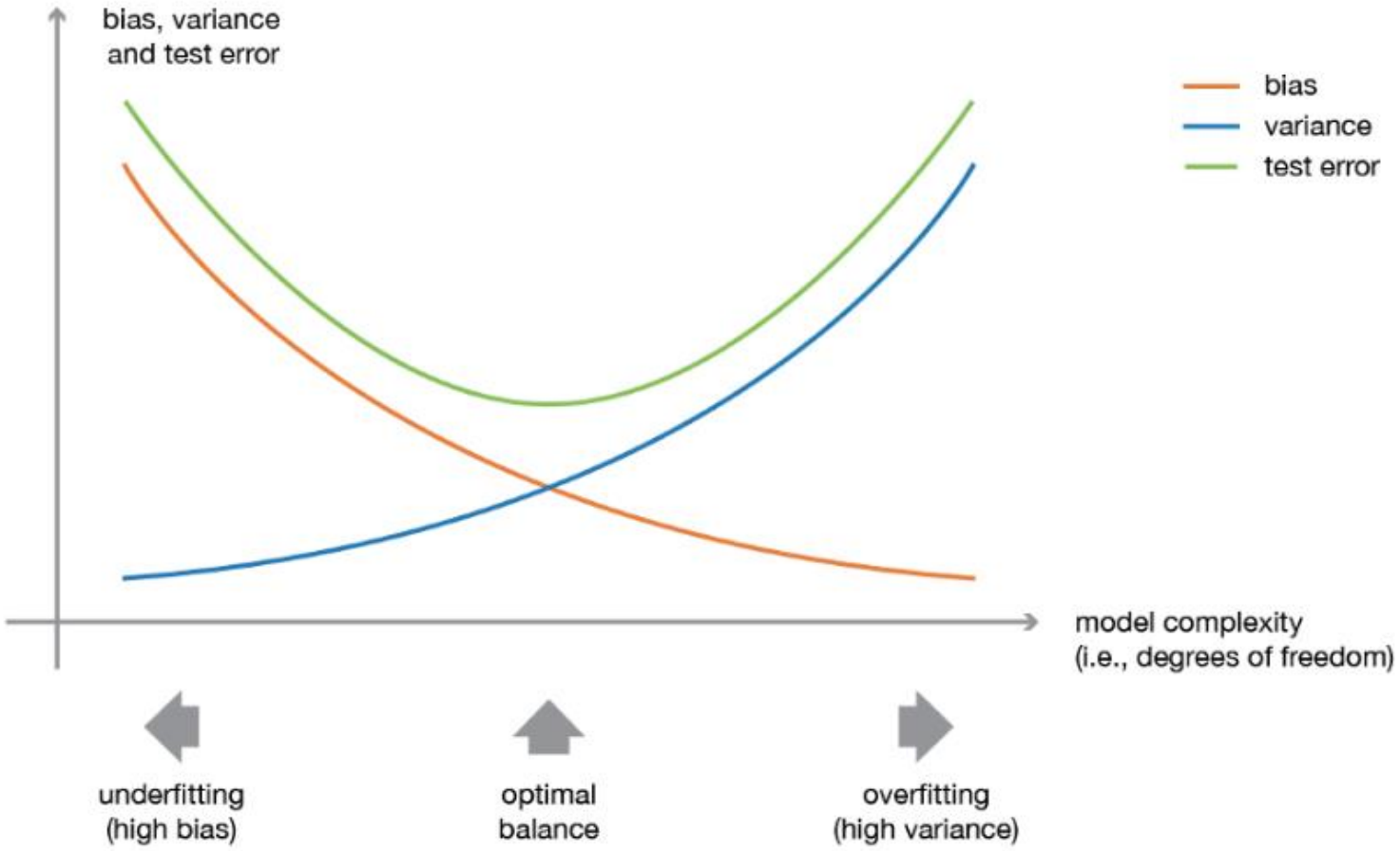
Ensemble Learning

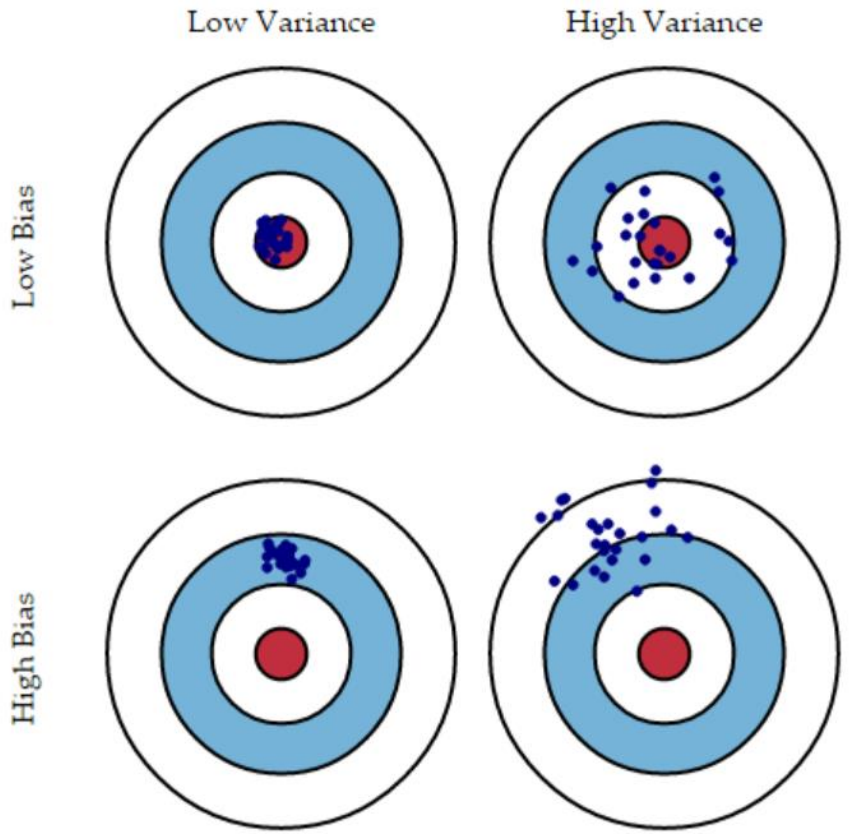
Ensemble learning

Combine multiple weak algorithms