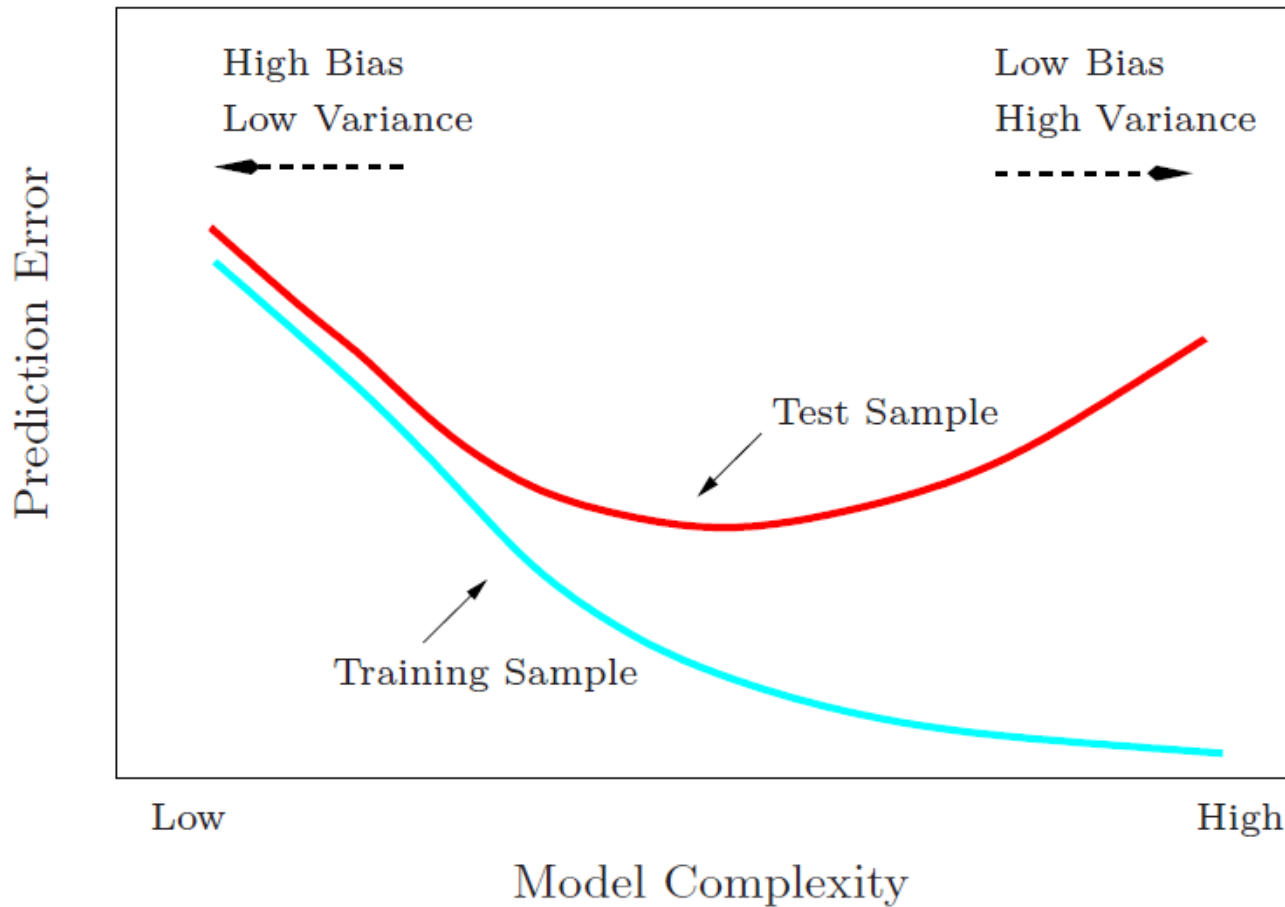


# **Supplementary Materials of Lecture 15**

# Bias/Variance Tradeoff



Hastie, Tibshirani, Friedman "Elements of Statistical Learning" 2001

# Reduce Variance Without Increasing Bias

- **Averaging** reduces variance:

$$Var(\bar{X}) = \frac{Var(X)}{N} \quad \text{(when predictions are independent)}$$

Average models to reduce model variance

One problem:

