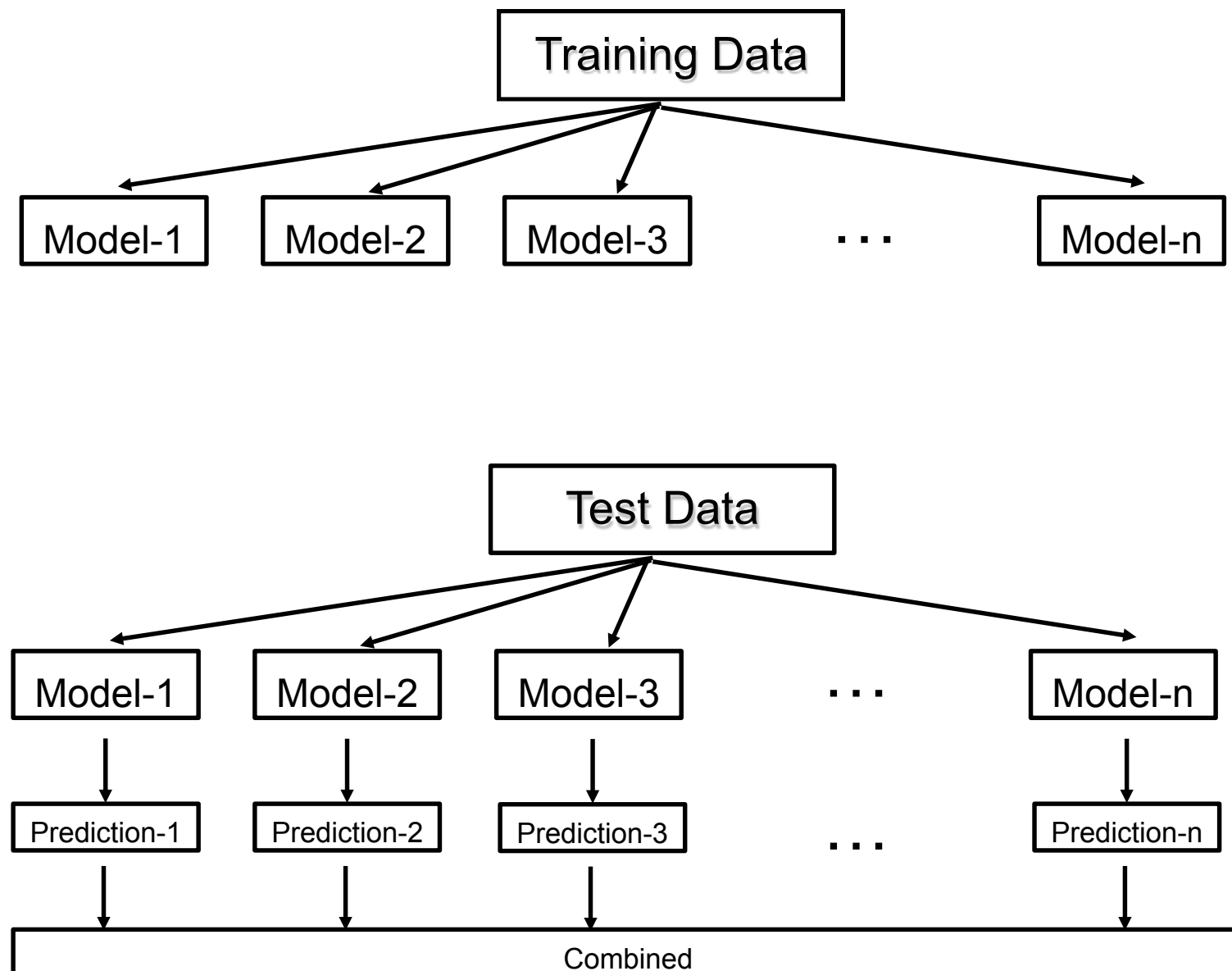


# Ensemble Methods

- Ensembles are machine learning methods for combining predictions from multiple separate models.
- The central motivation is rooted under the belief that a committee of experts working together can perform better than a single expert.

