1.
    1. 有无监督、能否增量学习、基于实例还是基于模型
    2. 有监督、无监督、半监督、强化学习
        1. 前五个有监督，后四个无监督
        2. 监督学习：训练样本和标签的集合；强化学习：计算机与环境的互动
    3. 在线学习、离线、批量学习
    4. 实例

2.
    1. 收集数据-输入数据-数据预处理-训练和测试模型-模型的评估
    2. 混淆矩阵

	实际正例	实际负例
预测正例	TP	FP
预测负例	FN	TN

    标准化 $z=\frac{x-\mu}{\sigma},\mu=\frac{1}{N}\sum_{i=1}^{N}(x_i),\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2}$

    归一化 $x_{norm}=\frac{x-x_{min}}{x_{max}-x_{min}}$

        1. 特征幅度较大，让特征归一到同一个范围中。提升模型的收敛速度（加快梯度下降的求解速度）；提升模型的精度（消除量级和量纲的影响）；简化计算

    3. 准确率

    $Accuracy=\frac{TP+TN}{TP+TN+FP+FN}$

    精确率

    $Precision=\frac{TP}{TP+FP}$

    召回率（查全率）

    $Recall=\frac{TP}{TP+FN}$

    F1分数

    $F1=\frac{2\times Precision\times Recall}{Precision+Recall}$

        1. 样本不平衡
        2. 不可以。二者两难全，才会出现F1指数。
        3. 留出法：按固定比例将数据集**静态**地划分为训练集、验证集、测试集

        留一法：每次的测试集都只有一个样本，要进行 m 次训练和预测

        k折交叉验证：将数据集分为训练集和测试集；将训练集分为 k 份；每次使用 k 份中的 1 份作为验证集，其他全部作为训练集。

        4. 可以

3.
    1. 模型中可被学习和调整的参数，通常是通过训练数据来`自动学习`的，以`最小化损失函数或优化目标`。

    `超参数`则是在算法运行之前手动设置的参数，用于`控制模型的行为和性能`。

    2. 训练速度、收敛性、容量和泛化能力等

3. 网格搜索、随机搜索；随机搜索。

---

4. 预测时，选取K个最近的邻居
5. 大，小，大，易，交叉验证法
6. 闵可夫斯基距离：$L_p(x_i,x_j)=(\sum_{l=1}^{n}|x_i^{(l)}-x_j^{(l)}|^p)^{\frac{1}{p}}$

   $p=1$，曼哈顿距离

   $p=2$，欧氏距离

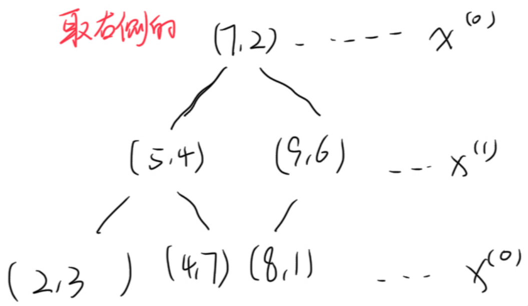
   $p=\infty$，切比雪夫距离。各个坐标距离的最大值。原式会变为$\max\limits_{l}|x_i^{(l)}-x_j^{(l)}|$

7.

$T=\{(2,3),(5,4),(9,6),(4,7),(8,1),(7,2)\}$.

在第0维上，中位数为7 以(7,2)为根结点.

取右侧的



搜索：(3,4.5)

维度0，3<7，进左子树.

维度1，4.5>4，进右子树

dist((3,4.5),(4,7)) = 2.69

返回(5,4)，在维度1，|4-4.5|=0.5<2.6

∴进左子树.

dist = 1.80

返回至(5,4). dist=2.06.

返回至(7,2).在维度0.

有 |7-3|=4>1.80

结束.

