

## 第 3 章：MATLAB 矩阵的操作课后习题

（本文档节选自数学建模清风老师的 MATLAB 入门课程）

b 站观看地址：<https://www.bilibili.com/video/BV1dN4y1Q7Kt>

### 第 3 章小节

本章我们主要介绍了 MATLAB 对于矩阵的相关操作，大家需要有线性代数这门科目的基础。

由于本章内容很多，我为大家准备了一套思维导图，大家可以照着思维导图来回顾本章的知识点。

本章涵盖了 MATLAB 中许多常用的内置函数。为了加深理解，我专门设计了一些题目，分为三个部分来帮助大家巩固和掌握本章的知识（不要使用循环、判断等本章没学的命令）：

- ✧ **基础篇**：帮助大家再次复习基础的内容，题目的答案大部分都能直接在本章的知识点中找到。
- ✧ **提高篇**：在本章课上练习题的基础上进行了一点变形，考验大家对于知识点的熟练程度。这部分的习题不算难，相信大多数的同学都能应对得了。
- ✧ **挑战篇**：这部分的习题有一定的难度，主要训练大家的编程思维。这部分的习题如果你能独立地做出来，那就说明你非常有编程天赋，大家可以尝试挑战一下。

- ✓  1. MATLAB中的向量.pdf
- ✓  2. MATLAB中的矩阵.pdf
- ✓  3. 矩阵的运算.pdf
- ✓  4. 线性代数相关的函数.pdf

作业参考答案请见文档最后一页

### 课后习题

#### 基础篇

#### Q1. 填空题

- (1) MATLAB 中矩阵的元素应包括在\_\_\_\_括号中，矩阵的同行元素之间用\_\_\_\_分隔，行与行之间用\_\_\_\_分隔。
- (2) 命令 `0:3:10` 生成的向量是\_\_\_\_；命令 `10:-2:5` 生成的向量是\_\_\_\_；命令 `10:5` 生成的向量是\_\_\_\_。
- (3) `A` 是一个向量，要计算 `A` 中包含的元素个数，使用的命令是\_\_\_\_；如果 `A` 是一个矩阵，要计算 `A` 中包含的元素个数的命令是\_\_\_\_。
- (4) `A` 是一个矩阵，命令\_\_\_\_可返回 `A` 的行数；命令\_\_\_\_可返回 `A` 的列数。
- (5) 要生成一个包含 50 个元素的等差数列，数列的第一个数是 0，最后一个数是  $2\pi$ ，我们可以使用命令\_\_\_\_。
- (6) 提取向量 `A` 中第三个位置的元素到最后一个位置的元素，可以使用命令\_\_\_\_；提取向量 `A` 第三个位置的元素到倒数第二个位置的元素可以使用命令\_\_\_\_。

- (7) 提取矩阵 A 中奇数行的元素可以使用命令\_\_\_\_\_。
- (8) 删除 A 矩阵的第二列和最后一列可以使用命令\_\_\_\_\_。
- (9) A 矩阵的大小为  $3 \times 3$ ，其线性索引为 4 的元素位于第\_\_\_\_行第\_\_\_\_列，如何使用代码得到这个结果：\_\_\_\_\_。
- (10) 命令\_\_\_\_\_可以将矩阵 A 中的所有元素按照线性索引的方式重构成一个列向量。
- (11) 命令\_\_\_\_\_可以生成一个大小为  $3 \times 4$  的全为 0 的矩阵；\_\_\_\_\_可以生成一个大小为  $4 \times 4$  的全为 1 的矩阵；\_\_\_\_\_可以生成一个大小为  $3 \times 3$  的单位矩阵。
- (12) 命令\_\_\_\_\_用来创建一个 100 行 2 列的随机矩阵，矩阵中的每个元素都在区间 0 和 1 内均匀分布。
- (13) 要模拟投掷有 6 个面的均匀骰子 100 次，那么我们可以使用命令\_\_\_\_\_得到一个长度为 100 的行向量，向量中的每个元素都是随机的取自[1,6]中的整数。
- (14) 命令\_\_\_\_\_用来创建一个 3 行 3 列的随机矩阵，矩阵中的每个元素都随机取样自标准正态分布。
- (15) 如果 A 是一个向量，那么命令 `diag(A)` 的作用是\_\_\_\_\_；如果 A 是矩阵，则作用是\_\_\_\_\_。
- (16) 如果矩阵 A 和 B 的行数相同，命令\_\_\_\_\_可以对 A 和 B 横向拼接；如果矩阵 A 和 B 的列数相同，命令\_\_\_\_\_可以对 A 和 B 纵向拼接。
- (17) 命令 `repelem(1:3,1:3)` 得到的结果是\_\_\_\_\_。
- (18) 将一个向量 a 沿着行方向进行重复的堆叠 5 次，可以使用命令\_\_\_\_\_。
- (19) 矩阵 A 大小是 4 行 5 列，将其形状变成 2 行 10 列可以使用命令\_\_\_\_\_。
- (20) 命令\_\_\_\_\_可以将矩阵 A 进行左右翻转，命令\_\_\_\_\_可以将 A 上下翻转。
- (21) 要将矩阵 A 顺时针（注意不是逆时针）旋转 90 度可以使用命令\_\_\_\_\_。
- (22) 将向量 v 按照从大到小的顺序进行降序排列可以使用命令\_\_\_\_\_。
- (23) 计算 A 矩阵每一行的最大值的命令为\_\_\_\_\_。
- (24) 计算累积和、乘积、差分的函数分别是\_\_\_\_\_。
- (25) A 是一个方阵， $A^3$  的作用是\_\_\_\_\_， $A.^3$  的作用是\_\_\_\_\_。
- (26) MATLAB 中关系运算符有六个，分别是\_\_\_\_\_。
- (27) 运算优先级非常重要，我们可以通过\_\_\_\_\_来改变运算的先后顺序。
- (28) 要返回数组中出现的唯一值，我们可以使用函数\_\_\_\_\_。
- (29) 判断一个数组的元素是否在另一个数组内，我们可以使用函数\_\_\_\_\_。
- (30) 要计算方阵 A 的特征值和特征向量，命令为\_\_\_\_\_。
- (31) 交集、并集、差集和对称差集四个函数分别为\_\_\_\_\_。
- (32) 返回矩阵的上三角部分和下三角部分的函数分别为\_\_\_\_\_。
- (33) 将一个普通数组转换成逻辑值数组的函数是\_\_\_\_\_。
- (34) 判断数组中的元素是否为不定值 NaN 的函数是\_\_\_\_\_。
- (35) 计算方阵 A 的逆矩阵的命令为\_\_\_\_\_。
- (36) 对 A 矩阵的所有元素求倒数的命令为\_\_\_\_\_。
- (37) 命令 `0 == 0 == 0` 返回的结果为\_\_\_\_\_；命令 `-1 < 0 < 1` 返回的结果为\_\_\_\_\_。
- (38) `xor(3,4)` 返回的结果为\_\_\_\_\_。
- (39) 计算矩阵的行列式、秩和迹的函数分别为\_\_\_\_\_。
- (40) 计算向量 x 的 1-范数、2-范数和 p-范数的命令为\_\_\_\_\_。

