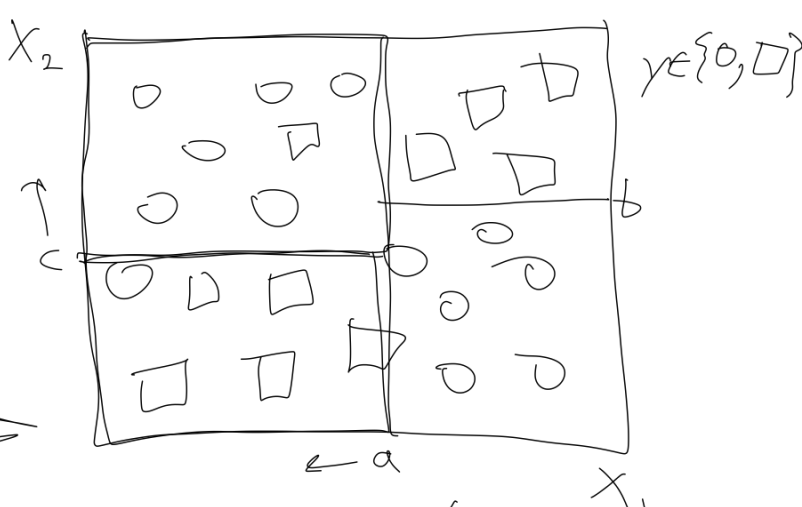
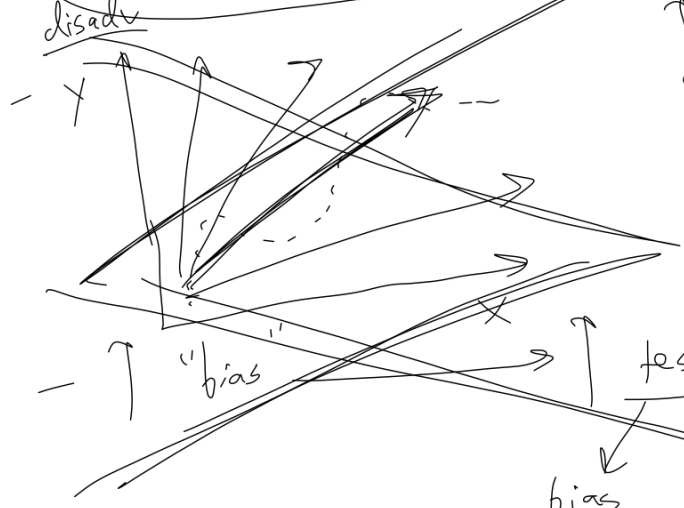


# Decision Trees

linear models

- adv
- easy to implement
  - easy to interpret

disadv



Bad: Finding a min-size tree  $\rightarrow$  NP-hard

Good: Approx.

greedy  
"impurity function"

Goal

1. Maximally compact
2. Only has pure leaves

Quiz: Is it always possible (for any dataset) to find a tree where each leaf is pure?

