

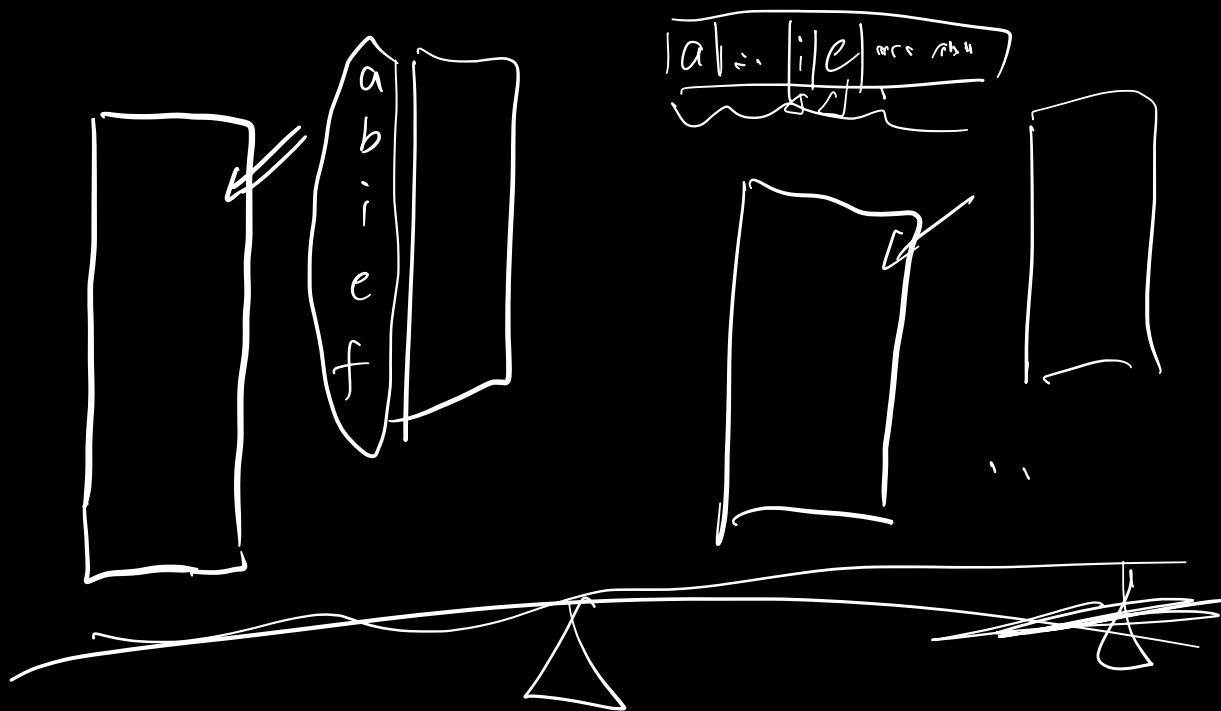
1. 预测报告结果数量

单词属性 $\xrightarrow{\text{影响}}$ 报告分数

2. 预测报告结果分数

3. 按难度分类单词

4. 列出数据集其它属性



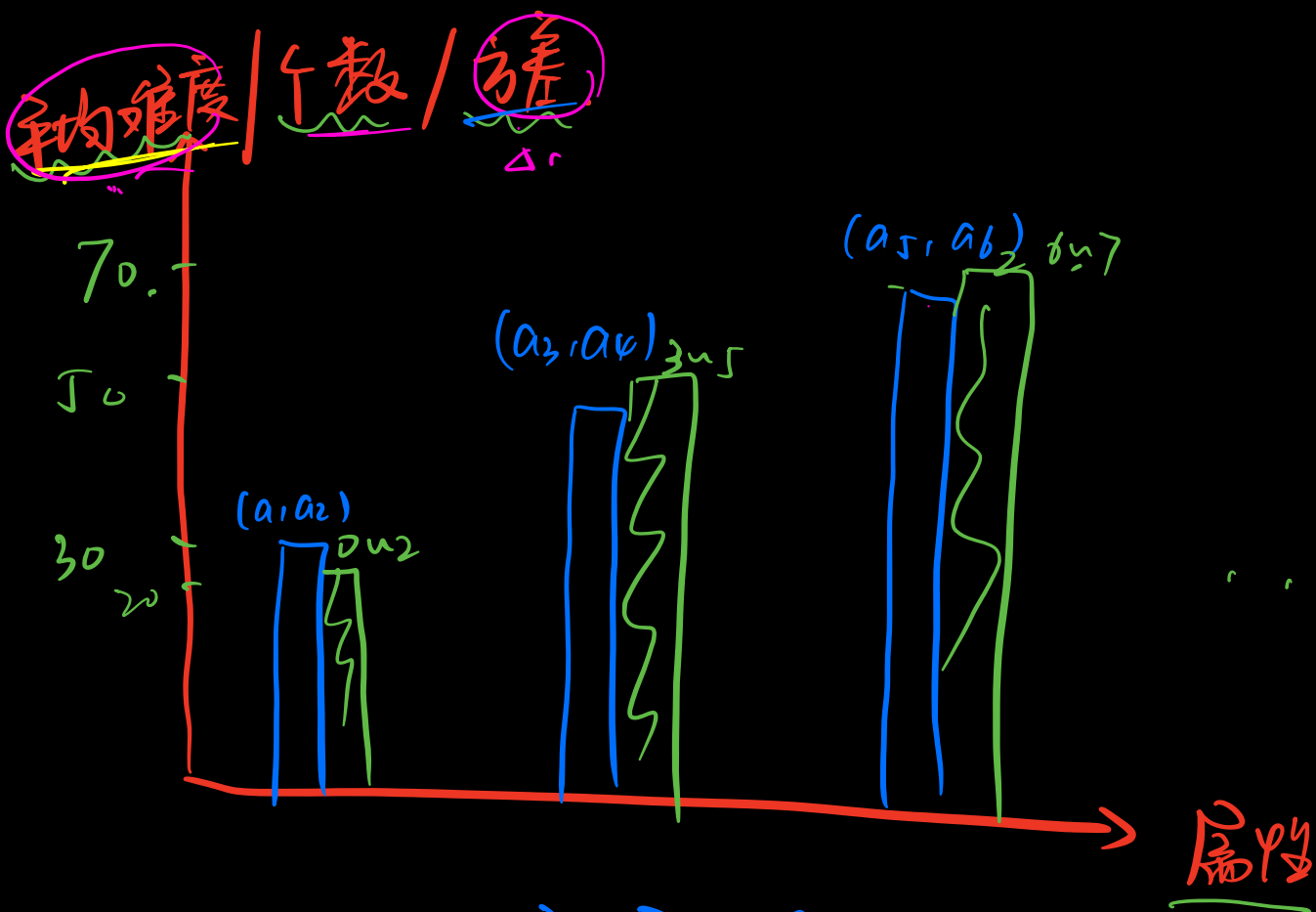
预处理: 难度百分比相加 = 100

剔除异常值

参与人数少则意义不大

属性: 重复字母个数

单词数字化