Q2.请完成下面的一系列任务：

- (1) 生成一个 6 行 3 列的随机矩阵 A，矩阵中每个元素都是位于区间[50,100]之间的随机整数，下面我们假设矩阵 A 的每一行代表一名学生，这六名同学的三门科目的成绩对应着三列；
- (2) 将第一门科目六名同学的成绩赋值给变量 B，对 B 进行降序排列，排序后的向量记为 BB，并返回 BB 中的每个元素在 B 中的索引向量 ind；
- (3) 计算 A 中所有成绩的自然对数；
- (4) 请基于第二科的成绩按升序对这六名同学进行排序，当第二科成绩相同时，请保持其在矩阵中出现的先后顺序；
- (5) 计算六名同学的总分以及每门科目的平均分；
- (6) 计算每门科目的最低分，并返回是第几位同学取得的分数；
- (7) 计算每名同学在哪门科目上分数最高、并返回最高分；
- (8) 假设这三门科目的权重分别是 0.2, 0.5, 0.3，请计算每名同学的加权平均分（即三门科目的成绩分别乘以对应的权重，然后再求和）；
- (9) 判断这三门科目是否有同学不及格（低于 60 分），如果有任意一名同学在某个科目中不及格就返回逻辑值 1，否则返回逻辑值 0；
- (10) 统计每门科目不及格的人数；
- (11) 将 A 中低于 60 分的成绩全部改成 60 分；
- (12) 在上一问的基础上重新计算六名同学的总分，并找出总分最高的同学。

Q3.简答题

- (1) 如何将一个向量倒序？例如原来的向量是[1 5 8 4]，倒序后是[4 8 5 1]。原向量的方向（行向量和列向量）会影响你的结果吗？（本题答案有至少两种方法）
- (2) 请说明&&和&的用法和主要区别。
- (3) MATLAB 中浮点数的计算可能存在误差，应该怎样判断两个浮点数相等？
- (4) 简述 MATLAB 中\*和.\*的作用，并介绍加法运算中所支持的几种兼容模式。
- (5) 简述 sum、mean、median 这三个函数使用方法的共同特点，它们和 max 函数在沿行和列方向计算时的命令有什么区别？

### 提高篇

Q1.如果 x 是一个常数, A 是一个矩阵, 请给出至少两种方法来判断 A 矩阵中是否存在 x? 存在则返回逻辑值 1, 否则返回逻辑值 0.

Q2.生成包含 3 个元素的行向量 A, A 中每个元素都是位于 1-10 之间的随机整数; 再生成包含 10 个的元素的行向量 B, B 中的每个元素也都是位于 1-10 之间的随机整数。

(1) A 中的这三个元素哪些出现在 B 中? 请返回下表中的结果列。例如:

A	B	结果	解释
[3 10 6]	[2 6 6 7 10 7 9 4 4 1]	2, 3	A 中的第二个元素和第三个元素出现在了 B 中
[5 4 6]	[7 10 8 8 2 2 9 3 5 9]	1	A 中只有第一个元素出现在了 B 中
[9 8 7]	[1 6 5 4 4 10 3 1 10 2]	[]	A 中没有元素出现在了 B 中

(2) A 中每个元素是否都包含在 B 中? 请返回下表中的结果列。例如:

A	B	结果	解释
[2 7 6]	[2 6 6 7 10 7 9 4 4 1]	逻辑值 1	A 中的所有元素都在 B 中找到
[3 5 6]	[1 8 8 4 2 7 9 3 5 9]	逻辑值 0	A 中的 6 在 B 中不能找到

Q3.如何创建一个包含  $n$  个元素的等比数列,其第一项为  $a$ ,最后一项为  $b$ 。例如  $a=2, b=1024, n=10$  时,创建的等比数列为: [2 4 8 16 32 64 128 256 512 1024].

Q4.命令 `rand(1)`可生成一个位于 0 和 1 之间均匀分布的随机数,那么  $10*\text{rand}(1)$ 生成的随机数所在的范围是\_\_和\_\_之间;  $2+3*\text{rand}(1)$ 生成的随机数所在的范围是\_\_和\_\_之间。考虑一般情况,命令\_\_\_\_\_可以创建一个  $m$  行  $n$  列的随机矩阵,矩阵中的每个元素都在区间  $[a, b]$  上( $a < b$ )均匀分布。

Q5.本题需要用到概率论中正态分布的一个性质:假设  $X \sim N(u, \sigma^2)$ , 且  $a, b$  是两个实数,那么  $aX + b \sim N(au + b, (a\sigma)^2)$ , MATLAB 中的\_\_\_\_\_函数可生成标准正态分布的随机数,其均值为 0 方差为 1。利用上面的性质,命令\_\_\_\_\_可以创建一个  $m$  行  $n$  列的随机矩阵,矩阵中的每个元素都服从均值为  $p$ ,方差为  $q^2$  的正态分布。

Q6.A 是一个包含至少两个元素的向量,请你判断 A 中是否所有元素都互不相同。例如:  $A = [6\ 5\ 0\ 3\ 6\ 2]$  时返回逻辑值 0,因为里面包含了重复的元素 6;  $A = [1\ 2\ 0\ 4\ -2\ 3\ 7]$  时返回逻辑值 1,因为所有的元素都互不相同。(你能给出两种甚至更多种解答方法吗?)

Q7.找到向量  $x$  中最大值对应的位置索引,如果有多个最大值,将它们的索引全部返回。例如  $x = [2\ 4\ 8\ 1\ 4\ 8\ 3]$ ,那么你需要返回[3, 6],因为第 3 个位置和第 6 个位置都是最大值 8。

Q8.给出一个实矩阵 A,请判断 A 是否为对称矩阵。(实矩阵是指矩阵中的每一个元素都是实数,不含复数;若 A 和 A 的转置相同,则 A 为对称矩阵)

Q9.给定两个同型方阵 A 和 B,请判断 A 和 B 是否互为对方的逆矩阵。(提示:如果  $A*B$  或者  $B*A$  的结果是单位矩阵,则 A 和 B 互逆)

Q10.对于任意一个实对称矩阵 A,判断 A 是否为正定矩阵。(提示:如果实对称矩阵 A 的所有特征值都大于 0,那么这个实对称矩阵是正定矩阵)

Q11.怎样将一个方阵 A 的主对角线元素重新赋值为 0。例如:

原来的 A:				现在的 A:			
3	3	3	8	0	3	3	8
4	8	7	7	4	0	7	7
2	1	9	8	2	1	0	8
5	3	7	9	5	3	7	0

Q12.下面是斐波那契数列的通项公式，请使用该通项公式计算  $n=1,2,\dots,10$  的前 10 项。

$$a_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1+\sqrt{5}}{2} \right)^n - \left( \frac{1-\sqrt{5}}{2} \right)^n \right]$$

Q13.拉马努金是印度历史上最著名的数学家之一，他没有接受过正规的数学教育，但有着令人惊异的数学天赋，似乎可以感知到大量数字关系背后的规律。请你验证他发现的下面两个等式是否成立（考虑浮点数计算的误差）。

$$(1) \sqrt[3]{\sqrt[3]{28}-3} = \frac{\sqrt[3]{98}-\sqrt[3]{28}-1}{3}$$

$$(2) -\sqrt[3]{\cos\left(\frac{\pi}{9}\right)} + \sqrt[3]{\cos\left(\frac{2\pi}{9}\right)} + \sqrt[3]{\cos\left(\frac{4\pi}{9}\right)} = \sqrt[3]{\frac{3\sqrt[3]{9}-6}{2}}$$

Q14.下面是拉马努金发现的计算圆周率的公式，请你计算等式右侧级数前三项的和，并将等式右侧取倒数来计算一个近似圆周率。(提示：factorial 函数也可用来计算阶乘)

