**自动调参**

汉融通评分模型的构建重点关注机器学习模型的自动化建设。随着机器学习算法的广泛应用，在模型构建过程中，人工调参和模型迭代占用了建模人员许多时间，为了简化建模过程，减少建模人员的重复工作，让建模人员能够更多地深入实际业务，将业务问题更好地与技术融合，我们致力于使用高级的控制系统去操作机器学习模型，使得模型可以自动化地学习到合适的参数和配置，减少人工干预。这可以极大地提升建模效率，取得更好的模型效果。

通常模型的调参工作由建模人员根据经验反复调整，耗时耗力，占用了技术人员的宝贵时间。我们采用自动调参技术，在给定的搜索空间范围内，获取使模型达到最佳性能的参数，提高模型的预测效果，自动调参方法包括随机搜索、网格搜索和贝叶斯优化。

作为中诚信征信万象智慧平台使用的专利授权方法，贝叶斯优化在超参数调优中得到了充分的运用。对于机器学习模型，在观测数据D已知的情况下，其目的就是寻找一组超参数，使得模型表达式f(x)在现有数据上的预测效果尽可能好，而贝叶斯优化就是在函数方程f(x)不知的情况下根据已有的采样点预估函数最大值的一个算法。

假设目标函数f(x)服从高斯过程，通过随机采样少量数据，就可以确定一个对应的高斯分布，然后通过这个高斯分布的某些指标来采样新的点x\*，把（x\*,f(x\*))添加到观测数据D中，就可以描述更精确的高斯分布，其中选择新的采样点x\*就要使用到提取函数。

提取函数的使用考虑到两个方面：exploit和explore。exploit用于强化已有的结果，在现有最大值的附近进行探索，基本思路是贪心算法的思想，认为最大值的附近更可能存在更大的值；而explore用于尽可能的探索未知的空间，这样对f(x)描述的高斯分布才会更接近f(x)，避免exploit陷入局部优化。所以，贝叶斯优化的整个过程就是不停的采样并计算高斯过程，采样时考虑explore 和exploit之间的权衡。

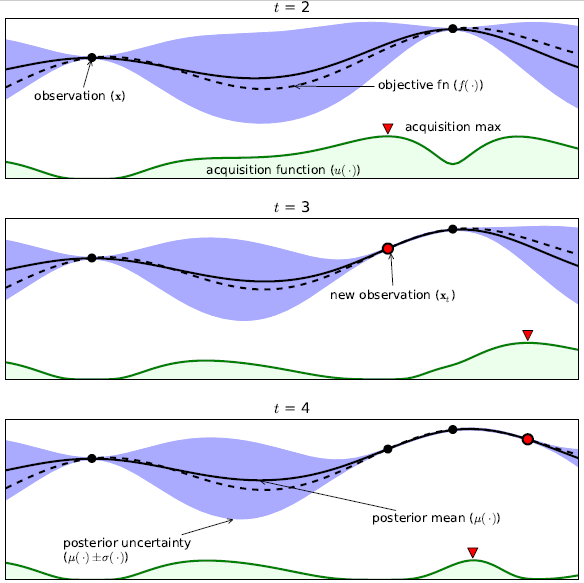


图2 贝叶斯优化在一维问题上的例子[1]

如图2是贝叶斯优化在一维问题上的例子，用于解释贝叶斯优化的简要过程。我们的目标函数f(x)是图中的黑色虚线，首先随机采样2个点，生成的高斯分布如图中黑色实线，明显黑色实线和虚线仍旧存在一定差距。接下来我们选择使提取函数（图中绿色曲线）达到最大值的位置作为新的采样位置x\*，提取新的采样点（x\*,f(x\*))作为第3个数据点加入，此时黑色实线更加逼近虚线。以此类推，当采样的样本点足够多，那么高斯分布就会收敛到类似f(x)的样子，将最终采样的数据点中最好的那组随机变量作为参数结果输出，此时的参数即为贝叶斯优化获取的最优参数。

贝叶斯优化用在参数调优过程中，我们需要确立目标变量，然后构建算法超参数和目标变量之间的关系，最后采用贝叶斯优化算法，确定使目标变量达到最大值的一组超参数，这组参数就是我们寻找的最优参数。

当模型的调参工作由贝叶斯优化操作后，可以为建模人员节省大量时间，使得建模人员可以深入信用评级和反欺诈业务中，挖掘更多有助于提升模型效果的特征和信息。