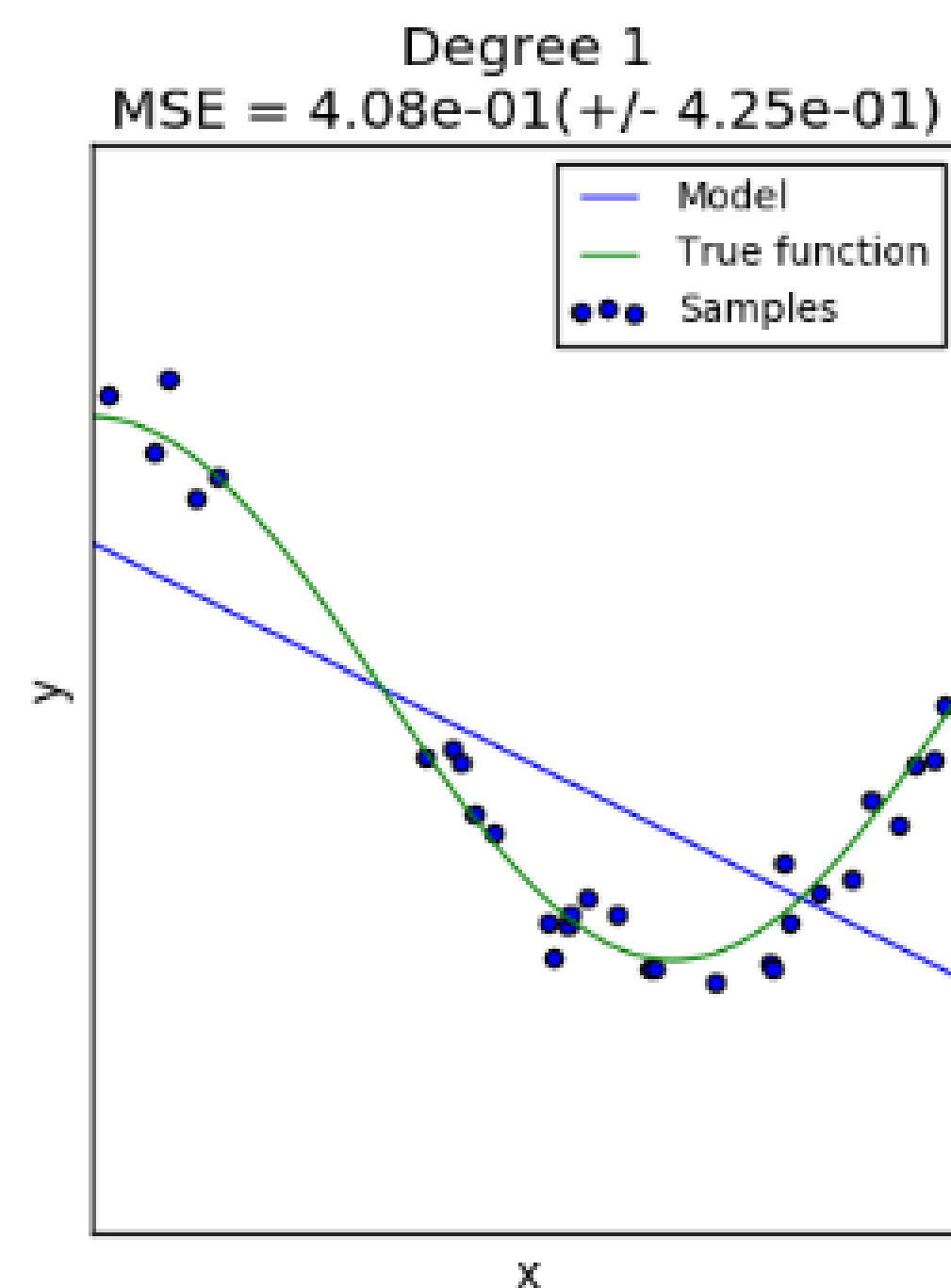
第11节 过拟合与欠拟合

主 讲: Robin

