IMMUNE
CODING INSTITUTE

Supervised Machine Learning

Module 3

IMMUNE
CODING INSTITUTE

Module 3 Summary

SESSION	TITLE	TEACHER
1	ML Foundations	Juan
2	Regression Introduction and Practice	Juan
3	Classification Introduction and Practice	Carlos
4	Feature Engineering and Selection for ML	Carlos
5	Advanced Supervised Models 1	Carlos
6	**Advanced Supervised Models 2**	**Carlos**
7	Hands-on Practice	Carlos

IMMUNE
CODING INSTITUTE

Ensemble Algorithms

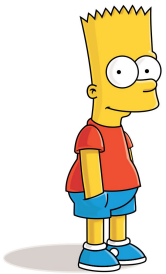
The Wanaka Tree, New Zealand

Outline

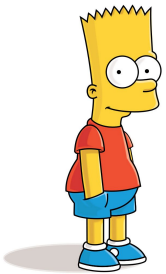
- What is Ensemble Learning?
- Types of Ensemble learning
- Bagging
  - Random Forest
- Boosting
  - Adaboost
  - XGBoost

Ensemble Learning

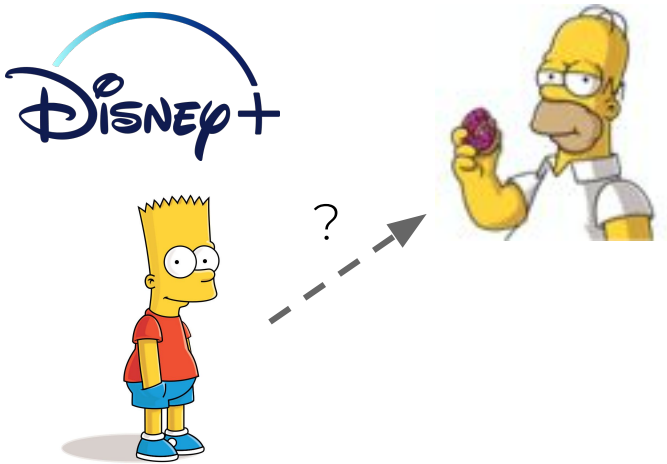
NETFLIX



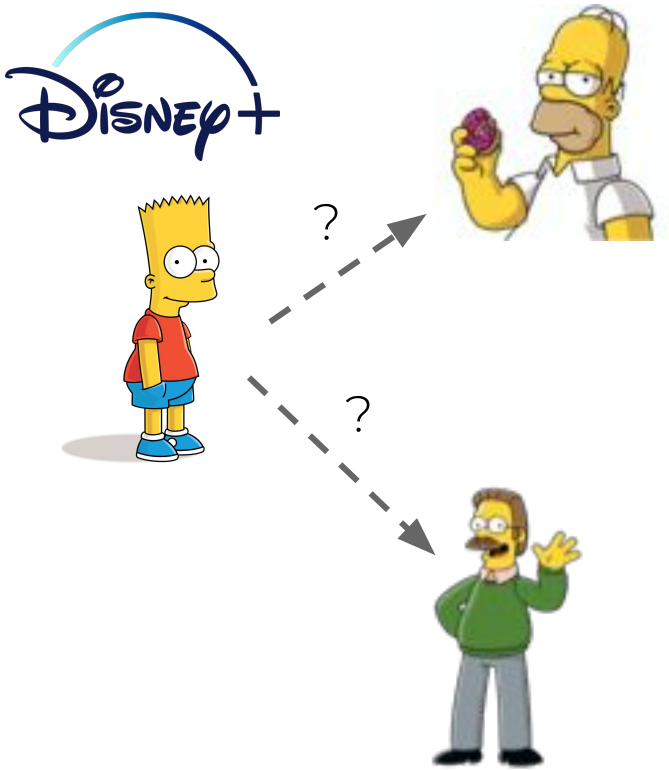
Ensemble Learning



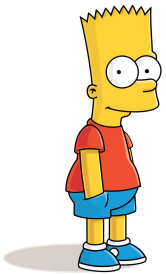
Ensemble Learning



Ensemble Learning

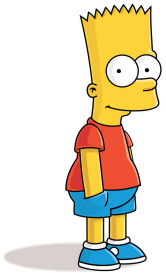


Ensemble Learning



?

Ensemble Learning



M1

M2

M3

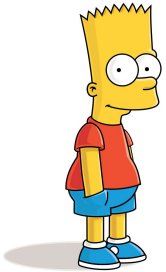
M4

M5

Final Prediction

Combine Model

Ensemble Learning



M1

M2

M3

M4

M5

Weak Learners

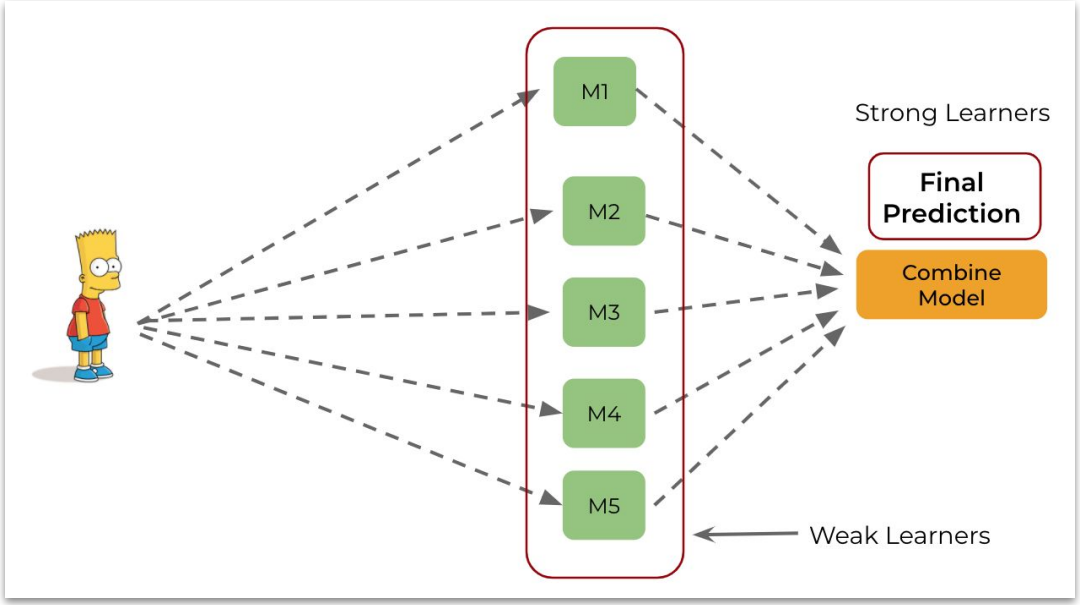
Strong Learners

Combine Model

Final Prediction

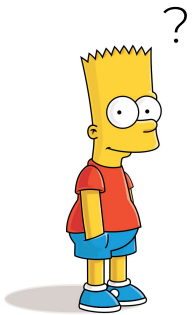
Ensemble Learning Advantages

- Turn weak classifier into strong
- Single model is biased
- Variance error reduces
- Model accuracy increases
- Less overfitting
- High computational cost