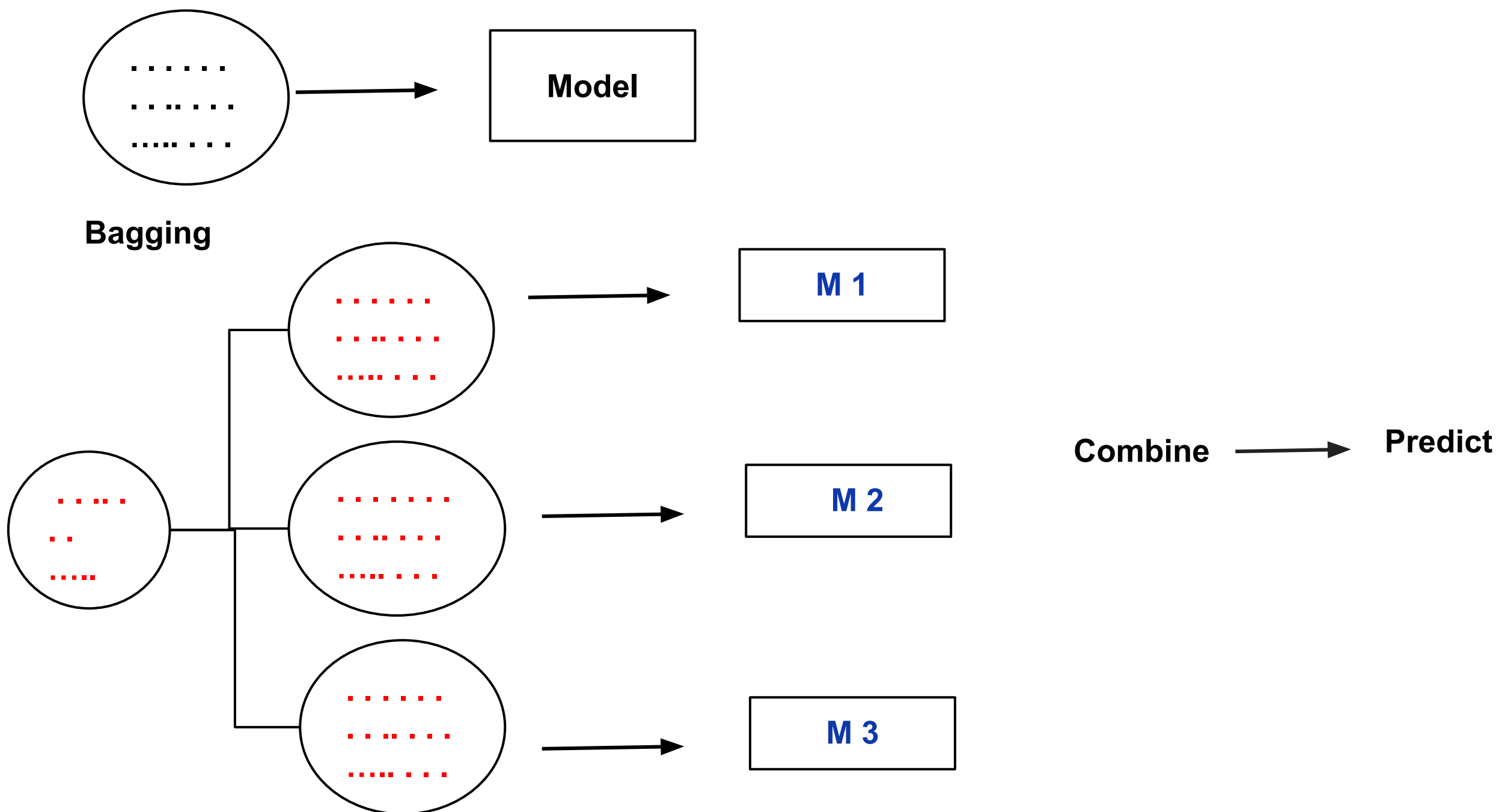


**Prediction**

