Section 1 Introduction

资料介绍

频率派 ⟶ 统计机器学习

贝叶斯派 ⟶ 概率图模型.

书籍:

① 统计学习方法(李航)

偏向于频率派.

内容:感知机、K-近邻、朴素贝叶斯、决策树、逻辑回归

SVM、boosting、EM、隐马尔可夫、条件随机场

② 机器学习(周志华)

都包括,但都不深入

③ PRML 模式识别与机器学习

贝叶斯派

内容:线性回归、分类、神经网络、核方法、稀疏核机

概率图模型、混合模型、近似算法、采样、连续型随机变量

顺序数据(序列) 组合模型.

③ MLAPP：Machine Learning a Probability Perspective
类似于百科全书
概率图为主

④ ESL 统计学习基本元素
频率派为主

⑤ Deep Learning
中译版，张志华
主要讲深度学习。

视频：
① 台大：林轩田老师
1) 机器学习基石：VC Theory、正则化、线性模型
2) 〃 〃 技法：SVM、决策树、随机森林、神经网络、deep learning

② 张志华老师：（公式推导）
1) 机器学习导论：以频率角度
2) 统计机器学习：以贝叶斯角度

③ Ng: CS229 吴恩达

④ 徐亦达 → 概率模型
github → notes

⑤ 台大: 李宏毅老师
深度学习相关

频率派与贝叶斯派.

X: data → $X \in R^{N\times D}$ 其中 N 为样本数, D 为特征维数

θ: parameter.

$x \sim p(x|\theta)$

频率派: θ 是未知的常量, 是一个随机变量 (random variable)

优化问题 用最大似然估计 (MLE) 求解.

① 建立模型 即 $\theta_{MLE} = \underset{\theta}{\operatorname{argmax}} \underbrace{\log P(X|\theta)}_{}$ $\mathcal{L}(\theta)$.

② loss function

③ algorithm 为什么要加 log? $P(X|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$ 其中 $x_i \overset{iid}{\sim} p(x|\theta)$

求解. (梯度下降 牛顿法) 对于连乘求解较为困难, 加上 log 后. 连乘变为连加. 易于求解

$\log P(X|\theta) = \sum_{i=1}^{N} \log p(x_i|\theta)$.

贝叶斯派：θ是随机变量，但是服从概率分布 $\theta \sim p(\theta)$、先验

积分问题。(MCMC等)

根据贝叶斯公式

$$\underbrace{p(\theta|X)}_{\text{后验}} = \frac{\underbrace{p(X|\theta)}_{\text{似然}}\,\underbrace{p(\theta)}_{\text{先验}}}{\underbrace{p(X)}_{\text{数据分布}}}$$

数据分布 → $\int_\theta p(X|\theta)p(\theta)\,d\theta$.

用最大后验估计(MAP)求解。

即 $\theta_{MAP} = \arg\max_\theta p(\theta|X) = \arg\max_\theta p(X|\theta)p(\theta)$

因为 $p(X)$ 可以看作是一个常量，和θ无关

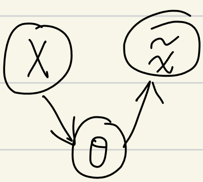
贝叶斯估计：$p(\theta|X) = \dfrac{p(X|\theta)\,p(\theta)}{\int p(X|\theta)p(\theta)\,d\theta}$ 很难算.

⇓

贝叶斯预测：已知 X.

求 $\tilde{x}$ 的概率.

即求 $p(\tilde{x}|X)$



$$p(\tilde{x}|X) = \int_\theta p(\tilde{x},\theta|X)\,d\theta$$
$$= \int_\theta p(\tilde{x}|\theta)\,p(\theta|X)\,d\theta.$$