

法律声明

■ 本课件包括演示文稿、示例、代码、题库、视频和声音等内容，北风网和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

■ 课程详情请咨询

◆ 微信公众号：北风教育

◆ 官方网址：<http://www.ibeifeng.com/>



人工智能之机器学习

KD Tree

主讲人：Gerry

上海育创网络科技有限公司



课程要求

■ 课上课下 “九字” 真言

- ◆ 认真听，善摘录，勤思考
- ◆ **多温故，乐实践**，再发散

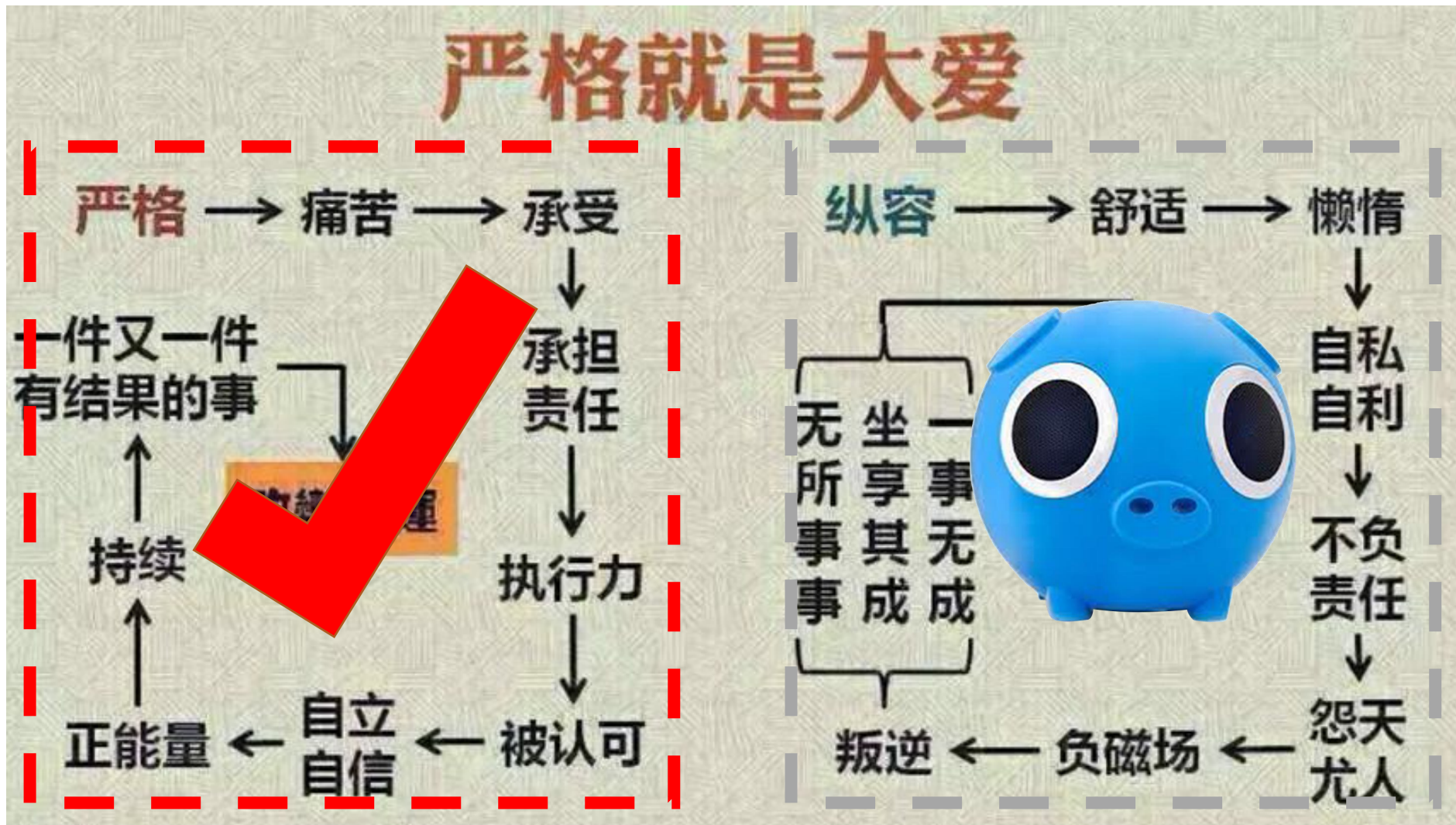
■ 四不原则

- ◆ **不懒散惰性，不迟到早退**
