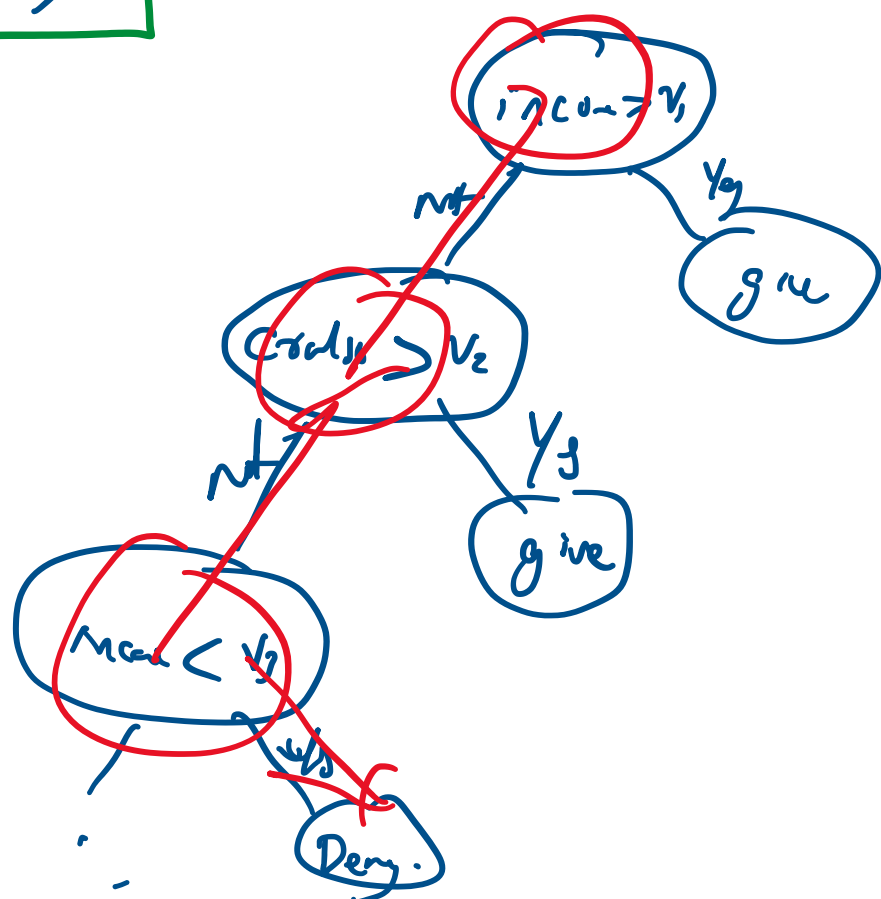
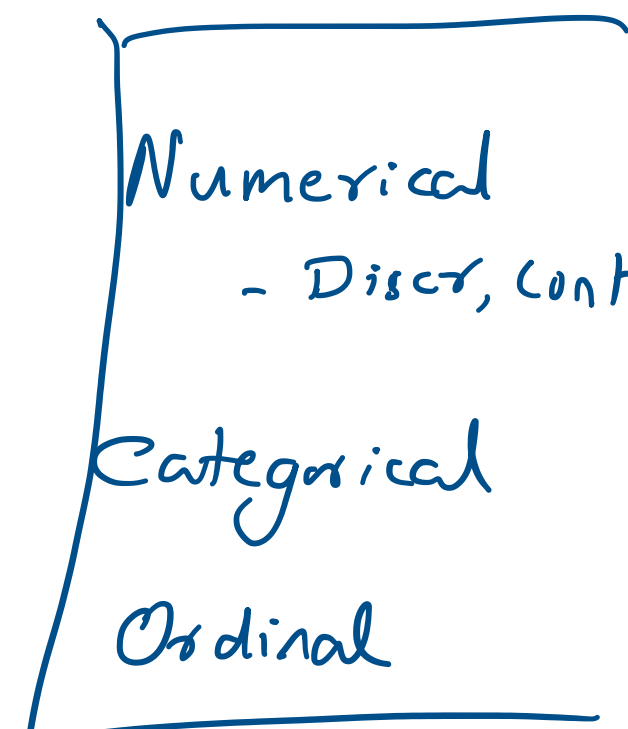


