

第二章：简单分类-察异辨花

从这一章开始，我们就正式进入了机器学习相关模型的介绍。这一章的主题是‘简单分类’，我们之所以选择这一主题作为机器学习模型的最初介绍，是因为在人工智能领域，有超过一半的任务是处理分类问题，比如信用卡违约识别是从消费人群中分别出哪些消费者更可能会违约，再比如较为复杂的人脸识别就是在千千万万个照片中将不同的人脸进行分类。

本章节中我们会重点学习两个模型：

1. 简单的线性分类 (感知器分类)
2. 支持向量机¹²分类 (可以进行非线性分类)

¹² *Support Vector Machines*

按照惯例，本章的学习仍然有三部分组成：课前预习，课堂讲义，和课后练习。在课前预习中，我们将会讨论什么是分类，在什么情况下可以分类；在课堂讲义中，我们会在理解了分类的意涵后，来定义测量和优化，明白了测量和优化之后，我们便可以比较好得掌握简单的线性分类模型和支持向量机模型。

2.4 课前预习 C1

此次课前预习重点强化同学们对分类的理解，课前预习照旧包括两个部分：

- 生活引导题：思考你是如何分类的
- 简单的数学练习题

第一部分：生活引导题

从生活中找一个分类的例子，然后描述下你是如何分类的 (不超过 60 字)，描述必须包含下面三个部分：

- 分类的主体对象是什么？(比如，汽车，花草，等等)
- 分类的依据和标准是什么？
- 分类后如何衡量分类得好坏？

第二部分：简单的数学演练

Question C1-Q1. 下面有一个数组 A:

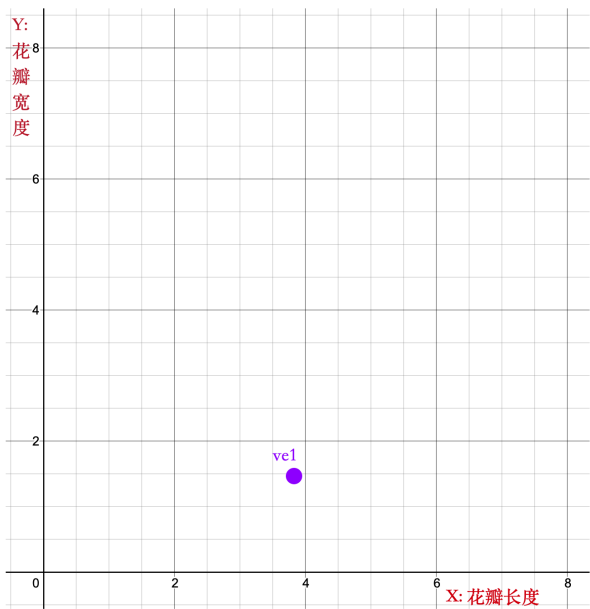
$$A = [1, 3, 2, 5, 4, 7, 9, 11, 10, 16]$$

你是否可以设计一个方程来将数组 A 中的奇数和偶数进行分类。如果有兴趣，你也可以尝试写一个 Python 小程序来完成该任务。

Question C1-Q2. 下面表格是我们从安德森鸢尾花卉数据集 (Anderson's Iris data set) 截取的一部分数据。

属种	代码	花瓣长度	花瓣宽度
山鸢尾	s1	1.4	0.2
山鸢尾	s2	1.7	0.4
变色鸢尾	ve1	3.9	1.4
变色鸢尾	ve2	4.9	1.5
维吉尼亚鸢尾	vig1	6.9	2.3
维吉尼亚鸢尾	vig2	6.1	1.9

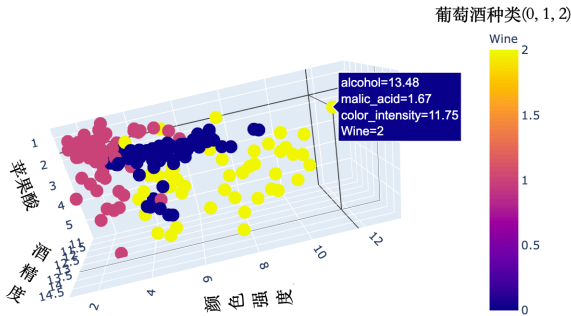
下面的坐标中 X 轴代表花瓣长度，Y 轴代表花瓣宽度，请将以上六个花朵定位到坐标系中去 (用相应代码代表)。



Question C1-Q3. 在生活中，我们经常需要用不同的数据来帮助我们记录一定的信息。比如下面这个数据阵中包含的就是不同葡萄酒的相关信息。

葡萄酒	代码	酒精度	苹果酸	颜色强度 (1-10)
品种 1	0	14.23	1.71	5.64
品种 1	0	13.20	1.78	4.38
品种 2	1	12.17	1.45	2.95
品种 2	1	12.37	1.21	4.60
品种 3	2	13.71	5.65	7.7
品种 3	2	13.17	2.59	9.3

按照同样的道理，我们也可以把这些普通酒的种类在三维图中定位，如果 X 代表酒精度， Y 代表苹果酸， Z 代表颜色强度，那么下图就是不同葡萄酒在这三个数据描述下的定位。



为什么要将数据可视化

我们通过几个预习的题目，是想通过数据的可视化让同学们有对分类有更直观的认识。也就是说，尽管不同种类的物体可能在某些数据维度上有所交叉，但是只有某一个维度有明显的差异，那么表现成**集群**后，其可分类的可能性就会大大提升。另外，需要提醒同学们的是，当对一个物体的描述超过 3 个时，就很难再进行可视化演示，因为我们只能进行三维构图，超出三维的同学们只能借助三维的图片在进一步想象。

