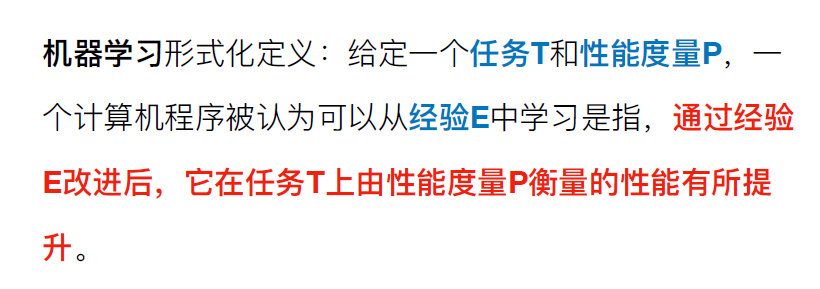
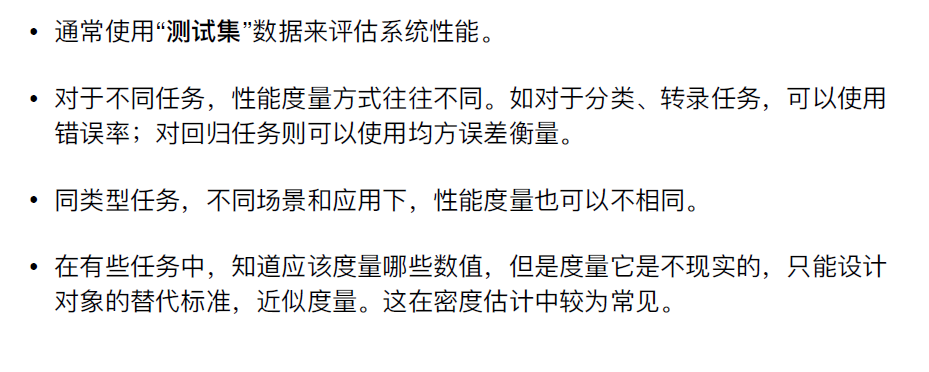
1. 机器学习的形式化定义是什么？



1. 机器学习中模型的性能度量一般使用什么数据集？为什么？

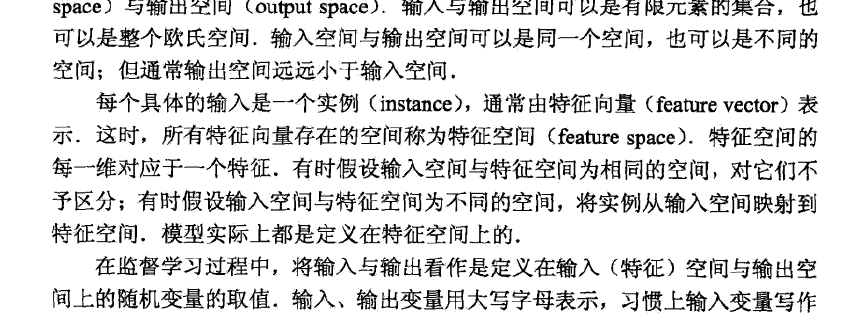
测试集



1. 解释“特征”与“特征空间”。“样本”，“特征”，“特征空间”三者之间的关系是什么？

特征（属性）：用于表征样本的观测信息。

特征空间：分别以每个特征作为一个坐标轴，所有特征所在坐标轴张成的一个用于描述不同样本的空间。



1. 解释“假设”、“假设空间”、“版本空间”。

假设：每个具体的模型就是一个假设

假设空间：所有假设构成的空间

版本空间：基于有限规模训练集进行的匹配搜索，会存在多个假设与训练集一致的情况，称这些情况组成的集合为版本空间。

1. 机器学习的三要素是什么？请用自己的话语简单描述一下三要素的含义。

机器学习的三要素：模型、策略（选取模型的准则）、算法（优化算法）

模型——机器学习的目的：

模型就是那个用来描述客观世界的数学模型，模型是从数据里抽象出来的。模型可以是确定性的，也可以是随机的，无所谓，总之用数学可以描述，只要数学可以描述的，就可以进行预测分析。所以，我们的根本目的，是找一个模型去描述我们已经观测到的数据。

策略——如何构造模型：

在挑选模型的时候，需要有一系列的标准来证明一个模型比另一个模型好，这就是策略。不同的策略，对应不同的模型的比较标准和选择标准。

算法——模型的实现：

我们有了数据，有了学习模型的策略，然后就要开始去构造模型了，如果模型的基本形式有了，就是一个优化模型参数的问题了。目标是让算法尽量高效，更少的计算机内存代价，更快的运算速度，更有效的参数优化结果。

1. 说说“经验风险”与“期望风险”之间有什么不同。

经验风险：模型在给定的训练集上的代价。

期望风险：模型在真实数据分布上的代价。

经验风险是局部的，基于训练集所有样本点损失函数最小化的。

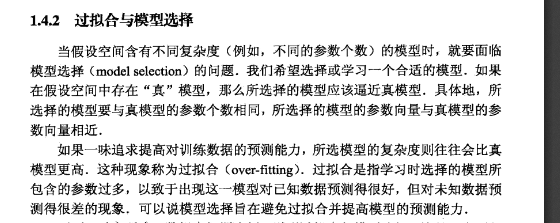
期望风险是全局的，是基于所有样本点的损失函数最小化的。

经验风险函数是现实的，可求的；

期望风险函数是理想化的，不可求的；

- 经验风险是局部的, 基于训练集中所有样本点损失函数最小化, 经验风险是局部最优, 是现实可求的.  
- 期望风险是全局的, 基于所有样本点损失函数最小化. 期望风险是全局最优, 是理想化的不可求的.

7. 简述“经验风险最小化”为什么可能导致过拟合。



经验风险越小，模型决策函数越复杂，其包含的参数越多，当经验风险函数小到一定程度就出现了过拟合现象。

8. 我们知道：在一个特定任务中，模型泛化性能往往随着模型容量的提升而逐步提升，再达到一定性能后，继续提升模型容量，反而会使得模型泛化性能降低，这意味着我们需要找到一个最佳容量的模型完成此任务才能得到最佳性能，而不同的任务下模型的容量是很难确定的。那么，思考：使用什么样的方法能得到性能较好的模型？