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{99^2} \sum_{k=0}^{\infty} \frac{(4k)! 26390k + 1103}{(k!)^4 396^{4k}}$$

Q15.清风老师每年年初都会在银行存 1000 元私房钱，银行每年支付 2% 的利息。

- (1) 请分别计算第 1 年、第 2 年和第 3 年的年末清风老师银行账户的余额。
- (2) 根据上一小问的规律，尝试使用一行代码计算清风老师前 10 年每一年年末的银行账户余额。

Q16.在机器学习算法中，我们经常会将一些数据进行特征缩放(Feature Scaling)来加快算法的收敛速度。下面我们介绍特征缩放中使用最多的两个方法：

**方法 1: Min-Max Scaling:**  $x^* = \frac{x-x_{\min}}{x_{\max}-x_{\min}}$ ，其中  $x_{\min}$  和  $x_{\max}$  分别是这一组数据的最小值和最大值。可以看到，该公式对原始数据进行线性变换，使结果映射到  $[0, 1]$  的范围，来实现对原始数据的等比缩放。因此，该方法中文常被翻译为“线性函数归一化”。

**方法 2: Z-score Normalization:**  $x^* = \frac{x-u}{\sigma}$ ，这里的  $u$  和  $\sigma$  分别是这一组数据的均值和标准差（通常为样本标准差），该方法常被翻译为“Z-score 标准化”。

以方法 1 给大家举例，假设原来的数据为  $[3 \ 5 \ 6 \ 2 \ 1]$ ，最小值为 1，最大值为 6，因此将数据中的每个数都减去最小值 1，然后再除以 5(由  $6-1$  计算得到)，就能够得到特征缩放的结果： $[0.4 \ 0.8 \ 1 \ 0.2 \ 0]$ 。

请大家解决下面两个问题：

- (1) 生成包含 30 个元素的随机列向量  $a$  ( $a$  中每个元素都是 0 到 100 间均匀分布的随机数)，分别使用上面两种方法对  $a$  进行特征缩放；
- (2) 生成一个 30 行 3 列的随机矩阵  $A$  ( $A$  中第一列每个元素都是 0 到 100 间均匀分布的随机数，第二列和第三列每个元素都是均值为 50，标准差为 20 的正态分布随机数)，分别使用上面两种方法对  $A$  的每一列进行特征缩放，得到的结果是一个和  $A$  大小相同的矩阵。



## 挑战篇

Q1.A 是一个矩阵, A(:) 可以按照线性索引的顺序 (沿着行方向依次遍历各列元素) 将 A 中所有元素重构成一个向量, 请问怎样沿着列方向依次遍历各行将 A 中的所有元素重构成一个向量。例如,  $A = \begin{bmatrix} 1 & 4 & 2 \\ 3 & 6 & 8 \end{bmatrix}$ , 需要输出的向量为 [1 4 2 3 6 8]。(输出的结果为行向量和列向量都可以)

Q2.利用 MATLAB 模拟随机丢一枚骰子 N 次, 骰子有均匀的六个面。

- (1) 取  $N = 6000$ , 将这 6000 次的结果保存到向量 A 中
- (2) 统计 A 中出现 6 点的次数, 记为 n
- (3) 利用公式  $p = \frac{n}{N}$  计算 6 点出现的频率, 并计算频率与实际概率之间的偏差:  

$$\text{bias} = \left| p - \frac{1}{6} \right|$$
- (4) 请百度“伯努利大数定律”这个概念, 这是概率论与数理统计中的知识点。如何利用本题来直观的说明伯努利大数定律的成立。

Q3.清风班上有 20 名同学, 这 20 名同学的编号分别是 1、2、…、20。假设现在清风老师要随机请班上同学去吃饭、唱歌或者去看电影, 已知每个同学被抽中去吃饭的概率是 30%, 被抽中去唱歌的概率为 20%, 剩下 50% 的概率是去看电影。请帮助清风老师随机抽取这些同学, 并分别返回吃饭、唱歌和看电影的同学的编号。

Q4.接上一题的题干, 请学过概率论与数理统计的同学接着做本题。将上一题中抽取同学去吃饭作为一个随机事件, 记随机变量 X 为每次抽出来去吃饭的同学人数, 显然: 随机变量 X 的期望值为 6 人, 即用 20 乘以 0.3 计算得到。请大家验证辛钦大数定律的成立。提示: 将上述抽取同学去吃饭这个随机事件重复 N 次 (N 取得尽量大一点, 例如 1000 次), 并求出这 N 次独立事件抽出来的去吃饭的同学人数的平均值, 计算该平均值和真实的期望的偏差有多大, 增加 N 会出现怎样的情况。

Q5. (这道题非常重要!) MATLAB 中有一个非常有用的函数: randperm 函数, 它能够将一个数字序列进行随机打乱。它有两种常见的用法:

用法 1: randperm(n) 可以将向量 1:n 中元素的顺序随机打乱, 生成一个长度仍为 n 的新向量, 所有可能出现的情况共 n! 种 (全排列)。例如, 当你运行 randperm(4) 时, 你可能得到 [1 4 3 2], 也可能得到 [3 2 4 1]。

用法 2: randperm(n,k) 表示从打乱的 1:n 序列中随机的选择 k 个数出来, 显然这 k 个数都不相同, 且 k 要小于等于 n。例如, 当你运行 randperm(10,3) 时, 你可能得到 [5 3 10], 也可能得到 [6 5 8]。(randperm 函数在后续章节也会经常用到)

请回答下面的问题 (做不出来的同学可以看本书的最后一页, 有讲解视频):

- (1) 根据上面的介绍, 请你在 MATLAB 中测试 randperm 函数的功能。特别地, 如果 n 是负数或者小数会出现怎样的情况? 如果 k 大于 n 会出现怎样的情况?
- (2) 假设一个商品推销员要去 10 个不同的城市推销商品, 该推销员随机选择一个城市出发, 依次经过其他所有的城市后, 回到出发的城市 (中途经过的城市不重复), 为了方便, 这 10 个城市就用数字 1 至 10 表示。请你为该推销员随机的生成一条路线 (例如 2 10 8 9 1 5 7 6 3 4 2)。

- (3) 使用代码模拟下列场景：假设你是一名数学老师，你正在给同学们讲不定积分的计算。这时候你的 PPT 上出现了你备课时准备的 4 道练习题，你需要随机抽取 4 名幸运同学到黑板上进行计算。已知你的班上共有 50 名同学，他们的学号分别是 2023001 至 2023050，你在 MATLAB 中运行了你写的这个程序，这四名同学的学号在 MATLAB 中被随机地抽取出来。
- (4) 假设某公司在年会上设置了抽奖环节。主办方准备了一个抽奖用的不透明盒子，盒子内有 10 张奖券，其面值分别为[1 2 5 10 20 50 100 200 500 1000]，每名员工从中随机地抽取 3 张，将这 3 张奖券的面值相加就是他能获得的现金奖励。请设计一个程序，模拟清风老师在该抽奖环节中抽取一次能获得多少钱。
- (5) 一副扑克牌有 54 张，其中大王和小王各一张，A,2,3,4,5,6,7,8,9,10,J,Q,K 各有 4 张。假设我们不考虑桃杏梅方这四种花色，请你设计一个随机的发牌程序，为地主发 20 张牌，两个农民各发 17 张牌。为了方便，A,2,3,4,5,6,7,8,9,10,J,Q,K 分别用数字 1 至 13 代替，小王用 14 代替，大王用 15 代替。进一步地，请你判断地主的牌是否有炸弹（有炸弹是指手上有双王或者有四张相同的牌例如 4 张 3）？