Question C1-Q4. 我们有下面两个表格，很明显如果只给你第一个表格的话，你是无法进行分类的。但是增加了数据描述后，我们发现，分类就变得很容易。

车型	动力	座位数	车型	动力	座位数	油耗
A	169	4	A'	169	4	6
A	169	4	A'	169	4	6
B	169	4	B'	169	4	9
B	169	4	B'	169	4	9

现在如果我们把相关的数据统计在向量中表达我们可以发现:

$$\begin{aligned} A - B &= [169, 4] - [169, 4] = [0, 0] \\ A' - B' &= [169, 4, 6] - [169, 4, 9] = [0, 0, -3] \end{aligned}$$

向量和矩阵本身就是一种可视化

通过这个例子我们可以发现, 当我们对一个物体的描述大于 3 时, 虽然不能进行可视化来进行快速分类, 但是我们可以通过向量运算来进行分类, 比如下面这三个向量。

$$\begin{aligned} A &= [0, 1, 9, 17, 100] \\ B &= [0, 2, 9, 17, 100] \\ C &= [0, 1, 9, 10, 56] \end{aligned}$$

Question C1-Q5 解方程. 我们再做几个解方程题目, 求解 x, y , 可以使用计算器。

$$\begin{aligned} 2x + 3y &= 1 \\ x + y &= -1 \end{aligned}$$

求解 x, y, z

$$\begin{aligned} 56x + 8y &= 36 \\ 32x + 4y + z &= 20.5 \\ 48x + 3y + z &= 29.8 \end{aligned}$$

Question C1-Q6, 向量点积运算. 向量的点积是一个比较重要的概念, 所以需要同学们提前了解下, 并且做一下相关计算。这里我们先不做具体的定义介绍, 只是把向量点积的运算公式告诉大家, 然后举例后大家进行练习。
相同长度的向量 A 和 B 的点积等于其同位元素求积后再求和, 比如:

$$\begin{aligned} A &= [3, 5]; \quad B = [2, 1]; \quad A \cdot B = 3 \times 2 + 5 \times 1 = 6 + 5 = 11 \\ A' &= [-2, 1]; \quad B' = [2, -5]; \quad A' \cdot B' = -2 \times 2 + [1 \times (-5)] = -4 - 5 = -9 \end{aligned}$$

现在请同学们计算下列向量的点积下面, 我们来进行几个简单的点积运算, 为了帮助同学们了解点积的几何意义, 我们固定一个向量 w :

$$w = [-1, 2]$$

然后让下面的几个向量依次与 v 进行点积:

$x_1 = [4, 2]$
 $x_2 = [-4, -2]$

$a = [2, -1]$
 $c = [-1, -3]$

$b = [1, 2]$
 $d = [-3, 2]$

结算可以得出以下结果:

$w \cdot x_1 = -1 \times 4 + 2 \times 2 = 0$

$w \cdot a =$

$w \cdot b =$

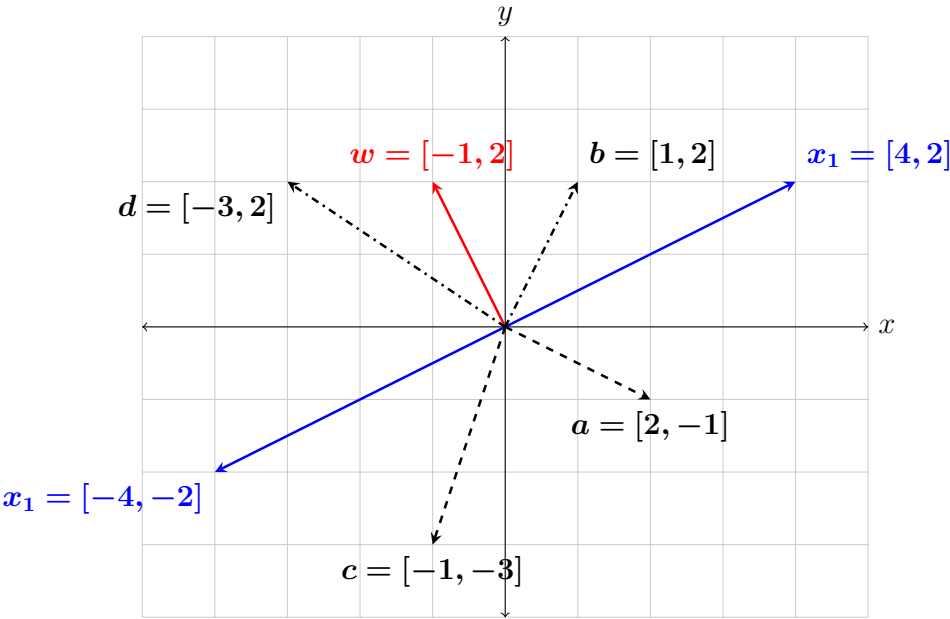
$w \cdot x_2 =$

$w \cdot c =$

$w \cdot d =$

计算结果完成后, 请填下下列表格:

与 w 点积	< 0	$= 0$	> 0
		x_1	
		x_2	



反馈统计: 请扫描右边二维码, 回答对此次课前预习的评估,30 秒就可以搞定。老师会表示非常感谢, 请同学多多配合。老师会根据你们的反馈对每一次 C1 进行提升和改进。填问卷的时候记得喝一杯茶或者咖啡 ♡。

