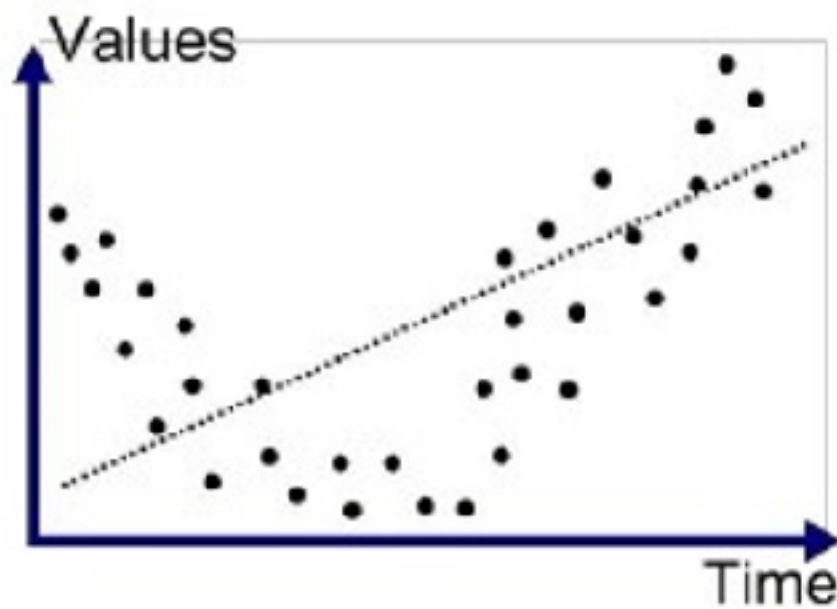


# Lecture 18:

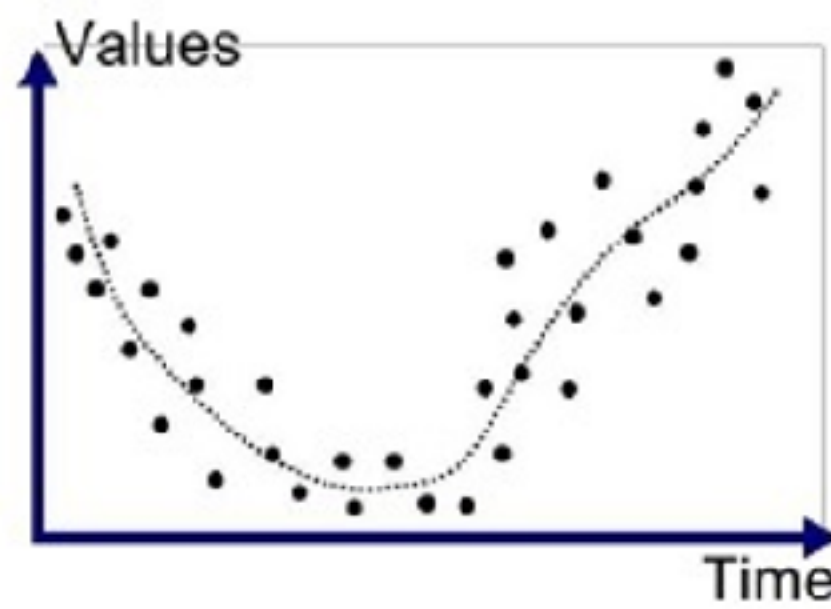
# Ensemble learning

Artificial Intelligence  
CS-UY-4613-A / CS-GY-6613-I  
Julian Togelius  
[julian.togelius@nyu.edu](mailto:julian.togelius@nyu.edu)

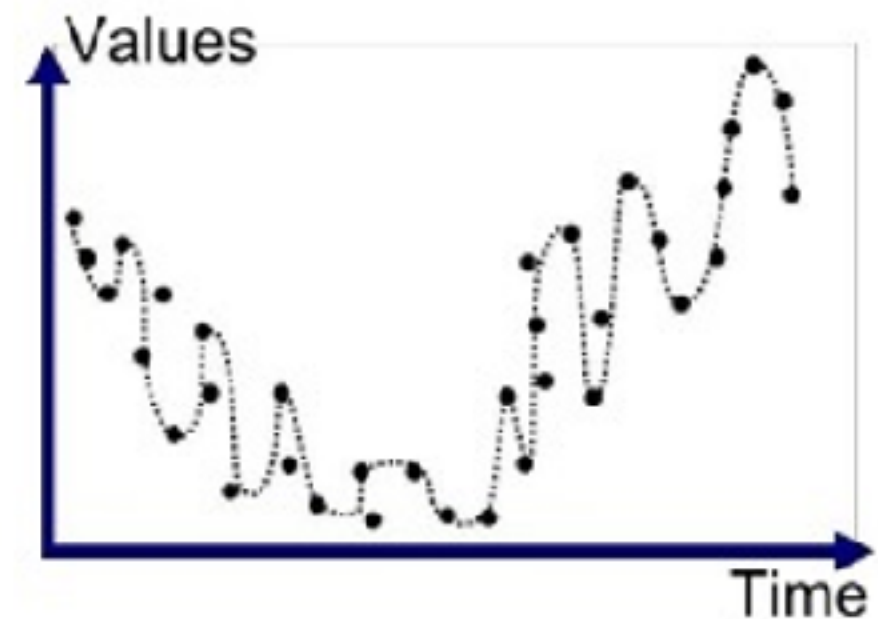
# Overfitting (in supervised learning)