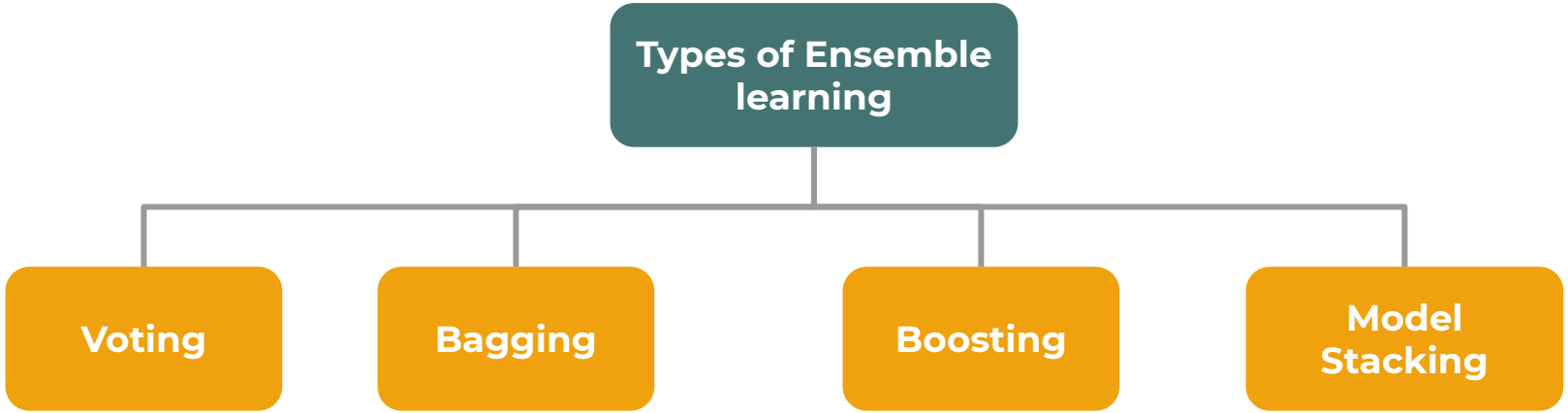


Ensemble Learning Questions

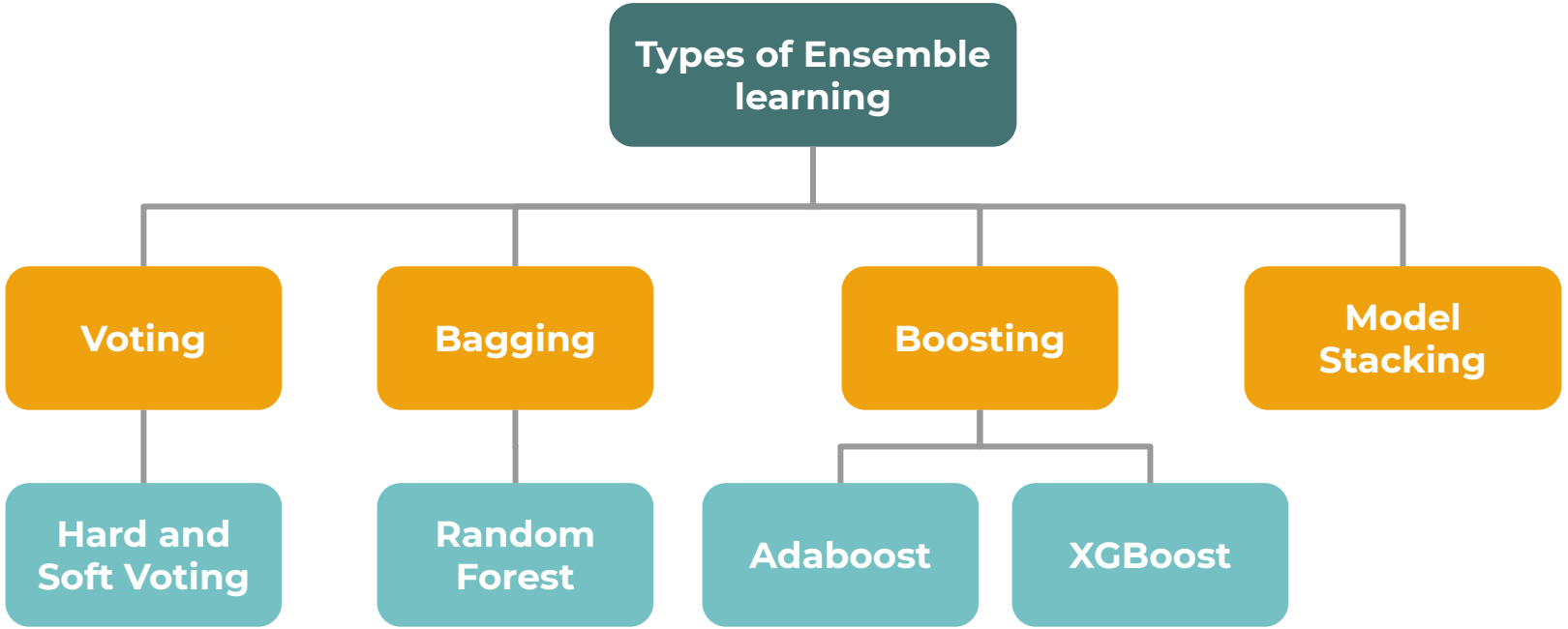
- Do we use all training data on every model?
- What is M1, M2, M3, M4 and M5?
- How to train all model?
- How to combine output from all models?
- Which series should I start watching?



