

《图解机器学习》阅读计划

——大数据与机器学习群阅读计划（第1期）

领读人：友人A

本书特色

- 用最通俗、最简洁的语言全面系统地讲解了几几乎所有主流的机器学习算法，并配有大量图例
- 侧重点也和其他图书有所区别，注重结论的分析和应用而不是机器学习原理的推导
- 读者在阅读的过程中要注意各种算法的特点和使用方法

适合读者

所有想学习和了解机器学习，以及刚接触机器学习的读者

总阅读时间长度（预估）：4周

每天阅读用时：2小时

答疑时间安排：每周一次，每周六晚大数据与机器学习群20:00—22:00

阅读建议

建议大家提前温习概率论（图书可以参考图灵的《程序员的数学2：概率统计》或者《概率论及其应用》等）。

这本书虽然只有200页出头，却没有想象中那么简单。要真正掌握这些算法，除了认真阅读之外，动手实践是关键，最好能把书中的Matlab代码用Python或者其他自己喜欢的语言重写一遍。

阅读规划

第一部分（第1—6章）

阅读时长：1周

重点内容

1. 监督学习、无监督学习、强化学习等基本概念
2. 机器学习中用到的各种模型

3. 和连续函数的近似问题相对应的各种回归算法

难点内容

1. 最小二乘学习法
2. 由最小二乘学习法进化出来的其他回归算法
3. 各种回归算法之间的差异和联系以及各自的使用场景

补充

全书中几乎所有的算法都是基于最小二乘学习法的，所以它非常重要！另外，由于本书中的示例源代码是Matlab编写的，而它又是一种商业软件，所以建议大家尝试用Python来自己写一遍，这样也有助于理解。

第二部分（第7—11章）

阅读时长：1周

重点内容

1. 各种分类算法

难点内容

1. 明确支持向量机分类器和最小二乘法之间的关系
2. 如何把向量机分类器向鲁棒学习扩展
3. 将多种弱学习器通过适当组合形成强学习器

补充

这一部分主要讲分类算法，其中最重要的是支持向量机分类，这部分需要概率论知识，我自己在确定领读这本书之后也在复习相关知识了。

第三部分（第12—14章）

阅读时长：1周

重点内容

1. 异常检测（支持向量机的检测和基于密度比的检测）
2. 数据降维处理
3. 基于相似度分类的聚类方法

难点内容

1. 线性降维的原理
2. 无监督机器学习方法：聚类

第四部分（第15—20章）

阅读时长：1周（不够可以延长一周）

重点内容&难点内容

1. 把训练样本逐次输入的逐次学习算法
2. 半监督学习算法
3. 有监督的降维算法
4. 迁移学习法
5. 多任务学习算法

补充

这部分已经不再局限于单个的机器学习系统，而是涉及了通过其他学习任务的信息来提升当前学习任务的学习精度，在多个学习任务之间实现信息共享等，很考验大家的学习能力。这部分内容需要大家灵活运用概率论的知识。