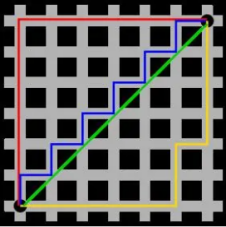
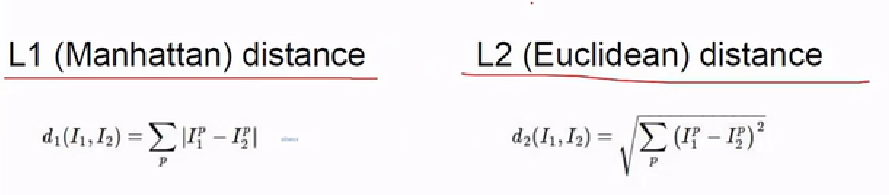
## K临近与交叉验证

1. 超参数：在机器学习的[上下文](https://baike.baidu.com/item/%E4%B8%8A%E4%B8%8B%E6%96%87/2884376" \t "https://baike.baidu.com/item/%E8%B6%85%E5%8F%82%E6%95%B0/_blank)中，超参数是**在开始学习过程之前设置值的参数**，而不是通过训练得到的参数数据。通常情况下，需要对超参数进行优化，给学习机选择一组最优超参数，以提高学习的性能和效果。

（参数难以理解，曼哈顿距离（红线），欧氏距离/直线距离（绿））参数含义





1. K最近邻(k-Nearest Neighbor，KNN)分类算法，是一个理论上比较成熟的方法，也是最简单的机器学习算法之一。该方法的思路是：在特征空间中，如果一个样本附近的k个最近(即特征空间中最邻近)样本的大多数属于某一个类别，则该样本也属于这个类别。物以类聚
2. 交叉验证

情景：选择超参数时

Test data是非常宝贵的，不能用它来验证参数，否则就没有真实数据测试了

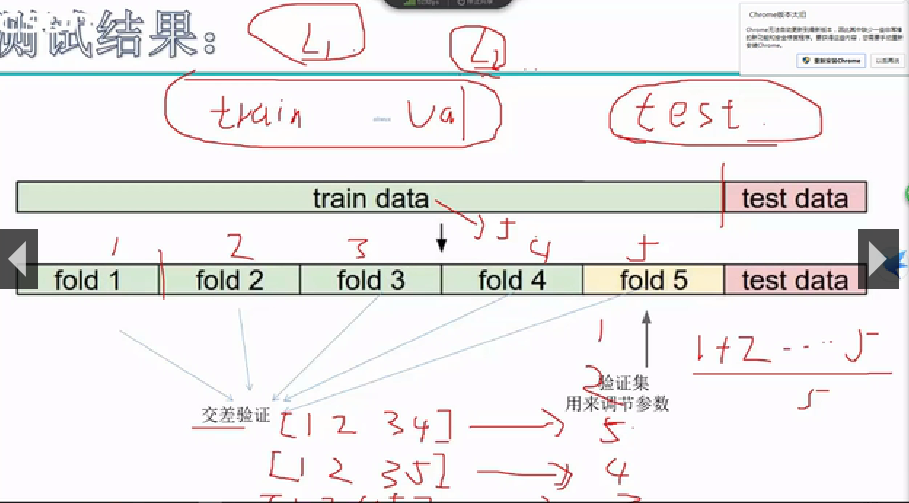
所以将train data划分为多组如下图示意，用fold1~4训练，用fold5验证，又产生问题：

如果fold5里有异常数据怎么办？怎样尽可能提高结果的准确性？

**交叉验证：用fold1235训练用fold4验证、用fold1245训练用fold3验证、用fold1345训练用fold2验证、用fold2345训练用fold1验证，得出结果相加除以五（取平均）**

假设对L1进行交叉验证的精度85%，L2精度95%，就选L2

所以选择超参数L2



调好参数，用test data进行最后的评估

## 得分函数

1.

2.