最近点(2,3).距离1.80

第2近点(5,4).距离2.06

8. $O(n);O(\log n)$

9.
   1. $P(Y|X)=P(Y)\frac{P(X|Y)}{P(X)}$
   2. $P(Y)$
   3. $P(Y|X)$
   4. $P(X)$
   5. $P(X|Y)$

10. 特征独立性假设

    1.

$$P(X|Y_i) = P(X^{(1)} = x^{(1)}, \ldots, X^{(n)} = x^{(n)} | Y = i)$$
$$= \prod_{j=1}^{n} P(X^{(j)} = x^{(j)} | Y = i)$$

11. $y = \arg\max_i P(Y = i) \prod_j P(X^{(j)} = x^{(j)} | Y = i)$

12. (1) 先验概率

$P(y=1) = \frac{10}{15}$, $P(y=-1) = \frac{5}{15}$

(2) 条件概率

$P(x^{(1)}=1 | y=1) = \frac{2}{10}$ $\quad$ $P(x^{(2)}=S | y=1) = \frac{1}{10}$

$P(x^{(1)}=2 | y=1) = \frac{4}{10}$ $\quad$ $P(x^{(2)}=M | y=1) = \frac{5}{10}$

$P(x^{(1)}=3 | y=1) = \frac{4}{10}$ $\quad$ $P(x^{(2)}=L | y=1) = \frac{4}{10}$

$P(x^{(1)}=1 | y=-1) = \frac{3}{5}$ $\quad$ $P(x^{(2)}=S | y=-1) = \frac{3}{5}$

$P(x^{(1)}=2 | y=-1) = \frac{1}{5}$ $\quad$ $P(x^{(2)}=M | y=-1) = \frac{1}{5}$

$P(x^{(1)}=3 | y=-1) = \frac{1}{5}$ $\quad$ $P(x^{(2)}=L | y=-1) = \frac{1}{5}$

(3) 后验概率

$$P(y=1 | x^{(1)}=2, x^{(2)}=S) = P(y=1) \cdot P(x^{(1)}=2 | y=1) \cdot P(x^{(2)}=S | y=1)$$
$$= \frac{10}{15} \cdot \frac{4}{10} \cdot \frac{1}{10} = 0.02$$

$$P(y=-1 | x^{(1)}=2, x^{(2)}=S) = P(y=-1) \cdot P(x^{(1)}=2 | y=-1) \cdot P(x^{(2)}=S | y=-1)$$
$$= \frac{5}{15} \cdot \frac{1}{5} \cdot \frac{3}{5} = 0.04$$

13. 极大似然估计

14. 拉普拉斯平滑

---

15. 损失函数：**单样本**预测的错误程度；

    代价函数：度量**全部样本集**的平均误差；

    目标函数：代价函数和正则化函数，最终要优化的函数

    1. 0-1损失函数、平方损失函数、绝对损失函数、对数损失函数
    2. 均方误差、均方根误差、平均绝对误差

16. $y = f(x) = w \cdot x + b$

17. $J(w) = \frac{1}{2} \sum_{i=1}^{N} (f(x_i) - y_i)^2 = \frac{1}{2} \sum_{i=1}^{N} (\sum_{k=1}^{n} w_k \cdot x_i^{(k)} + b - y_i)$

    $J(w) = \frac{1}{2} \sum_{i=1}^{N} (\sum_{k=0}^{n} w_k \cdot x_i^{(k)} - y_i)$

    也就是均方误差MSE

18.

    - 初始化参数w

- $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(w)$
- 重复直到收敛。其中，$\frac{\partial}{\partial w_j} J(w) = \sum_{i=1}^{N}(f(x_i) - y_i) \cdot x^{(j)}$

19. $w = (X^T X)^{-1} X^T y$
20. L2：岭回归：λw^2；L1：套索回归：$\lambda |w|$
21. L1;L2;L1

---

22. $f(x) = sign(w \cdot x + b)$,sign表示返回括号内的数的正负号
23. $|w \cdot x_i + b|$；$\frac{1}{||w||}|w \cdot x_i + b|$
24. $-\frac{1}{||w||} y_i(w \cdot x_i + b)$ 这里的符号和y_i相当于上式绝对值的作用
25. 函数间隔$\hat{\gamma}_i = y_i(w \cdot x_i + b)$
26. $L(w,b) = -\sum_{x_i \in M} y_i(w \cdot x_i + b)$其中，M为误分类点的集合。
27. 任意选择一个超平面w_0, b_0

随机选取一个误分类点(x_i, y_i)，对w, b进行更新。

条件终止：损失函数为0

28.