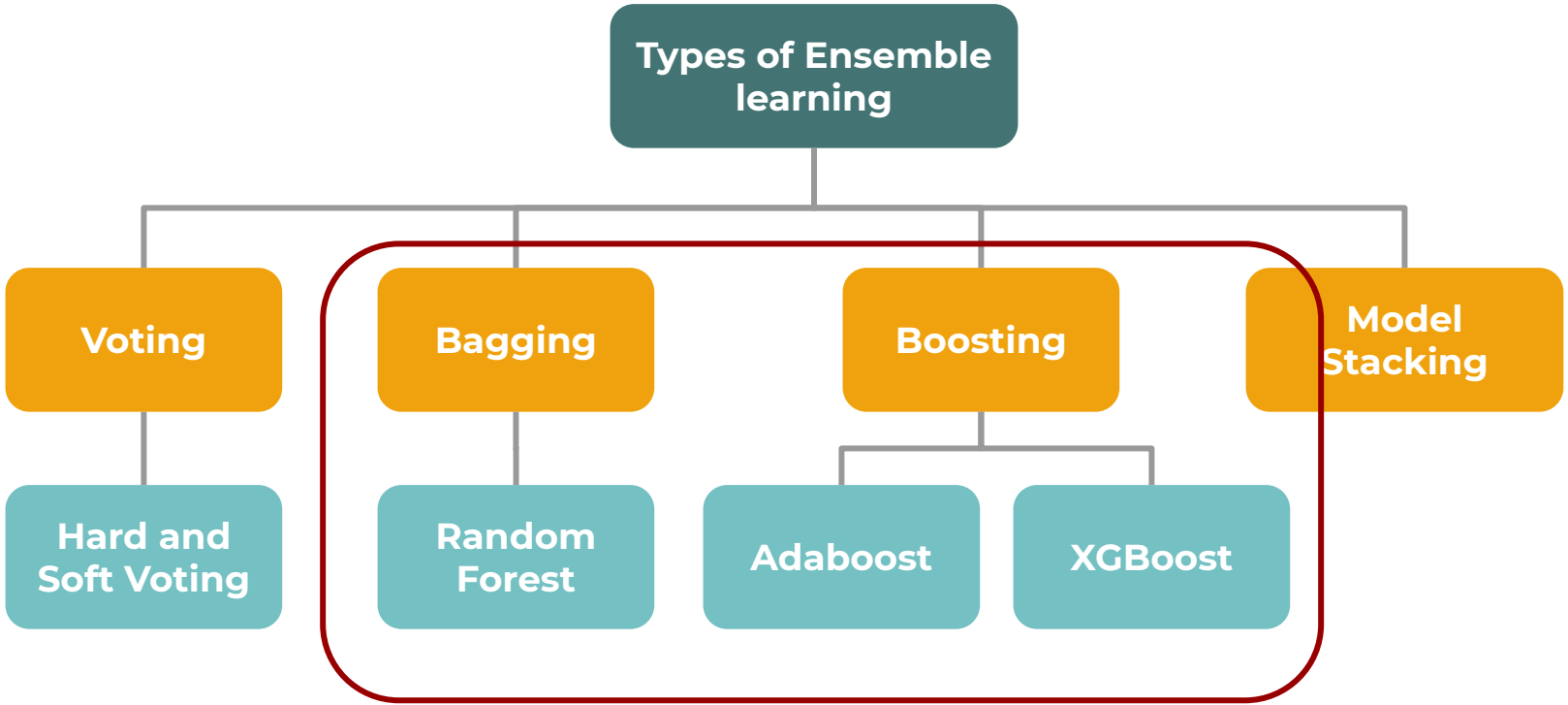
Types of Ensemble learning



Types of Ensemble learning



Types of Ensemble learning

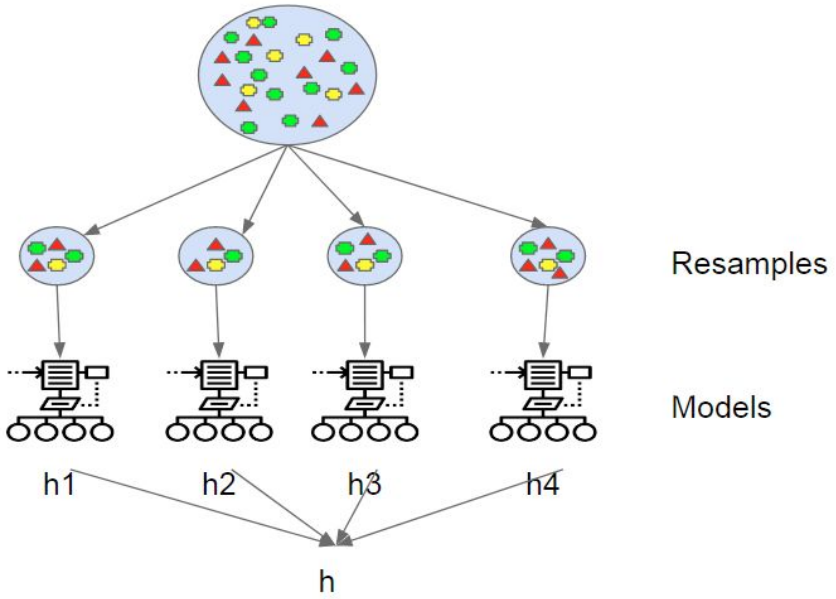


This helps to decrease **variance** i.e. reduce the **overfit**.

Bagging

It refers to (Bootstrap Aggregators).

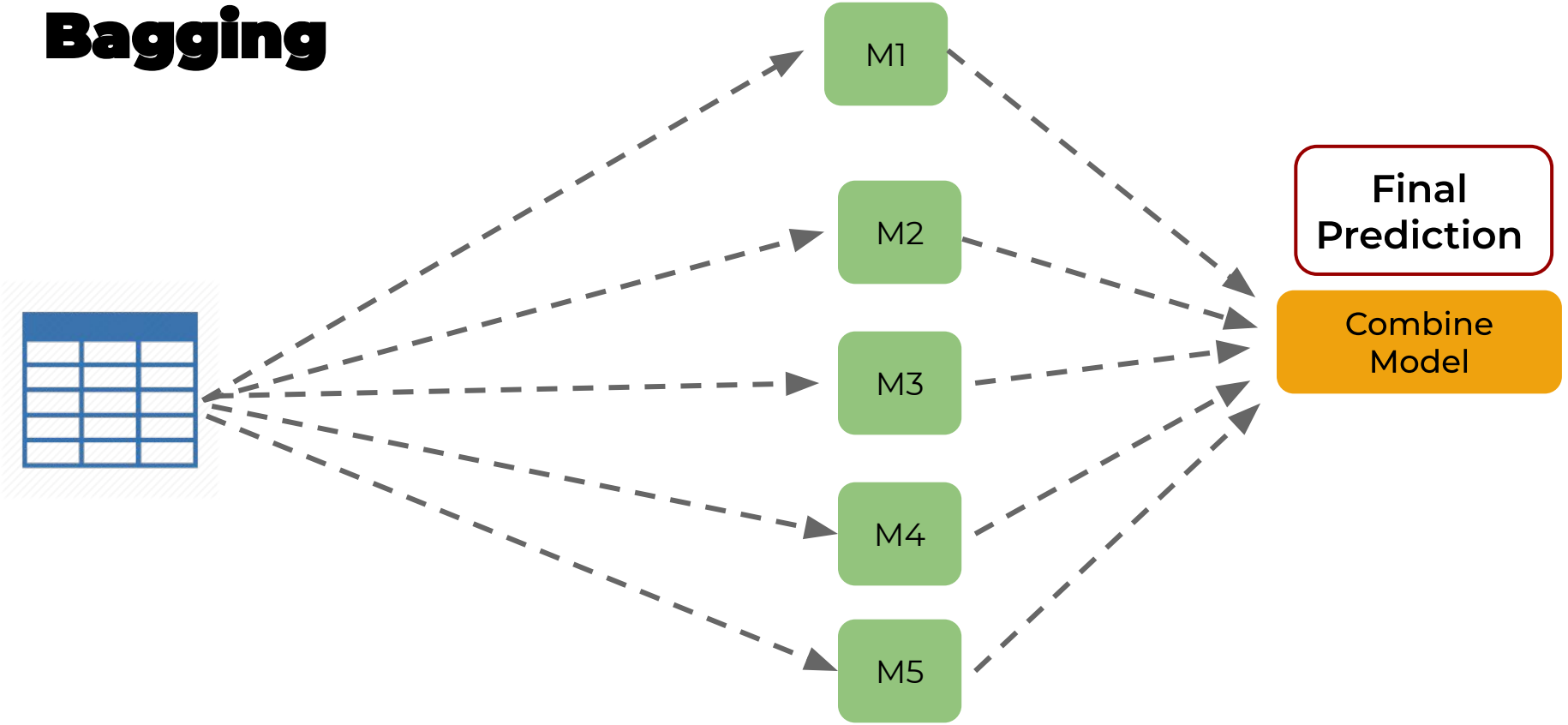
“Bagging predictors is a method for generating multiples versions of a predictor and using these to get an aggregated predictor”.



