> 本文主要是在阅读过程中对本书的一些概念摘录，包括一些个人的理解，主要是思想理解不涉及到复杂的公式推导。若有不准确的地方，欢迎留言指正交流

统计学习

统计学习包括： 监督学习，非监督学习，半监督学习与强化学习。 本书内容主要以监督学习为主来进行阐述：

统计学习的三要素： **模型 + 策略 + 算法**

模型：由输入空间到输出空间的映射的集合，这个集合就是假设空间。(例如一个多层感知器需要拟合一个3阶的函数，那么所有3阶的函数构成的集合就是我们的假设空间，，线性函数就不属于，假设空间一般是无穷的。)

策略：指损失函数。这里有一个小的概念区别，损失函数与风险函数。损失函数用来度量预测错误的程度，常用的有0-1损失函数、平方损失函数，绝对损失函数，对数损失函数等。而损失函数的期望就称为风险函数， **学习的目标就是风险函数最小的模型。**

由**模型+策略+训练集**，共同决定了经验风险函数，最后监督学习就成了 **经验风险函数最小化（ERM）**问题，为了防止过拟合，在结构风险函数中可能加入正则化项，提出了**结构风险最小化（SRM）**问题。

$$SRM = ERM + 正则化项$$

过拟合是因为网络参数过多，这里的正则化项可以理解为对模型复杂度进行了限制，这也符合 **奥卡姆剃刀原理： 在所有可选择模型中，能够很好地解释已知数据并且十分简单才是最好的模型。** 极大似然估计是经验风险结构化的一个例子，最大后验概率是结构风险最小化的一个例子。

算法：指用什么样的方法求解最优模型。如何保证找到全局最优解（当然了大部分时候是不可能的），并且求解过程高效，成了一个重要问题。

模型估计与模型选择

模型估计

评估主要是靠训练时的**训练误差(training error)**和模型的**测试误差(test error)**。 训练误差的大小，对判定给定的问题是不是一个容易学习的问题是有意义的，但本质上不重要。测试误差则是反应了学习方法对未知的测试数据集的预测能力，是学习的重要概念。

模型选择

模型选择的方法主要是 正则化 与 交叉验证。
正则化： 符合奥卡姆剃刀原理，是结构风险最小化问题。
交叉验证： 分为

- 简单交叉验证： （训练集：测试集 = 7 : 3 分配数据，直接在训练集上调参，得到不同模型，最后在选择在测试集上性能最好的一个）
- S折交叉验证：随机地将已给数据切分为 S 份，然后利用 S-1 个子集进行训练，余下的子集做测试集，将这一过程重复进行 S 次，最后选出 S 次测评中平均测试误差最小的模型
- 留一交叉验证：是S 折交叉验证的特殊情况，S=N，N为样本量，这往往是在数据缺乏的情况下使用。

泛化能力

通过在未知数据上的期望风险作为泛化误差，来评价模型的泛化能力。这里有一个模型**泛化误差上界** 的概念，即上述期望风险的表达式的概率上界，其随着**样本量**的增加而下降，随着 **假设空间**的增加而增加。

生成模型与判别模型

生成模型(generative model)：由数据学习联合概率分布 P(X,Y),然后求解条件概率分布 P(Y|X)作为预测的模型。
判别模型(discriminative model)：由数据直接学习决策函数 f(X)或者条件概率分布 P(Y|X)作为预测的模型。

生成模型收敛速度更快，当含有隐变量（后面章节会提到）时，生成模型可用，判别模型就失效了。 判别方法则往往准确率更高。

分类、回归、标注问题

监督学习主要分为以下三类：分类，回归和标注问题。

- **分类问题**：

  评价分类问题常用的指标是准确率(precision)和召回率(recall)。
  TP——将正类预测为正类数
  FN——将正类预测为负类数
  FP——将负类预测为正类数
  TN——将负类预测为负类数

  精确率定义为：

$$P = TP / (TP+FP)$$

  召回率定义为：

$$R = TP / (TP+FN)$$

  还有 F_1 值，用来调和它们，P 和 R都高时，F_1 值也会高：

$$2 / F_1 = 1 / P + 1 / R$$

- **标注问题**：

  标注问题的目的在于学习一个模型，使它可以对观测序列给出标记序列作为预测。评价指标同分类问题，也是精确率与召回率。 常用方法为： 隐马尔可夫模型（HMM）、条件随机场（CRF）。

- **回归问题**：

  回归问题等价于函数拟合。
  按输入变量的个数分为： 一元回归 和 多元回归;
  按输入与输出的关系分为： 线性回归 与 非线性回归;
  常用损失函数： 平方损失函数（由最小二乘法求解）