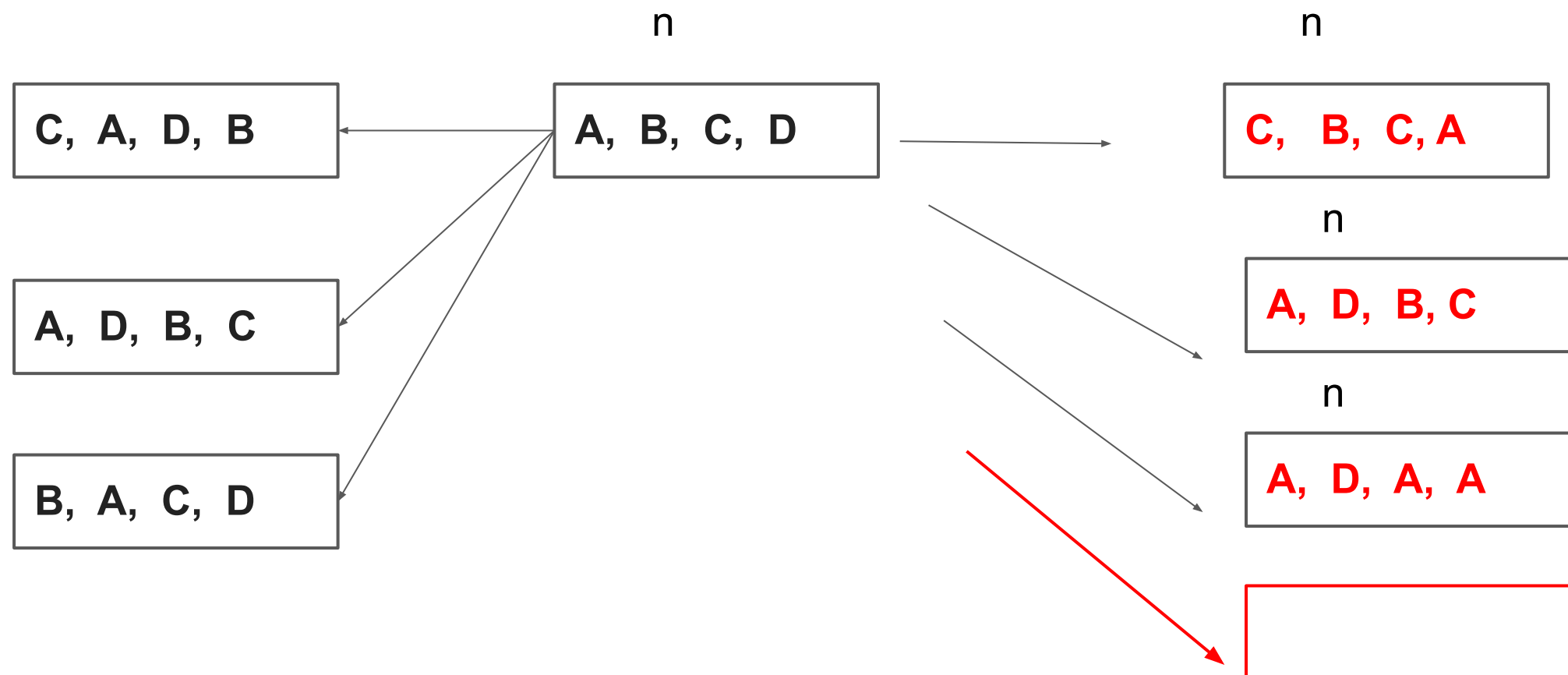
