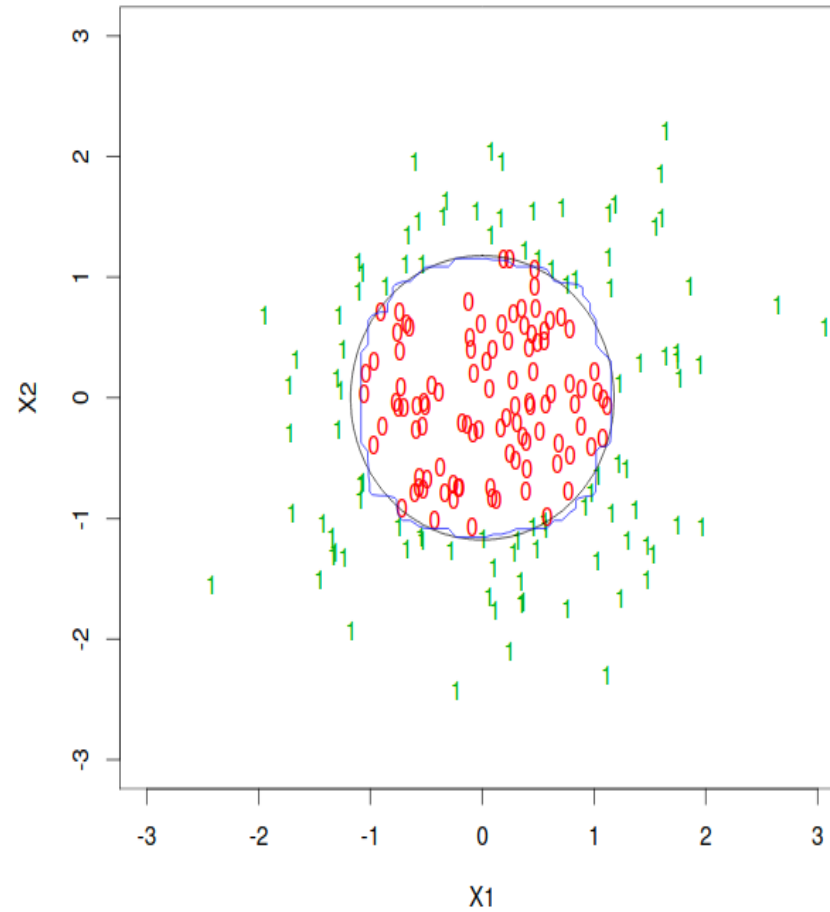


# Improved Classifiers

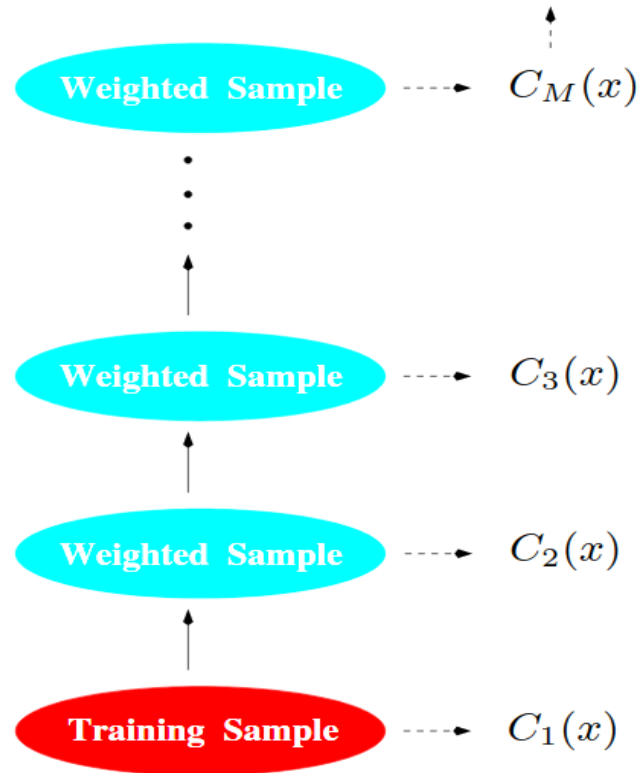
1. Bagging
2. Boosting
3. Stacking

# Bagging



Bagging averages many trees, and produces **smoother** decision boundaries.