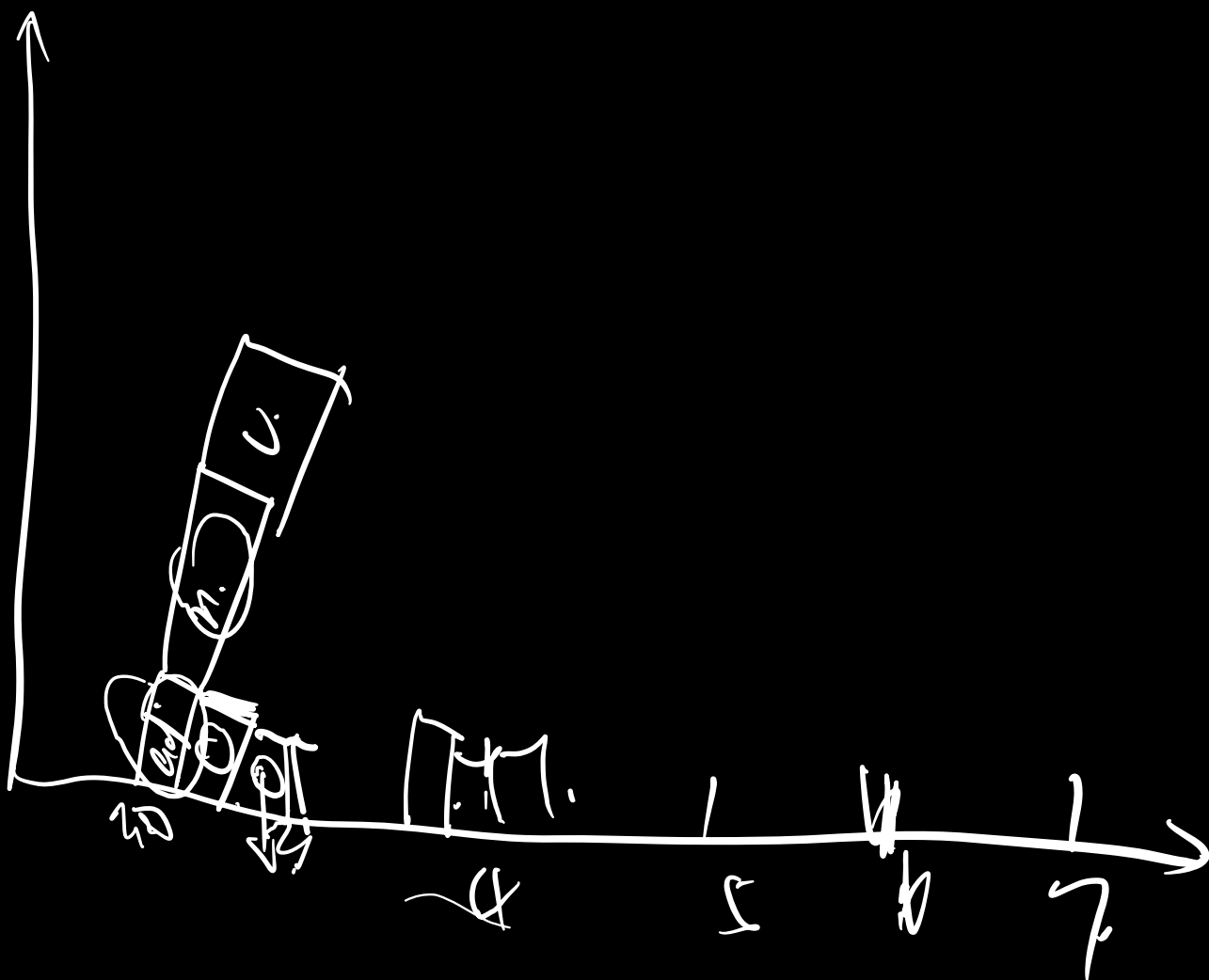
 字母频率加和 (x10)

 相似单词数-7: 8

 词性 n. v. a.

 常见词义

292, 频率和.



频率和: $(0.14289, 0.41605) \Delta = 0.054623$

$(0.14289, 0.197513, (0.252136, (0.306759, (0.361382,$
 $0.197513) 0.252136) 0.306759) 0.361382) 0.41605)$