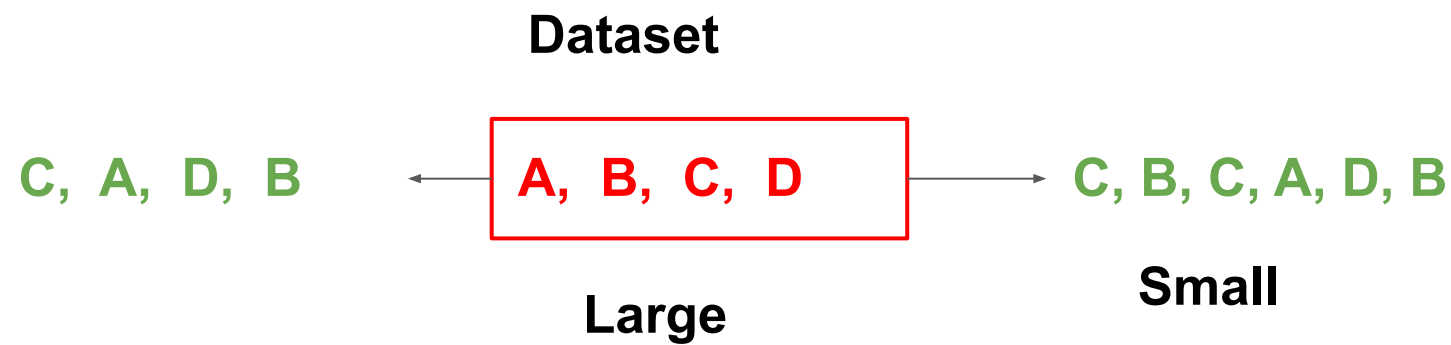
# Ensemble Methods

