### 3.1.欧式距离

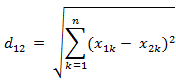
* **1. 欧氏距离**，最常见的两点之间或多点之间的距离表示法，又称之为欧几里得度量，它定义于欧几里得空间中，如点 x = (x1,...,xn) 和 y = (y1,...,yn) 之间的距离为：http://img.my.csdn.net/uploads/201211/20/1353398777_7638.png
* (1)二维平面上两点a(x1,y1)与b(x2,y2)间的欧氏距离：

http://img.my.csdn.net/uploads/201211/20/1353399552_4107.png

* (2)三维空间两点a(x1,y1,z1)与b(x2,y2,z2)间的欧氏距离：

http://img.my.csdn.net/uploads/201211/20/1353399601_9675.png

* (3)两个n维向量a(x11,x12,…,x1n)与 b(x21,x22,…,x2n)间的欧氏距离：



* 也可以用表示成向量运算的形式：

http://img.my.csdn.net/uploads/201211/20/1353399664_2255.png

**2. 标准化欧氏距离 (Standardized Euclidean distance )**，标准化欧氏距离是针对简单欧氏距离的缺点而作的一种改进方案。标准欧氏距离的思路：既然数据各维分量的分布不一样，那先将各个分量都“标准化”到均值、方差相等。至于均值和方差标准化到多少，先复习点统计学知识。

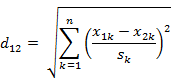
假设样本集X的数学期望或均值(mean)为m，标准差(standard deviation，方差开根)为s，那么X的“标准化变量”X\*表示为：(X-m）/s，而且标准化变量的数学期望为0，方差为1。

即，样本集的标准化过程(standardization)用公式描述就是：

http://img.my.csdn.net/uploads/201211/21/1353468927_4645.png

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(Weighted Euclidean distance)。