# Boosting



## Boosting

- Average many trees, each grown to re-weighted versions of the training data.
- Final Classifier is weighted average of classifiers:

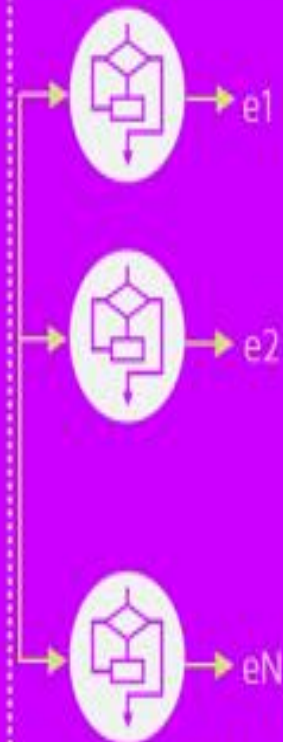
$$C(x) = \text{sign} \left[ \sum_{m=1}^M \alpha_m C_m(x) \right]$$

single



single estimate

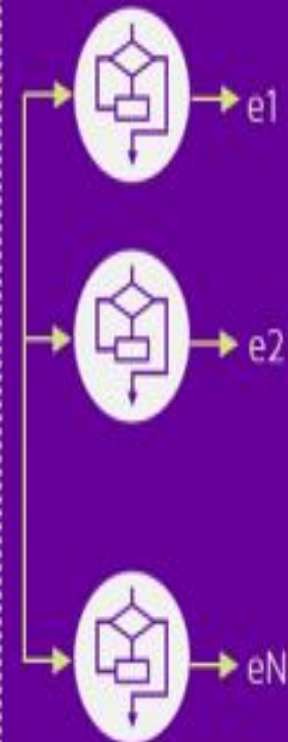
bagging



$$e = \frac{1}{N} \sum_{i=1}^N e_i$$

simple average

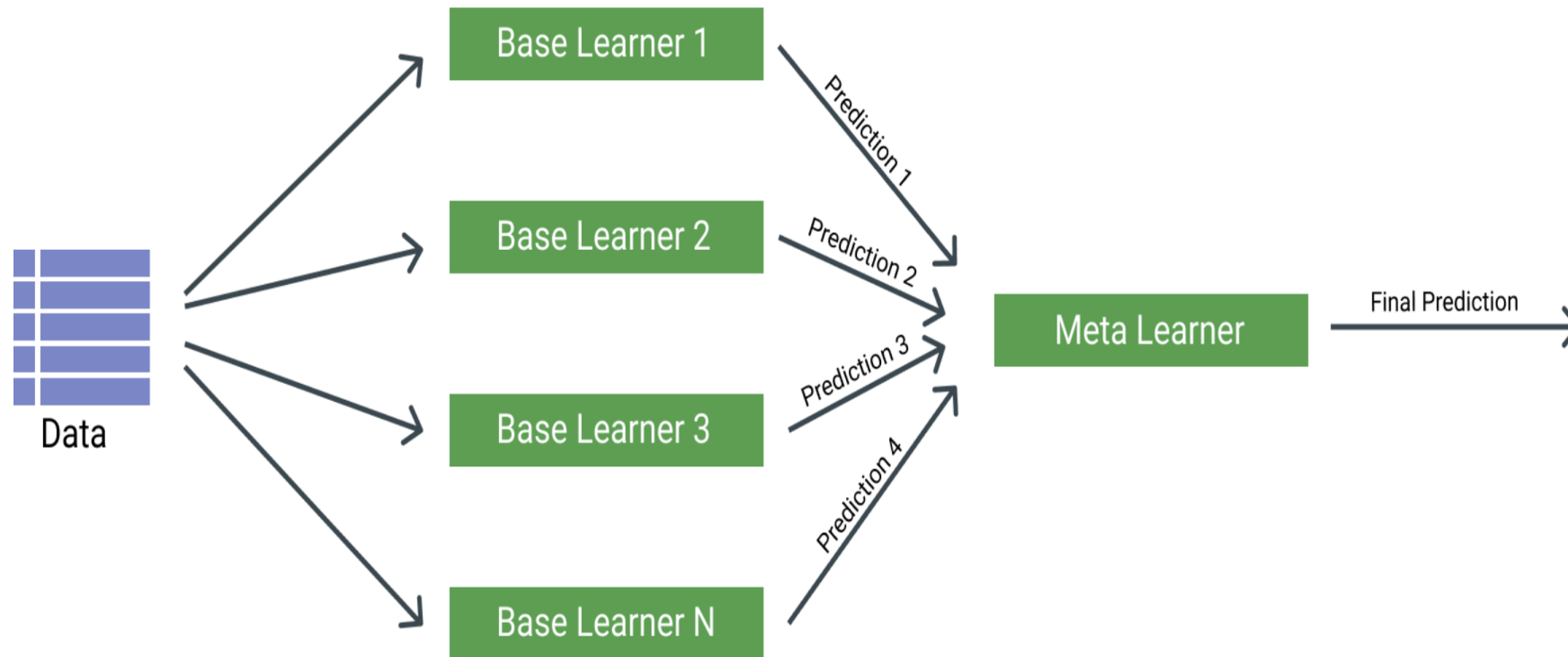
