ENSEMBLE LEARNING

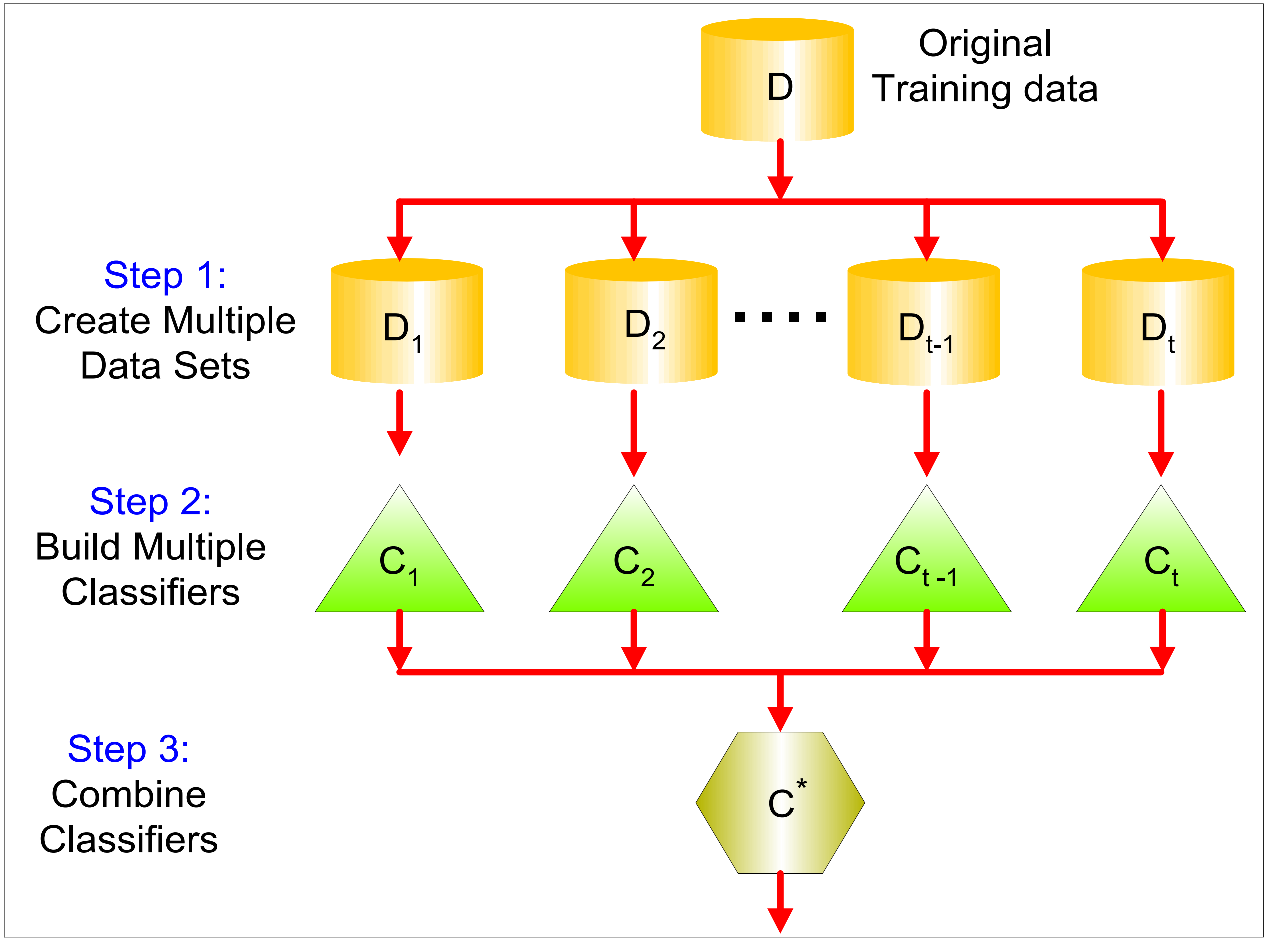
SYRACUSE UNIVERSITY
School of Information Studies

ENSEMBLE METHODS

Construct a set of classifiers from the training data.

Predict class label of previously unseen records by aggregating predictions made by multiple classifiers.

GENERAL IDEA



WHY DOES ENSEMBLE WORK?

Suppose there are 25 base classifiers.