to form a strong model.

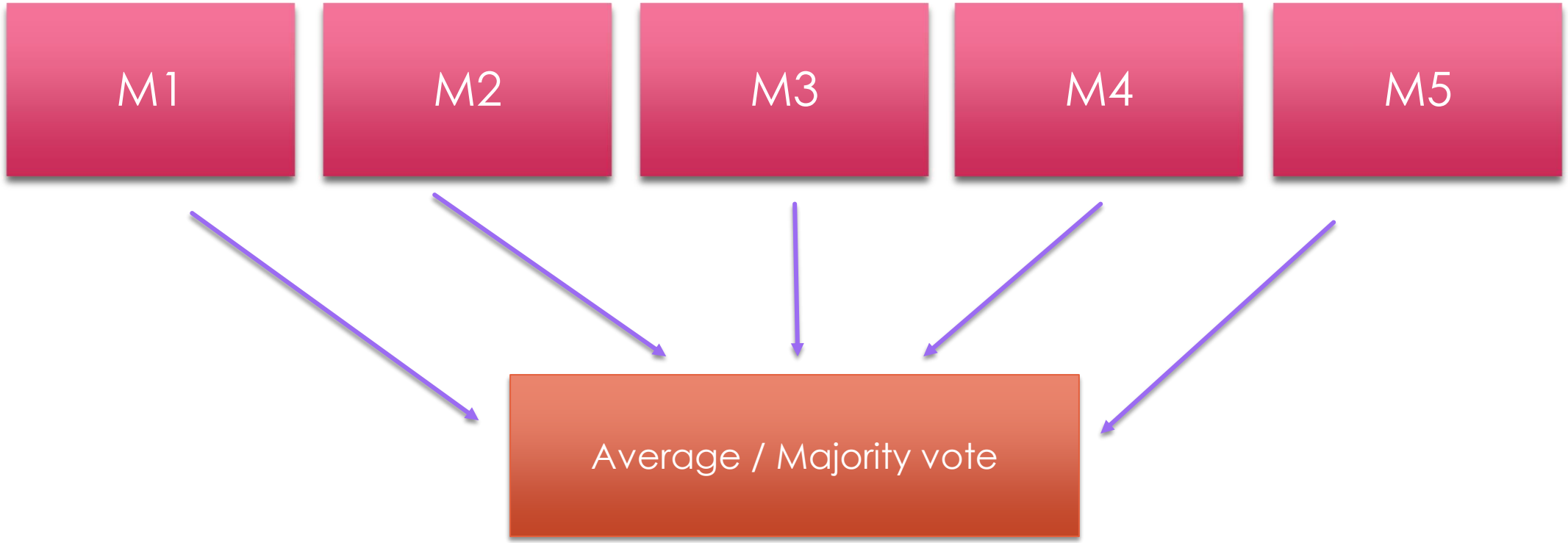
Using ensemble methods allows to produce better predictions compared to a single model



