

李航《统计学习方法》笔记 by kageyamaa

Chapter1 统计学习三要素

方法=模型+策略+算法

1. 模型

模型就是所要学习的条件概率分布或决策函数。模型的假设空间包含所有可能的条件概率分布或决策函数。有决策函数表示的模型称为~~非~~概率模型，由条件概率表示的模型为概率模型。

2. 策略

损失函数 度量模型一次预测的好坏

风险函数 度量平均意义下模型预测的好坏

输出的预测值 $f(X)$ 与真实值 Y 之间存在差异，可以采用损失函数(loss function)或者代价函数(cost function)来度量这种错误或者称差异的程度。损失函数是预测值 $f(X)$ 和 Y 的非负实值函数记作 $L(Y, f(X))$

**统计学习中常用的损失函数有:

(1) 0-1损失函数

$$L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$$

(2) 平方损失函数

$$L(Y, f(X)) = (Y - f(X))^2$$

(3) 绝对损失函数

$$L(Y, f(X)) = |Y - f(X)|$$

(4) 对数损失函数或对数似然损失函数(最重要!)

$$L(Y, P(Y|X)) = -\log P(Y|X)$$

tips :之所以会出现 $P(Y|X)$ 和 $f(X)$ 的差异是因为二者所使用的模型不同，前者使用的是概率模型，而后者使用的是非概率模型

损失函数的意义：损失函数值越小，说明模型越好，由于模型的输入、输出是随机变量，遵循联合分布 $P(X, Y)$ ，所以损失函数的期望是

$$R_{exp}(f) = E_P[L(Y, f(X))] = \iint_{xy} L(Y, f(X)) P(x, y) dx dy$$

这表示模型 $f(X)$ 关于联合分布 $P(Y|X)$ 的平均意义下的损失,称为风险函数或期望损失

⇒接下来引入经验风险的概念：

因为我们不知道联合分布函数 $P(Y, X)$ 不能直接求出 $R_{emp}(f)$ 因此可以引入经验风险的概念。

对给定的训练数据集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

模型 $f(X)$ 关于训练数据集的平均损失成为经验风险或者经验损失，记作 R_{emp} ：

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i))$$

根据大数定律,当样本容量很大的时候,经验风险 R_{emp} 趋近于期望风险 R_{exp} 所以自然而然地考虑用经验风险去预估期望风险。但是由于现实的训练样本往往存在问题,因此需要对经验风险进行矫正,这就关系到两个基本策略:经验风险最小化和结构风险最小化。

已知经验风险函数是确定的 $\frac{1}{N} \sum_{i=1}^N (L(y_i, f(x_i)))$ 。经验风险最小化的策略认为,经验风险最小的模型就是最优的模型,根据这一策略,按照经验风险最小化求最优模型就是求解最优化问题:

$$\min \frac{1}{N} \sum_{i=1}^N (L(y_i, f(x_i)))$$

样本容量足够大的时候,经验风险最小化的策略被广泛使用。*e.g*:极大似然估计,当模型为条件概率分布,而损失函数为对数损失函数 $(L(Y, P(Y|X)) = -\log P(Y|X))$ 时,经验风险最小化就等价于极大似然估计。

而当样本容量较小时,经验风险最小化会导致过拟合现象的发生,为了解决过拟合,就出现了结构风险最小化的策略,结构风险最小化等价于正则化。结构风险是指在经验风险上加上表示模型复杂度的正则化项或者罚项。结构风险的定义是:

$$R_{SRM}(f) = \frac{1}{N} \sum_{i=1}^N (L(y_i, f(x_i))) + \lambda J(f)$$

其中 $J(f)$ 为模型的复杂度,模型越复杂,则 $J(f)$ 越大,反之模型越简单,复杂度 $J(f)$ 就越小, $\lambda > 0$ 。结构风险需要经验风险和模型复杂度同时很小,也就是要寻找足够简单同时经验风险最小的模型。

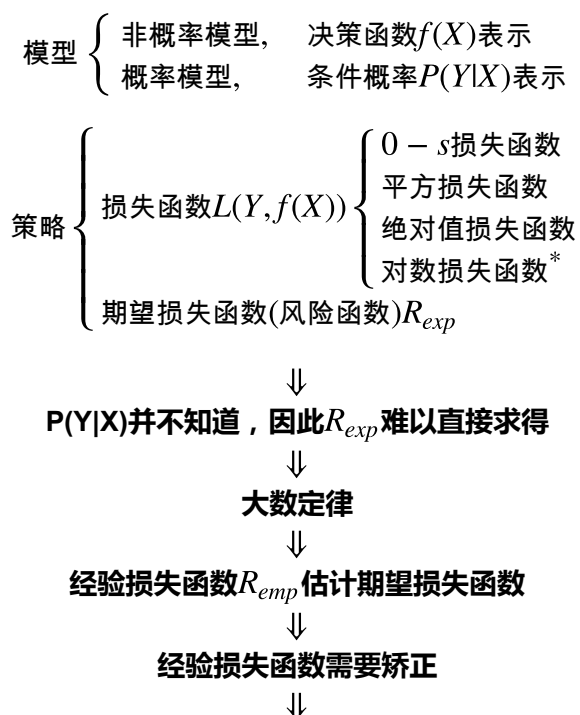
e.g:贝叶斯估计中的最大后验概率估计就是结构风险最小化的例子,当模型是条件概率分布、损失函数是对数损失函数、模型复杂度由模型的先验概率表示时,结构风险最小化就等价于最大后验概率估计。(*这一段还不是很明白等看完贝叶斯再回来看看)

3. 算法

统计学习问题归结为最优化问题,统计学习的算法成为求解最优化问题的算法。

Chapter1总结

这一章主要是三个问题:模型、策略和算法。其中最重要的是模型。



$$\begin{cases} \text{经验损失最小化} \Rightarrow \min_f \frac{1}{N} \sum_{i=1}^N (L(Y_i, f(X_i))) \\ \text{结构损失最小化} \Rightarrow \min_f \frac{1}{N} \sum_{i=1}^N (L(Y_i, f(X_i))) + \lambda J(f) \end{cases}$$

本章重要概念：假设空间、概率模型、非概率模型、损失函数、期望损失函数、经验损失函数、极大似然估计、结构损失函数、正则化、罚项