在线学习

介绍：

有监督学习中，不知道样本的真实分布，这种情况下的学习目标是最小化损失函数。

统计视角下的在线学习：

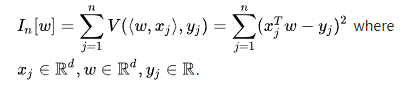
在统计学习模型中，假设样本从真实分布中抽取，目标是最小化经验风险：

1555335309(1)

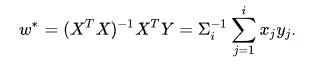
其中表示预测值与真实值的误差。这种情况下一个常见的例子是通过经验风险最小化或者岭回归来估计一个函数。这里损失函数的选择引出了几种常见的学习算法，如正则化最小二乘，支持向量机。一种单纯的在线学习模型只基于新输入的样本，当前的最佳预测算子，以及一些额外存储的信息。对于很多公式，比如非线性的核方法，真正的在线学习是不可能的。尽管当被允许依赖于以及之前的所有数据点时，可以使用一种伴有递推算法的混合在线模型。在这种情况下，需要的空间不再保证是常数，因为需要存储之前所有数据点的信息。但是因为添加了一个新的数据点，相比于批处理学习，计算出解可能需要更少的时间。

批处理学习：

有监督学习中，在均方误差损失函数的设置下，目标是最小化经验损失：



最优解为：



计算协方差矩阵的复杂度为，的矩阵求逆复杂度为。

在线学习：迭代最小二乘