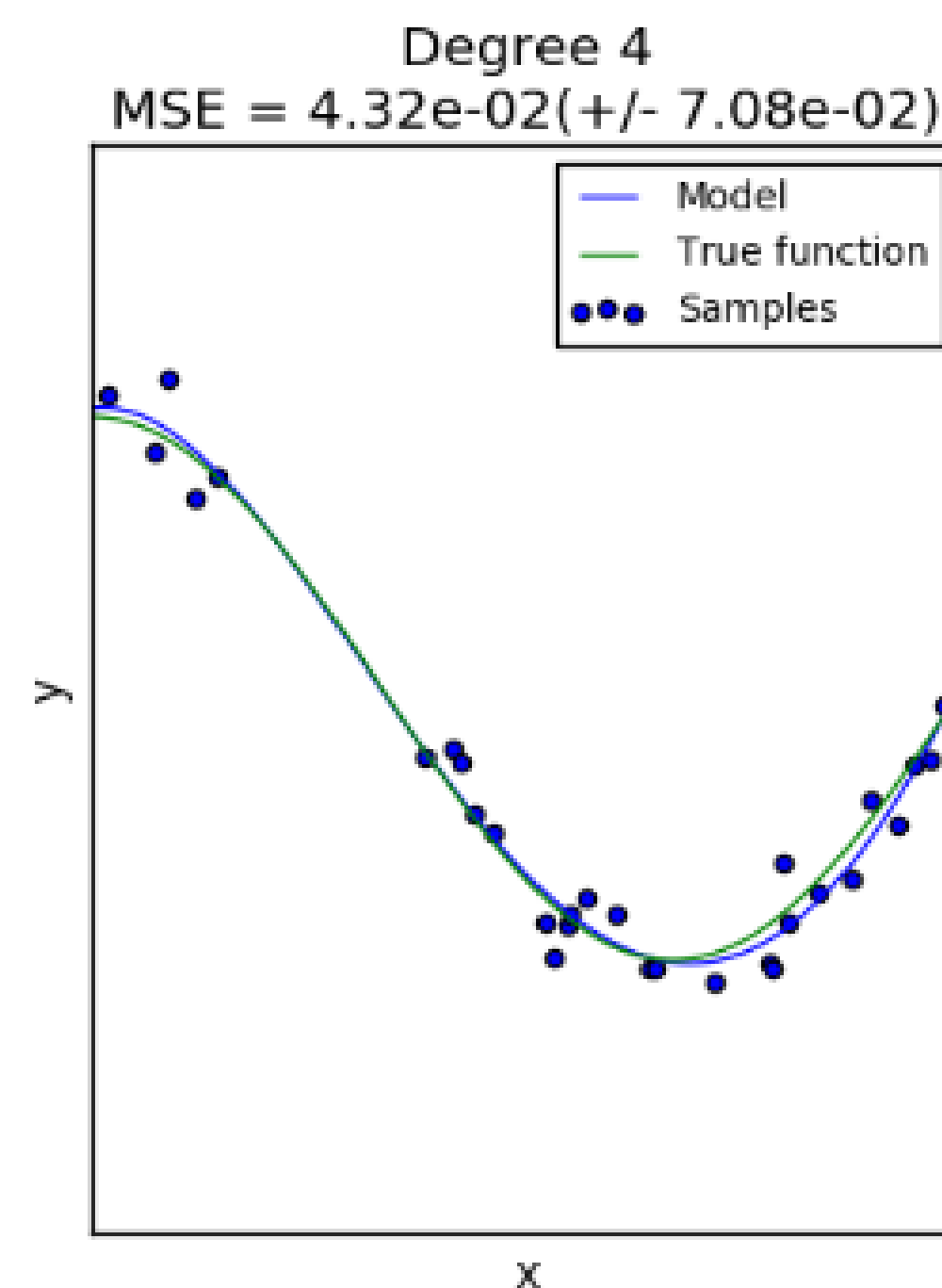
过拟合与欠拟合

- 过拟合
 - 模型对于训练数据拟合程度过当，以致太适应训练数据而非一般情况。
 - 在训练数据上表现非常好，但是