

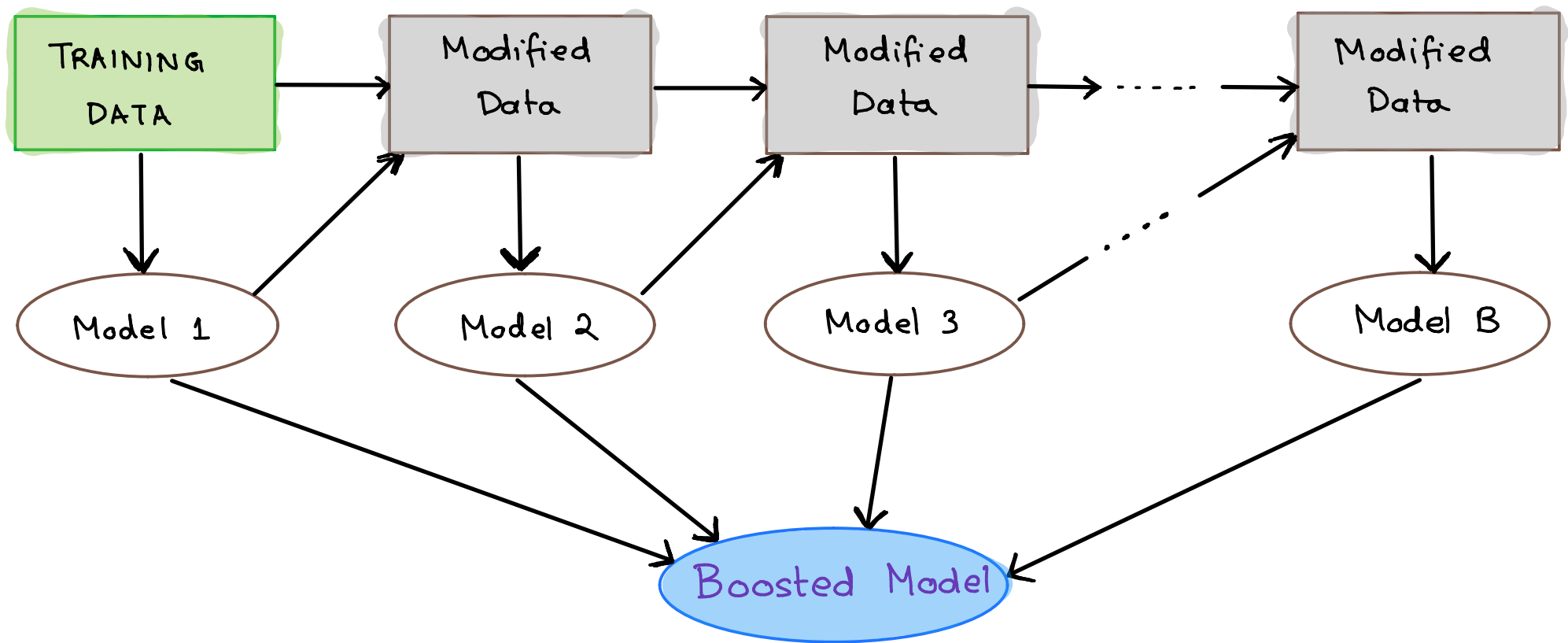
Boosting

- In bagging, we created an ensemble for reducing the variance in high-variance-low-bias (strong) base models
- Boosting is another ensemble method used for reducing the bias in high-bias-low-variance (weak) base models
- Intuition :
 - Even a simple (weak) model can typically describe some aspects of the input-output (I/O) relationship
 - Can we then learn an ensemble of "weak models", where each weak model describes some part of the I/O relationship, and combine these models into one "strong model" ?

- Boosting shares some similarities with bagging
 - Both use an ensemble of models for combining predictions
 - Both can be used with any regression or classification algorithm
- Difference between bagging and boosting lies in how the base models are being trained
 - In bagging, 'B' identically distributed models are constructed **parallelly**
 - In boosting, the ensemble members are constructed **sequentially**.

