**机器学习的一些概念**

**·有监督、无监督**

根据训练数据是否拥有标记信息，学习任务大致划分为两类：监督学习和无监督学习。分类和回归是监督学习的代表，聚类是无监督学习的代表

**·泛化能力**

机器学习的目标是使得学得的模型能很好地适用于新样本。学得模型适用于新样本的能力，称为泛化能力

**·过拟合、欠拟合**

过拟合是学习器把训练样本学的“太好了”，把训练样本自身的一些特点当作了所有潜在样本都会具有的一般性质，导致泛化性能下降。与之对应的是欠拟合，指对训练样本的一般性质尚未学好

**·方差**

期望输出和真实标记之差平方的均值，度量了同样大小的训练集的变动所导致的学习性能的变化，刻画了数据扰动所造成的影响

**·偏差**

期望输出和真实输出的差别，度量学习算法的期望预测与真实结果的偏离程度。

**·交叉验证**

先将数据集D划分为k个大小相似的互斥子集，每个子集都尽可能保持数据分布的一致性，即从D中分层采样得到。然后每次用k-1个子集的并集作为训练集，余下的那个子集作为测试集，从而可以进行k次训练和测试，最终返回这k个测试结果的均值。

**·线性回归的知识**

<https://www.cnblogs.com/hiyoung/archive/2018/10/09/9763599.html>

·sklearn教程

<https://blog.csdn.net/yuanshuaipeng/article/details/80399863>