Approaches for construct classifier ensembles

- Bagging (Random Forest).
- Boosting (AdaBoost).

**Best method → AdaBoost**
(for small ensemble sizes)

**Why? → its large diversity**
(Accuracy-diversity dilemma)

**Proposal → Rotation Forest**
- Based on PCA and decision trees.
- Achieves both, accuracy and diversity.

**Experimental setting**
- Compared to Bagging, AdaBoost and Random Forest.
- 33 benchmark datasets from UCI repository.

---

**Rotation Forest Algorithm**

1. The feature set is randomly split into K subsets.
2. PCA is applied to each subset.
3. All principal components are retained.
4. Arrange the PCA coefficients in a matrix (rotation matrix).
5. Apply the rotation matrix to the data features.
6. Build each decision tree on the rotated training data.

**Diversity heuristics**

- Different possible feature subsets.
- Rotation by PCA.
- Bootstrap samples.
- Random subset of X.
- Random selection of classes.