Q6.最近短视频上有一个有趣的街头抽奖游戏，规则如下：摆摊的店家准备了 24 个大小相同的玻璃球，其中红黄蓝各 8 个装进一个不透明的袋子里。玩家从袋子中随机的抓出 12 个球，然后计算每种颜色球的个数，颜色数量多的球放在前面。比如 5 个红色 4 个蓝色 3 个黄色，这样就属于 543。玩家抓出的 12 个球的颜色分布情况一定在下表所示的 13 种情况中。假设参与这个游戏是免费的，如果抽中了相应的情况，店家需要向玩家支付表中第二行所对应的金额。注意，除了 543 这一种情况玩家要赔给店家 30 元外，其余的 12 种情况玩家都是赚钱的，如果你运气好抓到了 840 这种情况（例如抽出了 8 个黄球 4 个红球），你可以赚 300 元。请模拟这个游戏，代码应返回你每次玩这个游戏获得的金额。多次运行代码，观察你是否能赚钱，你能从理论上解释原因吗？

情况	840	831	822	750	741	732	660	651	642	633	552	543	444
金额	300	200	200	80	60	15	100	10	5	5	5	-30	5

Q7.层次分析法是数学建模中一个常用的模型，它主要用于解决评价类问题，也可以为评价体系中的指标确定权重。在层次分析法中会用到判断矩阵，判断矩阵具有以下三个特点。特点 1：它是一个方阵，即行数和列数相同；特点 2：它的每个元素只能是 1,2,...,9 或者它们的倒数；特点 3：若它是一个  $n$  阶的方阵，那么对于任意的  $i, j = 1, 2, \dots, n$ ，均满足  $a_{ij} \times a_{ji} = 1$ 。

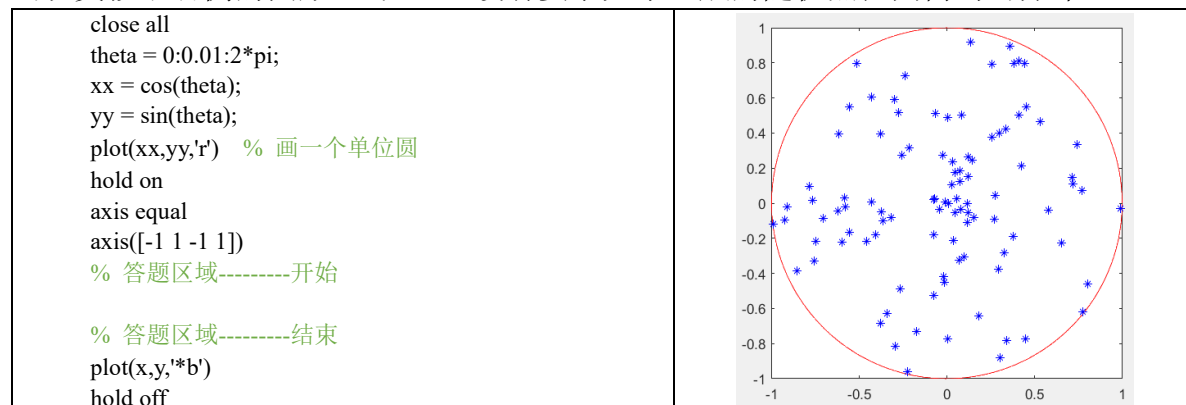
现在给你一个矩阵 A，请判断 A 是否符合层次分析法中判断矩阵的三个特点。

Q8.在上一题的基础上，请写出一段代码能够随机生成一个  $n$  阶（例如  $n=5$ ）的判断矩阵，该判断矩阵要满足上一问中的三个特点。

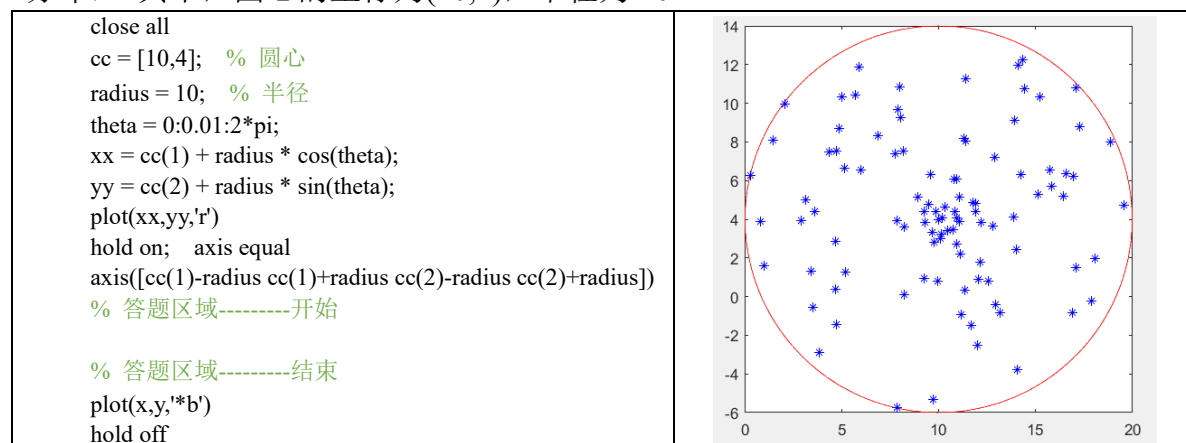
提示：（1）判断矩阵的主对角线元素一定是 1，为什么？（2）本章 3.5 节中我们学会了如何生成一个随机的对称矩阵，一个  $n$  阶的对称矩阵的上三角部分需要生成  $n*(n-1)/2$  个随机数，你需要思考本题中这个  $n$  阶的判断矩阵需要生成多少个随机数？

Q9.生成 100 个随机的点，这些点都位于单位圆内： $x^2 + y^2 \leq 1$ 。注意，要求返回两个长度为 100 的向量  $x$  和  $y$ ， $x$  和  $y$  分别表示这 100 个点的横坐标和纵坐标。

提示：我们可以借助圆的极坐标公式来生成随机数。下方我为大家准备了一段绘图的代码，大家将自己生成随机数的核心代码放入答题区域内，然后运行整段程序，就会出现类似于右侧的图形。（注意：没有要求大家生成的随机点在圆内均匀分布）

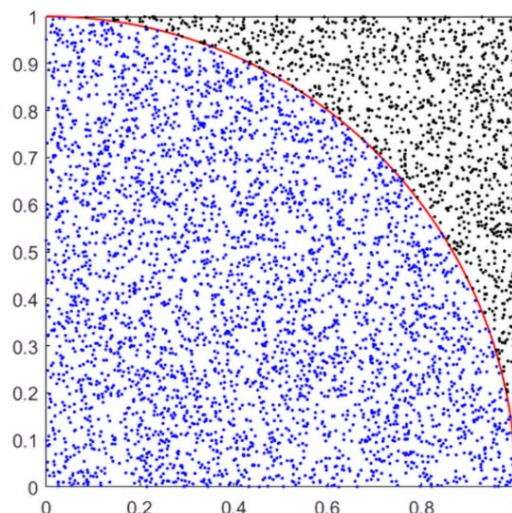


Q10.和上一题题干类似，请生成 100 个随机的点，这些点都位于圆内（不要求在圆内均匀分布）。其中，圆心的坐标为(10,4)，半径为 10。



Q11.蒙特卡罗模拟是一种以概率和统计理论为基础的计算方法，它通过随机数来解决许多计算问题。蒙特卡罗模拟将所求解的问题同一定的概率模型相联系，用计算机实现统计模拟或抽样，以获得问题的近似解。本题将利用蒙特卡罗模拟计算圆周率 $\pi$ 。

