欧氏距离（Euclidean distance）也称欧几里得度量、欧几里得度量，是一个通常采用的距离定义，它是在m维空间中两个点之间的真实距离。在二维和三维空间中的欧氏距离的就是两点之间的距离。

缺点：就大部分统计问题而言，欧氏距离是不能令人满意的。（每个坐标对欧氏距离的贡献是同等的。当坐标表示测量值时，它们往往带有大小不等的随机波动，在这种情况下，合理的方法是对坐标加权，使变化较大的坐标比变化较小的坐标有较小的权系数，这就产生了各种距离。当各个分量为不同性质的量时，“距离”的大小与指标的单位有关。它将样品的不同属性（即各指标或各变量）之间的差别等同看待，这一点有时不能满足实际要求。没有考虑到总体变异对距离远近的影响。

马氏距离(Mahalanobis distance)是由印度统计学家马哈拉诺比斯提出的，表示数据的协方差距离。为两个服从同一分布并且其协方差矩阵为Σ的随机变量与的差异程度:如果协方差矩阵为单位矩阵,那么马氏距离就简化为欧氏距离,如果协方差矩阵为对角阵,则其也可称为正规化的欧氏距离。它是一种有效的计算两个未知样本集的相似度的方法。对于一个均值为μ，协方差矩阵为Σ的多变量向量，样本与总体的马氏距离为(dm)^2=(x-μ)'Σ^(-1)(x-μ)。 在绝大多数情况下，马氏距离是可以顺利计算的，但是马氏距离的计算是不稳定的，不稳定的来源是协方差矩阵，这也是马氏距离与欧式距离的最大差异之处。

$$d_A(x_i, x_j) = \sqrt{(x_i - x_j)^T A (x_i - x_j)}\ ,$$

where $A \in \mathbb{R}^{m \times m}$ is a positive semi-definite matrix [51]. This is equivalent to projecting each input x to a new space $\mathbb{R}^m$ in which their euclidean distances obey the desired constraints. There are many different ways to create such a metric. The most original approach was to globally solve the following convex optimization problem:

$$\begin{aligned} &\min_A \textstyle\sum_{(x_i,x_j)\in\mathcal{S}} d_A(x_i,x_j)^2 \\ &\text{s.t. } \textstyle\sum_{(x_i,x_j)\in\mathcal{D}} d_A(x_i,x_j)^2 \geq 1 \text{ and } A \succeq 0. \end{aligned}$$

优点：它不受量纲的影响，两点之间的马氏距离与原始数据的测量单位无关。（它考虑到各种特性之间的联系（例如：一条关于身高的信息会带来一条关于体重的信息，因为两者是有关联的）并且是尺度无关的(scale-invariant)，即独立于测量尺度）；由标准化数据和中心化数据(即原始数据与均值之差）计算出的二点之间的马氏距离相同。马氏距离还可以排除变量之间的相关性的干扰。

缺点：夸大了变化微小的变量的作用。受协方差矩阵不稳定的影响，马氏距离并不总是能顺利计算出。