第十二章

统计学习方法总结

袁春 清华大学深圳研究生院
李航 华为诺亚方舟实验室

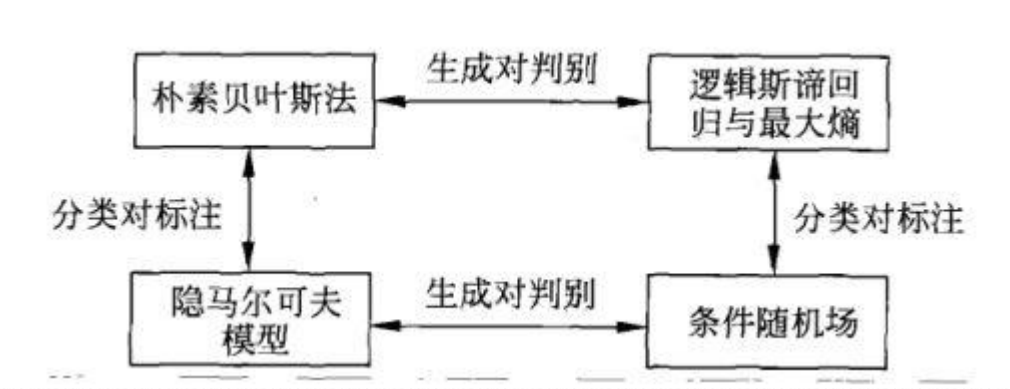
统计学习方法总结

- 感知机
- K近邻法
- 朴素贝叶斯
- 决策树
- 逻辑斯蒂回归与最大熵模型
- 支持向量机
- 提升方法
- EM算法
- 隐马尔科夫模型
- 条件随机场

表 12.1　10 种统计学习方法特点的概括总结

方法	适用问题	模型特点	模型类型	学习策略	学习的损失函数	学习算法
感知机	二类分类	分离超平面	判别模型	极小化误分点到超平面距离	误分点到超平面距离	随机梯度下降
k 近邻法	多类分类，回归	特征空间，样本点	判别模型			
朴素贝叶斯法	多类分类	特征与类别的联合概率分布，条件独立假设	生成模型	极大似然估计，极大后验概率估计	对数似然损失	概率计算公式，EM 算法
决策树	多类分类，回归	分类树，回归树	判别模型	正则化的极大似然估计	对数似然损失	特征选择，生成，剪枝
逻辑斯谛回归与最大熵模型	多类分类	特征条件下类别的条件概率分布，对数线形模型	判别模型	极大似然估计，正则化的极大似然估计	逻辑斯谛损失	改进的迭代尺度算法，梯度下降，拟牛顿法
支持向量机	二类分类	分离超平面，核技巧	判别模型	极小化正则化合页损失，软间隔最大化	合页损失	序列最小最优化算法（SMO）
提升方法	二类分类	弱分类器的线性组合	判别模型	极小化加法模型的指数损失	指数损失	前向分步加法算法
EM 算法[2]	概率模型参数估计	含隐变量概率模型		极大似然估计，极大后验概率估计	对数似然损失	迭代算法
隐马尔可夫模型	标注	观测序列与状态序列的联合概率分布模型	生成模型	极大似然估计，极大后验概率估计	对数似然损失	概率计算公式，EM 算法
条件随机场	标注	状态序列条件下观测序列的条件概率分布，对数线性模型	判别模型	极大似然估计，正则化极大似然估计	对数似然损失	改进的迭代尺度算法，梯度下降，拟牛顿法

- 从生成与判别、分类与标注两个方面描述了几个统计学习方法之间的关系



学习策略

在二类分类的监督学习中，支持向量机、逻辑斯谛回归与最大熵模型、提升方法各自使用合页损失函数、逻辑斯谛损失函数、指数损失函数．3 种损失函数分别写为

$$[1-yf(x)]_+ \tag{12.1}$$

$$\log[1+\exp(-yf(x))] \tag{12.2}$$

$$\exp(-yf(x)) \tag{12.3}$$

这 3 种损失函数都是 0-1 损失函数的上界，具有相似的形状，如图 12.2 所示．所以，可以认为支持向量机、逻辑斯谛回归与最大熵模型、提升方法使用不同的代理损失函数（surrogate loss function）表示分类的损失，定义经验风险或结构风险函数，实现二类分类学习任务．学习的策略是优化以下结构风险函数：

$$\min_{f\in H}\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))+\lambda J(f) \tag{12.4}$$

这里，第 1 项为经验风险（经验损失），第 2 项为正则化项，$L(y,f(x))$ 为损失函数，$J(f)$ 为模型的复杂度，$\lambda\geqslant 0$ 为系数．

支持向量机用 L_2 范数表示模型的复杂度．原始的逻辑斯谛回归与最大熵模型没有正则化项，可以给它们加上 L_2 范数正则化项．提升方法没有显式的正则化项，通常通过早停止（early stopping）的方法达到正则化的效果．

学习策略

概率模型的学习可以形式化为极大似然估计或贝叶斯估计的极大后验概率估计．这时，学习的策略是极小化对数似然损失或极小化正则化的对数似然损失．对数似然损失可以写成

$$-\log P(y|x)$$

极大后验概率估计时，正则化项是先验概率的负对数．

决策树学习的策略是正则化的极大似然估计，损失函数是对数似然损失，正则化项是决策树的复杂度．

逻辑斯谛回归与最大熵模型、条件随机场的学习策略既可以看成是极大似然估计（或正则化的极大似然估计），又可以看成是极小化逻辑斯谛损失（或正则化的逻辑斯谛损失）．

朴素贝叶斯模型、隐马尔可夫模型的非监督学习也是极大似然估计或极大后验概率估计，但这时模型含有隐变量．

- END
- Q&R