初值 $w_0=(0,0)$, $b_0=0$, $\eta=0.1$

$y_1(w_0 x_1+b_0)=0\le 0$

$w_1=w_0+\eta x_1 y_1=(-0.3,0.3)^T$

$b_1=b_0+\eta y_1=0.1$

$y_3(w_1 x_3+b_1)=-0.7\le 0$

$w_2=w_1+\eta x_3 y_3=(-0.5,-0.1)^T$

$b_2=b_1+\eta y_3=0$

$y_1(w_2 x_1+b_2)=1.2>0$

$y_2(w_2 x_2+b_2)=2.3>0$

$y_3(w_2 x_3+b_2)=1.4>0$

$y_4(w_2 x_4+b_2)=1.7>0$

不存在误分点，迭代结束

得到模型 $f(x)=\mathrm{sign}(-0.5x^{(1)}+(-0.1)x^{(2)})$

29.

初值 $\alpha_i=0,\ i=1,2,3.\quad b=0,\ \eta=0.1$

$$\text{Gram}=\begin{bmatrix} 18 & 21 & 6 & -3 \\ 21 & 29 & -2 & -11 \\ 6 & -2 & 20 & 14 \\ -3 & -11 & 14 & 13 \end{bmatrix}$$

取点 x_1,

$$y_1[(0\cdot y_1\cdot 18+0\cdot y_2\cdot 21+0\cdot y_3\cdot 6+0\cdot y_4\cdot(-3))+0]=0$$

$\alpha_1=0.1,\ b=0.1$

$$y_3[(0.1\times y_1\cdot 6+0\times y_2\times(-2)+0\times y_3\times 20+0\times y_4\times 14)+0.1]=-0.7<$$

$\alpha_3=0.1,\ b=0.$

$$y_1[(0.1\times y_1\times 18+0\times y_2\times 21+0.1\times y_3\times(6)+0\times y_4\times(-3))]=1.2>0$$

$$y_2[(0.1\times y_1\times 21+0\times y_2\times 29+0.1\times y_3\times(-2)+0\times y_4\times(-11))]=2.3>0$$

$$y_3[(0.1\times y_1\times 6+0\times y_2\times(-2)+0.1\times y_3\times 20+0\times y_4\times(-11))]=1.4>0$$

$$y_4[(0.1\times y_1\times(-3)+0+0.1\times y_3\times 14+0)]=1.1>0$$

无误分点

$$w=0.1\times 1\times(-3,3)+0.1\times(-1)\times(2,4)$$

$$=(-0.5,-0.1)$$

$b=0$

30. $z=\sigma(f(x))=\sigma(w\cdot x+b)=\frac{1}{1+e^{-(w\cdot x+b)}}$

31. 极大似然估计

32. $H(X)=-\sum\limits_{i=1}^{n}p(x_i)\log(p(x_i))$

33. 0.5

34. $loss=-\sum\limits_{i=1}^{n}y_i\log(\hat{y}_i)$

35. 交叉熵

36. $P(Y=1|x)=\pi(x), P(Y=0|x)=1-\pi(x)$

$\pi(x)=\frac{1}{1+e^{-(w^Tx+b)}}$

则似然函数为

$$\prod_{i=1}^{N}[\pi(x_i)]^{y_i}[1-\pi(x_i)]^{1-y_i}$$

为了得到使似然函数最大的参数，对上式取对数，得到**对数似然函数**

$$L(w)=\sum_{i=1}^{N}[y_i\log\pi(x_i)+(1-y_i)\log(1-\pi(x_i))]$$

对$L(w)$求极大值，转换为求极小值

$$J(w)=-L(w)=-\sum_{i=1}^{N}[y_i(w\cdot x_i)-\log(1+e^{w\cdot x_i})]$$

求偏导

$$\frac{\partial J(w_j)}{\partial w_j}=-\sum_{i=1}^{N}(y_i-\pi(x_i))\cdot x_i^{(j)}$$