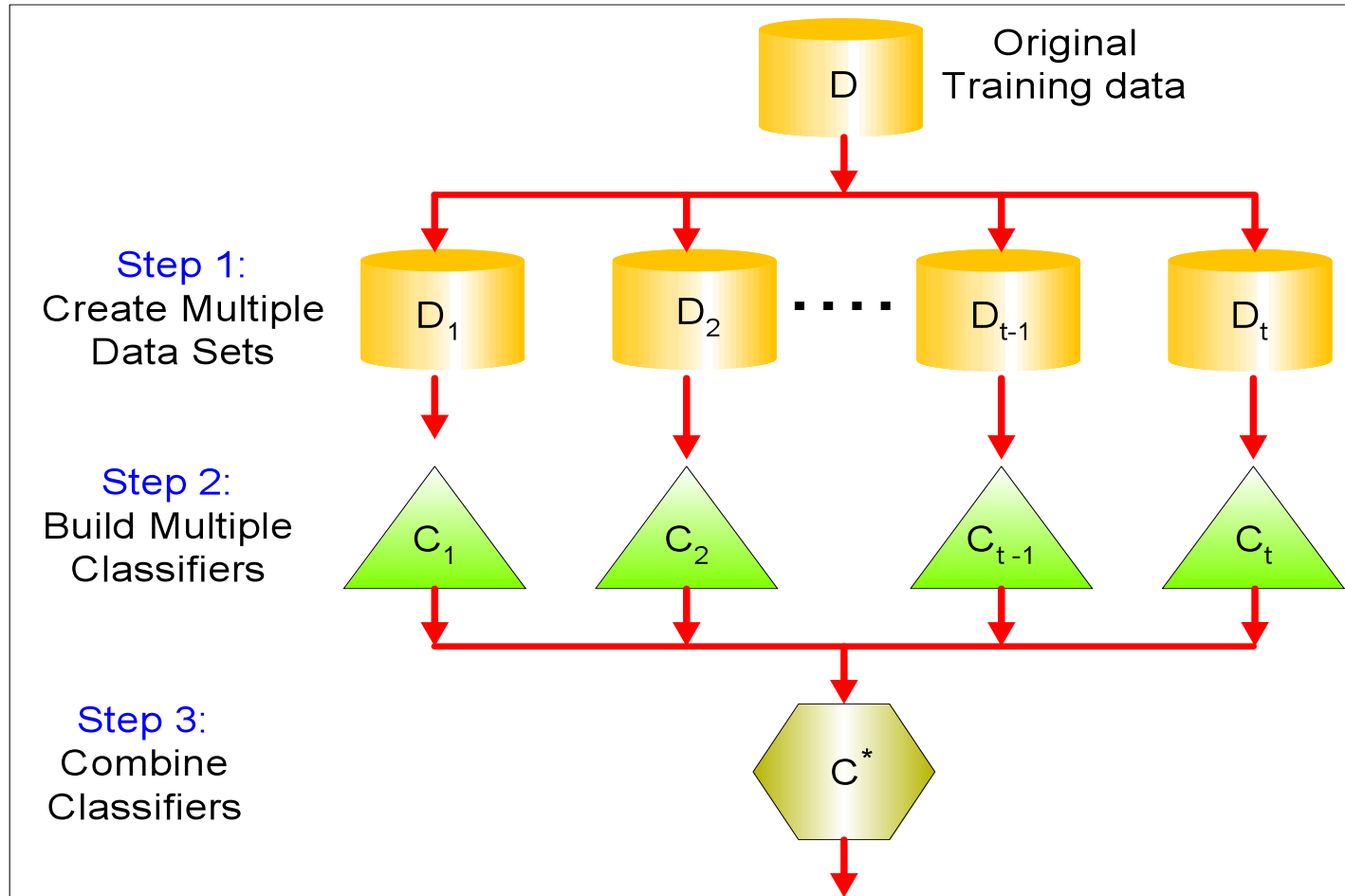
only one training set

where do multiple models come from?

# Bagging: Bootstrap Aggregation

- Leo Breiman (1994)
- Take repeated **bootstrap samples** from training set  $D$
- *Bootstrap sampling*: Given set  $D$  containing  $N$  training examples, create  $D'$  by drawing  $N$  examples at random **with replacement** from  $D$ .
- Bagging:
  - Create  $k$  bootstrap samples  $D_1 \dots D_k$ .
  - Train distinct classifier on each  $D_i$ .
  - Classify new instance by majority vote / average.

# General Idea



# Bagging

- Sampling with replacement

Training Data  
↙

Data ID	1	2	3	4	5	6	7	8	9	10
Original Data	1	2	3	4	5	6	7	8	9	10
Bagging (Round 1)	7	8	10	8	2	5	10	10	5	9
Bagging (Round 2)	1	4	9	1	2	3	2	7	3	2
Bagging (Round 3)	1	8	5	10	5	5	9	6	3	7

- Build classifier on each bootstrap sample
- Each data point has probability  $(1 - 1/n)^n$  of not being selected as training data
- Training data =  $1 - (1 - 1/n)^n$  of the original data