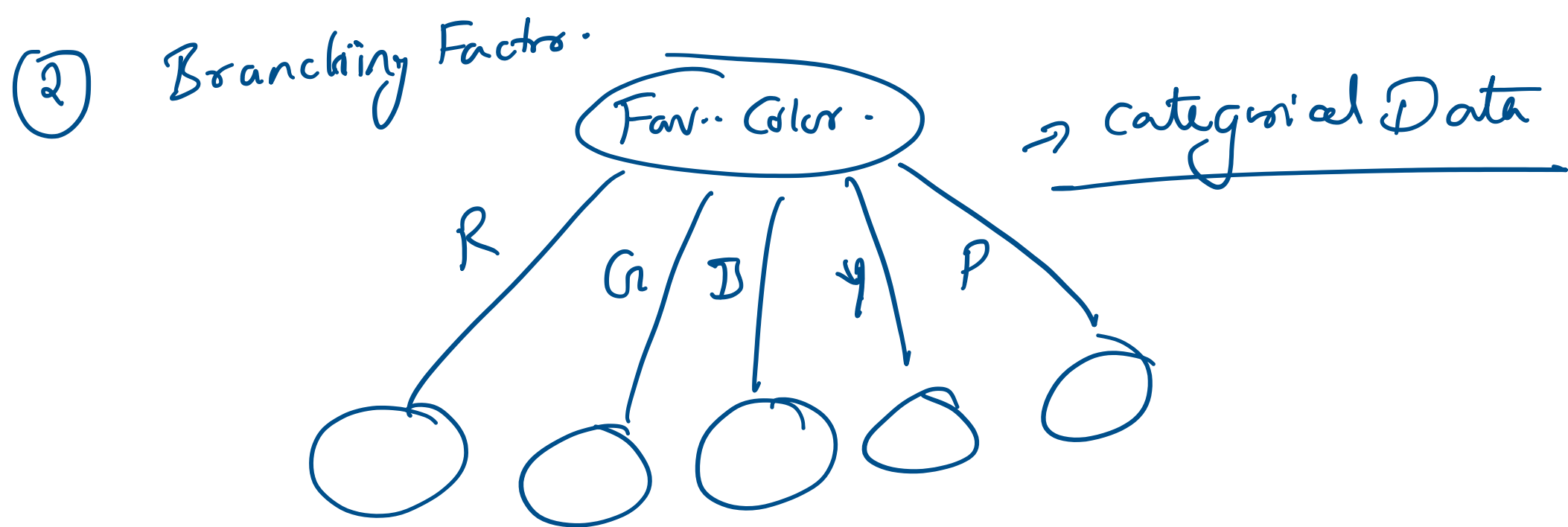
I Decision Tree Strengths.

- DTs can handle categorical data
- Explainability.
- Speed of Classification : $O(\log_2 d)$
| Try: $O(n^2 d \log d)$ |



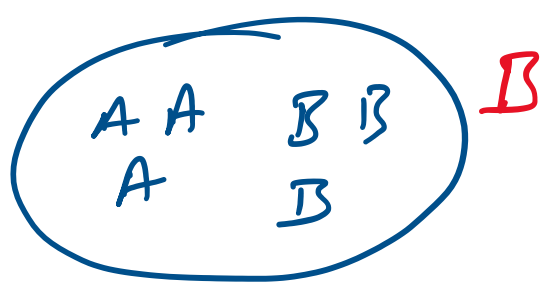
II D.T. Properties

- Prone to overfitting.
A. Early stopping.
① Control the depth.
② impurity.
B. Pruning.