对每个维度j进行更新

$$w_j=w_j-\alpha\frac{\partial J(w_j)}{\partial w_j}$$

37. 随机森林、朴素贝叶斯
38. 一对其余(OvR)或一对全部(OvA)/一对一(OvO)
    1. N $\frac{N(N-1)}{2}$
    2. N $\frac{N(N-1)}{2}$
    3. 置信度最大的类别/被预测最多的类别
    4. OvR相比于OvO：存储开销小、测试时间短、测试时间长

---

39. $H(Y|X)=\sum_{i=1}^{n}p_iH(Y|X=x_i),p_i=P(X=x_i)$，不确定性
40. 计算特征1和特征2对分类的交叉熵，发现特征2的交叉熵为0，表示特征2使得分类的不确定性为0（或者通过信息增益，显然特征2的信息增益会更大），故特征2更有价值。
41. 信息增益：$g(D,A)=H(D)-H(D|A)$

    信息增益比：$g_R(D,A)=\frac{g(D,A)}{H_A(D)}$。其中$H_A(D)$是数据集$D$关于特征$A$的值的熵。（特征熵，看这一列而不再是预测目标列的信息熵）

    基尼系数：$Gini(p)=\sum_{k=1}^{K}p_k(1-p_k)=1-\sum_{k=1}^{K}p_k^2$，$Gini(D,A)=p_1Gini(D_1)+p_2Gini(D_2)$
42. 没算完，累死了，看一下重点吧

ID3: 以 A、B、C 代表性别、车型、尺码

$$H(D) = -\left(\frac{1}{2} \times \log \frac{1}{2} + \left(1-\frac{1}{2}\right) \times \log \frac{1}{2}\right) = 1.0$$

$$H(D|A) = \frac{10}{20} \times \left[-\left(\frac{6}{10} \times \log \frac{6}{10} + \frac{4}{10} \times \log \frac{4}{10}\right)\right]$$

$$+ \frac{10}{20} \times \left[-\left(\frac{6}{10} \times \log \frac{6}{10} + \frac{4}{10} \times \log \frac{4}{10}\right)\right]$$

$$= 0.970$$

$$H(D|B) = \frac{4}{20} \times \left[-\left(\frac{1}{4} \times \log \frac{1}{4} + \frac{3}{4} \times \log \frac{3}{4}\right)\right] \quad 0.811$$

$$+ \frac{8}{20} \times \left[-\left(\frac{8}{8} \times \log \frac{8}{8} + 0\right)\right] \quad 0$$

$$+ \frac{8}{20} \times \left[-\left(\frac{1}{8} \times \log \frac{1}{8} + \frac{7}{8} \times \log \frac{7}{8}\right)\right] \quad 0.543$$

$$= 0.379$$

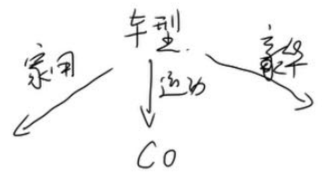
$$H(D|C) = \frac{5}{20} \times \left[-\left(\frac{3}{5} \times \log \frac{3}{5} + \frac{2}{5} \times \log \frac{2}{5}\right)\right] \quad 0.970$$

$$+ \frac{7}{20} \times \left[-\left(\frac{3}{7} \times \log \frac{3}{7} + \frac{4}{7} \times \log \frac{4}{7}\right)\right] \quad 0.985$$

$$+ \frac{4}{20} \times \left[-\left(\frac{2}{4} \times \log \frac{2}{4} + \frac{2}{4} \times \log \frac{2}{4}\right)\right] \quad 1.0$$

$$+ \frac{4}{20} \times \left[-\left(\frac{2}{4} \times \log \frac{2}{4} + \frac{2}{4} \times \log \frac{2}{4}\right)\right] \quad 1.0$$

$$= 0.987$$



$g(D, A) = 0.03$

$g(D, B) = 0.621$

$g(D, C) = 0.013$

选B做根结点.

车型=豪华作为D_1，家用=D_2.

