

- 1、 从市场上采购一批橘子和香蕉，下面列出了一批采购价格（单价为浮动）

橘子重量（单位：斤）	香蕉重量（单位：斤）	采购价格（单位：元）
1	2	8.6
2	3	13.8
4	1	11.4
7	7	36
3	6	24.2
5	4	22.9

求采购 6 斤橘子，5 斤香蕉需要多少钱？

要求：

- (1) 利用线性回归求得回归方程，其中最优化方法使用梯度下降法。
- (2) 不需要写代码，文字列式推导过程即可。
- 2、 利用朴素贝叶斯分类垃圾邮件与非垃圾邮件(需要详细计算过程)。“训练”过程：首先，解析所有邮件，提取每一个词。然后，计算每个词语在正常邮件和垃圾邮件中的出现频率。比如，我们假定"sex"这个词，在 4000 封垃圾邮件中，有 100 封包含这个词，那么它的出现频率就是 2.5%；而在 4000 封正常邮件中，只有 2 封包含这个词，那么出现频率就是 0.05%。现在，我们收到了一封信邮件。在未统计分析之前， 我们假定它是垃圾邮件的概率为 50%、然后，对这封邮件进行解析，发现其中包含了 sex 这个词，请问这封邮件属于垃圾邮件的概率有多高？

- 3、 网球运动受天气影响较大，某同学的若干次决定如下表所示，

请用决策树方法确定到底哪些天气情况是影响其决定的因素？

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

- 4、 给定 8 个数据点, $(1,1)$, $(2,1)$, $(1,3)$, $(2,2)$, $(4,4)$, $(4,3)$, $(5,4)$, $(5,3)$, 现对其使用 k-means 聚类方法进行聚类, 要求聚为 2 类, 请给出其详细算法过程 (文字列式推导过程即可)。
- 5、 一家电子公司生产电子元件, 他们对元件的电阻值进行了大量测试。以下是一组电子元件的电阻值 (以欧姆为单位):
- 100,98,101,99,102,97,103,100,96,104
- 公司质量控制团队希望你使用三西格玛准则来判断是否有任何异常的电子元件, 并说明, 如果有, 哪些元器件有异常。
- 6、 某汽车销售公司希望对他们的客户数据进行分箱分析, 以更好地理解客户的购车偏好。以下是一些客户的汽车购买价格数据 (单位: 万元):

客户编号	购买价格
1	10

2	12
3	15
4	12
5	9
6	22
7	18
8	8
9	17
10	14
11	13
12	30
13	9
14	11
15	14

请设计一个数据分箱方案，并回答以下问题：

- (1) 选择合适的分箱数目，并将客户的购车价格分成不同的箱。
- (2) 计算每个分箱的边界（最小值和最大值）。
- (3) 在每个分箱中，计算客户的数据量。
- (4) 计算每个分箱的平均购车价格。

7、某学校收集了学生的考试成绩数据，如下表所示：

姓名	年龄	性别	班级	语文	数学	英语
赵一凡	21	男	一班	88	90	92
钱一凡	22	男	二班	76	68	88
孙一凡	21	女	三班	80	92	66
李一凡	21	男	二班	82	70	70
周一凡	21	女	三班	86	95	90
武一凡	22	女	二班	72	64	68
郑一凡	21	女	一班	77	75	83
王一凡	21	男	一班	92	87	84

请针对以下要求给出数据变换方案。

(1) 语数英进行数据标准化，

(2) 性别和班级进行数据编码。

8、一家美国房地产公司收集了大量房屋属性数据，包括卧室数量、浴室数量和房屋的市场价值（以千美元为单位）。期望通过卧室数量和浴室数据预测房屋的市场价格。为方便计算，假设只有下面4个数据样本（请忽略合理性）。

房屋 1: [1, 2, 220]

房屋 2: [2, 3, 350]

房屋 3: [3, 4, 500]

房屋 4: [4, 5, 700]

请使用主成分分析（PCA）对上述数据进行降维，要求采用零均值规范化，并给出详细计算过程。

作业要求：

(1) 作业本答题，严禁抄袭！

(2) 截止时间：2025年11月17日