如右图所示，在边长为 1 的正方形内随机地生成  $N$  个随机点（图中  $N$  取的是 5000），由于每一个随机点都有横纵坐标，因此我们可以将它们绘制在图中，并用图中的小点表示。如果小点落在半径为 1 的  $1/4$  圆的范围内，其颜色为蓝色；否则颜色为黑色。因此蓝色小点的个数和黑色小点的个数加起来恰好等于  $N$ 。我们不妨记蓝色小点的个数为  $n$ ，那么根据蒙特卡罗模拟的思想， $\frac{1/4 \text{ 圆的面积}}{\text{正方形的面积}} \approx \frac{\text{蓝色小点的个数}}{\text{所有点的个数}}$ ，替换成数学符号则有以下的关系式成立： $\frac{\pi/4}{1} \approx \frac{n}{N}$ ，将其变形后可得： $\pi \approx \frac{4n}{N}$ 。





有同学可能会好奇为什么这个关系式会成立,实际上这里用到了概率论与数理统计中几何概型和大数定律的思想,我们这里不深入探究。

现在请大家使用 MATLAB 生成  $N$  个随机点 ( $N$  取 1000), 每个随机点的横纵坐标都分别在区间  $[0,1]$  上均匀分布, 接下来统计出位于  $1/4$  圆内的随机点的数量  $n$ , 然后套用上面的公式计算出一个近似的圆周率。将你的  $N$  分别变成之前的 10 倍、100 倍和 1000 倍, 你计算出来的圆周率的精度有何变化?

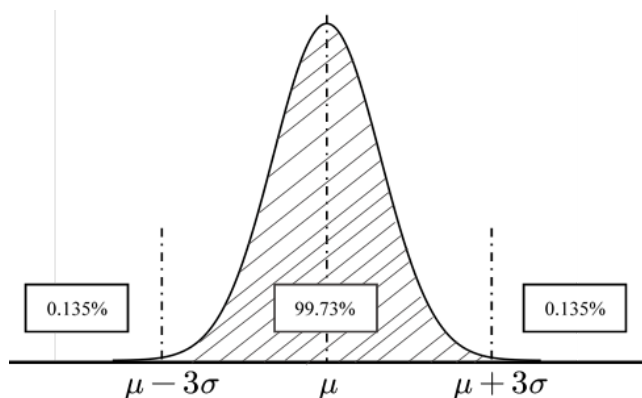
Q12. 学过不定积分的同学应该知道, 有一些不定积分的原函数无法用初等函数表示, 例如  $\int \frac{x}{e^x-1} dx$ , 对于这种类型的不定积分, 如果我们要计算其在特定区间的定积分, 只能通过某些数值分析的方法, 例如梯形法或辛普森积分法。蒙特卡罗模拟也能帮助我们计算定积分, 其基本的思路类似于上一题。现在请大家根据上一题的求解步骤来计算定积分  $\int_0^2 \frac{x}{e^x-1} dx$ , 在此之前, 你可以先在草稿纸上画出被积函数在积分区间上的草图。注意: 一个较为精确的答案是 1.2139, 大家可以自己对答案看看你计算的精度如何。

Q13. 清风订了一份报纸, 送报人可能在早上 6:30 至 7:30 之间把报纸送到他家门口, 而清风出门的时间在早上 7:00 到 8:00 之间 (假设送报人到达的时间和清风出发的时间在对应的时间区间内都是均匀分布的)。请使用蒙特卡罗模拟来估计清风在离开家之前能拿到报纸的概率。注意: 这是概率论中几何概型的一道经典题目, 准确的概率是  $7/8$ 。(思路: 你可以模拟这个送报纸的过程  $N$  次, 其中能拿到报纸的次数为  $n$ , 那么频率  $n/N$  就能看成概率的近似值, 理论依据是伯努利大数定律)

Q14. 异常值或离群值是指在一组数据中与其他数值相比差异较大的一个或几个数值。举个极端一点的例子,  $[4600, 0, 5000, 5200, 4700, 4300, 6000, 5400, 100000, 6200]$  这一组数据中, 我们可以认为 0 和 100000 这两个数就是异常值, 因为剩下的数都集中在 5000 附近。异常值的识别和处理是数据清洗的重要环节, 异常值的存在可能会导致后续的数据分析和建模工作出现偏差。本题将教大家两种识别异常值的方法。

#### 方法一: $3\sigma$ 原则识别异常值

学过概率论的同学应该知道, 正态分布的概率密度函数图像是关于均值点处对称的, 假设总体服从均值为  $\mu$ , 标准差为  $\sigma$  的正态分布, 那么从该总体中随机抽取一个样本点, 该点落在区间  $[\mu - 3\sigma, \mu + 3\sigma]$  上的概率约为 99.73%, 而超出这个范围的可能性仅占不到 0.3%, 是典型的小概率事件, 所以这些超出该范围的数据可以认为是异常值。这就是  $3\sigma$  原则识别异常值的理论基础。

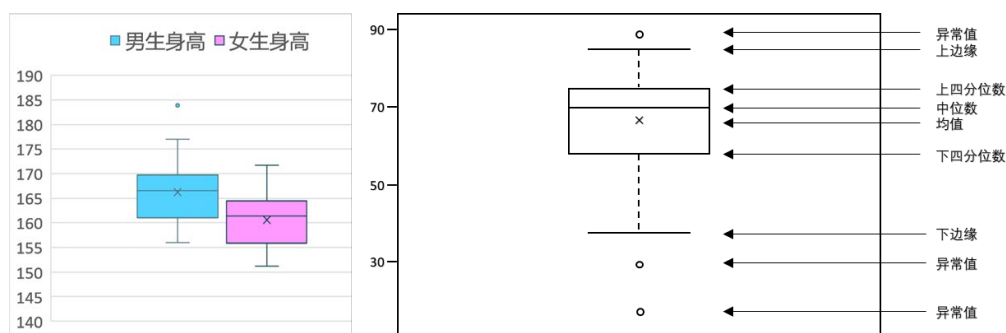


下面总结 $3\sigma$ 原则识别异常值的步骤：（1）计算这组数据的均值 $\mu$ 和标准差 $\sigma$ （注意：我们得到的数据一般是样本数据，因此这里的标准差通常为样本标准差。如果总体标准差是已知的，那么就用总体的标准差）。（2）判断这组数据中的每个值是否都位于 $[\mu - 3\sigma, \mu + 3\sigma]$ 这个区间内，如果不在这个区间内就标记为异常值。

注意事项：使用 $3\sigma$ 原则确定异常值时，样本数据要来自正态分布总体或者近似于正态分布总体，这一点需要根据历史经验或统计检验来进行判断。

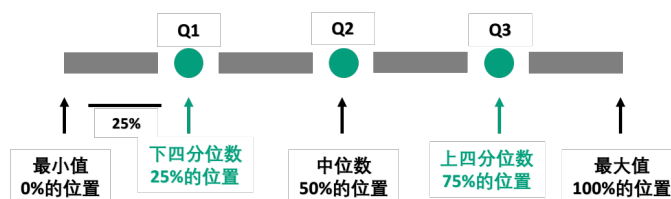
## 方法二：箱线图识别异常值

箱线图又称为盒须图、盒式图或箱形图，是一种用于显示数据分散情况资料的统计图，因形状如箱子而得名。下方左侧给出了一个用来反映某班男女同学身高分布情况的箱线图，右侧是箱线图上各元素所代表的含义。可以看到，箱线图可以反映数据的许多统计信息，例如均值、中位数、上四分位数和下四分位数。另外，箱线图中规定了数据的异常值，因此我们可以借助箱线图来识别数据的异常值，下面我们来介绍箱线图中异常值的定义方法。（注意：箱线图的画法不唯一，下图是一种典型的画法）