	90%	90%	90%	90%	90%	
Truth	M1	M2	M3	.....	M10	
Y	✓	✓	✓	✓	X	✓
Y	X	X	X	X	X	✓
N	✓	✓	✓	X	X	✓
..	✓	X	✓	X	X	✓
...	✓	✓	✓	✓	.....	✓
...	✓	✓	✓	✓	✓	...
Y	✓	✓	✓	✓		...
N	✓	✓	✓	✓	✓	...

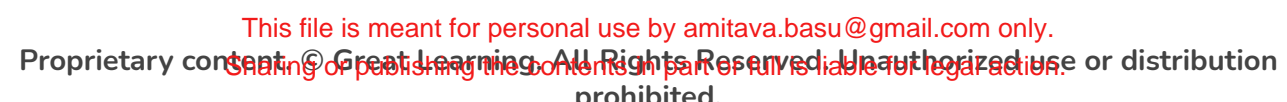
# Bagging



# Why Sampling with Replacement?



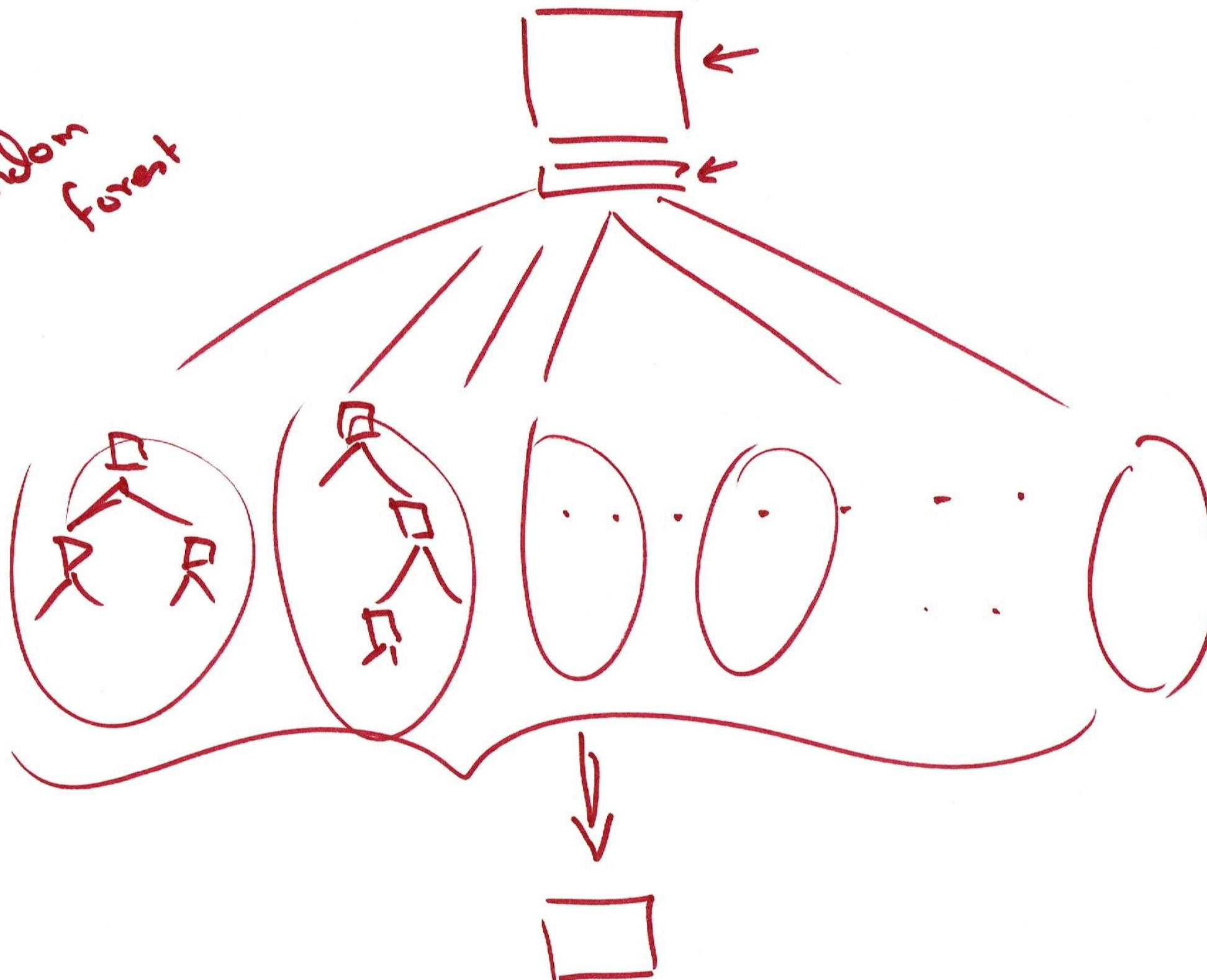
# Tree



# Tree to a Forest

- Decision trees are very sensitive to even small changes in the data - usually called unstable.
- Can we get a whole bunch of decision trees to work together to yield a better and more robust prediction?
- Then for prediction we could use the mean for regression trees and mode for classification trees
- While individual trees are tend to over-fit training data, averaging corrects this.

