SNS COLLEGE OF TECHNOLOGY

Coimbatore – 35

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with ‘A++’ Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

MACHINE LEARNING

Ensemble methods bagging and boosting in machine learning

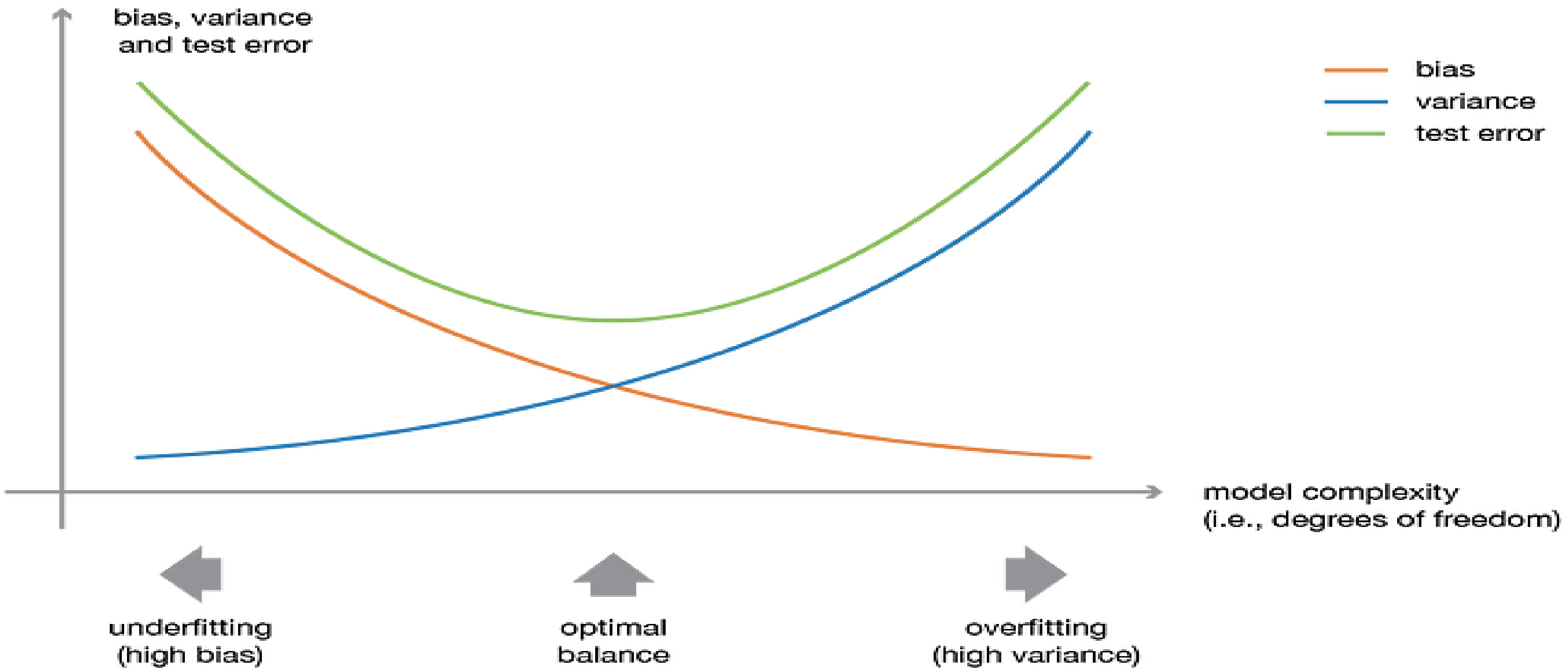
Ensemble learning is a machine learning paradigm where multiple models (often called "weak learners") are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.

•

Single weak learner

- In machine learning, no matter if we are facing a classification or a regression problem, the choice of the model is extremely important to have any chance to obtain good results. This choice can depend on many variables of the problem: quantity of data, dimensionality of the space, distribution hypothesies.

A low bias and a low variance, although they most often vary in opposite directions, are the two most fundamental features expected for a model. Indeed, to be able to “solve” a problem, we want our model to have enough degrees of freedom to resolve the underlying complexity of the data we are working with, but we also want it to have not too much degrees of freedom to avoid high variance and be more robust. This is the well known **bias-variance tradeoff.**


bias, variance
and test error
bias
variance
test error
model complexity
(i.e., degrees of freedom)
underfitting
(high bias)
optimal
balance
overfitting
(high variance)