Underfitted



Good Fit/Robust



Overfitted

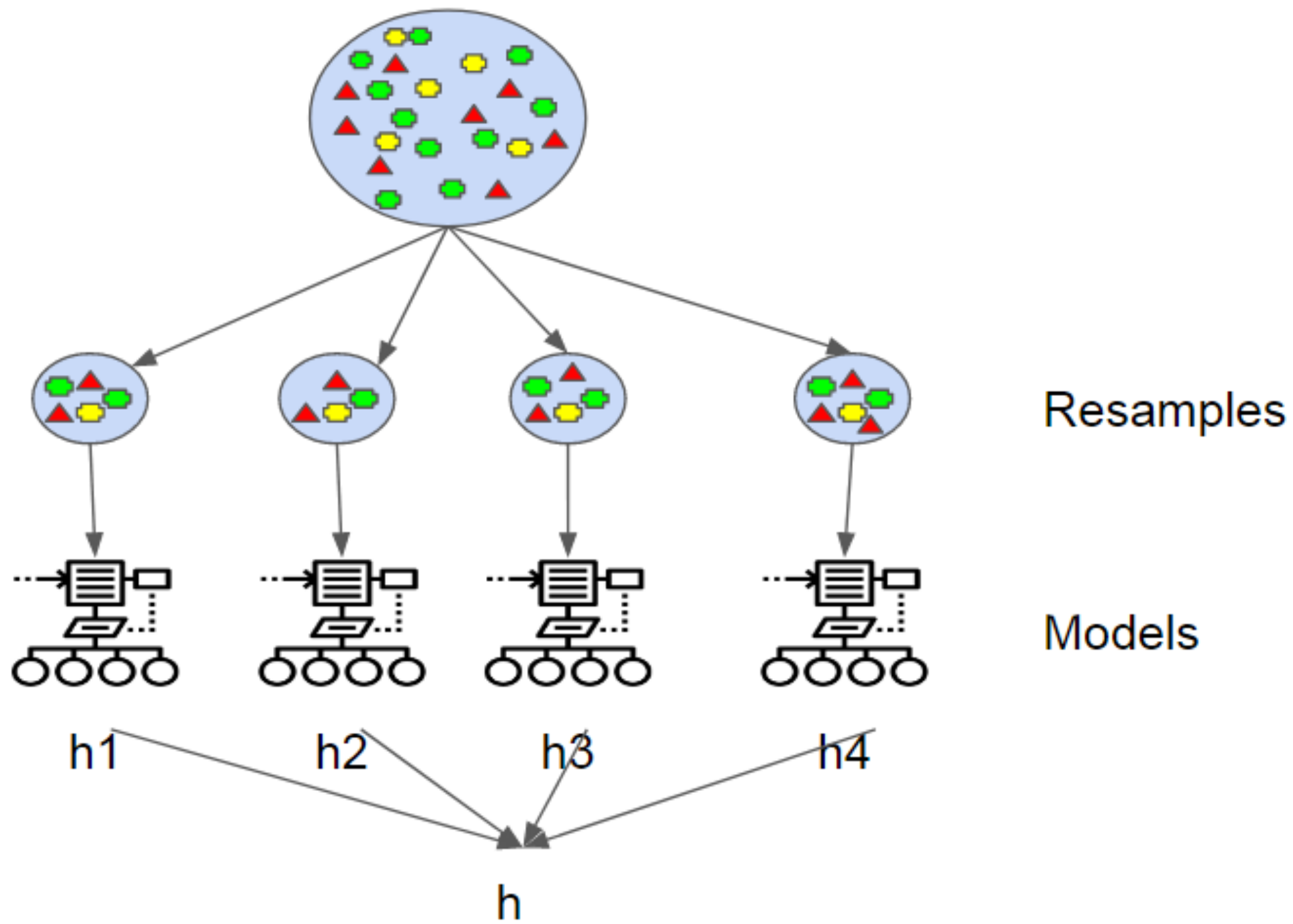
- Fitting the data too well, including the noise
- Reduces accuracy on unseen data