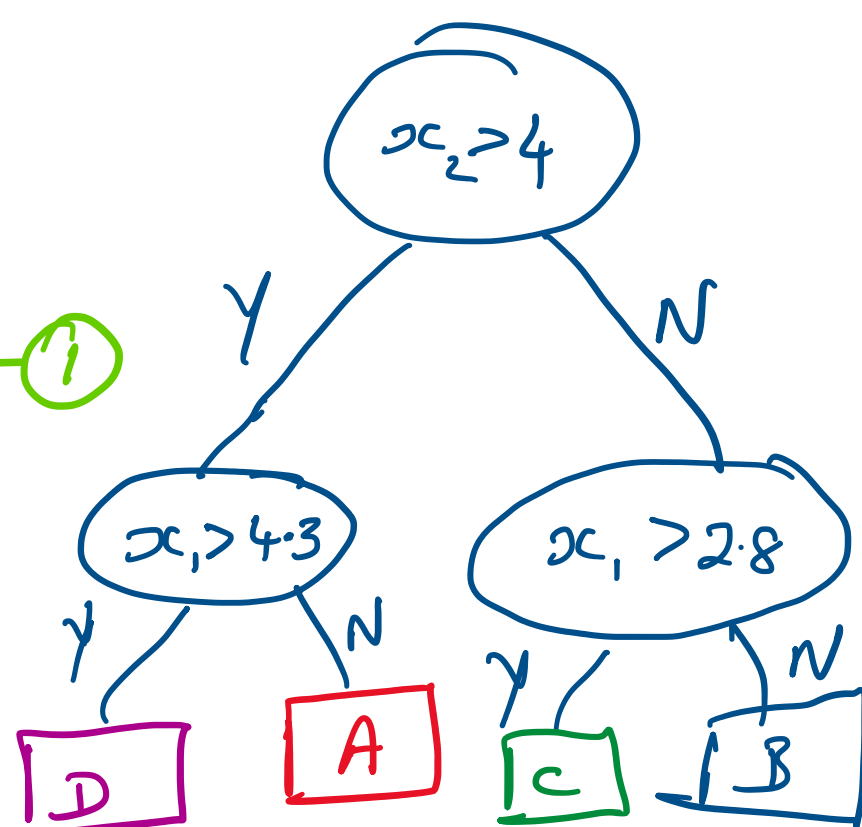
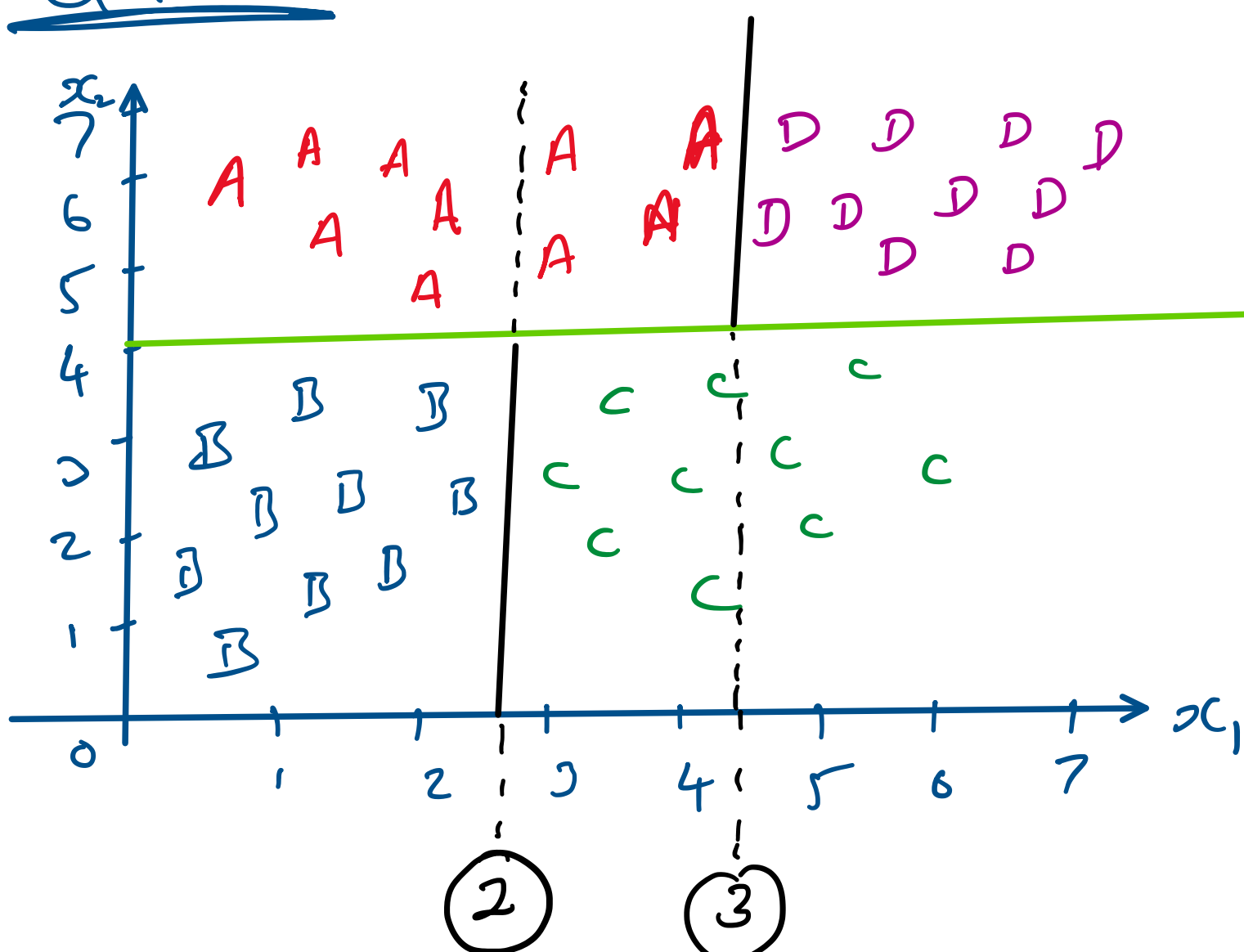


- Labelling of leaf nodes.
 - Most frequent class in the node
 - Can output a probability.
 - Class Imbalance in training data.