boosting



$$e = \sum_{i=1}^N w e_i$$

weighted average

# Stacking



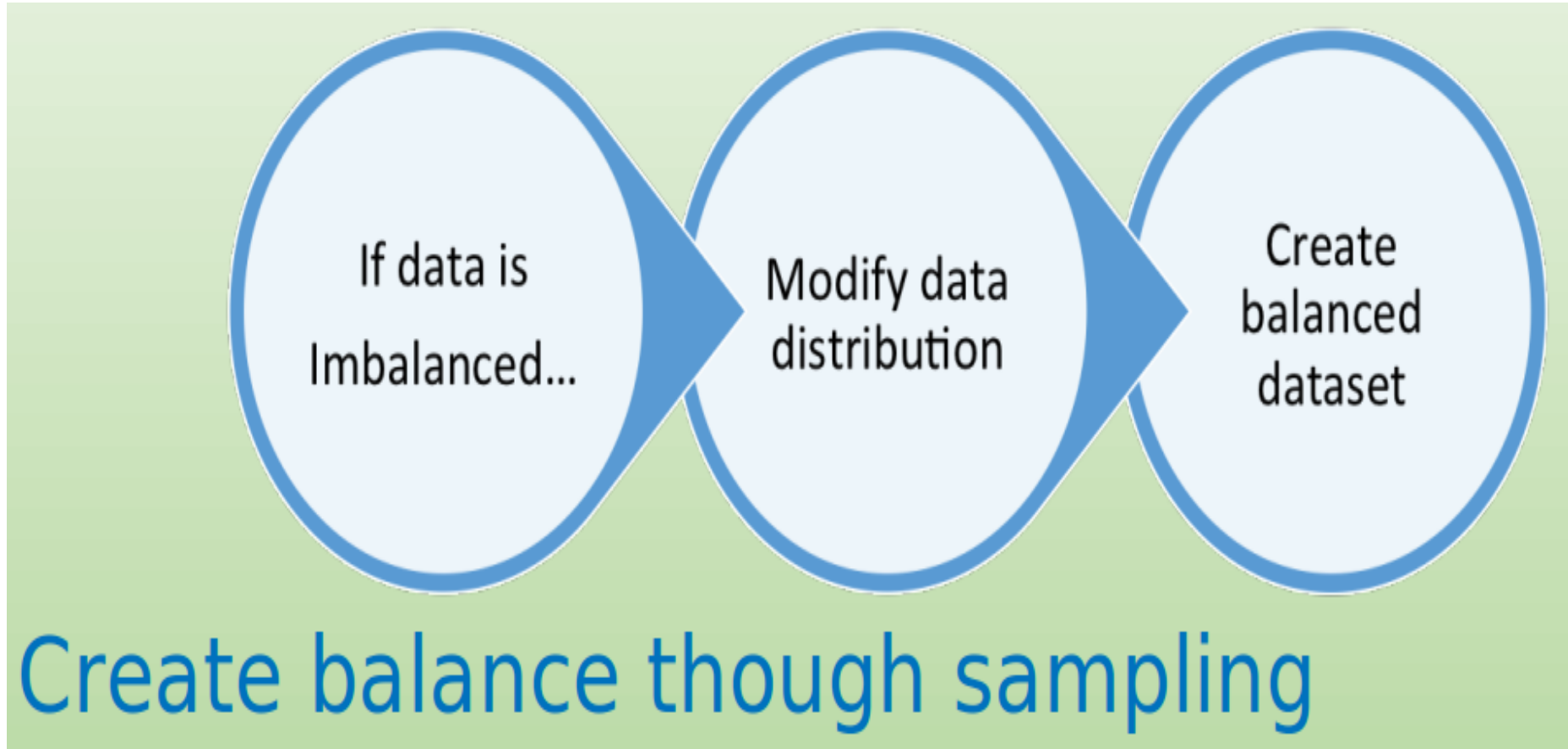
# Learning from Imbalanced Data

1. Sampling Methods
2. Cost-sensitive Methods
3. Kernel and Active learning Methods

# Learning from Imbalanced Data

1. Sampling Methods
2. Cost-sensitive Methods
3. Kernel and Active learning Methods (e.g: One-class SVM)

# Sampling Methods





# Cost-sensitive Methods