# Can we do better?

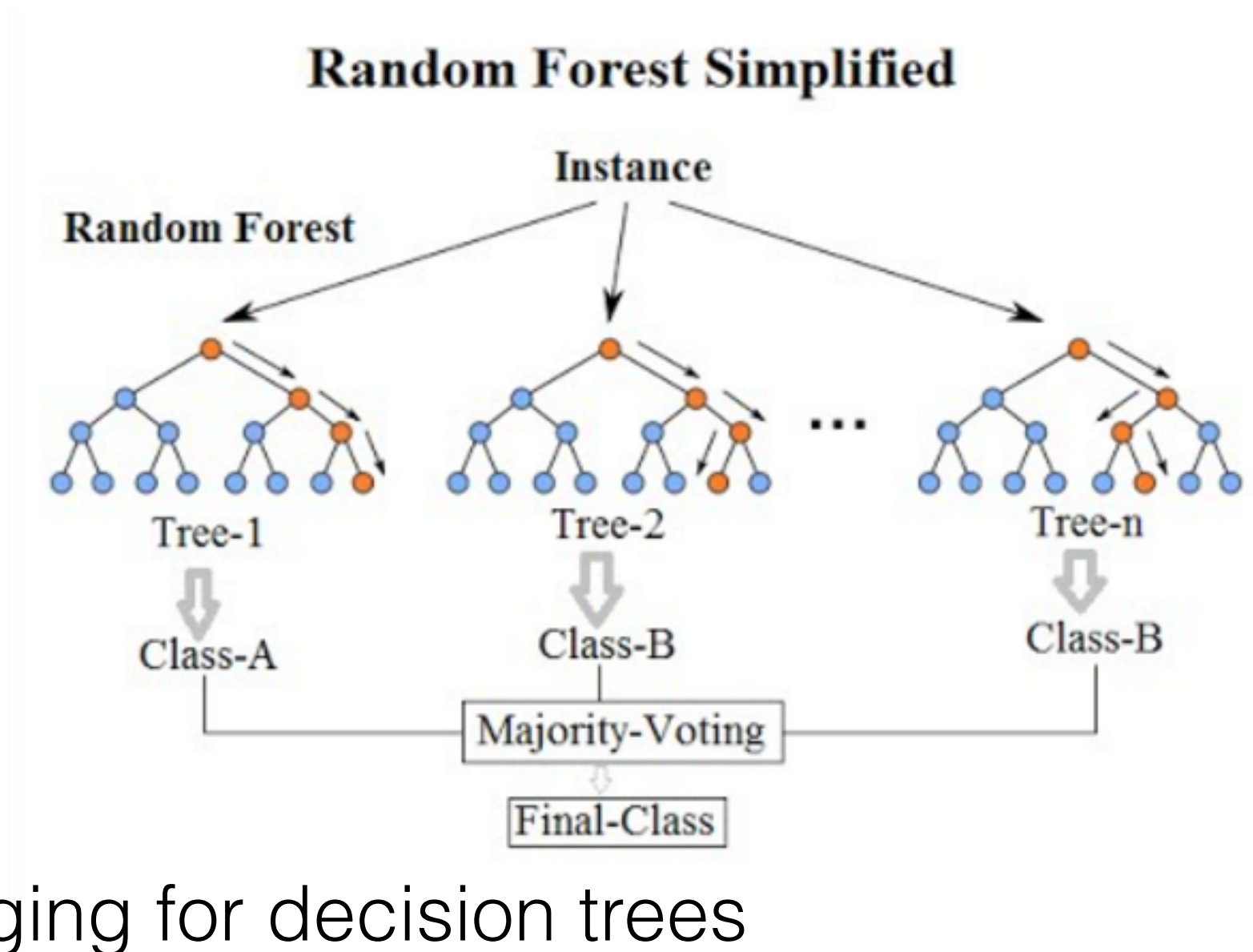
- It's hard to train good classifiers
- Can we replace quality with quantity?
  - Train many classifiers and combine them somehow
  - Called *ensemble learning*

