**1.欧氏距离**

 简单来说就是两点间距离

二维平面上两点a(x1,y1)与b(x2,y2)间的欧氏距离

IMG_256

三维空间两点a(x1,y1,z1)与b(x2,y2,z2)间的欧氏距离

扩展到--IMG_257两个n维向量a(x11,x12,…,x1n)与 b(x21,x22,…,x2n)间的欧氏距离

**2.曼哈顿距离**

      从一个十字路口开车到另外一个十字路口，驾驶距离是两点间的直线距离吗？显然不是，除非你能穿越大楼。实际驾驶距离就是这个“曼哈顿距离”。

IMG_256

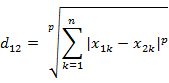
1. **切比雪夫距离**

IMG_256

      形象地来说，就是棋盘上，格子(x1,y1)走到格子(x2,y2)最少需要多少步。

**4. 闵可夫斯基距离**

 两个n维变量a(x11,x12,…,x1n)与b(x21,x22,…,x2n)间的闵可夫斯基距离定义为：



其中p是一个变参数。

当p=1时，就是曼哈顿距离

当p=2时，就是欧氏距离

当p→∞时，就是切比雪夫距离

       根据变参数的不同，闵氏距离可以表示一类的距离。

**———————————————————————————————**

**上面四个距离的缺点主要有两个：**

**(1)将各个分量的量纲(scale)，也就是“单位”当作相同的看待了。**

**(2)没有考虑各个分量的分布（期望，方差等)可能是不同的。**

**———————————————————————————————**

**5.标准化欧氏距离**

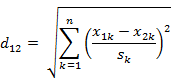
对欧氏距离进行优化，由于数据各维分量的分布不一样，先将各个分量都“标准化”到均值、方差相等

因此样本集的标准化过程(standardization)用公式描述就是：

IMG_259

　　标准化后的值 =  ( 标准化前的值  － 分量的均值 ) /分量的标准差

　　经过简单的推导就可以得到两个n维向量a(x11,x12,…,x1n)与b(x21,x22,…,x2n)间的标准化欧氏距离的公式：



　　如果将方差的倒数看成是一个权重，这个公式可以看成是一种**加权欧氏距离(WeightedEuclidean distance)**。

**6.马氏距离**

有M个样本向量X1~Xm，协方差矩阵记为S，均值记为向量μ，则其中样本向量X到u的马氏距离表示为：

**IMG_261IMG_261**

       而其中向量Xi与Xj之间的马氏距离定义为：

IMG_262

       若协方差矩阵是单位矩阵（各个样本向量之间独立同分布）,则公式就成了：

IMG_263

       也就是欧氏距离了。

　　若协方差矩阵是对角矩阵，公式变成了标准化欧氏距离。

1. **夹角余弦**

几何中夹角余弦可用来衡量两个向量方向的差异，机器学习中借用这一概念来衡量样本向量之间的差异。

IMG_256

余弦取值范围为[-1,1]。求得两个向量的夹角，并得出夹角对应的余弦值，此余弦值就可以用来表征这两个向量的相似性。夹角越小，趋近于0度，余弦值越接近于1，它们的方向更加吻合，则越相似。当两个向量的方向完全相反夹角余弦取最小值-1。当余弦值为0时，两向量正交，夹角为90度。因此可以看出，余弦相似度与向量的幅值无关，只与向量的方向相关。

1. **汉明距离**

两个等长字符串s1与s2之间的汉明距离定义为将其中一个变为另外一个所需要作的最小替换次数。例如字符串“1111”与“1001”之间的汉明距离为2。  
       应用：信息编码（为了增强容错性，应使得编码间的最小汉明距离尽可能大）。

1. **杰卡德距离& 杰卡德相似系数**

(1) 杰卡德相似系数

两个集合A和B的交集元素在A，B的并集中所占的比例，称为两个集合的杰卡德相似系数，用符号J(A,B)表示。

IMG_256

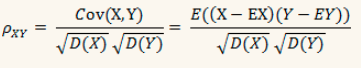
　杰卡德相似系数是衡量两个集合的相似度一种指标。

(2) 杰卡德距离

与杰卡德相似系数相反的概念是杰卡德距离(Jaccard distance)。杰卡德距离可用如下公式表示：

IMG_256

1. **皮尔逊相关系数& 相关距离**



前面提到的余弦相似度只与向量方向有关，但它会受到向量的平移影响，在夹角余弦公式中如果将 x 平移到 x+1, 余弦值就会改变。怎样才能实现平移不变性？这就要用到****皮尔逊相关系数****

相关系数具有平移不变性和尺度不变性，计算出了两个向量（维度）的相关性