$H(D_1)=0.543$

$H(D_1|A)=\frac{1}{8}\times 0+\frac{7}{8}\times\left[-\left(\frac{1}{7}\log\frac{1}{7}+\frac{6}{7}\times\log\frac{6}{7}\right)\right]=0.518$

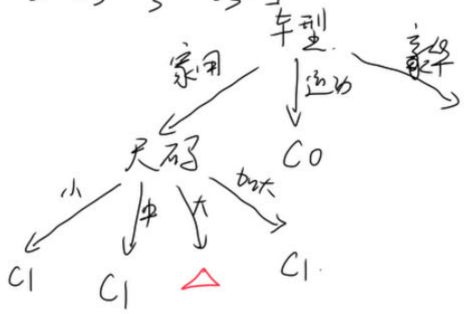
$H(D_1|C)=\frac{2}{8}\times 0+\frac{3}{8}\times 0+\frac{3}{8}\times\left[-\left(\frac{1}{3}\log\frac{1}{3}+\frac{2}{3}\times\log\frac{2}{3}\right)\right]$

$+\frac{2}{8}\times 0=0.344$

$g(D_1,A)=0.025$

$g(D_2,A)=0.199$

选C.



⋮

信息增益比:

$H_A(D)=-\left(\frac{10}{15}\log\frac{10}{15}+\frac{5}{15}\log\frac{5}{15}\right)=0.918$

$g_R(D,A)=\frac{0.03}{0.918}$

$H_B(D)=-\left(\frac{4}{15}\log\frac{4}{15}+\frac{8}{15}\log\frac{8}{15}+\frac{3}{15}\log\frac{3}{15}\right)$

$g_R(D,B)=\frac{\quad}{0.621}$

⋮

gini.

$Gini(D,A=\text{'男'})=\frac{10}{20}\times\left(2\times\frac{6}{10}\times\left(1-\frac{6}{10}\right)\right)+\frac{10}{20}\times\left(2\times\frac{4}{10}\times\left(1-\frac{4}{10}\right)\right)$

$Gini(D,B=\text{'家用'})=\frac{4}{20}\times\left(2\times\frac{1}{4}\times\frac{3}{4}\right)+\frac{16}{20}\times\left(2\times\frac{9}{16}\times\frac{7}{10}\right)$

$Gini(D,B=\text{'运动'})=\frac{8}{20}\times\left(2\times\frac{8}{8}\times\frac{0}{8}\right)+\frac{12}{20}\times\left(2\times\frac{2}{12}\times\frac{1}{12}\right)$

$Gini(D,B=\text{'豪华'})=\frac{8}{20}\times\left(2\times\frac{1}{8}\times\frac{7}{8}\right)+\frac{12}{20}\times\left(\frac{9}{12}\times\frac{3}{12}\right)$

⋮

选最小的.

43. 分类标记，所得的分类结果；测试的条件；测试的节点，对数据属性的测试。

44.

算法	支持模型	树结构	特征选择	连续值处理	缺失值处理	剪枝	特征属性多次使用
ID3	分类	多叉树	信息增益	不支持	不支持	不支持	不支持
C4.5	分类	多叉树	信息增益率	支持	支持	支持	不支持

算法	支持模型	树结构	特征选择	连续值处理	缺失值处理	剪枝	特征属性多次使用
CART	分类 回归	二叉树	基尼指数 均方差	支持	支持	支持	支持

45. 离分离超平面最近的点
46. 使支持向量距离分离超平面最远
47. 几何间隔
48.

$$
\max_{w,b} \gamma \\
s.t.\ y_i(\frac{w}{||w||} \cdot x_i + \frac{b}{||w||}) \geq \gamma, i = 1, 2, \cdots, N
$$

目标函数：最大化样本点到分离超平面的距离

约束条件：超平面到每个训练样本点的几何间隔至少为γ

49.

$$
\min_{w,b} \frac{1}{2}||w||^2 \\
s.t. y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \cdots, N
$$

50.

50.