首先回顾下中位数的定义：我们将数据按从小到大的顺序排列，在排列后的数据中居于中间位置的数就是中位数，我们用  $Q_2$  表示。

下四分位数则是位于排列后的数据 25% 位置上的数，我们用  $Q_1$  表示；上四分位数则是处在排列后的数据 75% 位置上的数，我们用  $Q_3$  表示。（MATLAB 中可以直接计算出  $Q_1$  和  $Q_3$ ，大家不用担心计算问题，请接着看后面的内容）



然后我们要定义一个叫做四分位距（IQR: interquartile range）的指标，它是上四分位数（ $Q_3$ ，即位于 75%）与下四分位数（ $Q_1$ ，即位于 25%）的距离，因此  $IQR = Q_3 - Q_1$ 。四分位距反映了中间 50% 的数据的离散程度，其数值越小，说明中间的数据越集中；其数值越大，说明中间的数据越分散。

接下来的工作和 $3\sigma$ 原则识别异常值类似，我们需要给出一个合理的区间，位于该区间内的值是正常的数值，而在区间外的值就是异常值。在箱线图中，该区间一般为  $[Q_1 - k \times IQR, Q_3 + k \times IQR]$ ， $k$  是控制区间长度的一个正数，通常  $k$  取为 1.5。因此，我们只需要判断这组数据中的每个值是否都位于  $[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$  这个区间内，如果不在这个区间内就将其标记为异常值。另外，如果我们将  $k$  取为 3，那么在区间  $[Q_1 - 3 \times IQR, Q_3 + 3 \times IQR]$  外的异常值被称为极端异常值。

和 $3\sigma$ 原则相比，箱线图并没有对数据服从的分布作任何限制性要求（ $3\sigma$ 原则要求数据服从正态分布或近似服从正态分布），其判断异常值的标准主要以四分位数和四分位距为基础。在总体分布未知的情况下，使用箱线图识别异常值的结果更加客观。

根据上面的介绍，请完成下面的问题：

- (1) 清风开了一家撸猫店，他统计了最近一个月以来每天进店撸猫的人数（本故事纯属虚构）。请大家用下面这三行代码来生成一个随机的向量  $x$ ，向量  $x$  就表示清风记录的最近一个月以来每天进店的人数。从下一问开始，请你假装不知道这个数据是我们随机生成的，就把它当成真实的数据。

```
x = fix(50+5*randn(30,1));
x(randi(30,2,1)) = randi([50,90],2,1);
x(randi(30,2,1)) = randi([20,50],2,1);
```

- (2) 计算这一个月来，撸猫人数最多的一天和最少的一天各有多少人？平均每天有多少人？标准差是多少？下四分位数、中位数和上四分位数各是多少？（提示：MATLAB 中可以使用函数 `prctile` 来计算分位数，请大家自己搜索它的用法）
- (3) 无论是使用 $3\sigma$ 原则还是箱线图识别异常值，均需要计算正常数值所在的区间，请分别算出这两种方法正常数值所在的区间，哪种方法的区间范围更大？
- (4) 分别用两种方法识别出  $x$  中存在的异常值，并返回哪些天是异常值。你可以多次运行第一小问的代码生成新的  $x$ ，观察哪种方法识别出来的异常值更多？
- (5) 以箱线图识别的异常值为例，分别完成下面两个任务：
- (a) 删除  $x$  中的异常值，将剩下的正常数值保存到向量  $y$  中；
- (b) 计算所有正常数值的平均值，并将  $x$  中的异常值替换成四舍五入后的平均值，将结果保存到向量  $z$  中。

Q15. 在本章 3.3.5 小节介绍 `sort` 函数时，我们留下了一个问题：如果存在同学的成绩相同的情况，那么课上讲的代码将会失效，我们算出来的排名无法区分相同成绩的同学。下表给出了两种不同的排名结果，成绩越高排名越靠前，成绩相同则排名一样，但普通排名的并列排名会占据名次的数字位置，而中国式排名中的并列排名不占用名次。现给定第一行的成绩向量，请大家分别算出第二行和第三行的两种排名。

（提示：这个题目可以使用 `ismember` 函数解决，大家不要想的过于复杂。对于每种排名，参考答案只需要两行代码就能解决）

成绩	20	60	80	50	90	90	60	80	80	70	50	30
普通排名	12	7	3	9	1	1	7	3	3	6	9	11
中国式排名	7	4	2	5	1	1	4	2	2	3	5	6

Q16. 行向量  $A$  中包含至少两个元素且  $A$  中不含元素 0，请判断  $A$  中能否找到两个元素互为相反数，能找到则返回逻辑 1，不能找到则返回逻辑 0。例如  $A = [3 \ 1 \ -1 \ 2]$  返回逻辑 1， $A = [-3 \ 4 \ 1 \ 2 \ 6 \ -5]$  返回逻辑 0。进一步思考，如果向量  $A$  中可能包含 0，你的代码还适用吗？如果不适用请修改你的代码。例如  $A = [3 \ 4 \ -2 \ 0 \ 1]$  返回逻辑 0， $A = [3 \ 4 \ 0 \ -2 \ 0 \ 1]$  和  $A = [-6 \ 3 \ 0 \ -5 \ 3 \ 2 \ 6 \ 1]$  返回逻辑 1。（你能使用两种不同的思路求解吗？）

Q17.请完成下面的一系列任务:

- (1) 随机生成两个长度为 6 的行向量  $s$  和  $v$ ,  $s$  和  $v$  中的元素都是位于区间 $[1,10]$ 之间的随机整数, 例如随机生成的  $s$  和  $v$  分别为 $[2\ 4\ 1\ 7\ 4\ 9]$ 和 $[7\ 8\ 3\ 2\ 4\ 5]$ ;
- (2) 如果  $s$  和  $v$  中相同位置的元素相同, 则将该位置的元素从  $s$  和  $v$  中同时剔除掉; 如果不存在这种情况则不剔除。例如: 上面  $s$  和  $v$  中第五个位置的元素都是 4, 那么将 4 剔除后, 新的  $s$  和  $v$  分别为 $[2\ 4\ 1\ 7\ 9]$ 和 $[7\ 8\ 3\ 2\ 5]$ ;
- (3) 经过上一步后, 记向量  $s$  和  $v$  的长度为  $m$ , 请判断  $s$  和  $v$  中是否存在  $s_i = s_j$  且  $v_i = v_j$ , 这里的  $i, j \leq m$  且  $i \neq j$ 。如果存在, 则从  $s$  和  $v$  中剔除掉重复的值, 仅保留一组即可。在上面的  $s$  和  $v$  的例子中,  $m$  等于 5, 不存在这样的情况。换个例子, 如果  $s$  和  $v$  分别为 $[2\ 4\ 9\ 7\ 9]$ 和 $[7\ 8\ 3\ 2\ 3]$ , 那么  $i$  和  $j$  分别取 3 和 5 时,  $s_i = s_j = 9$  且  $v_i = v_j = 3$ , 我们剔除后, 新的  $s$  和  $v$  分别为 $[2\ 4\ 9\ 7]$ 和 $[7\ 8\ 3\ 2]$ 。
- (4) 经过上一步后, 记向量  $s$  和  $v$  的长度为  $n$ , 请判断  $s$  和  $v$  中是否存在  $s_i = v_j$  且  $v_i = s_j$ , 这里的  $i, j \leq n$  且  $i \neq j$ 。例如  $s$  和  $v$  分别为 $[2\ 4\ 9\ 7]$ 和 $[7\ 8\ 3\ 2]$ 时,  $n$  等于 4, 可以找到  $i=1, j=4$  满足  $s_i = v_j = 2$  且  $v_i = s_j = 7$ 。

