

作者	生姜 DrGinger
脚本	生姜 DrGinger
视频	崔崔 CuiCui
开源学习资源	https://github.com/Visualize-ML
平台	https://www.youtube.com/@DrGinger_Jiang https://space.bilibili.com/3546865719052873 https://space.bilibili.com/513194466

13.4 距离度量



本节你将掌握的核心技能：

- ▶ 马氏距离：衡量样本点到均值的距离，考虑了特征的方差与协方差。
- ▶ 平面马氏距离，等高线为旋转椭圆。
- ▶ 马氏距离单位：标准差。
- ▶ 马氏距离几何变换：平移 → 旋转 → 缩放。
- ▶ 标准化欧氏距离：消除变量的尺度和单位影响，不考虑相关性。
- ▶ 平面标准化欧氏距离，等高线为正椭圆。
- ▶ 平面欧氏距离，等高线为正圆。

本节从二次型角度来观察几个常见的距离度量，请大家特别注意马氏距离和其他距离的联系和区别。前文中大家经常看到数据椭圆，本节的马氏距离将解释这些旋转椭圆背后的数学原理。

马氏距离

如图 1 所示， A 、 B 、 C 、 D 在以数据质心为圆心的同一个正圆上。很明显， A 、 B 、 C 、 D 距离数据质心 μ 有相同的直线距离（欧氏距离）。

但是，从“数据云”分布的紧密程度来看， C 、 D 距离数据质心 μ 更“近”； A 、 B 则离 μ 更“远”。

马氏距离 (Mahalanobis distance, Mahal distance) 正是考虑了这种分布的距离度量。马氏距离的定义为

$$d_{\text{Mahal}}(\mathbf{x}, \mu) = \sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)} \quad (1)$$

其中， \mathbf{x} 代表样本列向量； μ 是数据质心向量。 Σ 为协方差矩阵。



大家是否发现 (1) 也出现在多元高斯分布的概率密度函数中。

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

⚠ 注意，协方差矩阵 Σ 必须可逆。

马氏距离衡量样本点 x 偏离总体均值 μ 的程度，考虑了每个变量的离散程度以及变量之间的线性相关性。

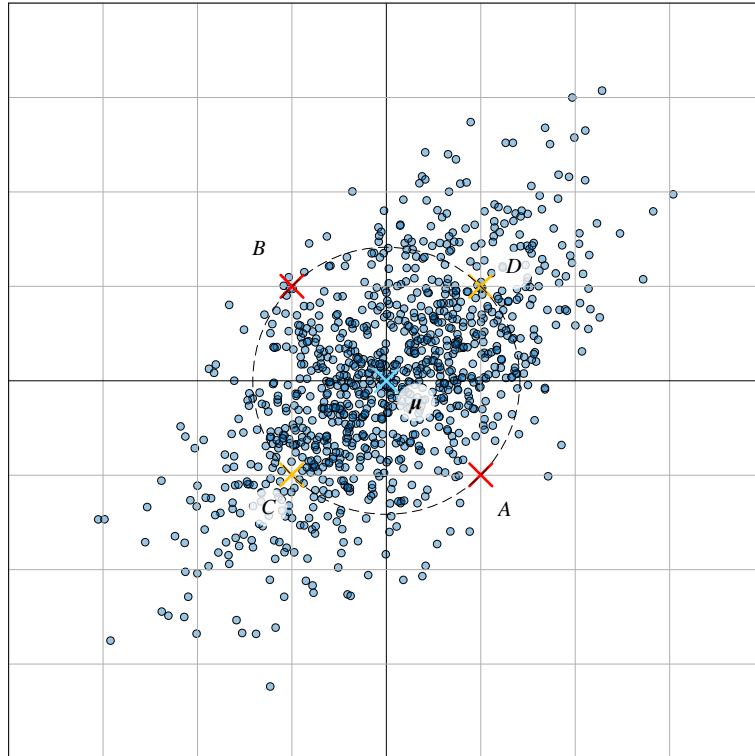


图 1. 马氏距离视角下的距离远近

马氏距离的平方为

$$d_{\text{Mahal}}^2(x, \mu) = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (2)$$

(2) 本质上就是一个二次型。

举个例子

给定如下协方差矩阵 Σ

$$\Sigma = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix} \quad (3)$$

这样协方差矩阵的逆为

$$\Sigma^{-1} = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} \quad (4)$$

假设数据的质心 μ 为

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (5)$$

图 1 中，A 点对应的列向量为

$$\mathbf{x}_A = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (6)$$

计算 A 点和质心 $\boldsymbol{\mu}$ 的马氏距离

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T @ \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} @ \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = 4 \quad (7)$$

