机器学习第三次作业

郭洲蕊 2021E8014682024

1. 随机森林为何比决策树 Bagging 集成的训练速度更快

决策树的生成过程中，最耗时的就是搜寻最优切分属性；随机森林在决策树训练过程中引入了随机属性选择，大大减少了此过程的计算量；因而随机森林比普通决策树Bagging训练速度要快。

1. 比较 Gradient Boosting 与 AdaBoost 的异同

Gradient Boosting 和其它 Boosting 算法一样，通过将表现一般的数个模型（通常是深度固定的决策树）组合在一起来集成一个表现较好的模型。抽象地说，模型的训练过程是对一任意可导目标函数的优化过程。通过反复地选择一个指向负梯度方向的函数，该算法可被看做在函数空间里对目标函数进行优化。

和 AdaBoost 一样，Gradient Boosting 也是重复选择一个表现一般的模型并且每次基于先前模型的表现进行调整。不同的是，AdaBoost 是通过提升错分数据点的权重来定位模型的不足而 Gradient Boosting 是通过算梯度（gradient）来定位模型的不足。因此相比 AdaBoost, Gradient Boosting 可以使用更多种类的目标函数。

1. 比较包裹式选择、过滤式选择与嵌入式选择的异同

过滤式方法**先对数据集进行特征选择，然后再训练学习器**。特征选择过程与后续学习器无关，这相当于先对初始特征进行“过滤”，再用过滤后的特征训练模型。

包裹式**从初始特征集合中不断的选择特征子集，训练学习器，根据学习器的性能来对子集进行评价，直到选择出最佳的子集。**包裹式特征选择直接针对给定学习器进行优化。

在过滤式和包裹式特征选择方法中，特征选择过程与学习器训练过程有明显的分别。而嵌入式特征选择**在学习器训练过程中自动地进行特征选择**。嵌入式选择最常用的是L1 正则化和L2正则化。正则化项越大，模型越简单，系数越小，当正则化项增大到一定程度时，所有的特征系数都会趋于0，在这个过程中，会有一部分特征的系数先变成0。也就实现了特征选择过程。逻辑回归、线性回归、决策树都可以当作正则化选择特征的基学习器，只有可以得到特征系数或者可以得到特征重要度的算法才可以作为嵌入式选择的基学习器。

1. 直接求解 L0 范数正则化会遇到的困难

L0范数是指向量中非0的元素的个数。用L0范数来规则化一个参数矩阵W的话，就是希望W的大部分元素都是0。L0范数非连续，不可导，很难优化求解。

在L0正则化时的求解方法是，分别设定ω中某些元素为零的情况下，求解无正则化的优化问题，最终比较各种情况下的目标函数值，确定最优解。设特征数为N，则按照上面的方法，要在2^N的情况下分别求解优化问题，这个次数随特征数指数级增长，当特征数较多时，比较困难。

1. 为什么基于 L1 范数可以进行特征选择

形状, 圆圈

描述已自动生成

等值线是J0的等值线，黑色方形是L1函数的图形，J0等值线与L1图形首次相交的地方就是最优解，易发现**黑色方形必然首先与等值线相交于方形顶点处**。可以直观想象，因为L1函数有很多“**突出的角**”，J0与这些角接触的概率远大于与其它部分接触的概率。而这些点某些维度为0，从而会使部分特征等于0，产生稀疏模型，进而可以用于特征选择。

1. 比较 K-SVD 与 K-means 方法的异同

图示

描述已自动生成

K-means算法可以看做是K-SVD的简单版，同样可以理解K-SVD是K-means算法的衍生版本。K-means算法中X类似one-hot形式，X中的每个列向量中只有一项不为0，其余均为0，即每个样本只由一个原子逼近。

K-means主要通过计算距离来分类，并不是利用DX来逼近原样本矩阵Y，从而主要用于聚类，而K-SVD则由一系列原子来线性组合逼近，因此相比K-means更适用于压缩，编码等应用。