Bias Variance Trade off


Low Variance
High Variance
Low Bias
High Bias

BAGGING



MG ANALYTICS
MG ANALYTICS

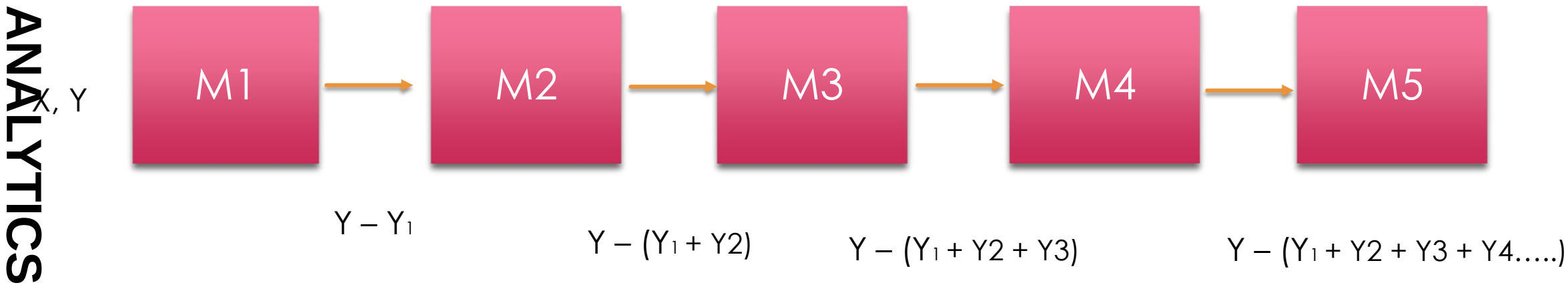
BAGGING

- Bagging consists of building different parallel models
- Each model has different set of input samples helping to create unique models.
- Result is generated by taking the average of those predictions.
- This is useful when you want to decrease the variance while keeping the bias same.
- It works this way because bagging is kind of an averaging technique.
- Bagging helps when it is applied to an over fitted model by decreasing variance error.
- It does not help much with models which have high bias.

BOOSTING

MG ANALYTICS

MG ANALYTICS



BOOSTING

- Boosting consists of building different sequential models one after another.
- Each model has same X as input.
- First model predicts Y.
- then onwards models predict the error value left prom previous model until the error is 0.
- **Decreases the bias error** and builds strong predictive models.
- They may sometimes over fit on the training data.
- For each iteration, boosting updates the weights of the samples, so that, samples that are misclassified by the ensemble can have a higher weight, and therefore, higher probability of being selected for training the new classifier.

- Bagging will mainly focus at getting an ensemble model with less variance than its components.
- Boosting and stacking will mainly try to produce strong models less biased than their components (even if variance can also be reduced).

Stacking

STACKING ALLOWS TO CREATE A LINEAR COMBINATION OF MULTIPLE NON LINEAR MODELS.

Stacking creates a hierarchy of models using the outputs from previous layers.

Base Models / Weak Learners

- Building blocks for designing more complex models.
- do not perform well :
  - high bias or too much variance.
- Ensemble methods try reduce bias and/or variance of such weak learners by combining several of them.
- Create a **strong learner** (or **ensemble model**) that achieves better performances.

Which model can be a WL?

- Linear model with high penalty
- Linear model with subset of variables
- Dtree stumps / Shallow trees.

MG ANALYTICS
MG ANALYTICS

Why to use WL?

Cannot learn niche patterns hence cannot overfit.

A combination of these will capture a general pattern.

As results would be combined will not be impacted by noise.