开根号得到 A 点和质心 $\boldsymbol{\mu}$ 的马氏距离为 2。

? 请大家计算点 B, [-1; 1], 距离 $\boldsymbol{\mu}$ 的马氏距离。

图 1 中，C 点对应的列向量为

$$\mathbf{x}_C = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad (8)$$

计算 C 点和质心 $\boldsymbol{\mu}$ 的马氏距离

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \left(\begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)^T @ \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} @ \left(\begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = \frac{4}{3} \quad (9)$$

开根号得到 C 点和质心 $\boldsymbol{\mu}$ 的马氏距离约为 1.155。

? 请大家计算点 D, [1; 1], 距离 $\boldsymbol{\mu}$ 的马氏距离。

马氏距离等高线：旋转椭圆

(3) 中协方差矩阵 $\boldsymbol{\Sigma}$ 对应的马氏距离等高线如图 2 所示。我们看到的是一组同心旋转椭圆。这个椭圆就是大家在本书前文看到的旋转椭圆的来源。图中，点 A、B 落在马氏距离为 2 的等高线上；点 C、D 的马氏距离则略大于 1。

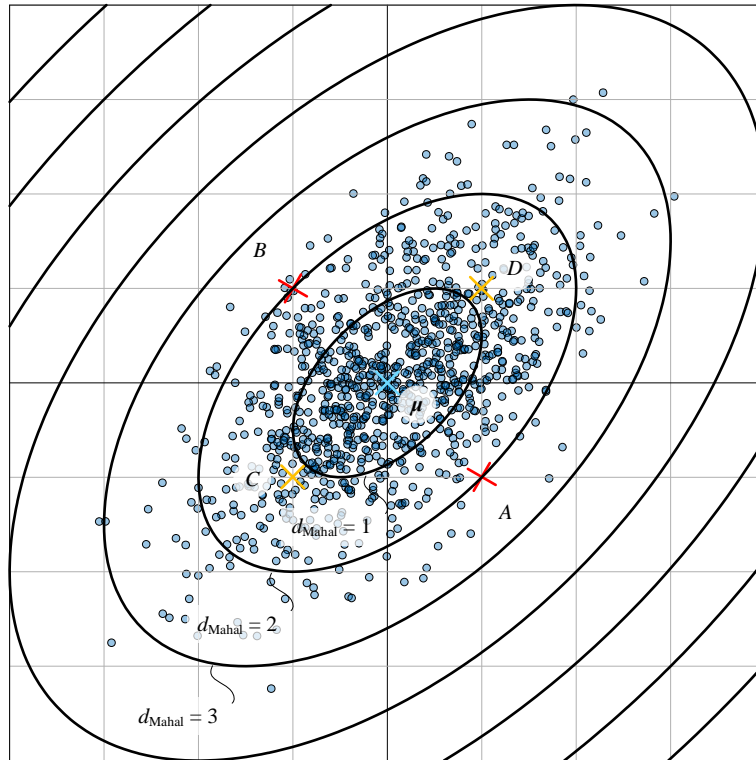


图 2. 马氏距离等高线

⚠ 注意，马氏距离的单位是“标准差”。比如，马氏距离计算结果为 2，应该称作 2 个标准差。

特征值分解：缩放 → 旋转 → 平移

Σ 谱分解得到：

$$\Sigma = V\Lambda V^T \quad (10)$$

其中， V 为正交矩阵。

Σ^{-1} 的特征值分解可以写成：

$$\Sigma^{-1} = V\Lambda^{-1}V^T \quad (11)$$

将 (11) 代入 (2) 得到：

$$\begin{aligned} d_{\text{Mahal}}^2(\mathbf{x}, \boldsymbol{\mu}) &= (\mathbf{x} - \boldsymbol{\mu})^T V\Lambda^{-1}V^T (\mathbf{x} - \boldsymbol{\mu}) \\ &= (\mathbf{x} - \boldsymbol{\mu})^T V\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}V^T (\mathbf{x} - \boldsymbol{\mu}) \\ &= \left(\Lambda^{-\frac{1}{2}}V^T (\mathbf{x} - \boldsymbol{\mu}) \right)^T \left(\Lambda^{-\frac{1}{2}}V^T (\mathbf{x} - \boldsymbol{\mu}) \right) \\ &= \left\| \Lambda^{-\frac{1}{2}}V^T (\mathbf{x} - \boldsymbol{\mu}) \right\|^2 \end{aligned} \quad (12)$$