Q18.本题带大家了解枚举法和网格搜索法。

枚举法在我们日常生活中使用的频率很高, 它的核心思想就是枚举所有可能来找到正确或者最优的情况。

举个简单的例子: 现在有一个两位数, 十位数是 8, 个位数未知, 但是知道这个数可以被 9 整除, 请回答它的个位数是什么? 显然, 个位上的数只能是  $0, 1, 2, \dots, 9$  中的一个, 因此你可以依次尝试, 这里面只有 81 符合条件。用枚举法解决问题, 通常可以从以下两个方面思考: (1) 找出枚举范围: 我们需要列出问题中所有可能的解。在上面的例子中, 枚举范围就是个位上可能存在的 0 到 9 这 10 个整数。(2) 找出约束条件: 我们需要找出问题的解应该满足的约束条件。在上面的例子中, 约束条件就是这个数能被 9 整除。下面使用 MATLAB 来完成这个过程, 仅需三行代码:

```
x = 0:9; % 枚举范围
tmp = (mod(x+80, 9) == 0); % tmp 是一个长度为 10 的逻辑向量
x(tmp)+80 % 输出符合约束条件的解
```

以上是一个非常简单的使用枚举法的例子, 那么枚举法有什么缺点呢? 最容易想到的缺点就是: 如果枚举范围中的情况非常的多, 那么枚举所有情况就非常耗费时间。例如我们要判断整数 10 到 1 亿里面哪些数是质数, 这个范围就非常的大。枚举法的另一个缺点就是可能会做很多无用功导致效率低下, 例如判断质数这个例子中, 我们没有必要列举 10 到 1 亿中为偶数的情况, 因为它们一定不是质数。另外, 如果枚举范围中有无数种可能, 那么我们就无法使用枚举法来解决这个问题, 例如使用计算机验证哥德巴赫猜想(任一大于 2 的偶数都可写成两个质数之和)是不可能的, 因为这里面有无数个待验证的数。

但有一种情况例外, 即使枚举范围中有无数种可能, 我们仍可以挑选出有限种方案进行求解, 这样能保证最后获得一个近似解。举个例子, 我们要求函数  $y = 2x - e^x$  在区间  $[0, 1]$  上的最大值。如果你学过高数的话, 应该很容易求出最大值在  $x = \ln 2$  处取得。现在假设你没有学过高数, 你应该怎样借助计算机求解这个问题?



如果我们使用枚举法，由于 $x$ 在区间 $[0,1]$ 上有无数种可能的取值，我们不可能全部列举出来，但我们可以在区间 $[0,1]$ 上按一定的步长取有限个点作为 $x$ （例如 $x = 0, 0.01, 0.02, \dots, 0.99, 1$ ），然后依次算出在这些点上 $y$ 的取值，从里面挑选出最大的那个 $y$ 所对应的 $x$ 来作为一个的近似解，这种策略被称为网格搜索法。

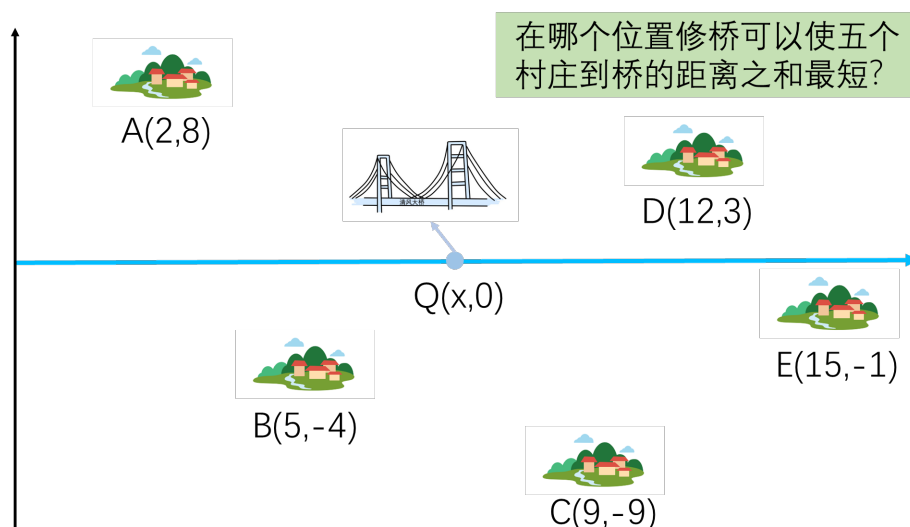
因为这里只有 $x$ 这一个参数，所以搜索起来比较简单。显然，选取的步长越小，找到的解会越精确，例如 $x$ 取为  $0:0.0001:1$  得到的解比  $0:0.01:1$  更精确。但要注意：步长设置的越小，网格搜索所需的时间会越长，因此这里有一个权衡取舍关系：我们需要根据要获得的解的精度和求解的时间综合考虑来设置一个合适的搜索步长。

另外，如果搜索的参数很多，例如我们要求出二元函数 $z = f(x, y)$ 在 $x$ 和 $y$ 都属于 $[0,1]$ 区间时它的最大值，那么我们使用网格搜索时就要对 $x$ 和 $y$ 可能的取值进行排列组合，假设 $x$ 和 $y$ 各有一万种可能的取值，那么我们的搜索次数将会高达一亿次。

如果有更多要搜索的参数，那么这个计算量将会非常的大，搜索次数随着要搜索的参数数量呈指数增长。因此，研究优化理论的学者提出了很多高效的优化算法，这些算法能够在可接受的计算时间下给出待解决的优化问题的一个可行解。做数学建模的同学可能听过启发式算法，常见的启发式算法有模拟退火、粒子群算法、遗传算法等。总之，优化算法的内容非常丰富，本题仅通过介绍枚举法和网格搜索法来引出相关的概念，有兴趣的同学可以在网上查阅相关的资料。

下面请大家完成下面的练习题：

- (1) 使用网格搜索法求出 $y = 2x - e^x$ 在区间 $[0,1]$ 上的最大值，并求出此时的 $x$ 。
- (2) 使用网格搜索法求出 $f(x, y) = x^3 - y^3 + 18x^2 + 12y^2 - 9x - 9y$ 在 $x$ 和 $y$ 都位于区间 $[-1,1]$ 上的最小值，并求出此时 $x$ 和 $y$ 的值。（提示：你可能需要使用 `meshgrid` 函数生成 $xy$ 平面上的二维网格坐标，该函数具体的用法可以查看 MATLAB 帮助文档。本题的参考答案为-2.854，最小值在(0.245,0.394)处取得）
- (3) 假设你知道 $y = 2x - e^x$ 在区间 $[0,1]$ 上的图形是先递增后递减，你将如何改进你的搜索策略来提高搜索效率。（不需要写代码，可以用文字描述这种策略）
- (4) 如下图所示，图上标出了五个村庄的坐标，现在要在横轴上（看成一条河流）选取某一点修建一座桥，请问在哪个位置修桥可以使这五个村庄的距离到桥的距离之和最短，这里的距离请用曼哈顿距离计算。



