

---

# 欠拟合、过拟合

---

# 欠拟合、过拟合

## 【实践】实践

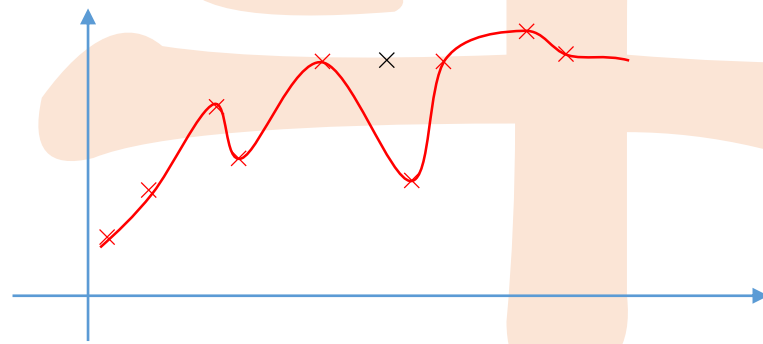
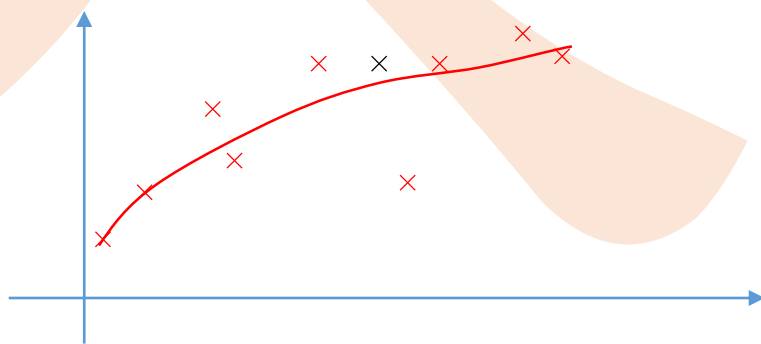
## 过拟合

- 过拟合 (Overfitting)

- 一个模型在训练集上表现很好，但实际预测中效果却大打折扣
- 大部分算法模型都存在该问题

- 产生原因:

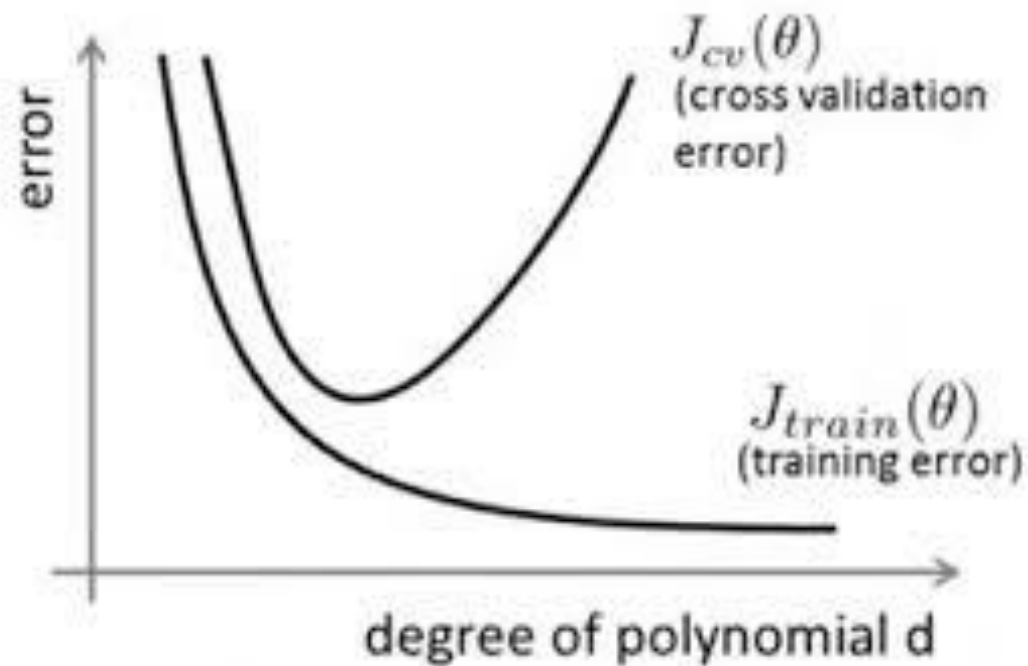
- 过度训练：将噪声一起学进了模型



- 训练数据相比特征数太少：学出来的特征虽然训练集由好的表现，但实际用处不大

## 过拟合

- 过拟合现象：



## 过拟合

- 避免方法：
  - 交叉验证 (cross validation) : 将训练数据分为2份, 一份用于训练, 一份用于测试
    - 模型通过训练的数据生成, 在测试上验证效果
  - 若特征太多, 进行特征筛选
    - 人工筛选
    - 机器筛选, 例如mRMR
  - 模型特有方法: 增加惩罚 (L1, L2) , 例如决策树控制树的高度等
  - 增加训练数据
  - 神经网络Dropout

### 欠拟合

- 模型没有很好捕捉到数据特征，不能很好拟合数据
- 学习能力低下造成
- 解决方法：
  - 决策树扩展分支
  - 增加训练轮数

# 欠拟合、过拟合

## 【实践】实践

---

# Q & A

@八斗数据

---