Once each model has develop a prediction, the models use voting for classification or averaging for regression

ref:
https://link.springer.com/content/pdf/10.1023/A:1018054314350.pdf

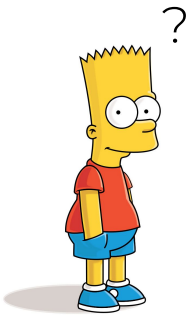
Bagging



Bagging Ensemble

What is M1, M2, M3, M4 and M5?

They are the same type of weak learner ML Algorithm.



How to train all model?

Train in parallel.

How to combine output from all models?

voting classifier or voting regression

Do we use all training data on every model?

The simplest answer is **no**.

Bagging: Select data

- Row Sampling
- Feature Sampling
- Combine above both
  (Row and Feature)

Bagging: Select data

- Row Sampling
- Feature Sampling
- Combine above both (Row and Feature)

Temp	Humidity	Outlook	Football	Rubén?
Mild	80	Sunny	No	Yes
Hot	75	Sunny	Yes	**No**
Hot	77	Overcast	No	Yes
Cool	70	Rain	No	Yes
Cool	72	Overcast	Yes	Yes
Mild	77	Sunny	No	**No**
Cool	70	Sunny	No	Yes
Mild	69	Rain	No	Yes
Mild	65	Sunny	Yes	Yes
Mild	77	Overcast	Yes	Yes
Hot	74	Overcast	No	Yes
MIld	77	Rain	Yes	**No**
Cool	73	Rain	Yes	**No**
Mild	78	Rain	No	Yes

IMMUNE
CODING INSTITUTE

Bagging: Select data

- **Row Sampling**
- Feature Sampling
- Combine above both (Row and Feature)

	Temp	Humidity	Outlook	Football	Rubén?
M1	Mild	80	Sunny	No	Yes
	Hot	75	Sunny	Yes	**No**
M2	Hot	77	Overcast	No	Yes
	Cool	70	Rain	No	Yes
	Cool	72	Overcast	Yes	Yes
	Mild	77	Sunny	No	**No**
M3	Cool	70	Sunny	No	Yes
	Mild	69	Rain	No	Yes
M4	Mild	65	Suny	Yes	Yes
	Mild	77	Overcast	Yes	Yes
	Hot	74	Overcast	No	Yes
M5	MIld	77	Rain	Yes	**No**
	Cool	73	Rain	Yes	**No**
	Mild	78	Rain	No	Yes

IMMUNE CODING INSTITUTE

Bagging: Select data

- Row Sampling
- **Feature Sampling**
- Combine above both (Row and Feature)

M1

M2

M3

Temp	Humidity	Outlook	Football	Rubén?
Mild	80	Sunny	No	Yes
Hot	75	Sunny	Yes	**No**
Hot	77	Overcast	No	Yes
Cool	70	Rain	No	Yes
Cool	72	Overcast	Yes	Yes
Mild	77	Sunny	No	**No**
Cool	70	Sunny	No	Yes
Mild	69	Rain	No	Yes
Mild	65	Sunny	Yes	Yes
Mild	77	Overcast	Yes	Yes
Hot	74	Overcast	No	Yes
MIld	77	Rain	Yes	**No**
Cool	73	Rain	Yes	**No**
Mild	78	Rain	No	Yes

What is weak Learner?

- Linear Regression
- Logistic Regression
- KNN
- Neural Network
- Polynomial Regression
- Naïve Bayes
- Decision Tree