Q19. 本题考察大家收集资料以及自学的能力。请大家查询相关资料并回答下面几个问题：

- (1) 本章中我们学过如何生成随机数，请大家搜索生成可重复的随机数的方法。
- (2) 如果大家学过概率论与数理统计，那么一定知道泊松分布和指数分布，请大家搜索生成这两个分布的随机数的方法。
- (3) 请大家自学 MATLAB 中排列组合的两个函数：nchoosek 和 perms。

Q20. 本题将带大家学习皮尔逊相关系数和斯皮尔曼相关系数的计算步骤。

(1) 皮尔逊相关系数(Pearson correlation coefficient)

皮尔逊相关系数用于度量两组数据  $X$  和  $Y$  之间的线性相关的程度，它的范围为  $-1$  到  $+1$ ，其中  $0$  代表无线性相关性，负值代表线性负相关，正值代表线性正相关。

在现实生活中，我们收集到的数据可以分为总体数据和样本数据，根据收集的数据类型不同，我们计算皮尔逊相关系数的公式也有所区别。通常我们收集到的数据都是样本数据，因此下面给出样本皮尔逊相关系数的计算公式。

假设我们搜集了两组样本数据  $X: [X_1, X_2, \dots, X_n]$  和  $Y: [Y_1, Y_2, \dots, Y_n]$ ，那么  $X$  和  $Y$  之间的样本皮尔逊相关系数  $r_{XY} = \frac{\text{Cov}(X, Y)}{S_X S_Y}$ ，其中  $\text{Cov}(X, Y)$  表示  $X$  和  $Y$  之间的样本协方差，计算公式  $\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$ ，这里的  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  和  $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$  分别表示  $X$  和  $Y$  的样本均值， $n$  为样本个数； $S_X$  和  $S_Y$  分别表示  $X$  和  $Y$  的样本标准差， $S_X = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$ ， $S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$ 。请你根据公式计算下表给出的  $X$  和  $Y$  的样本皮尔逊相关系数。

X	92	79	78	96	64	88	88	69	78	53	52	77	89
Y	58	29	41	47	19	40	35	42	34	27	17	41	38

(提示：答案是 0.7629，你也可以使用 MATLAB 内置的 corrcoef 函数计算  $X$  和  $Y$  的皮尔逊相关系数，根据公式算出的结果应和内置函数的返回结果相同)

(2) 斯皮尔曼等级相关系数(Spearman's rank correlation coefficient)

斯皮尔曼等级相关系数简称斯皮尔曼相关系数，或称为秩相关系数，它是一种非参数方法，用于衡量两个变量依赖性的强度和方向。与皮尔逊相关系数不同，斯皮尔曼相关系数不要求两个变量是线性相关的，它主要用于衡量两个变量之间单调关系的强度和方向。斯皮尔曼相关系数的范围为  $-1$  到  $+1$ ，如果当  $X$  增加时， $Y$  趋向于增加，则斯皮尔曼相关系数为正；如果当  $X$  增加时， $Y$  趋向于减少，则斯皮尔曼相关系数为负；斯皮尔曼相关系数为  $0$  表明当  $X$  增加时  $Y$  没有任何趋向性。

在介绍斯皮尔曼相关系数的计算公式之前，我们先介绍秩（或称为等级）的概念。一个数的秩就是将这个数所在的那组数据按照从小到大的顺序重新排列，这个数所处的位置下标；若出现相同的数，则将它们的位置下标计算平均值作为它们的秩（这里介绍的秩和矩阵的秩没有任何关系，大家不要弄混了）。

(例 1) 计算 18、23、9、11、3、20 这组数据中各元素的秩。

首先从小到大进行排序：3、9、11、18、20、23；那么，18 所处的位置下标是 4；23 所处的位置下标是 6；9 所处的位置下标是 2；11 所处的位置下标是 3；3 所处的位置下标是 1；20 所处的位置下标是 5；因此，这组数据的秩就是：4、6、2、3、1、5。

（例 2）计算 18、23、9、11、11、20 这组数据中各元素的秩。

首先从小到大进行排序：9、11、11、18、20、23；那么，18 所处的位置下标是 4；23 所处的位置下标是 6；9 所处的位置下标是 1；11 所处的位置有两个，下标分别是 2 和 3，将 2 和 3 求平均值等于 2.5；20 所处的位置下标是 5；因此，这组数据的秩就是：4、6、1、2.5、2.5、5。

介绍完秩的概念后，我们就能计算斯皮尔曼相关系数，它的计算方法有两种：（1）直接套用公式计算；（2）计算 X 和 Y 的秩之间的皮尔逊相关系数，将这个结果当成原始数据 X 和 Y 之间的斯皮尔曼相关系数。

第一种方法是基于以下公式直接计算： $r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$ ，式中的  $d_i$  为  $X_i$  和  $Y_i$  之间的秩差， $n$  为 X 或 Y 中包含的元素个数。

举个例子，假设要计算  $X = [3 \ 8 \ 4 \ 7 \ 2]$  和  $Y = [5 \ 10 \ 8 \ 10 \ 6]$  这两组数据的斯皮尔曼相关系数，那么我们可以先求出 X 和 Y 对应的秩，然后算出秩差和秩差的平方：

X	Y	X 的秩	Y 的秩	秩差	秩差的平方
3	5	2	1	1	1
8	10	5	4.5	0.5	0.25
4	8	3	3	0	0
7	10	4	4.5	-0.5	0.25
2	6	1	2	-1	1

接下来只需要套用公式算出  $r_s = 1 - \frac{6 \times (1+0.25+0+0.25+1)}{5 \times 24} = 0.875$ 。

第二种方法是计算 X 和 Y 的秩之间的皮尔逊相关系数，即计算数据  $[2 \ 5 \ 3 \ 4 \ 1]$  和数据  $[1 \ 4.5 \ 3 \ 4.5 \ 2]$  的皮尔逊相关系数，大家可以自行计算，得到的结果为 0.8721。

可以看出，两种方法得到的结果有细微差别，这是因为原来的 Y 数据中有重复的元素。可以证明的是：如果 X 中的元素各不相同，且 Y 中的元素也各不相同，两种方法得到的结果是相等的。

下面请大家编写代码求解下表中 X 和 Y 的斯皮尔曼相关系数，你需要使用上面介绍的两种方法分别计算。

X	92	79	78	96	64	88	88	69	78	53	52	77	89
Y	58	29	41	47	19	40	35	42	34	27	17	41	38

（提示：计算一组数据的秩需要用到 sort 函数，你需要思考如何处理存在相同数值的情况；当然，如果你绞尽脑汁也没有想出来怎么处理，你可以在 MATLAB 官网搜索 tiedrank 函数的用法。参考答案：第一种方法得到的结果为 0.6401，第二种方法得到的结果为 0.6386。另外，大家可以在 MATLAB 官网搜索 corr 函数，该函数可以用来计算斯皮尔曼相关系数）

**以上就是本章的课后习题，大家一定要认真做，这些题目都是按照本章知识点精心编写的，对大家编程能力的提高非常有帮助。**

**题目答案的讲解视频在 b 站观看（一定要先自己做，然后再对答案）：**

<https://www.bilibili.com/video/BV1bh4y1z7sa>