- ◆ **不请假旷课，不拖延作业**

■ 一点注意事项

- ◆ 违反 “四不原则”，不包就业和推荐就业

严格是大爱



寄语



做别人不愿做的事，
做别人不敢做的事，
做别人做不到的事。

课程内容

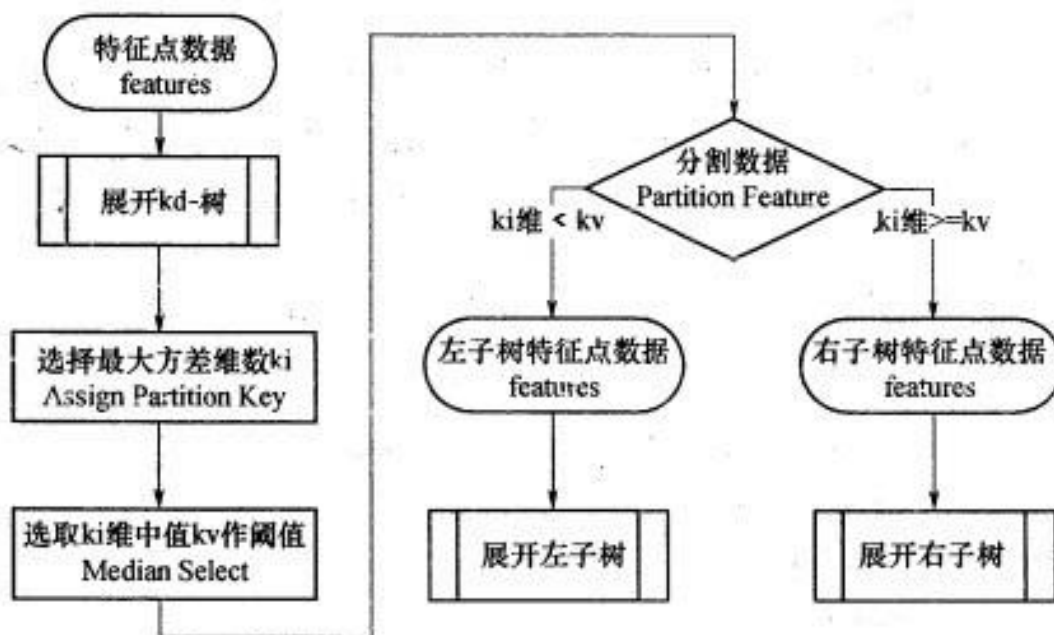
■ KD Tree

KD Tree

- KD Tree是密度聚类(DBSCAN)算法中计算样本和核心对象之间距离来获取最近邻以及KNN算法中用于计算最近邻的快速、便捷构建方式。
- 当样本数据量少的时候，我们可以使用brute这种暴力的方式进行求解最近邻，即计算到所有样本的距离。但是当样本量比较大的时候，直接计算所有样本的距离，工作量有点大，所以在这种情况下，我们可以使用kd tree来快速的计算。