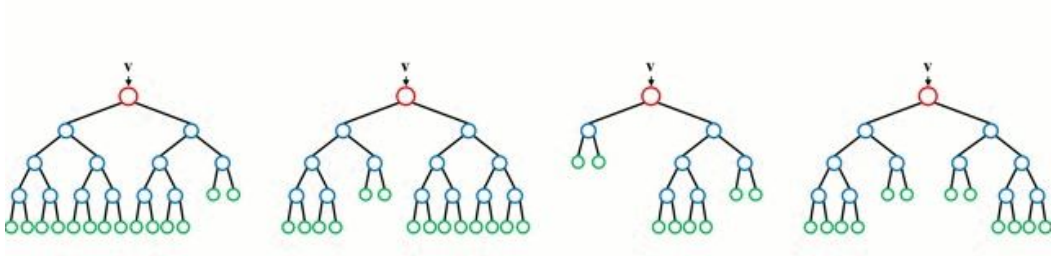
IMMUNE CODING INSTITUTE

Random Forest

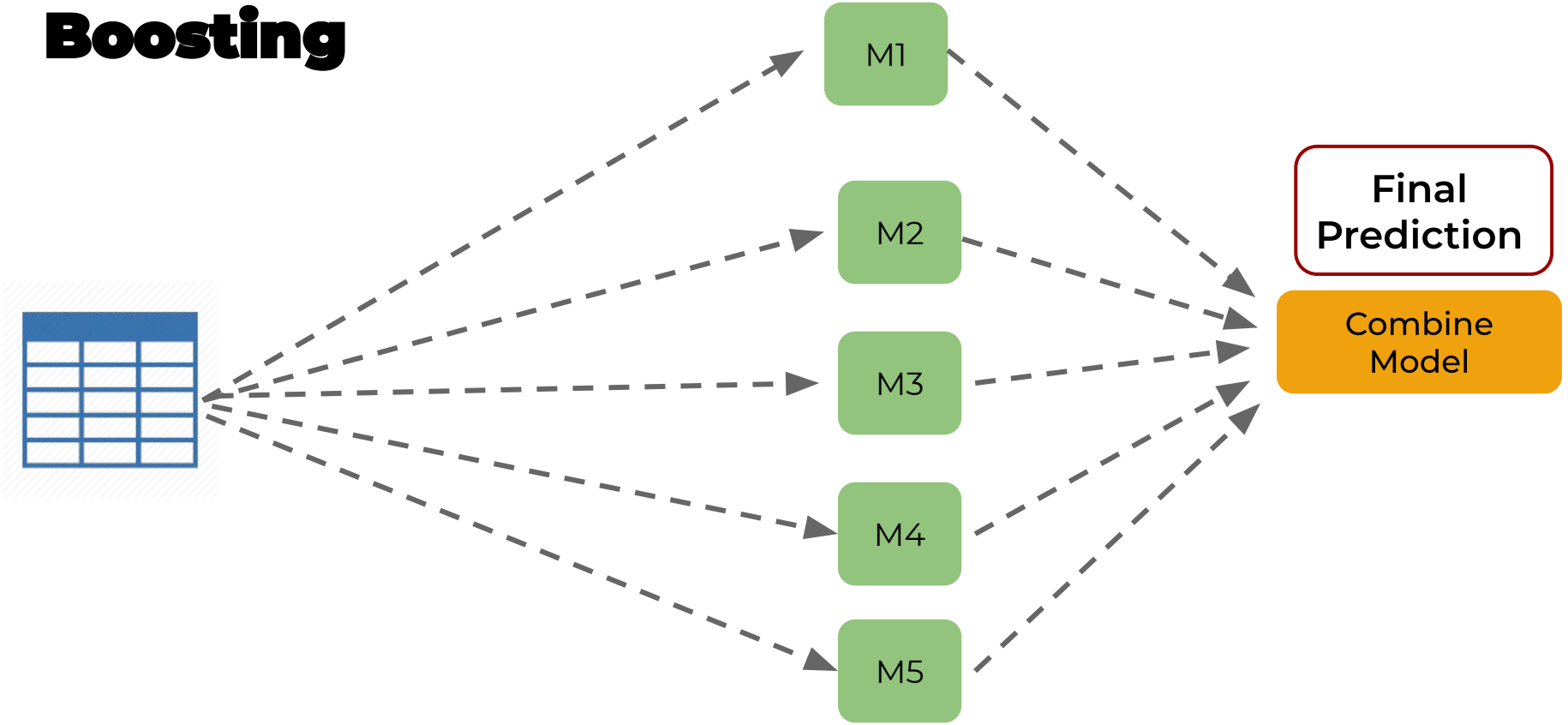
To improve performance, we can use many trees with a random sample of features chosen as the split.

- A new random sample of feature is chosen for **every single tree at every single split.**

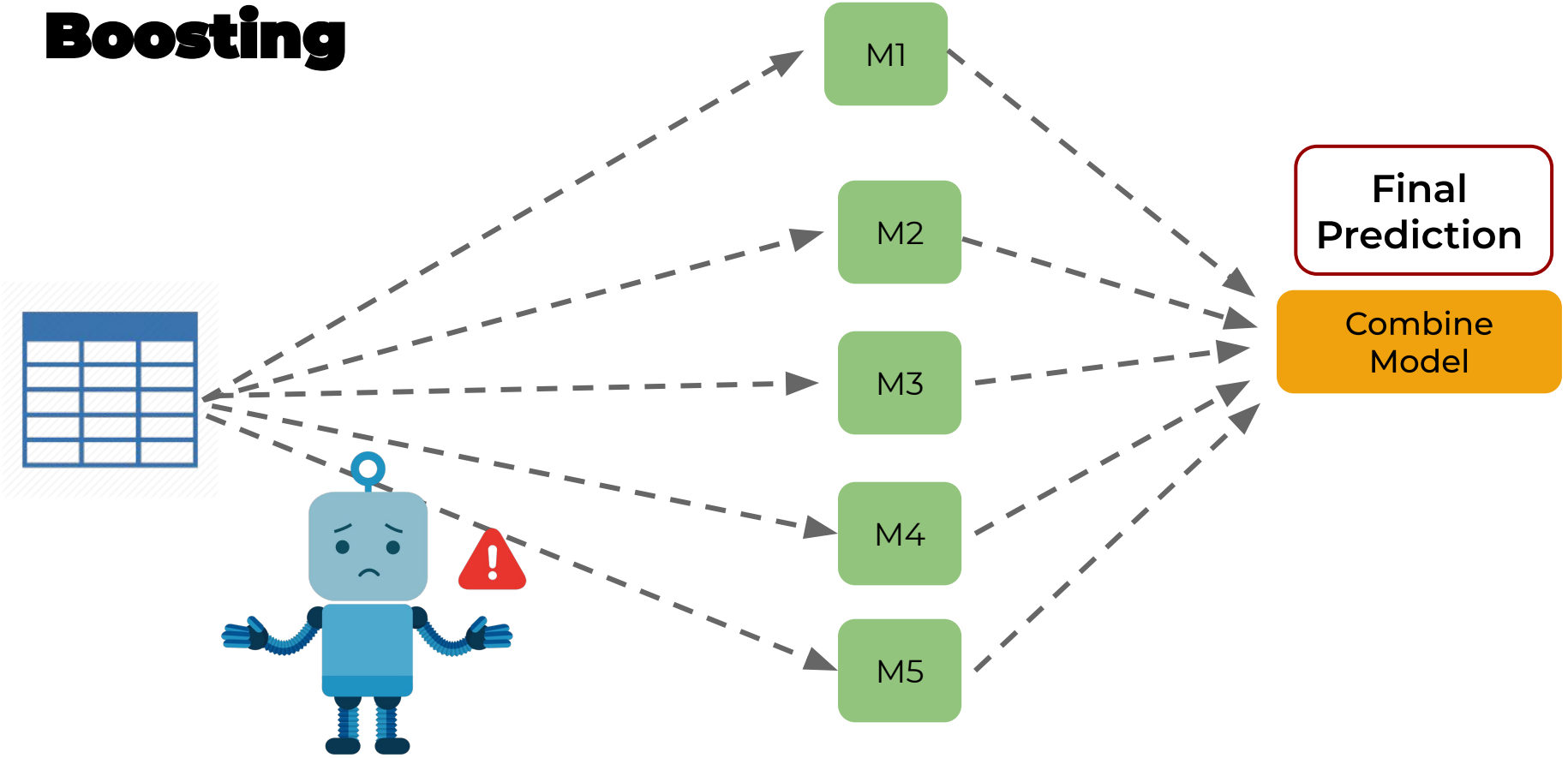
- For **classification,** *m* (sample of features) is typically chosen to be the square root of the total of features. .



