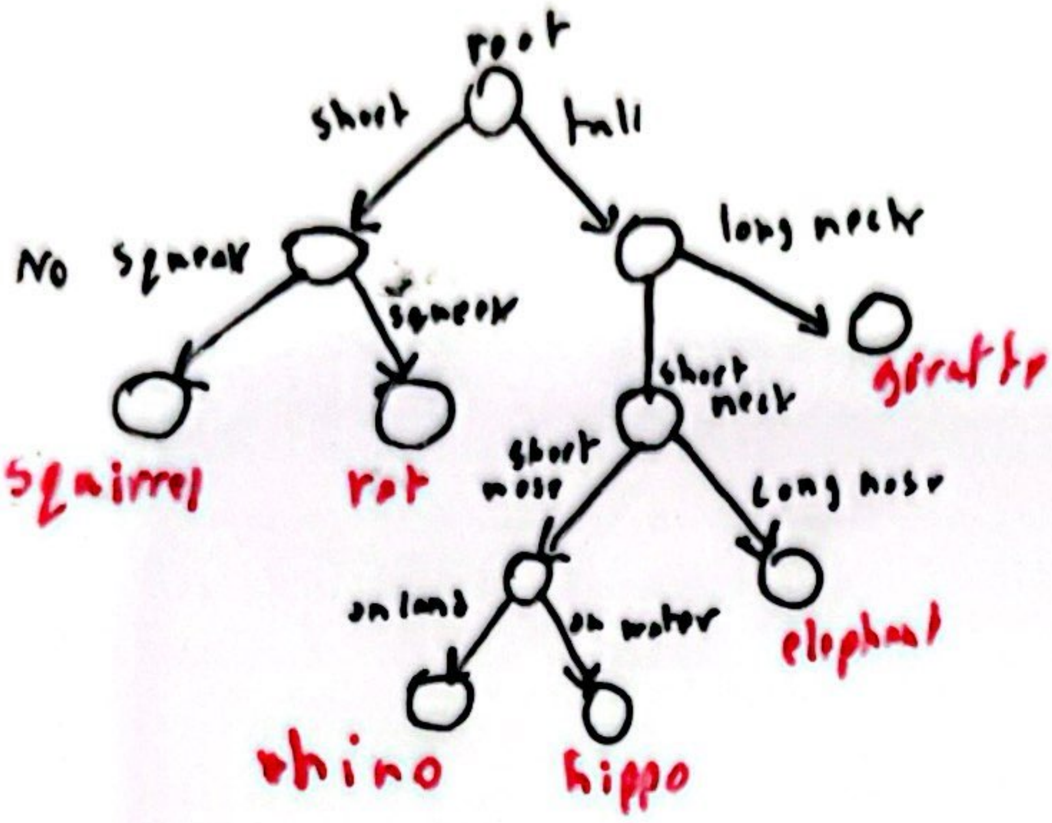
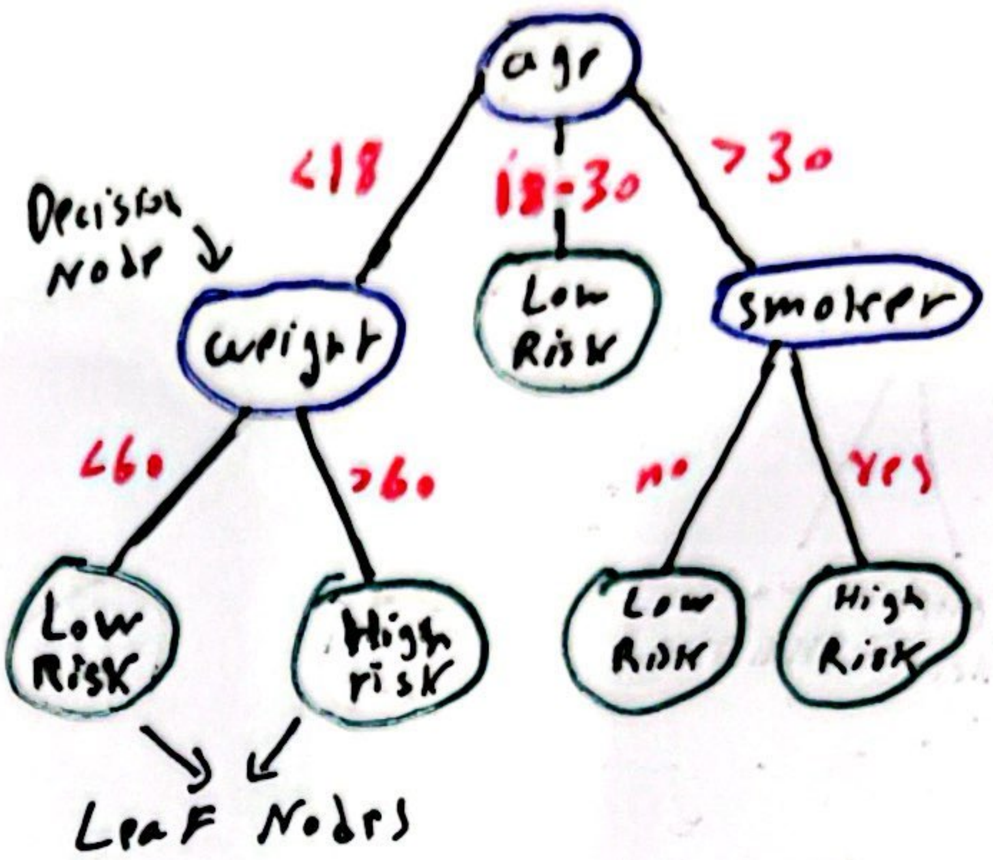
# Decision trees (- supervised - classification)