开平方得到

$$d_{\text{Mahal}}(\mathbf{x}, \boldsymbol{\mu}) = \left\| \underset{\substack{\text{Scale} \quad \text{Rotate} \quad \text{Translate}}}{\mathbf{A}^{\frac{-1}{2}} \mathbf{V}^T} (\mathbf{x} - \boldsymbol{\mu}) \right\| \quad (13)$$

令

$$\mathbf{z} = \mathbf{A}^{\frac{-1}{2}} \mathbf{V}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (14)$$

上式代表，列向量 \mathbf{x} 先用 $-\boldsymbol{\mu}$ 完成平移；然后，再用 \mathbf{V}^T 矩阵完成（绕原点）旋转；最后，用 $\mathbf{A}^{\frac{-1}{2}}$ 矩阵完成缩放。整个几何变换过程如图 3 所示。

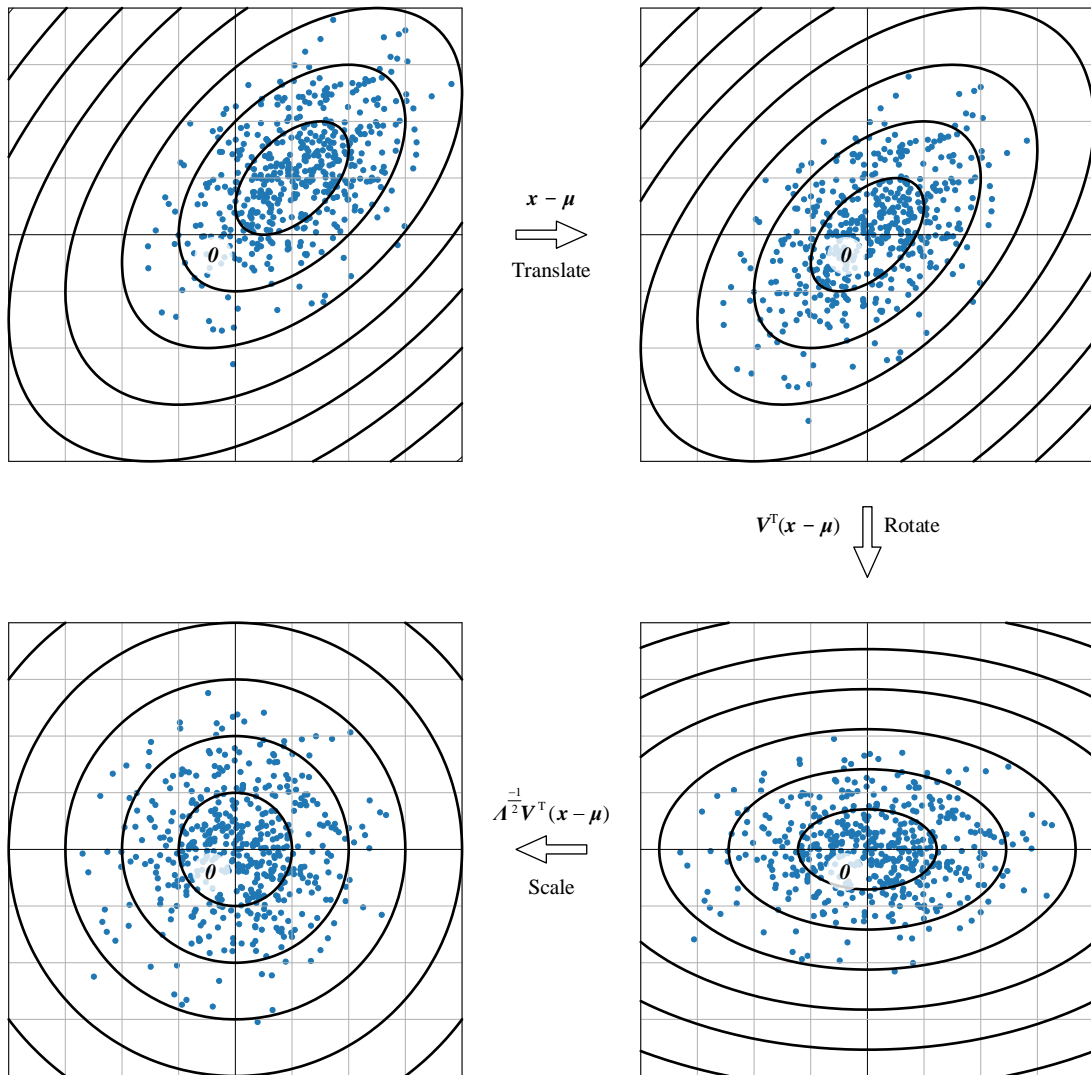


图 3. 马氏距离等高线“平移 → 旋转 → 缩放”