Random forest

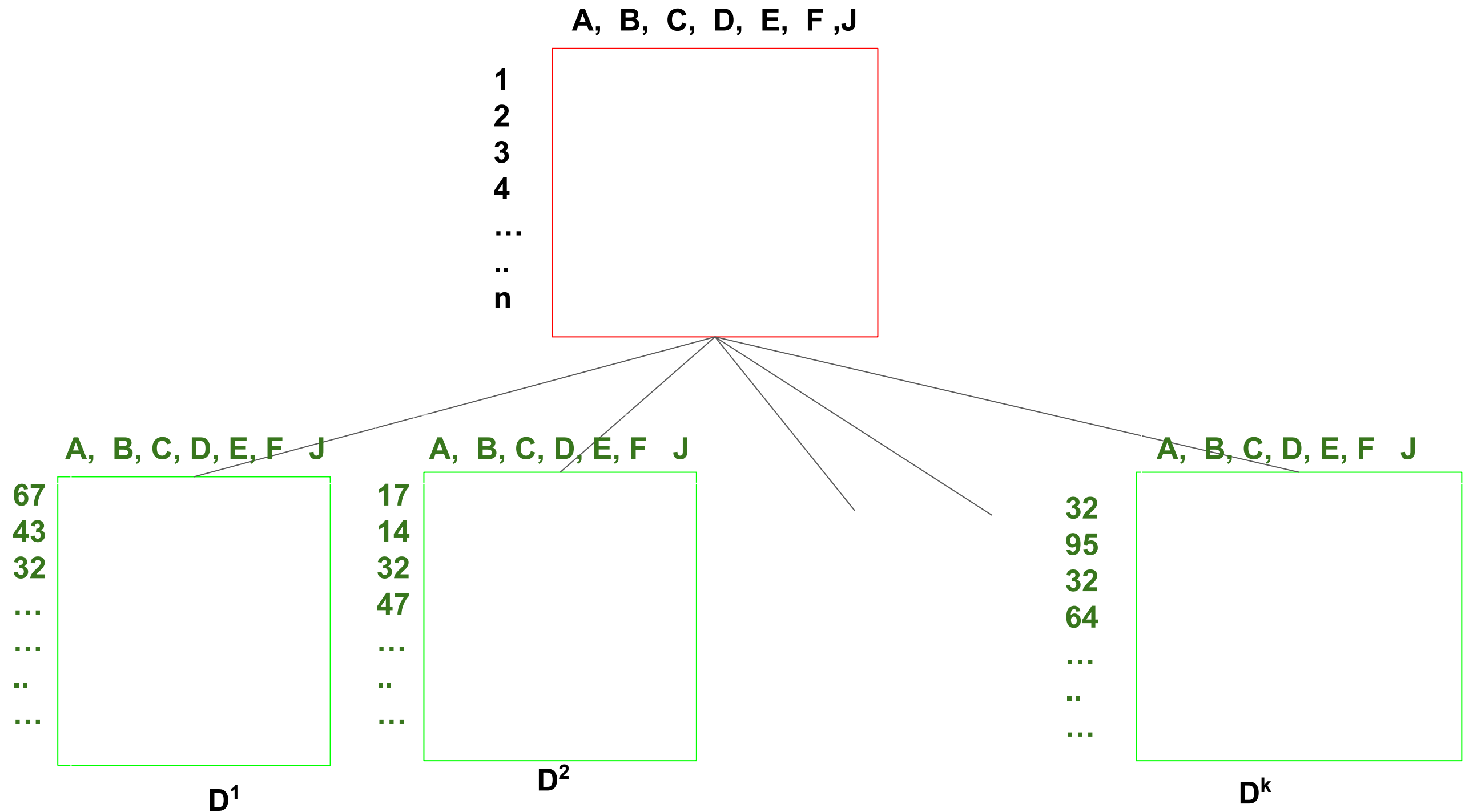


# The General Ideas

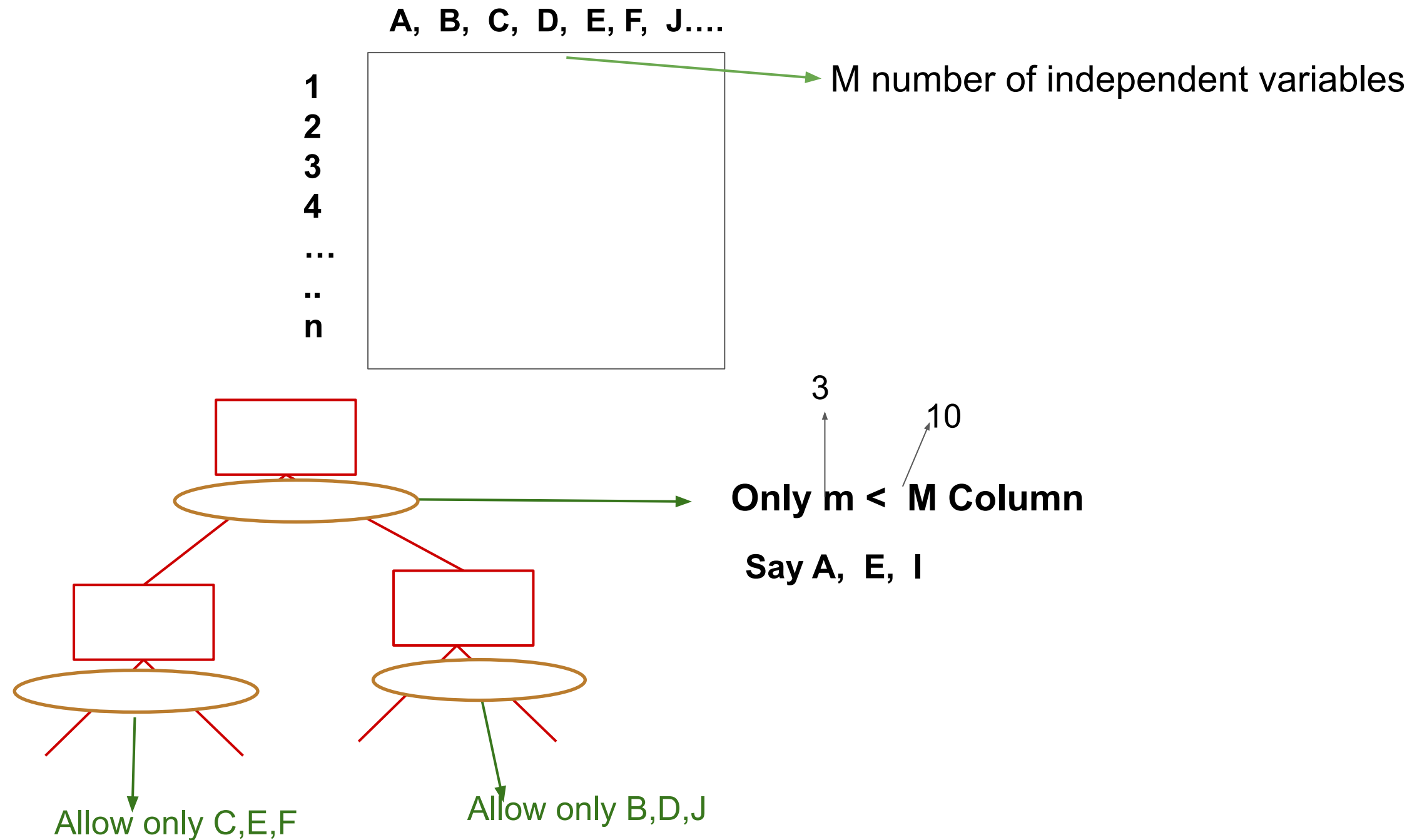
- The general procedure of using multiple models (trees, in this case) to obtain better predictive performance is called ensemble learning.
- Bootstrap aggregating. also called bagging:
  - Generate new training subsets of the original, each of the same size (usually the size of the data) by sampling with replacement.
  - By sampling with replacement, some observations may be repeated in each subset.







# Random Forest



# Random Forest



# Random forests

- 
- Random Sampling with replacement
  - For each subset build a decision tree. However, only use  $m$  randomly pick independent variables for each node's branching possibilities. 
  - Do not prune
  - While predicting:
    - Use each tree to make individual predictions
    - Combine predictions using voting:
      - Means for regression
      - Modes for classification
- 
- 

# Random Forest

Say  $M = 10 \Rightarrow A B C D \bigcirc \bigcirc \bigcirc \textcircled{J}$