- a tree with a series of Yes No style questions that allow us to partition a data set into several Dimensions

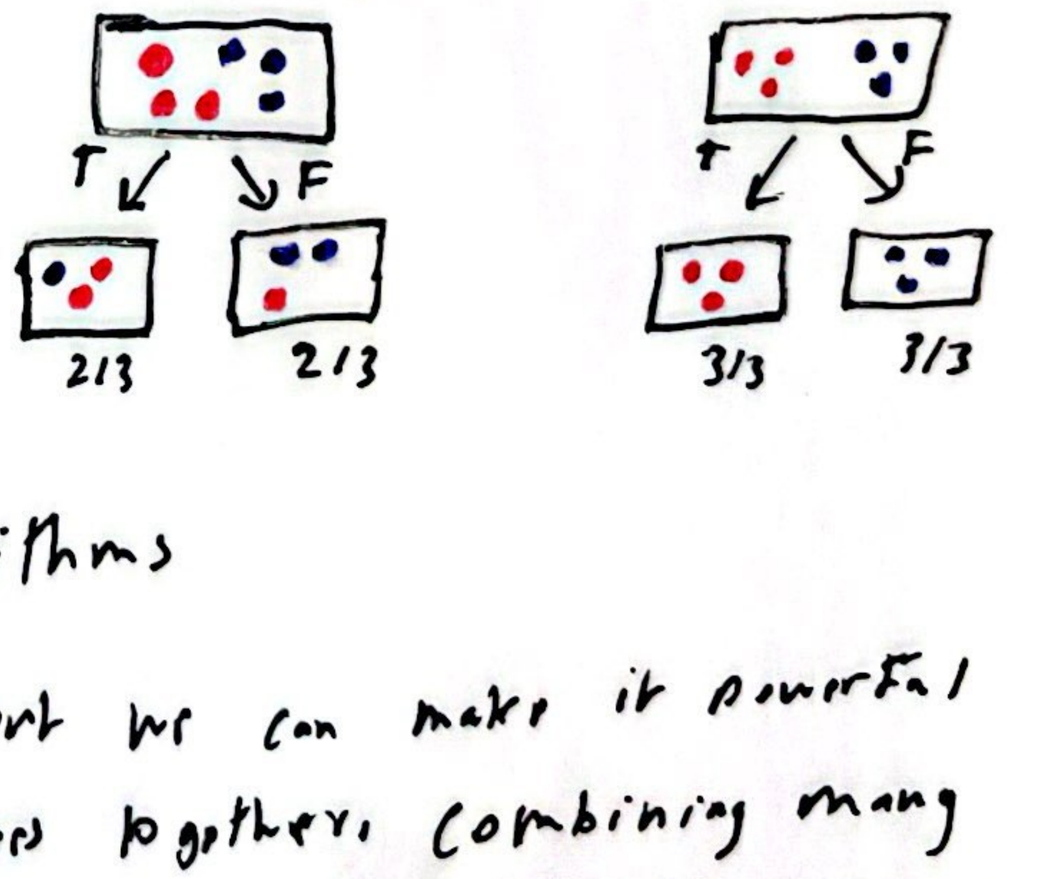
Ex's • Animal classifier




• Heart attack Risk




★ The goal of this algo is to have Leaf nodes that are as pure as possible i.e instead of randomly splitting the data we try to find splits that lead to the resulting groups of leafs to be as pure as possible i.e as few datapoints are misclassified




## Ensemble Algorithms

- a decision tree is simple but we can make it powerful by combining many decision trees together. Combining many simple ML models to a better model is called Ensemble Algorithm

★ types of Ensembles for Decision trees • more on next pages

| Ensemble method | How it works | Analogy |
|---|---|---|
| Bagging | trains many trees on random subsets of data. final results : avg (regression) or Majority (classification) | ask 100 ppl who saw parts of the story |
| Boosting | trains trees sequentially each tries to fix mistakes of previous one final prediction is weighted sum or vote | like a team each member corrects the last one got wrong |
| Random Forests | type of Bagging where each tree also uses a random subset of features | bagging but trees don't all use same attributes |