# Bagging

- Short for Bootstrap Aggregation
- Train a number of classifiers, each on a randomly selected subset of the data
  - Drawn with replacement (the same instance can be part of many subsets)
- Use the majority vote of the models as the output
- For prediction, a mean or weighted mean



# Random forests



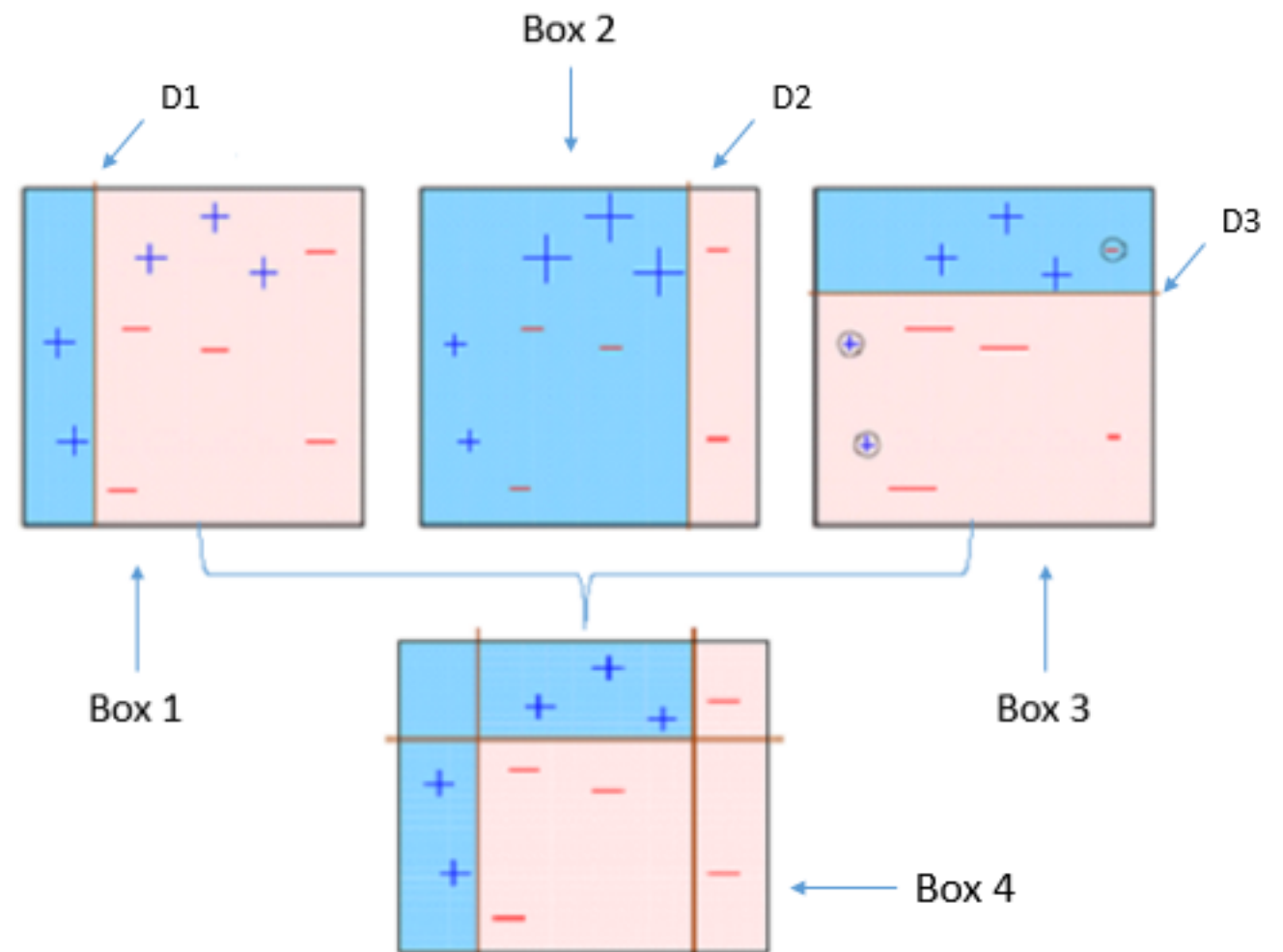
- Bagging for decision trees
- Also *randomly select features* for each subset

# Boosting

- Train many models and return the (weighted) average/majority classification, like bagging
- Unlike bagging, train each model to be good at where the previous models failed

# AdaBoost

- First train a classifier on the whole data set
- Then, train again while weighing the misclassified instances higher





# Gradient boosting

- See boosting as an optimization problem
- Optimize the parameters of each new classifier for minimum error using gradient descent