



# 测试驱动的机器学习

## 机器学习中的常见问题

- 欠拟合
- 过拟合
- 数据不稳定
- 不可预测性

## 测试驱动的解决办法

- 交叉验证
- 运行速度测试
- testing seams
- 指标追踪
- 验证拟合效果
- 奥卡姆剃刀原则，越简单越稳定
- 输入输出检查，测试数据稳定性
- 不断监测模型性能

## 机器学习与 SOLID 原则

- SRP — 耦合依赖
  - CACE (纠缠)
  - 胶水代码
- OCP—隐藏反馈
  - 预测-实施-证实预测
- LSP—关键特征
  - 用尽可能少的特征和数据，取得稳定结果
- ISP—数据变动
  - 数据源人为变化
  - 数据本身变化
- DIP—遗留代码与数据
  - 测试代码与中间数据累计

## 机器学习中的技术债务

- 边界模糊，数据彼此依赖关联
- 没有系统级代码分离，胶水代码处理一切
- 随外部世界的改变而彻底改变

## TDD 测试驱动开发

- 写测试来记录想要实现的功能
- 一开始测试会失败
- 编写代码修复测试
- 测试通过，按 SOLID 原则修复代码

## SOLID原则

- SRP 单一职责原则
  - 一片代码只做一件事
- OCP 开闭原则
  - (对象)对扩展开放，对修改关闭
- LSP Liskov 替换原则
  - 任何子类型应该轻松被同一对象树下其他对象替代
- ISP 接口隔离原则
  - 多个客户多个接口优于所有客户一个接口
- DIP 依赖倒置原则
  - 依赖抽象而非实体或实现，无论高层、底层或是细节

## 技术债务

- 情景
  - 加快软件开发速度
- 措施
  - 折中妥协，选择易于实现的方案
- 结果
  - 短期内加速软件开发
  - 未来带来额外开发负担，发布逐渐停滞

重构

