处理连续性特征：

离散化和二分类方法

离散化：

若在其特征上有m个取值，不妨假设他们为 的；不失一般性，假设他们满足，(若不然，进行一次排序操作即可)，那么依次选择作为界限对数据进行划分，其中构成等差数列，且：





p的选择视情况而定，一般而言，p会反比与‘深度’。这意味着对数据越分越细时，需要选择更粗的特征选取。

离散化的问题：

离散化到二类方法：

使用二类问题的解决方案处理连续型特征。具体而言，当二类问题和决策树结合起来时，在连续的情况下我们通常可以把它转述为：





相对应的，同样可以用处理二类问题的思想来处理离散型特征，此时：



更进一步，我们通常表示为：



我们通常称上式中的a1 为‘二分标准’。一般而言，如何处理连续型特征会归结于，如何选择‘二分标准’这个问题。