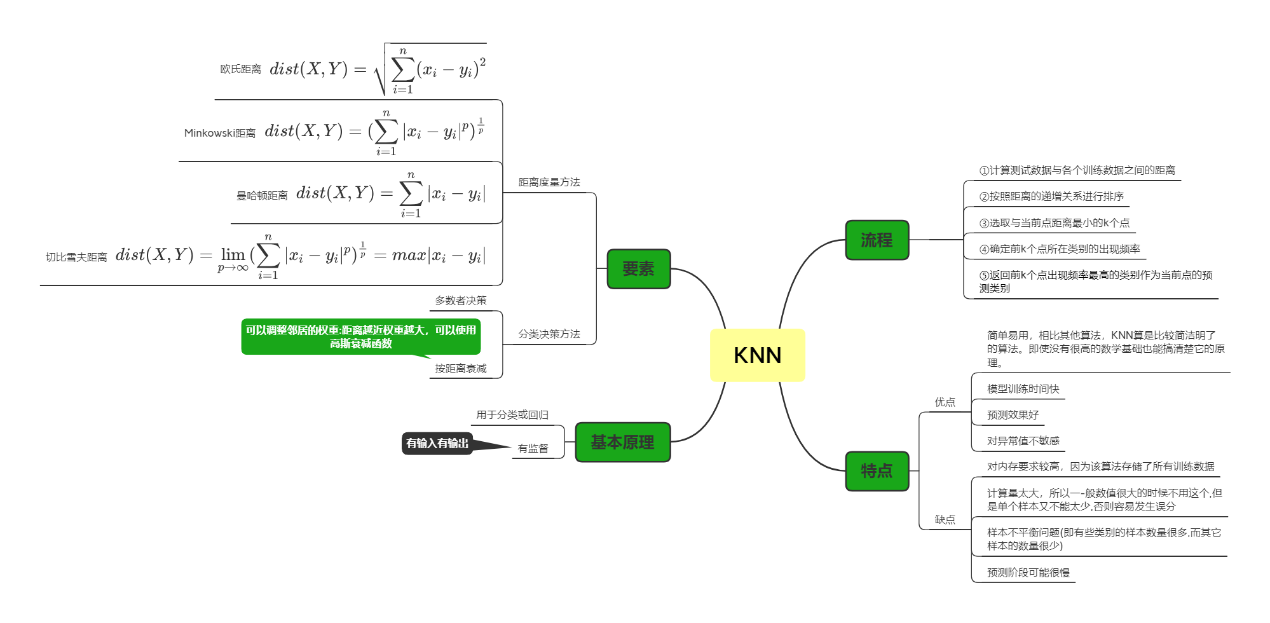
1.**思维导图**



2. **为什么要在模型训练之前对数据进行标准化?数据归一化有哪些方法?**

对数据进行归一化是为了提高结果的准确度。归一化的方法有：最值归一化、均值方差归一化。

3. **什么是交叉验证?伪代码如何表达?**

交叉验证(Cross-Validation): 有时亦称循环估计， 是一种统计学上将数据样本切割成较小子集的实用方法。于是可以先在一个子集上做分析， 而其它子集则用来做后续对此分析的确认及验证。 一开始的子集被称为训练集。而其它的子集则被称为验证集或测试集。在KNN算法里面就是为了选取最合适的K值。

伪代码：

Int sum,n;

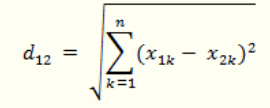
Int model=sum/n;

K=getK();//getK函数为求n个模型最终的验证集的分类准确率的平均数，这个K即为最终的K

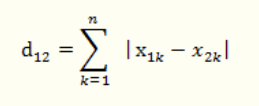
4. **距离的度量有哪些方法?如何自己设计一种距离度量方法,度量在Internet物理网络中两个路由器的距离?有没有办法度量两个人之间的社交距离?**

两个n维向量a(x11,x12,…,x1n)与 b(x21,x22,…,x2n)间的距离：

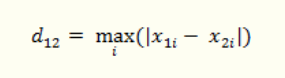
1. 欧氏距离



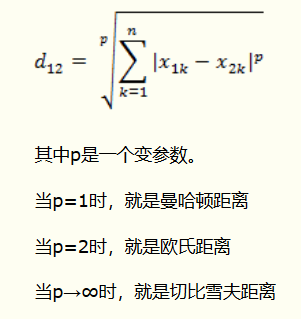
1. 曼哈顿距离



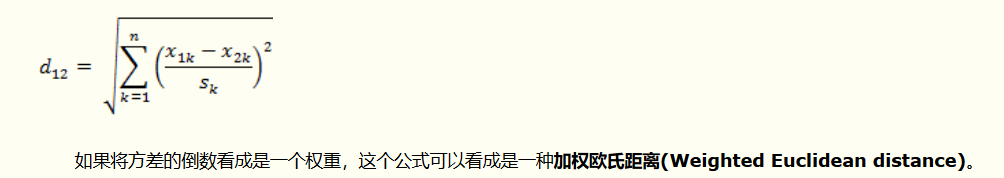
1. 切比雪夫距离



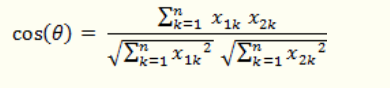
1. 闵可夫斯基距离



1. 标准化欧氏距离



1. 夹角余弦



路由器距离：数据在两个路由器间的数据传输速度\*数据总传输时间

社交距离：分别将两个人的家庭住址，学校，工作单位在地图上标注出来，算出它们的距离。

5**.请查资料了解KNN的选取k的方法有哪些?**

交叉验证