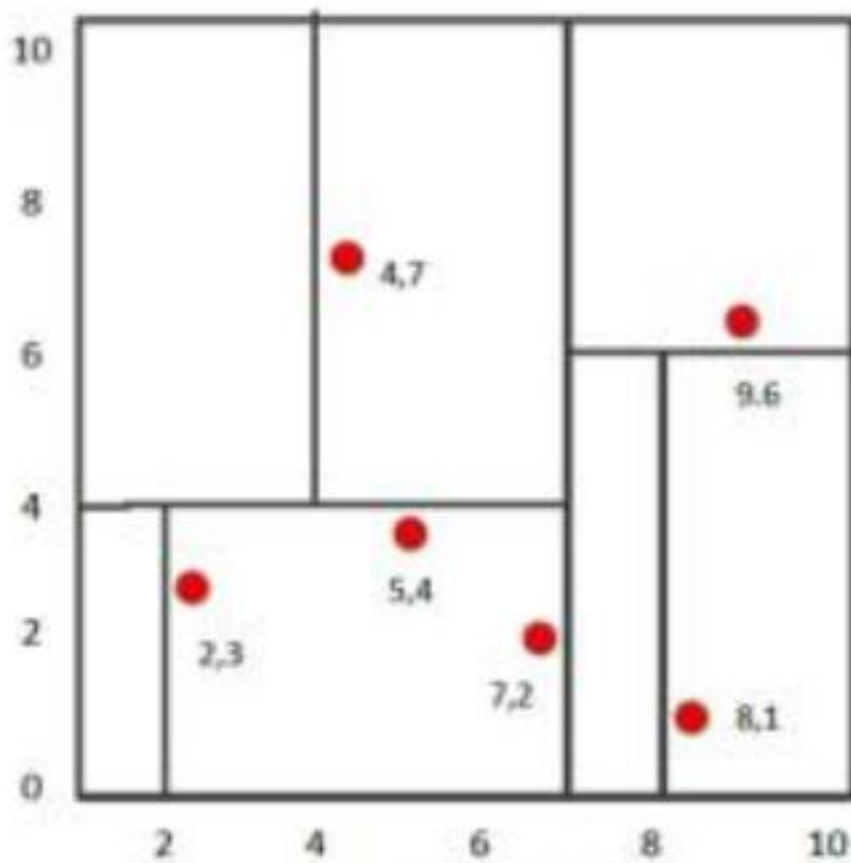
KD Tree构建方式

- KD树采用从m个样本的n维特征中，分别计算n个特征取值的方差，用方差最大的第k维特征 n_k 作为根节点。对于这个特征，选择取值的中位数 n_{kv} 作为样本的划分点，对于小于该值的样本划分到左子树，对于大于等于该值的样本划分到右子树，对左右子树采用同样的方式找方差最大的特征作为根节点，递归即可产生KD树。



KD tree

- 二维样本: $\{(2,3), (5,4), (9,6), (4,7), (8,1), (7,2)\}$

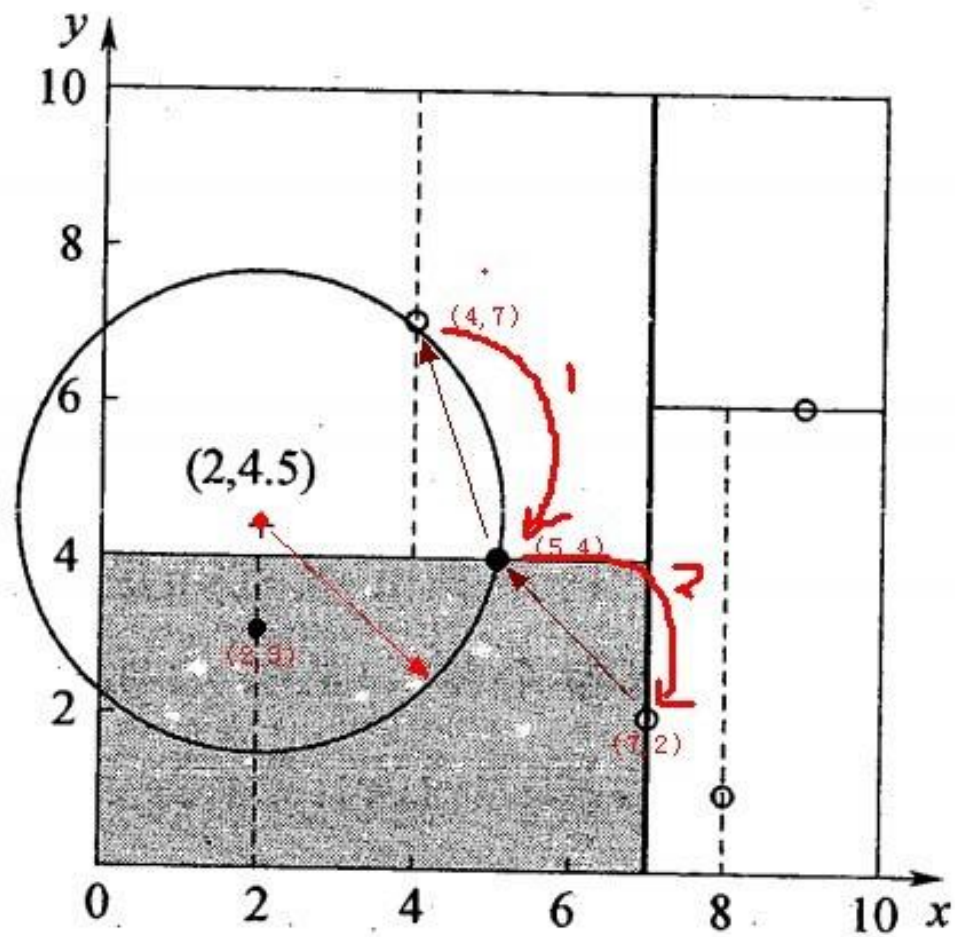


特征空间的划分

KD tree查找最近邻

- 当我们生成KD树以后，就可以去预测测试集里面的样本目标点了。对于一个目标点，我们首先在KD树里面找到包含目标点的叶子节点。以目标点为圆心，以目标点到叶子节点样本实例的距离为半径，得到一个超球体，最近邻的点一定在这个超球体内部。然后返回叶子节点的父节点，检查另一个子节点包含的超矩形体是否和超球体相交，如果相交就到这个子节点寻找是否有更加近的近邻，有的话就更新最近邻。如果不相交那就简单了，我们直接返回父节点的父节点，在另一个子树继续搜索最近邻。当回溯到根节点时，算法结束，此时保存的最近邻节点就是最终的最近邻。

KD tree查找最近邻





上海育创网络科技有限公司