图 3 反过来看，欧氏距离的网格为单位正方形，经过 $\mathbf{A}^{\frac{1}{2}} \mathbf{V}$ “逆向”线性变换后，单位正方形网格变为旋转长方形，具体如图 4 所示。

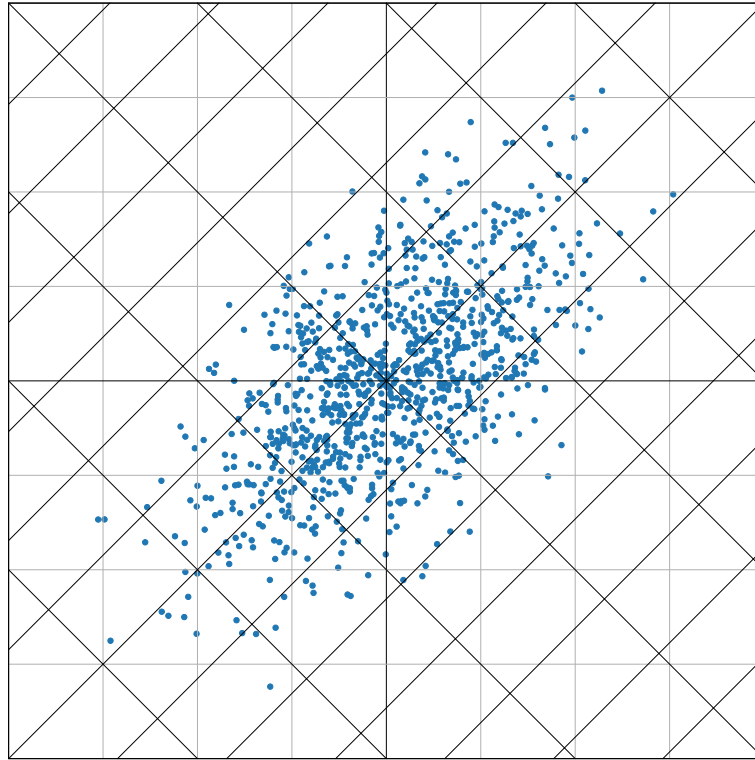


图 4. 马氏距离网格

标准化欧氏距离

把 (1) 协方差矩阵 Σ 替换成如下对角方阵

$$D^2 = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D^2 \end{bmatrix} \quad (15)$$

我们便得到另外一个距离度量——**标准化欧氏距离** (standardized Euclidean distance)，即

$$d(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T D^{-2} (\mathbf{x} - \boldsymbol{\mu})} \quad (16)$$

标准化欧氏距离能够消除不同特征之间的度量单位和尺度差异，从而减少距离计算结果偏差。

不同于欧氏距离，标准化欧氏距离没有考虑数据不同特征之间的相关性。这也体现在 (15) 上，这个矩阵为对角方阵，主对角线元素为各个特征的方差；非主对角线元素为 0。

对于 $D = 2$ ，两特征的情况，样本点 \mathbf{x} 和质心 $\boldsymbol{\mu}$ 的标准化欧氏距离可以写成：

$$d = \sqrt{\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2} \quad (17)$$