Sequential Construction in Boosting

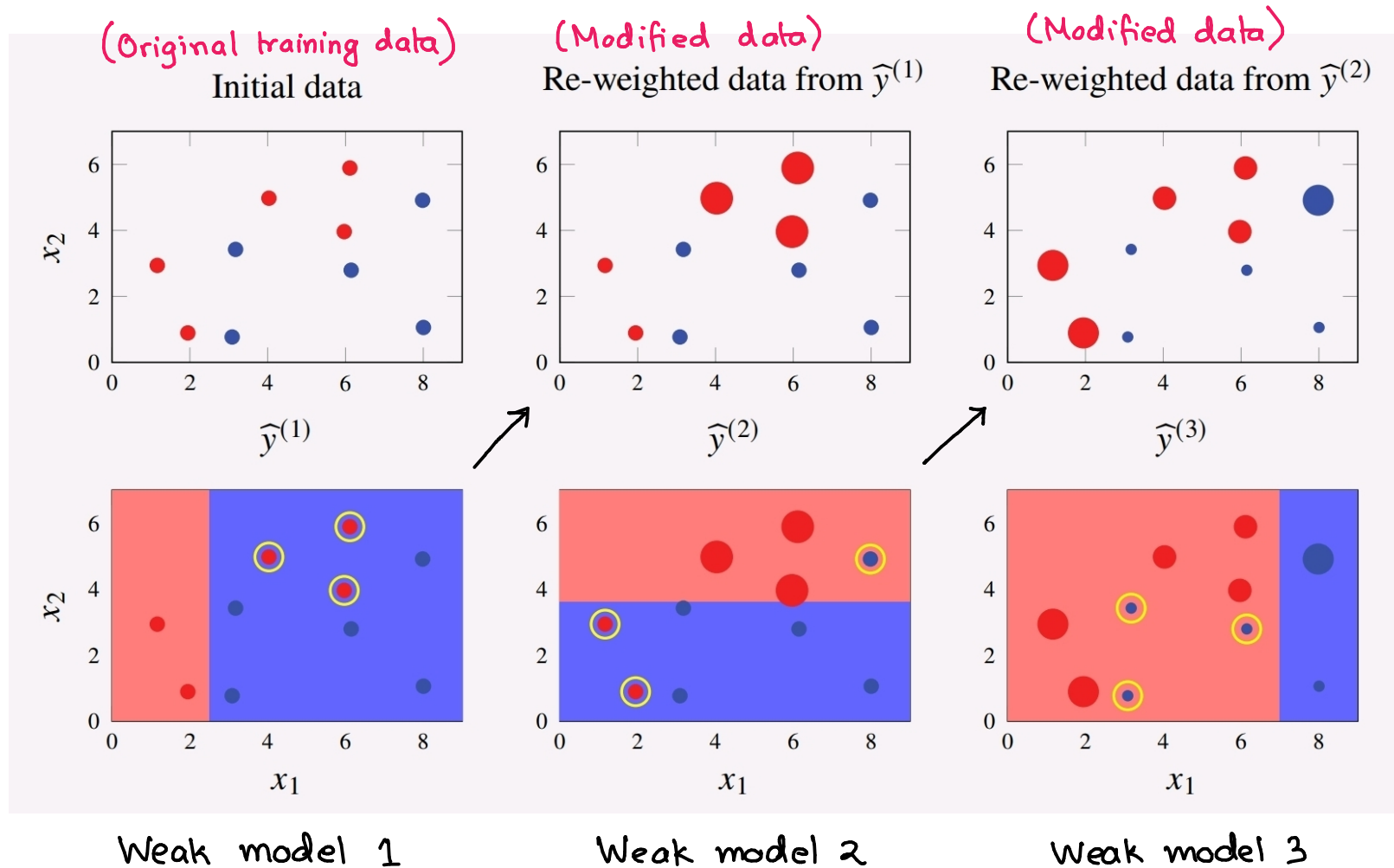
Informally, the sequential construction of ensemble members is done in such a way that each model tries to correct the mistakes made by the previous one

