x=8.5 的总损失：15.178；x=9.5 的总损失：21.328。在这些分割点中，x=5.5 有最小的总损失（3.359），因此它是最佳的分割点。在构建决策树时，第一步将会是以 x=5.5 作为根节点的分割条件。

c. 递归分割:

- 对于 x<=5.5 的左侧数据集，重复步骤 b。
- 对于 x>5.5 的右侧数据集，也重复步骤 b

d. 终止条件:

假设我们的停止条件是每个节点至少需要包含两个数据点，那么当数据不能再分时，递归停止。(根据具体停止划分情况而定)

6. 试析使用“最小训练误差”作为决策树划分选择准则的缺陷。

答：使用"最小训练误差"作为决策树划分选择准则，由于使用的是训练集数据，可能会将训练特征中的一些异常或者偶然作为模型的一部分，导致过度拟合的问题。

7. 试将 4.4.2 节对缺失值的处理机制推广到基尼指数的计算中去。

$$\mathrm{Gini}-\mathrm{index}(D,a)=\rho*\mathrm{Gini}-\mathrm{index}(\overline{D},a)=\rho*\sum_{v=1}^{V}\tilde{r}_v\mathrm{Gini}(D^v)$$

$$\mathrm{Gini}(D^v)=1-\sum_{k=1}^{|\gamma|}\tilde{p}_k^2$$

8. 某公司招聘职员考查身体、业务能力、发展潜力这 3 项。身体分为合格 1、 不合格 0 两级，业务能力和发展潜力分为上 1、中 2、下 3 三级。分类为合格 1、不合格-1 两类。已知 10 个人的数据，如表所示。假设弱分类器为决策树桩。试用 AdaBoost 算法学习一个强分类器。

应聘人员情况数据表

	1	2	3	4	5	6	7	8	9	10
身体	0	0	1	1	1	0	1	1	1	0
业务能力	1	3	2	1	2	1	1	1	3	2
发展潜力	3	1	2	3	3	2	2	1	1	1
分类	-1	-1	-1	-1	-1	-1	1	1	-1	-1

答：编程题，采用 sklearn 的 AdaBoostClassifier 分类器，构建并训练得到强分类器（https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html）

9. 给定如表所示训练数据。假设弱分类器由 x < v 或 x > v 产生，其阈值 v 使该分类器在训练数据集上分类误差率最低。试用 AdaBoost 算法学习一个强分类器。

序号	1	2	3	4	5	6	7	8	9	10
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	-1	-1	-1	1	1	1	-1

答:

计算步骤: