表 1 贷款申请样本数据表

ID	年龄	有工作	有自己的房子	信贷情况	类别
1	青年	否	否	一般	否
2	青年	否	否	好	否
3	青年	是	否	好	是
4	青年	是	是	一般	是
5	青年	否	否	一般	否
6	中年	否	否	一般	否
7	中年	否	否	好	否
8	中年	是	是	好	是
9	中年	否	是	非常好	是
10	中年	否	是	非常好	是
11	老年	否	是	非常好	是
12	老年	否	是	好	是
13	老年	是	否	好	是
14	老年	是	否	非常好	是
15	老年	否	否	一般	否

5. 已知如下表所示的训练数据，试用平方误差损失准则生成一个二叉回归树。

x	1	2	3	4	5	6	7	8	9	10
y	4.50	4.75	4.91	5.34	5.80	7.05	7.90	8.23	8.70	9.00

答:

a. 初始节点：mean= (4.50 + 4.75 + 4.91 + 5.34 + 5.80 + 7.05 + 7.90 + 8.23 + 8.70 + 9.00) / 10 = 6.618

总平方损失：$\sum_i^{10}(y_i\text{-mean})^2$ =27.63 （总损失较大，开始划分节点）

b. 寻找分割点:

- 遍历可能的分割点：1.5, 2.5, ..., 9.5。
- 对每个分割点，根据下面公式分别计算左侧和右侧的平均值和平方损失。

---以 x=5.5 为例：

第一组（x ≤ 5.5）: {4.50, 4.75, 4.91, 5.34, 5.80}

第二组（x > 5.5）: {7.05, 7.90, 8.23, 8.70, 9.00}

计算每组 y 的平均值，然后计算损失：

$$\text{第一组平均值 } \bar{y}_1$$
$$\text{第一组总损失 } L_1 = \sum(y_{1i} - \bar{y}_1)^2$$
$$\text{第二组平均值 } \bar{y}_2$$
$$\text{第二组总损失 } L_2 = \sum(y_{2i} - \bar{y}_2)^2$$
$$\text{分割点总损失 } L_{total} = L_1 + L_2$$

- 计算得到的每个分割点总损失，选择总损失最小的分割点作为划分节点：

---每个分割点的总损失如下：

x=1.5 的总损失：22.648；x=2.5 的总损失：17.702；x=3.5 的总损失：12.193；x=4.5 的总损失：7.379；**x=5.5 的总损失：3.359**；x=6.5 的总损失：5.074；x=7.5 的总损失：10.052；