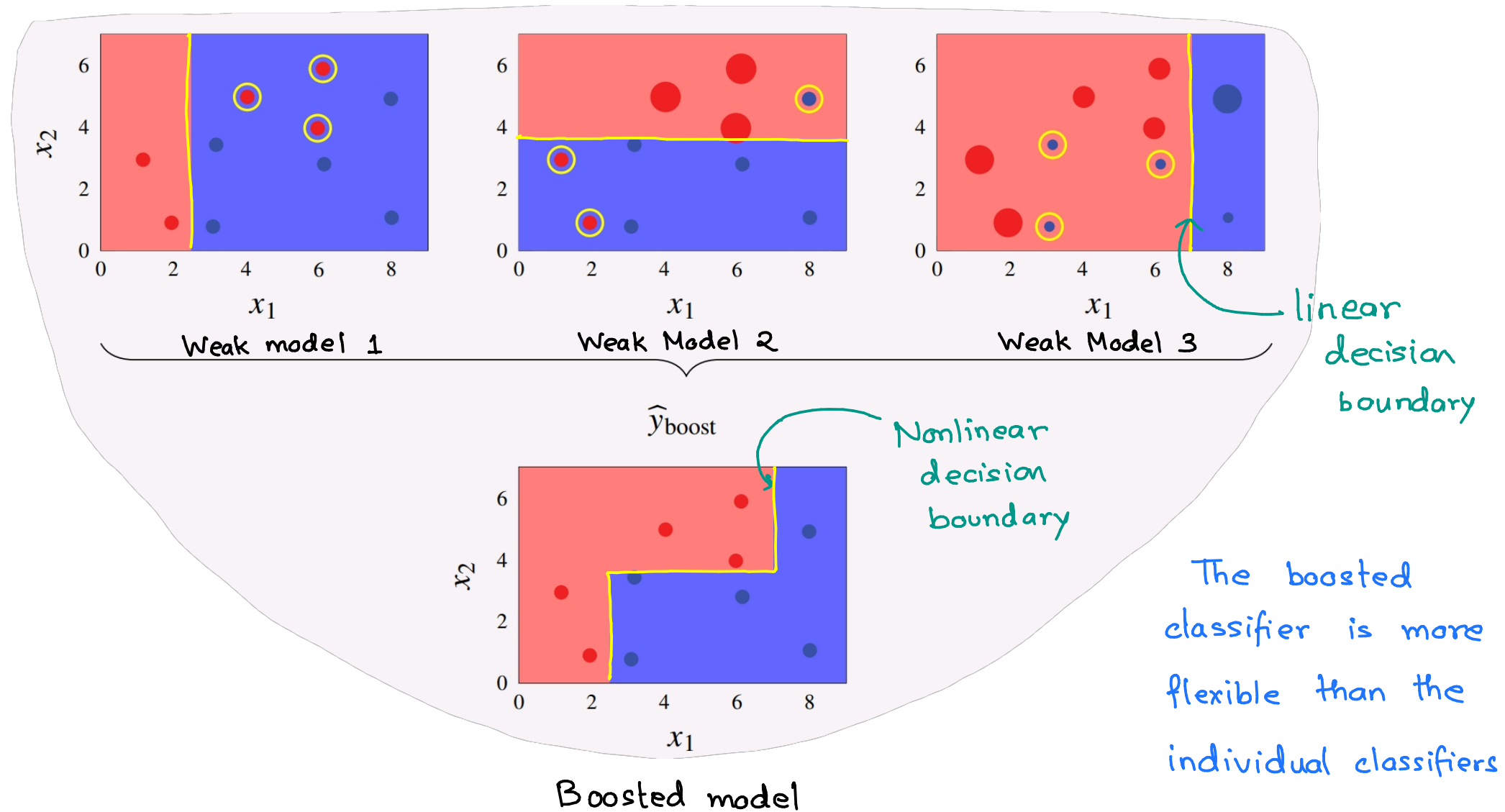


Example of sequential construction in boosting

Consider a binary classification problem with 2D input $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

- There are $N=10$ datapoints, 5 from each class
- A classification tree of depth one (weak model) is used as the **base classifier** (splits into two regions)





The final classifier $\hat{y}_{\text{boost}}(\underline{x}) =$ Weighted majority vote of the three weak decision trees