**中国科学院大学 模式识别与机器学习 2019-2020复习**

1. **概述**
2. 什么是模式？

广义的说存在于时间和空间中可观察的物体如果我们可以区分他们是否相同或者是否相似都可以称之为模式。模式所指的不是事物本身，而是从事务获得的信息，因此，模式往往表现为具有时间和空间分布的信息。

直观特性：可观察性、可区分性、相似性

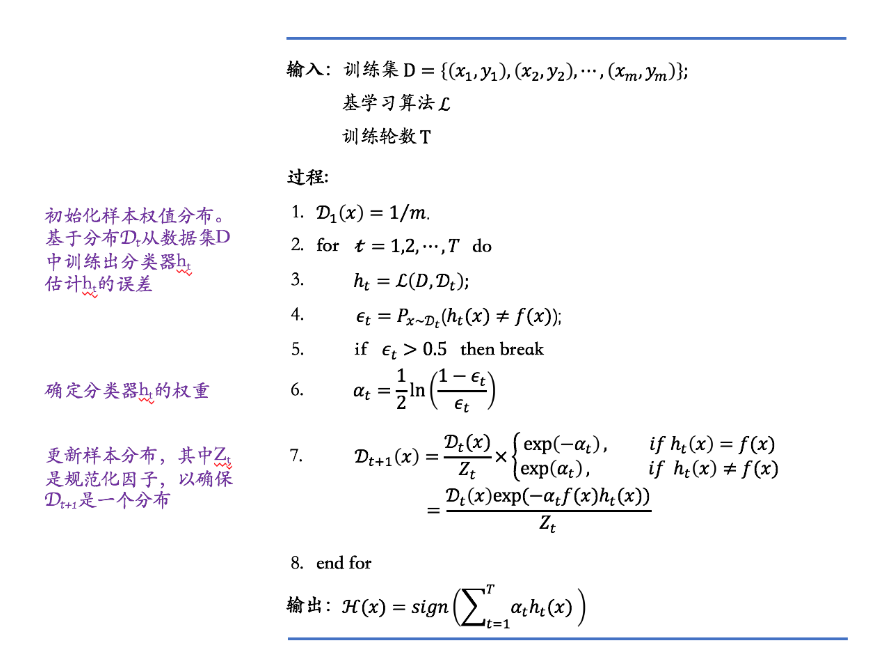
1. 统计判别

Fisher概念、fisher特点、fisher算例、贝叶斯算例、风险贝叶斯算例、感知器惩罚算例

1. 集成学习
2. **Adaboost原理**：先从初始训练集训练出一个学习器，再根据基学习器的表现来对训练样本分布进行调整，使得先前基学习器做错的训练样本在后续得到更多的关注，然后基于调整后的样本分布来训练下一个基学习器；如此重复进行，直到基学习器达到事先指定的值T，最终将这T个基学习器进行加权结合。
3. **Adaboost损失函数**：

使用指数损失函数：

1. **Adaboost算法流程：**



1. **为什么AdaBoost经常可以在训练误差为0后继续训练还可能带来测试误差的持续下降？**

在训练误差下降到接近0的时候，更多的训练，会增加分类器的分类margin，这个过程也能够防止测试误差的上升，随着Margin的变大，测试误差会逐渐收敛。

1. **AdaBoost优缺点：**

优点：实现快速简单、灵活、通用性高

缺点：AdaBoost性能取决于数据和弱学习器，如果弱分类器过于复杂，可能会产生过拟合情况，如果弱分类器太弱有可能造成欠拟合，还容易收到均匀噪声的影响。

1. **随即猜测作为一个分类算法是否一定比SVM差？对“No Free Lunch Theorem”“Occam’s razor”理解：**

不一定，根据No Free Lunch Theorem的前提：所有问题出现的机会相同，或者说所有问题都同等重要，所以如果脱离具体问题空泛谈论哪种学习算法更毫无意义，从模型的角度看，一个特定的模型必然会在解决某些问题时误差较小，在解决另一些问题时误差较大，从问题的角度看，在解决一个特定问题时，必然有某些模型具有较高的精度，而另一些模型精度就没那么理想。

Occam’s razor 简单有效原理, 不需要人为地把事情复杂化。比如在做决策树分析时，采用10个属性地预测性能和5个属性地预测性能相似，那么我们就会选择5个属性来预测。

1. 欠拟合、过拟合应对措施：

欠拟合：增加模型迭代次数、使用更多特征、降低模型正则化水平

过拟合：及早停止迭代、扩大训练集、减少特征数量、提高模型正则化水平

1. **Means算法流程：**

