(1) 采用 cart 方法，首先选择根节点：

A. 身材：

	一组	二组
矮	2	2
高	2	3

$gini(矮) = 1-(\frac{1}{2})^2 \cdot 2 = \frac{1}{2}$　$gini(高) = 1-(\frac{2}{5})^2-(\frac{3}{5})^2 = 0.48$

$gini(身材) = \frac{4}{9} \cdot \frac{1}{2} + \frac{5}{9} \cdot 0.48 = 0.489$

B. 发色：

	一组	二组
金	3	1
红	1	0
黑	0	4

$gini(金) = 1-(\frac{3}{4})^2-(\frac{1}{4})^2 = 0.375$

$gini(红) = 0$　$gini(黑) = 0$

$gini(发色) = \frac{4}{9} \cdot 0.375 = 0.167$

C. 年龄：

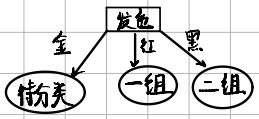
	一组	二组
老	3	2
成	1	1
儿	0	2

$gini(老) = 1-(\frac{3}{5})^2-(\frac{2}{5})^2 = 0.48$

$gini(成) = 1-(\frac{1}{2})^2 \cdot 2 = 0.5$　$gini(儿) = 0$

$gini(年龄) = \frac{5}{9} \cdot 0.48 + \frac{4}{9} \cdot \frac{1}{2} = 0.378$

$gini(发色) < gini(年龄) < gini(身材)$，故以发色为根节点



接着：

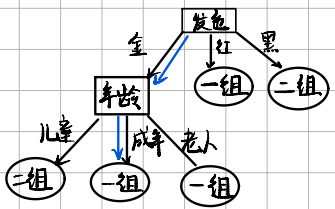
A. 身材：$gini(矮) = 1-(\frac{2}{3})^2-(\frac{1}{3})^2 = 0.44$　$gini(高) = 1-1^2 = 0$

$gini(身材) = \frac{3}{4} \cdot 0.444 = 0.333$

B. 年龄：$gini(老) = 1-1^2 = 0$　$gini(成) = 0$　$gini(儿) = 0$

$\therefore gini(年龄) = 0$

$gini(年龄) < gini(身高)$，第二个分类特征选年龄



(矮、金发、成年) 按上方决策树分类，应为第一组。

$P(\text{身材=矮} \mid \text{组次=第一组}) = 1/2$
$P(\text{身材=高} \mid \text{组次=第一组}) = 1/2$
$P(\text{身材=矮} \mid \text{组次=第二组}) = 2/5$
$P(\text{身材=高} \mid \text{组次=第二组}) = 3/5$
$P(\text{发色=金色} \mid \text{组次=第一组}) = 3/4$
$P(\text{发色=红色} \mid \text{组次=第一组}) = 1/4$
$P(\text{发色=黑色} \mid \text{组次=第一组}) = 0$
$P(\text{发色=金色} \mid \text{组次=第二组}) = 1/5$
$P(\text{发色=红色} \mid \text{组次=第二组}) = 0$
$P(\text{发色=黑色} \mid \text{组次=第二组}) = 4/5$
$P(\text{年龄=老人} \mid \text{组次=第一组}) = 3/4$
$P(\text{年龄=成年} \mid \text{组次=第一组}) = 1/4$
$P(\text{年龄=儿童} \mid \text{组次=第一组}) = 0$
$P(\text{年龄=老人} \mid \text{组次=第二组}) = 2/5$
$P(\text{年龄=成年} \mid \text{组次=第二组}) = 2/5$
$P(\text{年龄=儿童} \mid \text{组次=第二组}) = 1/5$
$P(\text{组次=第一组}) = 4/9$
$P(\text{组次=第二组}) = 5/9$

(1) 建立所得分类器如左图.

(2) $P(X, \text{一组}) = P(X|\text{一组}) \cdot P(\text{一组})$
$= P(\text{矮}, \text{金}, \text{成}|\text{一组}) P(\text{一组})$
$= P(\text{矮}|\text{一组}) P(\text{金}|\text{一组}) P(\text{成}|\text{一组}) \cdot P(\text{一组})$
$= \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{4}{9} = 0.042$

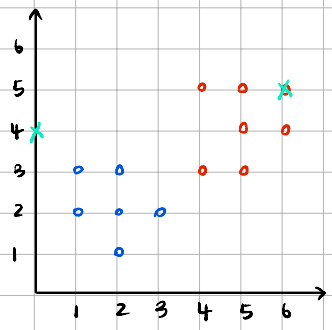
$P(X, \text{二组}) = P(X|\text{二组}) \cdot P(\text{二组})$
$= P(\text{矮}, \text{金}, \text{成}|\text{二组}) P(\text{二组})$
$= P(\text{矮}|\text{二组}) P(\text{金}|\text{二组}) P(\text{成}|\text{二组}) \cdot P(\text{二组})$
$= \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{2}{5} \cdot \frac{5}{9} = 0.009$

$\therefore P(X, \text{一组}) > P(X, \text{二组})$

故(矮,金,成人)分到一组

二. K-means. 以图示完成:

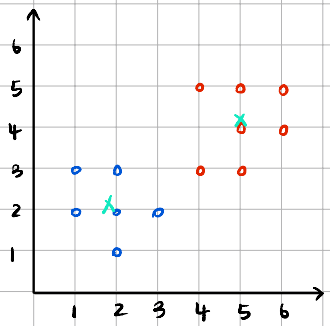
完成初始分类后,数据点之间分组如下图:



计算新的质心为:

质心1: $(1\times2+2\times3+3)/6, (1+2\times3+3\times2)/6) = (\frac{11}{6}, \frac{13}{6})$

同理:质心2为 $(5, \frac{29}{9})$,再次聚类:



再次分组,发现分组不变,故迭代结束,

最后质心为 $(\frac{11}{6}, \frac{13}{6})$ 与 $(5, \frac{29}{9})$ 分组如右上图:

一组:质心 $(\frac{11}{6}, \frac{13}{6})$: (1,3),(1,2),(2,1),(2,2),(2,3),(3,2)

二组:质心 $(5, \frac{29}{9})$: (5,3),(4,3),(4,5),(5,4),(5,5),(6,4),(6,5)