在测试数据或验证数据上表现很差。

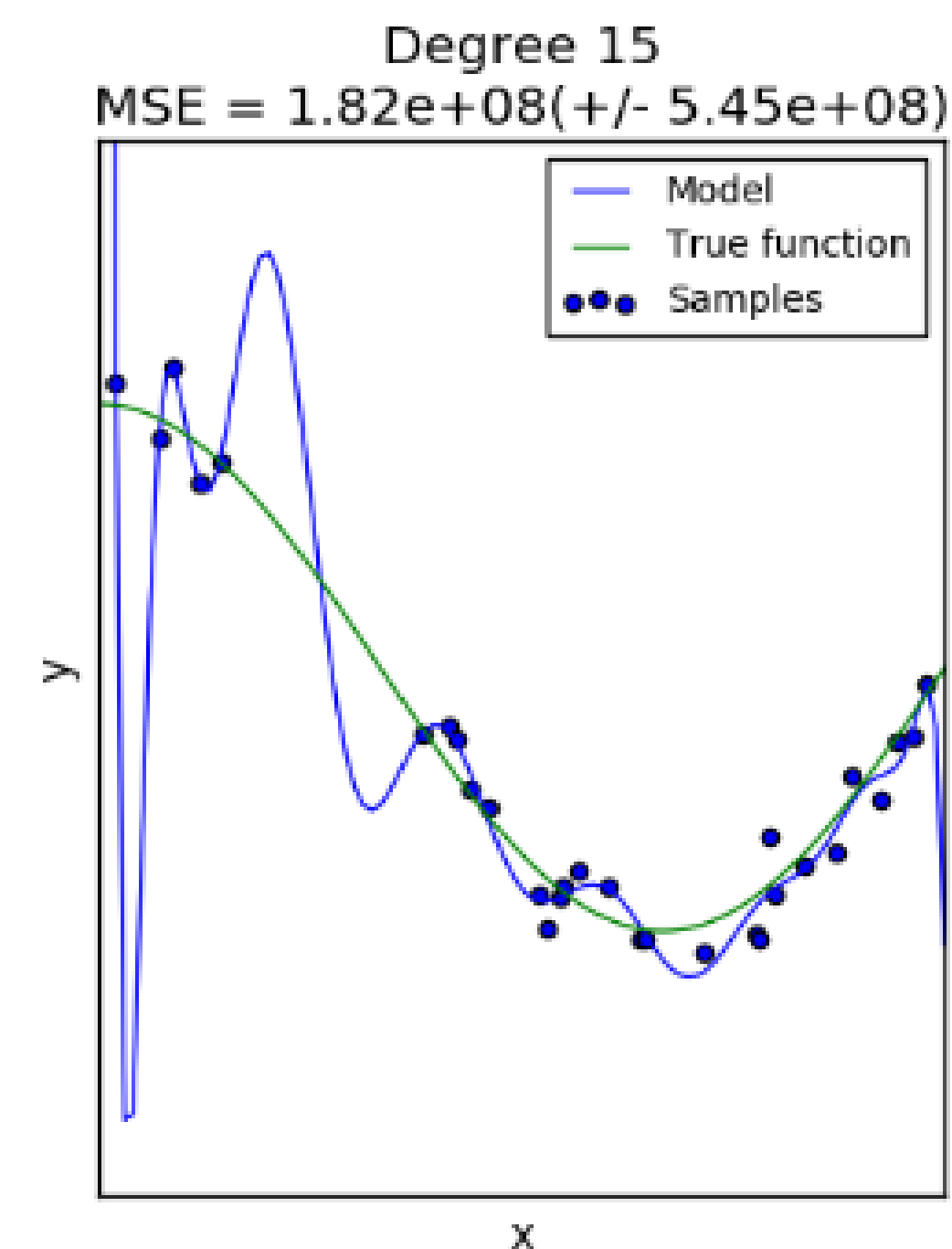
- 欠拟合则相反



欠拟合



“刚刚好”