Boosting

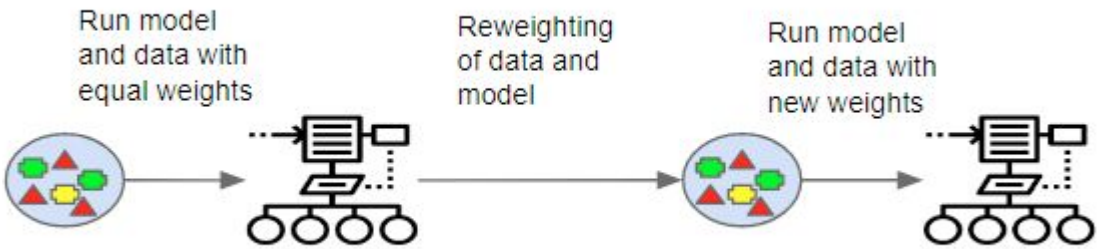


Boosting



Boosting

It refers to a group of algorithms that utilize weighted averages to make weak learners into stronger learners.

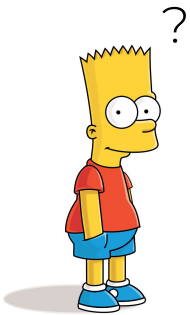


The models with better outcomes have a stronger pull on the final output.

Boosting Ensemble

What is M1, M2, M3, M4 and M5?

They are the same type of weak learner ML Algorithm.



How to train all model?

Train in sequential.

How to combine output from all models?

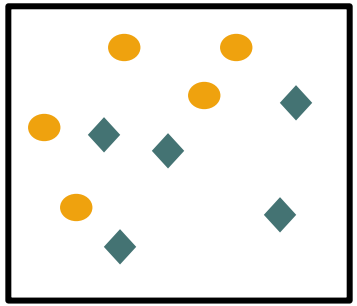
voting classifier or voting regression

Do we use all training data on every model?