- 训练数据集
  - 正实例点：$\boldsymbol{x_1} = (3,3)^{\mathrm{T}}$, $\boldsymbol{x_2} = (4,3)^{\mathrm{T}}$
  - 负实例点：$\boldsymbol{x_3} = (1,1)^{\mathrm{T}}$

$$
\begin{aligned}
\min_{w,b} \quad & \frac{1}{2}(w_1^2 + w_2^2) \\
\text{s.t.} \quad & 3w_1 + 3w_2 + b \geqslant 1 \\
& 4w_1 + 3w_2 + b \geqslant 1 \\
& -w_1 - w_2 - b \geqslant 1
\end{aligned}
$$

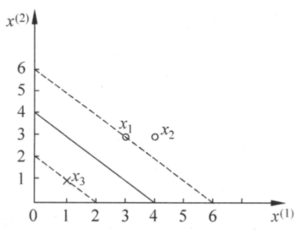


图 7.4 间隔最大分离超平面示例

求得此最优化问题的解 $w_1 = w_2 = \frac{1}{2}$，$b = -2$。于是最大间隔分离超平面为

$$
\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2 = 0
$$

其中，$x_1 = (3,3)^{\mathrm{T}}$ 与 $x_3 = (1,1)^{\mathrm{T}}$ 为支持向量。

51. 软间隔
52. 核函数解决低维度线性不可分问题。
53.

名称	表达式	参数
线性核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$	
多项式核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j)^d$	$d \geq 1$为多项式的次数
高斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\delta^2}\right)$	$\delta > 0$为高斯核的带宽(width)
拉普拉斯核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{\delta}\right)$	$\delta > 0$
Sigmoid核	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\beta \boldsymbol{x}_i^\top \boldsymbol{x}_j + \theta)$	tanh为双曲正切函数，$\beta > 0$, $\theta < 0$

---

54. B
55. D
56. C
57. B
58. A
59. 样本1：0.25；样本2：0.25；样本3：0.5
60. 更低
61. A

---

62. 类别数变小时，平均直径会增加。类别数变大超过某个值时，平均直径不变，这个值正是最优的k值。
63. 平坦、非层次化
64. 兰德指数/轮廓系数；[0,1],[-1,1]；都是越大越好
65. 可以用层次聚类对样本进行聚类，得到k个类时停止。然后从每个类中选取一个与中心距离最近的点。

---

66. 先升后降
67. 第一个新坐标轴选择是原始数据中方差最大的方向，第二个新坐标轴选取是与第一个坐标轴正交的平面中使得方差最大的，第三个轴是与第1,2个轴正交的平面中方差最大的。依次类推，可以得到n个这样的坐标轴。
68. 计算数据矩阵的**协方差矩阵**

    然后得到**协方差矩阵的特征值和特征向量**

    选择特征值最大(即方差最大)的k个特征所对应的特征向量组成的矩阵

    这样就可以将数据矩阵转换到新的空间当中，实现数据特征的降维
69.

1)因为X矩阵的每行已经是零均值，所以不需要去平均值。

2)求协方差矩阵：

$$C = \frac{1}{5}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ -1 & 0 \\ 0 & 0 \\ 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{6}{5} \end{pmatrix}$$

3)求协方差矩阵的特征值与特征向量。

求解后的特征值为：

$\lambda_1 = 2$, $\lambda_2 = \frac{2}{5}$

对应的特征向量为：

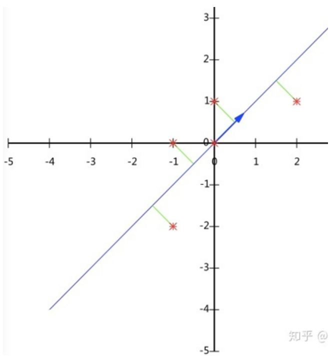
$c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $c_2\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

其中对应的特征向量分别是一个通解， c_1 和 c_2 可以取任意实数。那么标准化后的特征向量为：

$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$, $\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

4)矩阵P为：

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$



5)最后我们用P的第一行乘以数据矩阵X，就得到了降维后的表示：

$$Y = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} -1 & -1 & 0 & 2 & 0 \\ -2 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$