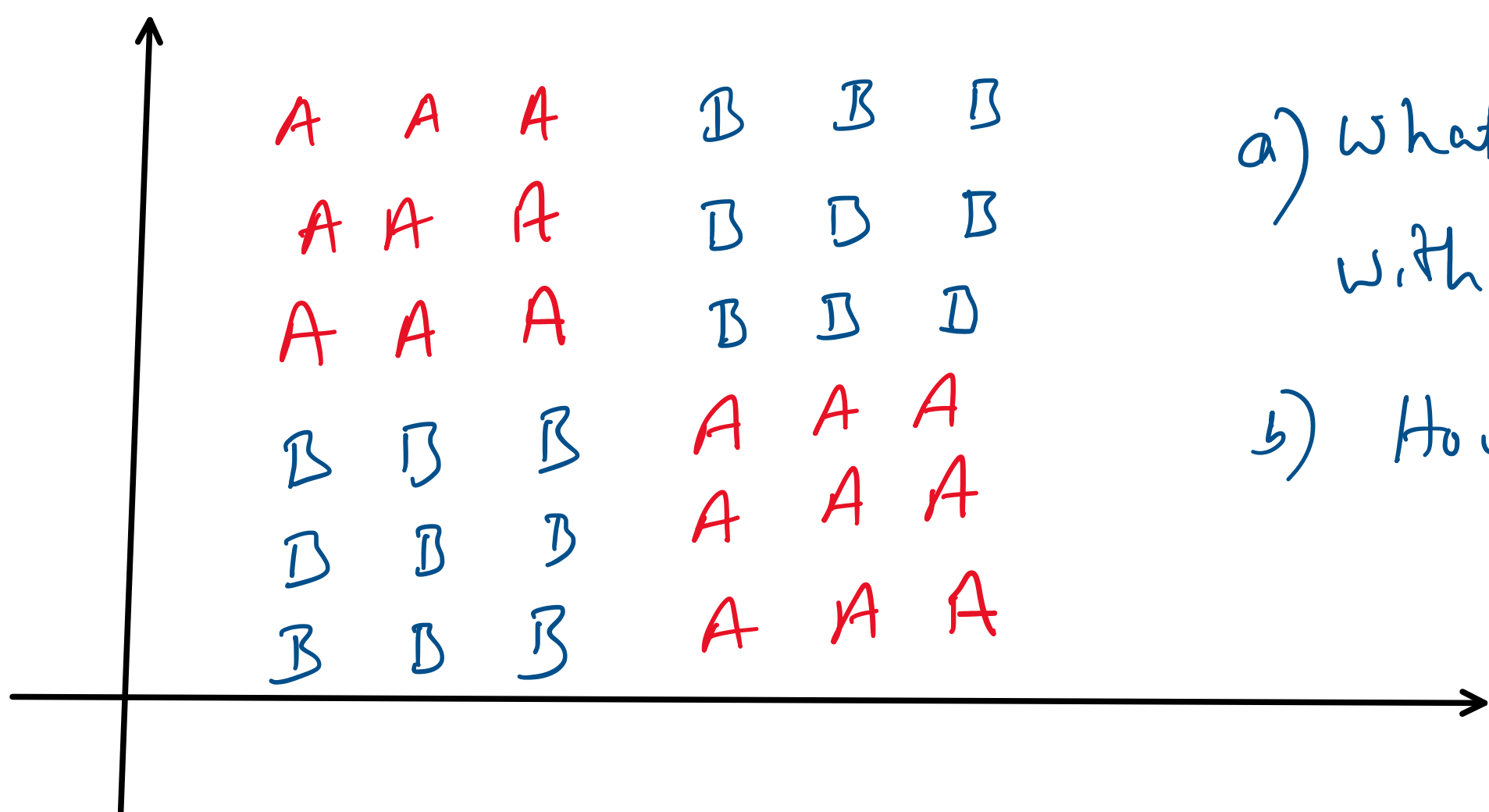
A	B
2/3	1/3



QUIZ



Information Gain. Δi ?



- What is the problem with this data?
- How can we overcome this?