Based on sampling and especially from previous mistakes

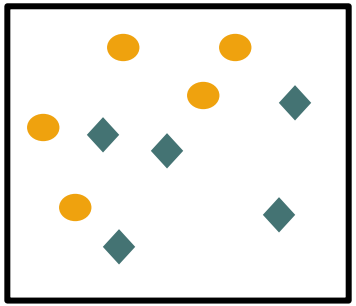
AdaBoost

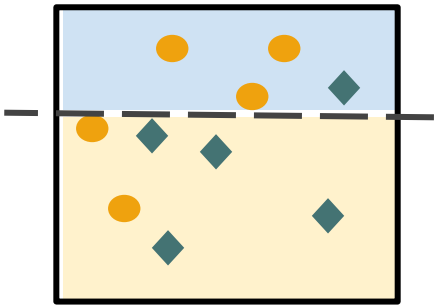
D1



AdaBoost

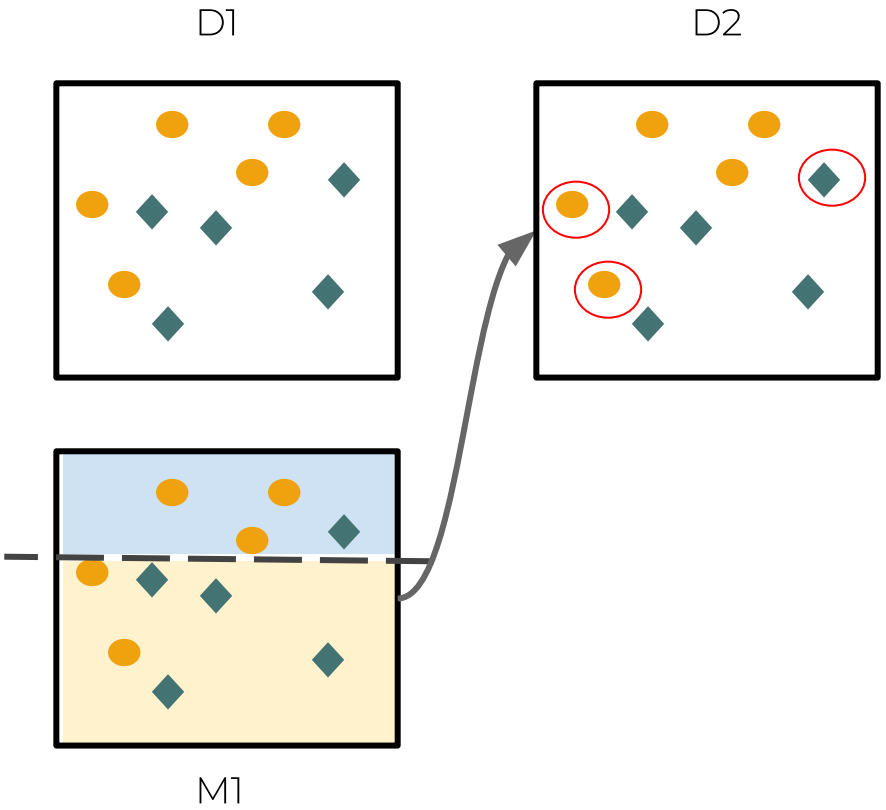
D1





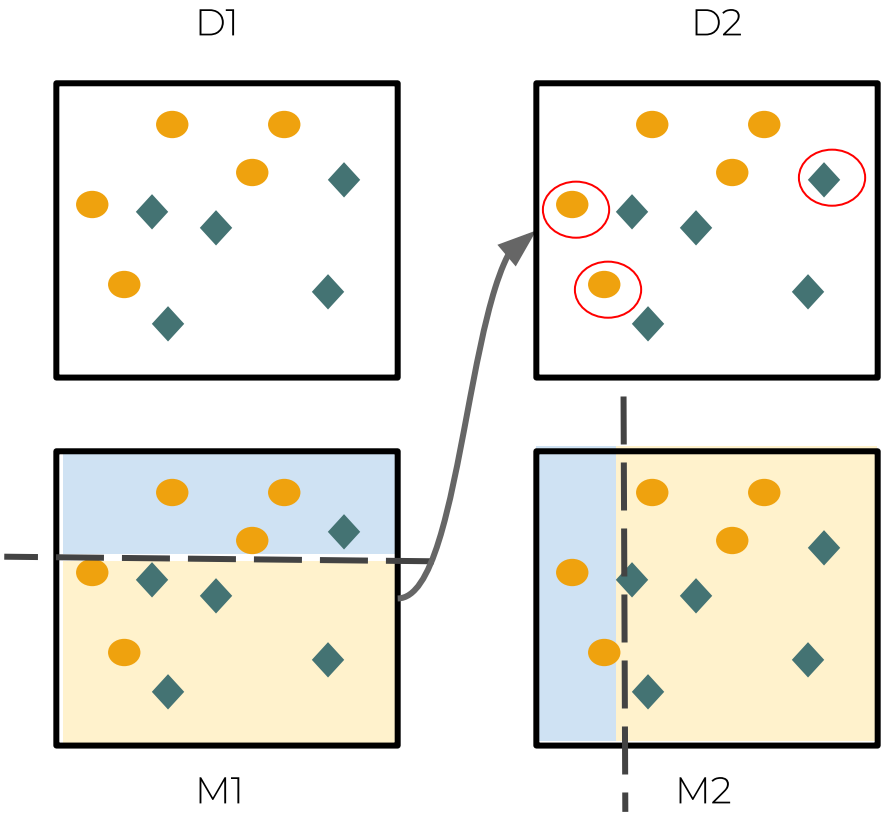
M1

AdaBoost

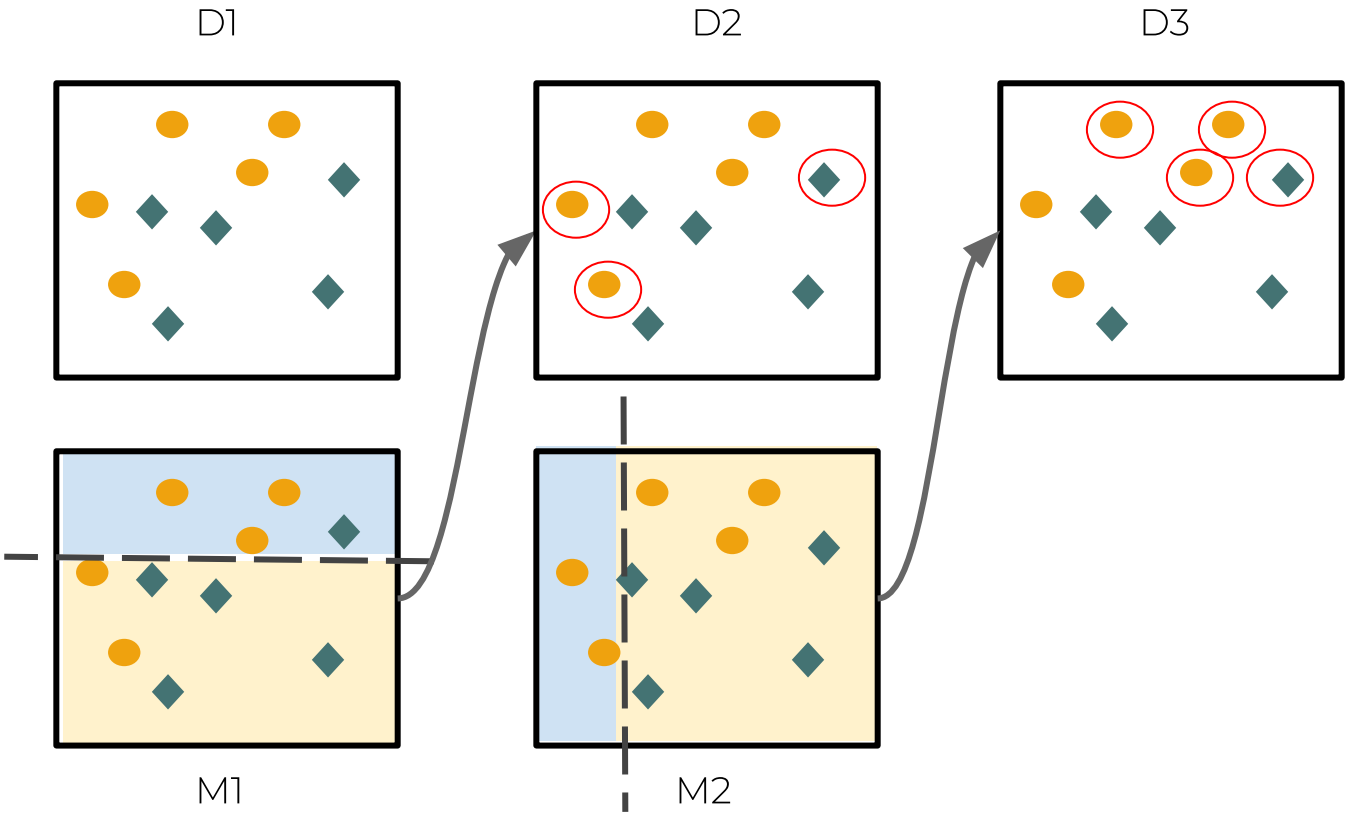


IMMUNE CODING INSTITUTE

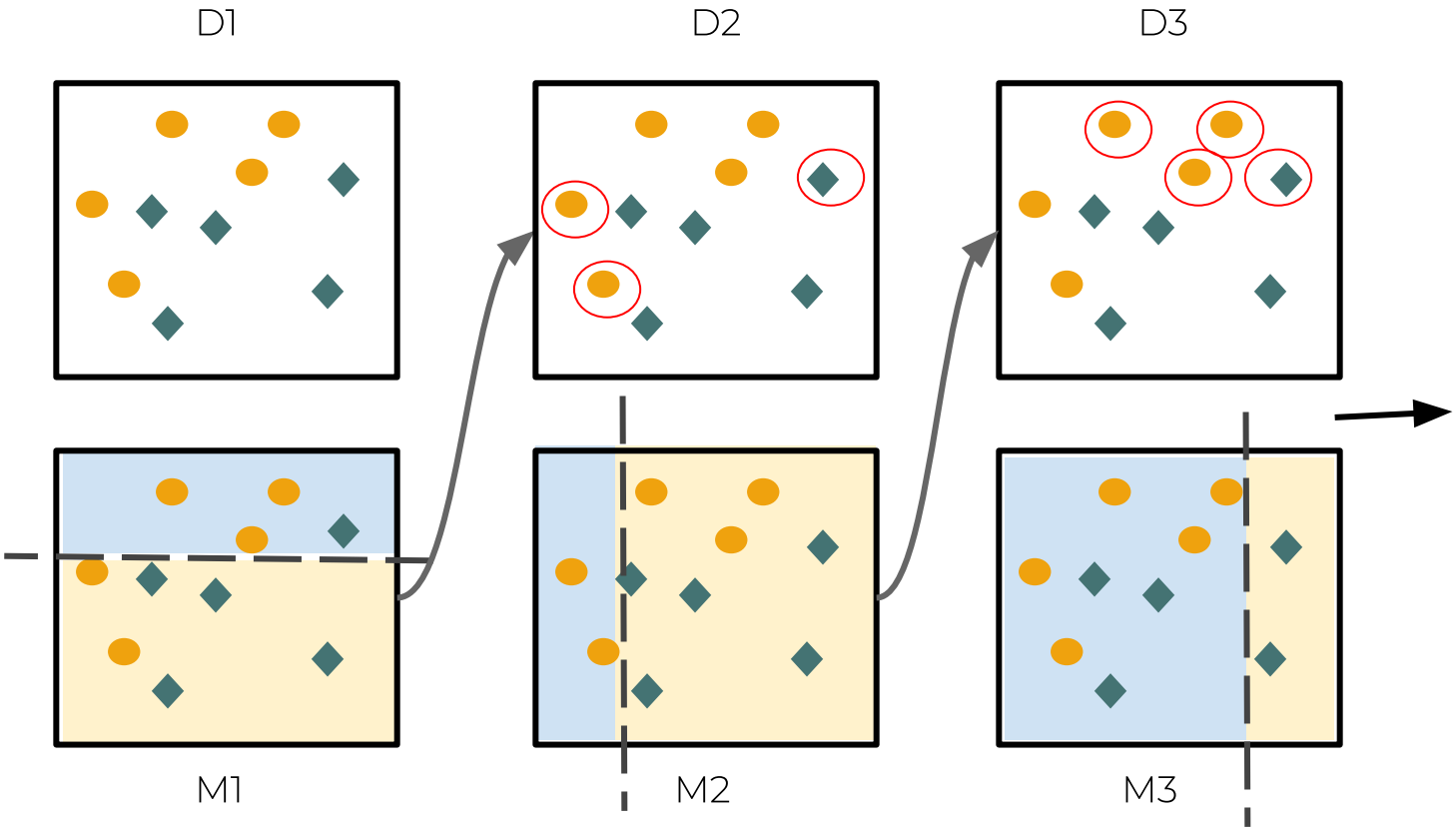
AdaBoost



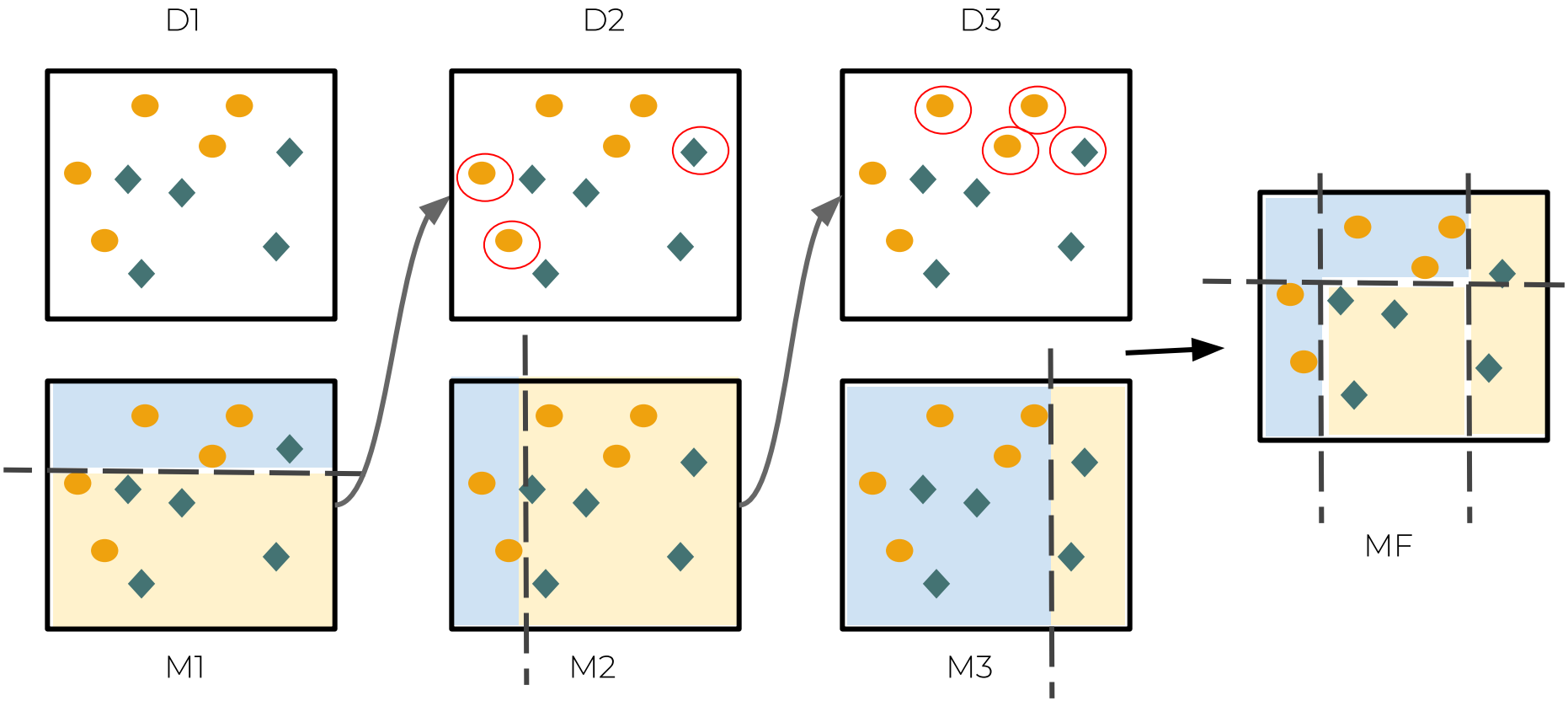
Adaboost



AdaBoost



Adaboost



AdaBoost

Pro

- Easy to implement
- The only parameter is the number of iterations
- Computationally fast

Cons

- Complex weak learners may overcome overfitting
- Simplex weak learners may overcome low margin and overfitting
- Noise sensitive

XGBoost
(Extreme Gradient Boosting Tree)

- Faster Executing Speed
- Better Performance
  - Many ML competitions (Kaggle) won XGBoost
- Parallel Processing
- XGBoost can handle missing value
- It has Built-in cross validation. Regularization
- It can do incremental training (for large Dataset)

XGBoost
Paramaters

Some key parameters to consider:

- **max_depth:** more complex problems, deeper the tree
- **subsample:** percentage of rows taken to build the tree. (Take values from 0.7 to 1)
- **objective:** classification or regression algorithm
- **n_estimators:** the number of runs XGBoost will try to learn
- **learning_rate:** learning speed
- **early_stopping_rounds:** overfitting prevention, stop early if improvement in learning
- **subsample:** ratio of the training instances. Setting it to 0.5 means that XGBoost would randomly sample half of the training data prior to growing trees. and this will prevent overfitting.
- **eval_set:** validation/test set to check for errors.

XGBoost
Paramaters

Some key parameters to consider (Continue):

- **gamma:** more complex classification task, deeper the tree
- **colsample_tree:** number of columns used by each tree
- **objective:** classification or regression algorithm
- **n_estimators:** the number of runs XGBoost will try to learn
- **learning_rate:** learning speed (usually between 0.1 and 0.01). One strategy could be to decrease incrementally the learning rate while increasing the number of trees.
- **early_stopping_rounds:** overfitting prevention, stop early if improvement in learning
- **alpha:** L1 regularization term on weights. Increasing this value will make model more conservative.
- **lambda:** L2 regularization term on weights. Increasing this value will make model more conservative.

XGBoost
Paramaters

The complete list of XGBoost parameters can be found here:

https://xgboost.readthedocs.io/en/latest/parameter.html


IMMUNE
CODING INSTITUTE