

MACHINE LEARNING LECTURE NOTES

yujie6@sjtu.edu.cn¹

2020 年 3 月 3 日

¹Yujie Lu, ACM class 18, ID is 518030910111

目录

第一章 Lecture 1 (Mar 2)	3
1.1 人工智能简介	3
1.1.1 什么是人工智能	3
1.1.2 人工智能方法	3
1.2 机器学习简介	3
1.2.1 机器学习定义	3
1.2.2 机器学习应用	4
1.2.3 机器学习基本思想	4

第一章 Lecture 1 (Mar 2)

1.1 人工智能简介

1.1.1 什么是人工智能

智能是实现世界目标的计算能力部分。

人工智能探讨对机器进行设计的方法论使得其可以去完成基于智能的任务。

1.1.2 人工智能方法

基于规则的方法

- 直接编程实现
- 借鉴人类启发式学习

基于数据的方法

- 专家系统：基于数据创造决策的规则
- 机器学习：基于数据进行预测或决策

1.2 机器学习简介

1.2.1 机器学习定义

学习是系统通过经验提升性能的过程。

Definition 1.2.1 (Tom Mitchell) 机器学习是一门研究学习算法的学科，这些算法 (非显式编程) 能在某些任务 T 上通过经验 E 来提升性能 P 。

机器学习分类

- 预测
 - 监督学习

- 无监督学习
- 决策
 - 在动态环境中采取行动（强化学习）

1.2.2 机器学习应用

预测

- 网页搜索（根据 profile 分析）
- 人脸识别（Computer vision）
- 推荐系统
- 在线广告
- 信息提取

决策

- 交互式内容推荐
- 机器人控制
- 自动驾驶
- 游戏智能
- 多智能体协作

1.2.3 机器学习基本思想

以监督学习为例，给定带 label 的数据集

$$D = \{(x_i, y_i)\}_{i=1,2,\dots,N}.$$

寻找 $\theta = (a, b, c, \dots)$ 使得函数映射

$$y_i \approx f_{\theta}(x_i).$$

我们用 loss function

$$\mathcal{L}(y_i, f_{\theta}(x_i)) = \frac{1}{2} (y_i - f_{\theta}(x_i))^2$$

来衡量预测的误差

$$\mathcal{L}(\theta) = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y_i, f_{\theta}(x_i)).$$

通过梯度下降求得 $\min_{\theta} \mathcal{L}(\theta)$

模型选择

模型选得不好会导致欠拟合或过拟合。为了防止这两种情况，我们可以采用

正则化

添加 θ 的惩罚项 $\Omega(\theta)$ ，一般选择 L2 正则化 $\lambda \|\theta\|_2^2$ （也称岭回归 Ridge）。当我们增加参数 θ_n 时，我们实际上是在学习前面的参数产生的残差

这样做的另一个解释是奥卡姆剃刀原则（Occam's Razor），这个原则是说能用简单的方法完成任务的就尽量不要复杂，在这里就是能用简单的模型去拟合就不用复杂的能把噪声都刻画出来的方法。

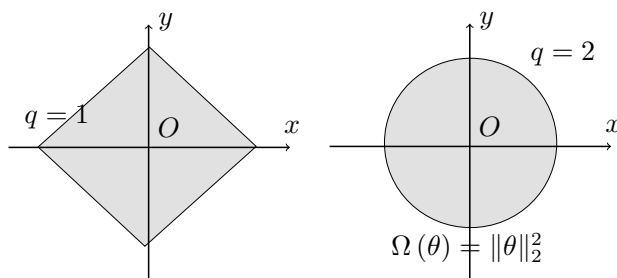


图 1.1: Ridge

有时候也会选择 $q \leq 1$ 进行稀疏性学习。

一个机器学习的解决方案的模型包含参数 θ 与超参数 λ 。

Definition 1.2.2 超参数为需要预先定义，无法直接从数据学习的参数。

交叉验证

K-fold, 将训练集分成 k 份。

泛化能力

Definition 1.2.3 泛化能力 (*generalization ability*) 指对未训练数据的预测能力。

为了描述这种能力我们引入泛化误差 (Generalization Error):

$$R(f) = \int_{X \times Y} \mathcal{L}(y, f(x)) p(x, y) \, dx \, dy.$$

其中 $p(x, y)$ 是潜在的联合数据分布 (joint probability distribution)。在已有数据集上也可进行经验估

计:

$$\hat{R}(f) = \frac{1}{N} \sum_{i=1}^n \mathcal{L}_i.$$