- Each classifier has error rate, $\varepsilon = 0.35$ (weak learner).
- Assume classifiers are independent.
- Use majority voting to combine results, so ensemble makes a wrong prediction only if over half of the base classifiers are wrong.
- Probability that the ensemble classifier makes a wrong prediction:

$$\sum_{i=13}^{25} \binom{25}{i} \varepsilon^{i} (1-\varepsilon)^{25-i} = 0.06$$

- Error rate is reduced from 0.35 to 0.06.
- In practice, the base classifiers may not be totally independent for a reduction in error rate to occur.

BAGGING

Sampling with replacement:

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 sample has probability $(1 – 1/n)^n$ of being selected.

BOOSTING

An iterative procedure to adaptively change distribution of training data by focusing more on previously misclassified records.

Initially, all N records are assigned equal weights.

Unlike bagging, weights may change at the end of boosting round.

BOOSTING

Records that are wrongly classified will have their weights increased.

Records that are classified correctly will have their weights decreased.

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

Example 4 is hard to classify.

Its weight is increased; therefore, it is more likely to be chosen again in subsequent rounds.

RANDOM FOREST

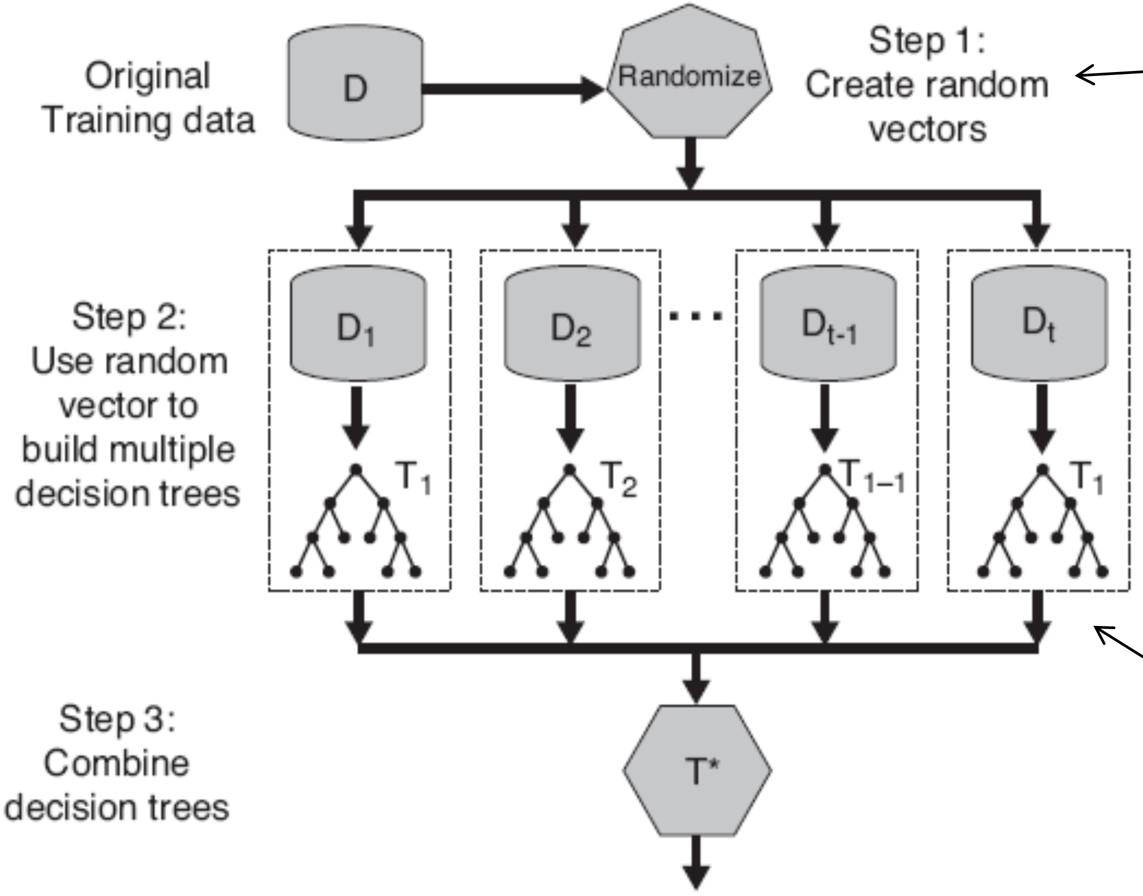


Figure 5.40. Random forests.

Randomly select subset of features.

Model with too many trees tends to overfit.