# The 0.632 bootstrap

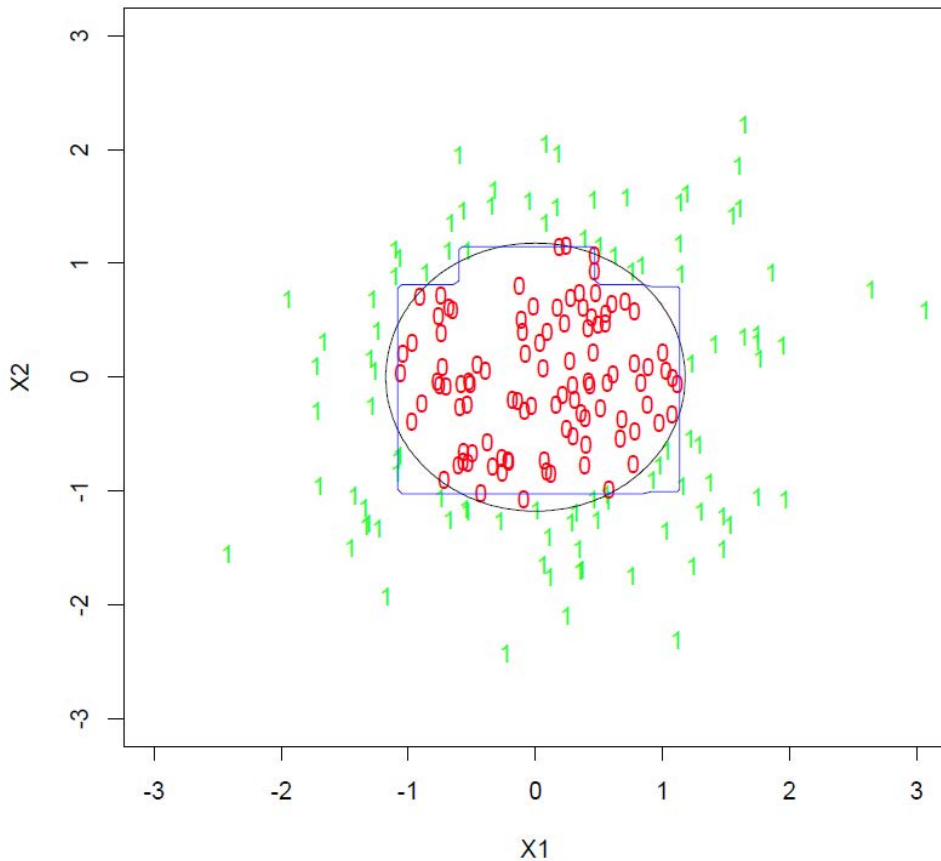
- This method is also called the *0.632 bootstrap*
  - A particular training data has a probability of  $1-1/n$  of *not* being picked
  - Thus its probability of ending up in the test data (not selected) is:

$$\left(1 - \frac{1}{n}\right)^n \approx e^{-1} = 0.368$$

- This means the training data will contain approximately 63.2% of the instances

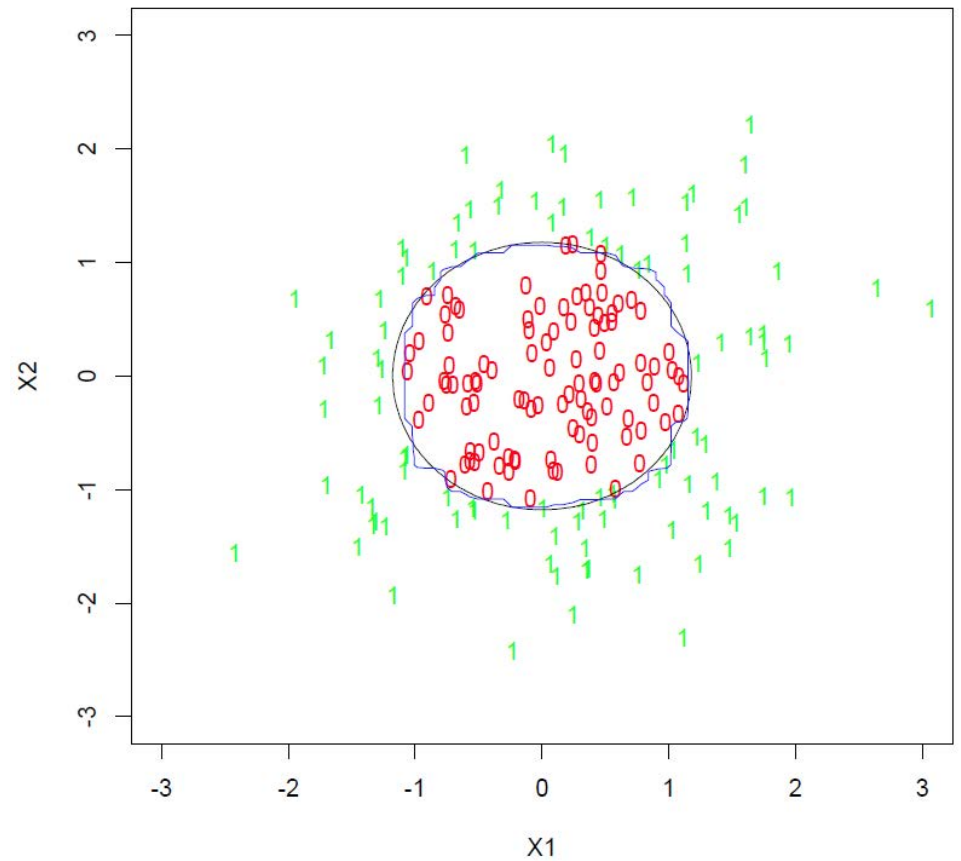
# Comparison on Decision Boundaries

Error Rate: 0.073



Single decision tree

Error Rate: 0.032



Bagging (decision tree)

Bagging produces smoother decision boundary



# Random Forests

- Ensemble method specifically designed for decision tree classifiers
- Introduce two sources of randomness:  
“Bagging” and “Random input vectors”
  - **Bagging method**: each tree is grown using a bootstrap sample of training data
  - **Random vector method**: **At each node**, best split is chosen from a random sample of  $m$  attributes instead of all attributes

# Random Forests

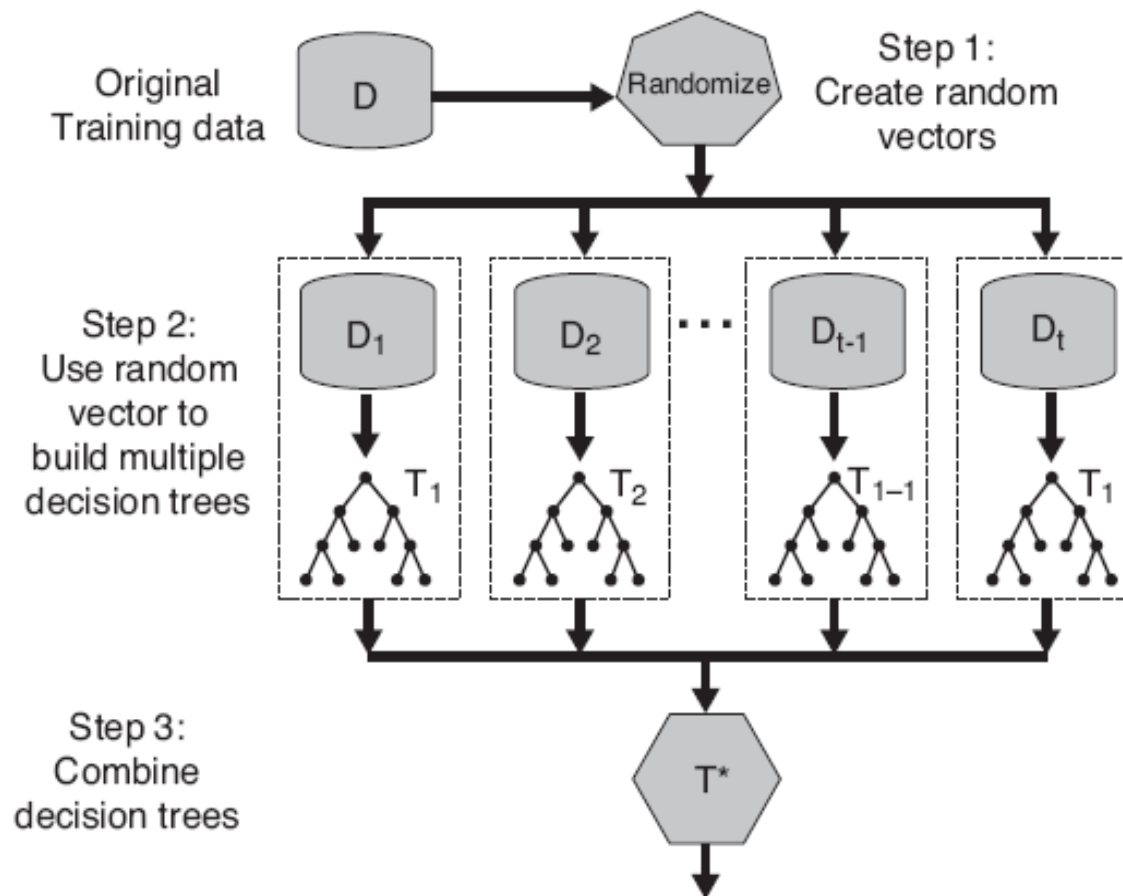


Figure 5.40. Random forests.

# Reduce Bias<sup>2</sup> and Decrease Variance?

- Bagging reduces variance by averaging
- Bagging has little effect on bias
- Can we average *and* reduce bias?
- Yes:

- Boosting

In general, Boosting > Bagging > Base learner  
(In performance)

# Bagging versus Boosting