如图 5 所示，标准化欧氏距离等高线为正椭圆。

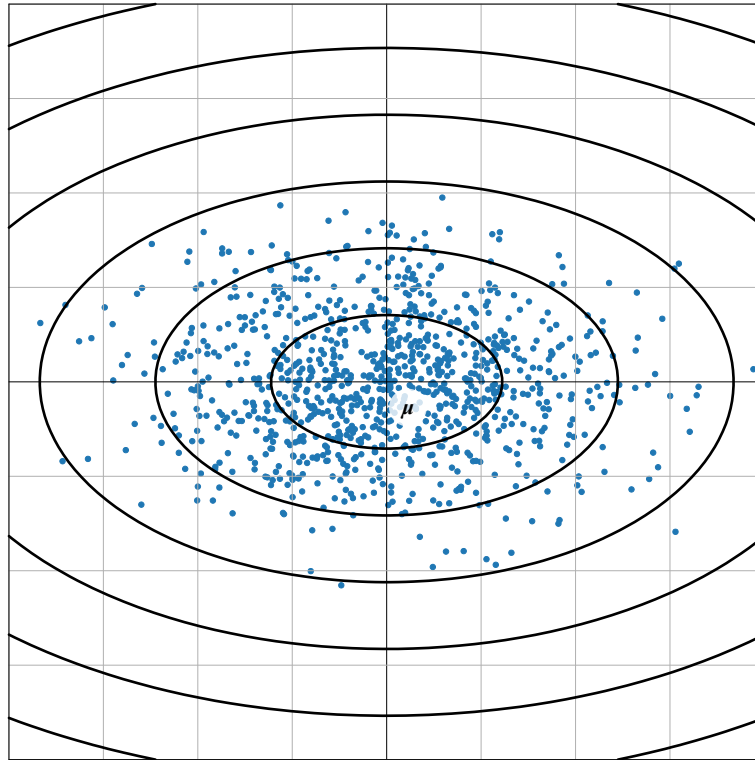


图 5. 标准化欧氏距离等高线

欧氏距离

把 (1) 中协方差矩阵换成单位矩阵，我们便得到欧氏距离

$$d(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{I}^{-1} (\mathbf{x} - \boldsymbol{\mu})} = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{x} - \boldsymbol{\mu})} \quad (18)$$

对于 $D = 2$ ，两特征的情况，样本点 x 和质心 μ 的欧氏距离可以写成：

$$d = \sqrt{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2} \quad (19)$$

如图 6 所示，欧氏距离的等高线为正圆。

欧氏距离是一种直观的几何距离，具有量纲，直接依赖于原始数据的测量单位，对变量间的差异一视同仁，且不考虑变量之间的相关性；因此，当不同变量的量纲或方差差异较大时，欧氏距离可能会失真。

相反，马氏距离是一种无量纲的距离度量方式，通过引入协方差矩阵，综合考虑了各变量之间的线性相关关系，从而消除了冗余信息和变量之间的相关性干扰，使得距离的计算更具统计意义。

马氏距离还能有效识别离群点，其数值越大，说明该样本越偏离总体分布，但其应用前提是变量需近似服从多元正态分布。



本书前文提过，欧氏距离是一种特殊的 L^p 范数。请大家回顾 L^p 范数这个概念，特别是 $p = 1, 2, \infty$ 。

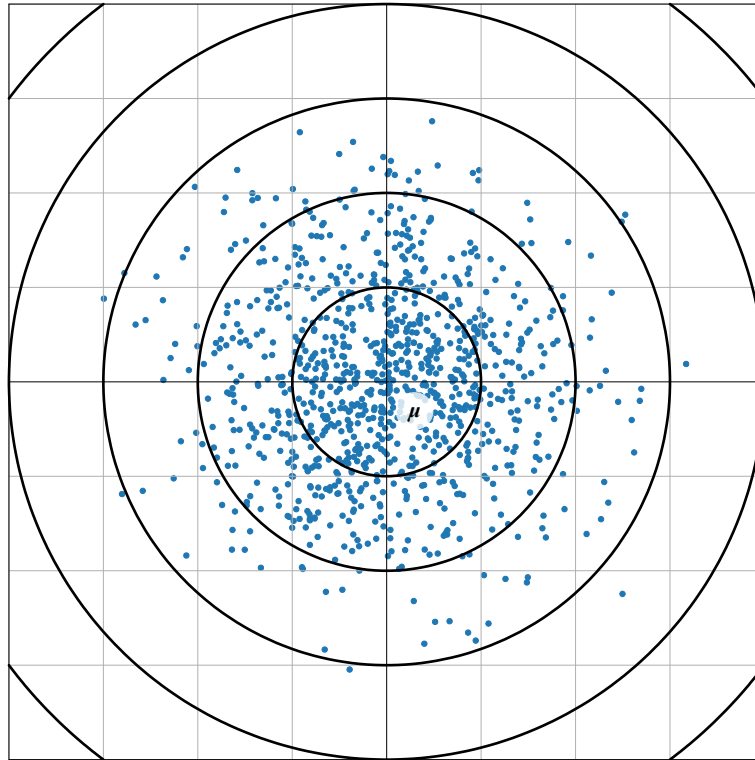


图 6. 欧氏距离等高线



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 请大家学习使用如下函数，计算 (7) 马氏距离

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.mahalanobis.html>

Q2. 请大家学习使用如何函数计算标准化欧氏距离

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.seuclidean.html>

Q3. 请了解曼哈顿距离 (城市街区距离)，并学习使用如下函数

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.cityblock.html>

Q4. 请了解切比雪夫距离，并学习使用如下函数

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.chebyshev.html>

Q5. 请了解闵可夫斯基距离 (闵氏距离)，并学习使用如下函数

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.minkowski.html>

Q6. 请了解余弦距离，并学习使用如下函数

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.cosine.html>