聚类算法

机器学习笔记 create by siwanghu v1.0

K均值算法，无监督学习算法，用于将相似的样本自动归到一个类别中

欧式距离：

曼哈顿距离：

切比雪夫距离：**d=Max(||，||，，，||)**

Jaccard相似系数：

相关系数：

选择k个点作为初始质心

repeat

将每个点指派到最近的质心，形成k个簇

重新计算每个簇的质心

until 簇不发生变化或达到最大迭代次数

假设使用欧式距离计算数据之间的离散程度，则优化损失函数为：

K表示簇数目，表示每个簇的质心，优化目标是使得最小，求解的位置。对求偏导，令导数等于0，求解的位置,代表每个簇的元素数目