过拟合

过拟合与欠拟合

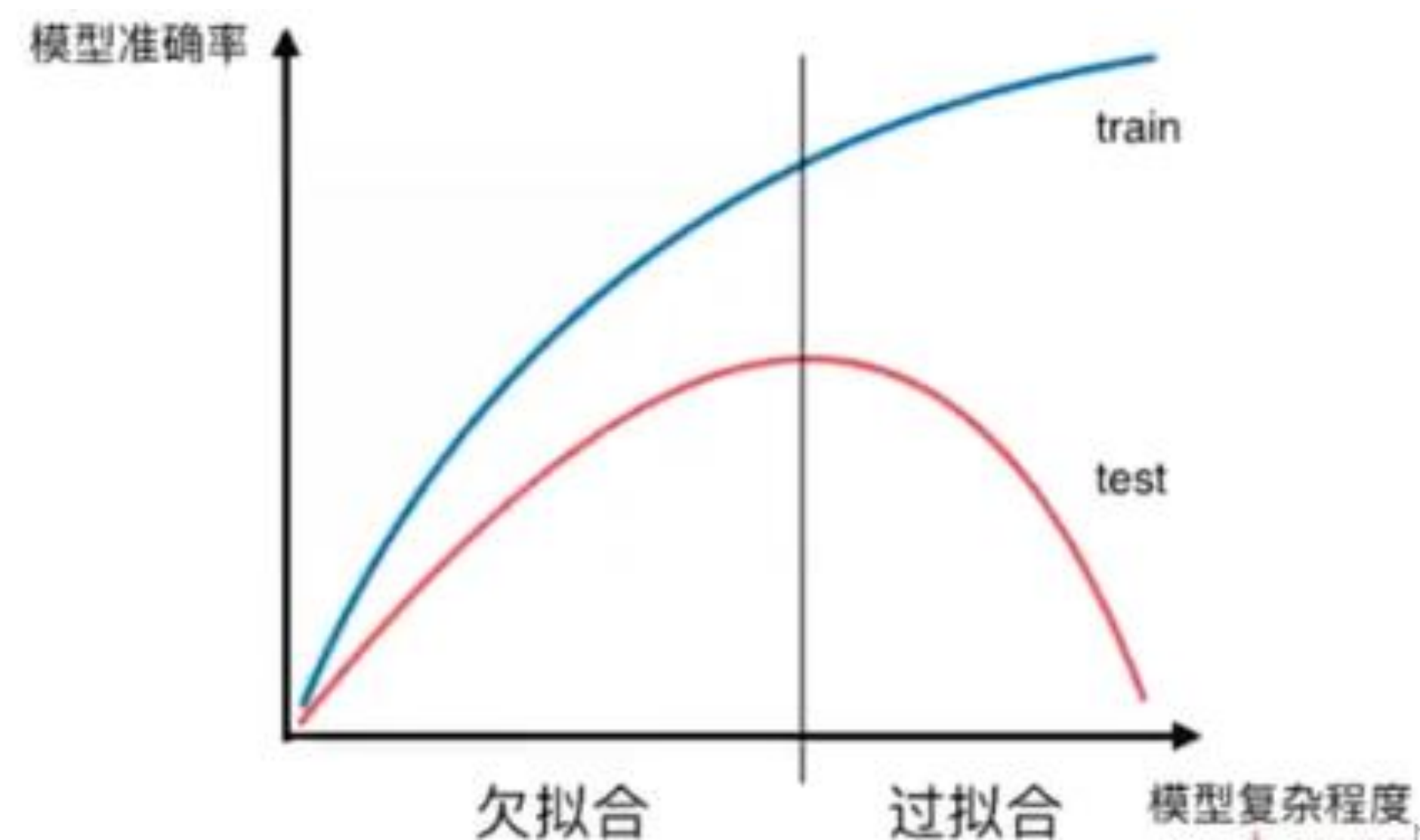
- 造成过拟合的原因：

1. 训练数据过少
2. 模型过于复杂
3. ...



- 解决方法：

1. 收集更多的数据
2. 降低模型复杂度 (正则化)
3. ...



过拟合与欠拟合

正则化

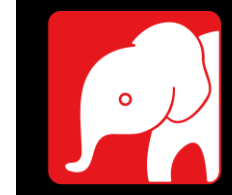
- 控制模型复杂度，模型复杂度越高，越容易过拟合
- 为**目标函数**添加**正则项**（描述**模型复杂度**）

$$\text{minimize: } \boxed{Loss(Data | Model)} + C * \boxed{complexity(Model)}$$

- 使用**参数C**平衡目标函数与模型复杂度
 - C是手工指定的，而非模型学习的

思考题：

- 如何为模型选择合适的参数？
 - 如kNN中的邻居个数k，线型SVM中的正则项系数C



只为遇见明天更优秀的你！