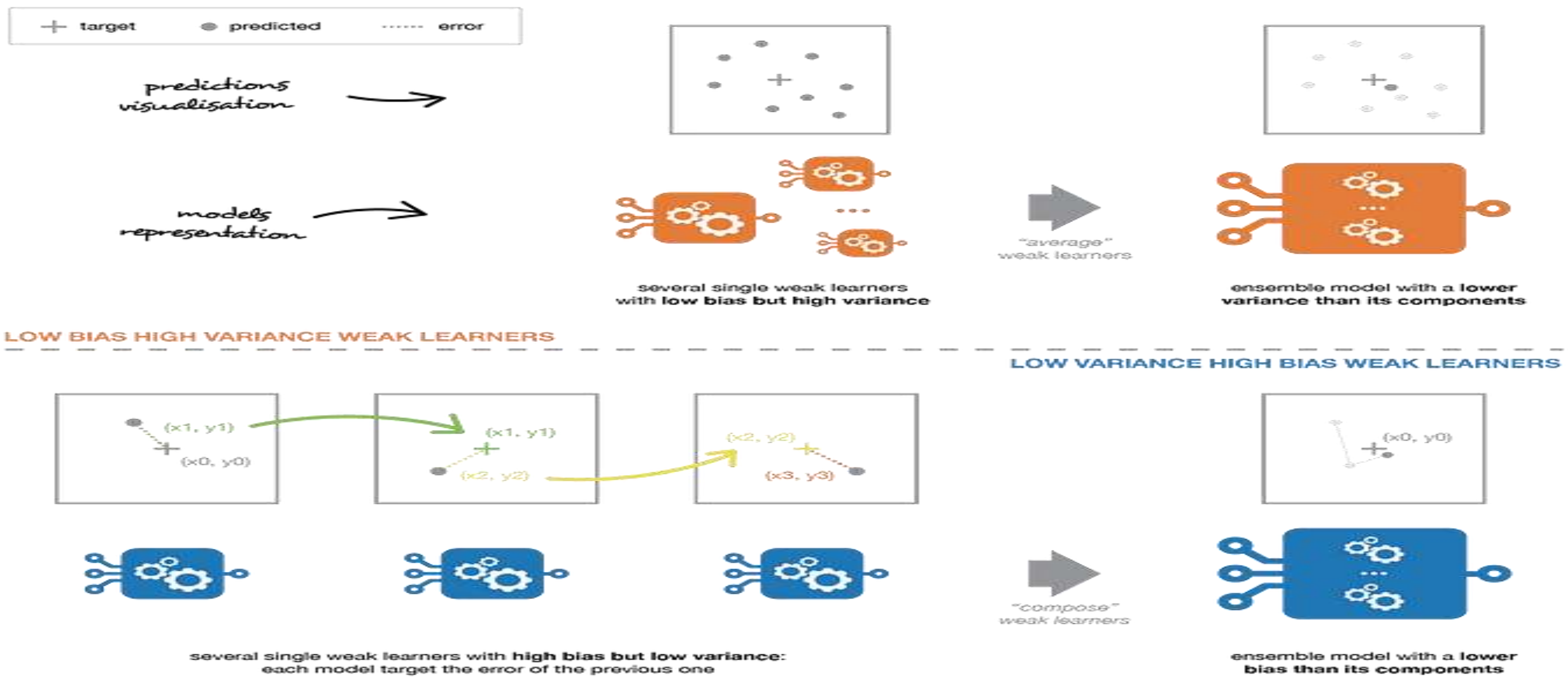
- In ensemble learning theory, we call **weak learners** (or **base models**) models that can be used as building blocks for designing more complex models by combining several of them. Most of the time, these basics models perform not so well by themselves either because they have a high bias (low degree of freedom models, for example) or because they have too much variance to be robust (high degree of freedom models, for example). Then, the idea of ensemble methods is to try reducing bias and/or variance of such weak learners by combining several of them together in order to create a **strong learner** (or **ensemble model**) that achieves better performances.

Combine weak learners

- In order to set up an ensemble learning method, we first need to select our base models to be aggregated. Most of the time (including in the well known bagging and boosting methods) a single base learning algorithm is used so that we have homogeneous weak learners that are trained in different ways. The ensemble model we obtain is then said to be “homogeneous”. However, there also exist some methods that use different type of base learning algorithms: some heterogeneous weak learners are then combined into an “heterogeneous ensembles model”.

- One important point is that our choice of weak learners should be **coherent with the way we aggregate these models**. If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.

- **bagging**, that often considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process
- **boosting**, that often considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy
- **stacking**, that often considers heterogeneous weak learners, learns them in parallel and combines them by training a meta-model to output a prediction based on the different weak models predictions


target
predicted
error
predictions visualisation
models representation
several single weak learners
with low bias but high variance
"average"
weak learners
ensemble model with a lower
variance than its components
LOW BIAS HIGH VARIANCE WEAK LEARNERS
LOW VARIANCE HIGH BIAS WEAK LEARNERS
(x1, y1)
(x0, y0)
(x1, y1)
(x2, y2)
(x2, y2)
(x3, y3)
(x0, y0)
"compose"
weak learners
several single weak learners with high bias but low variance:
each model target the error of the previous one
ensemble model with a lower
bias than its components

Bagging

- Bagging is used when our objective is to reduce the variance of a decision tree. Here the concept is to create a few subsets of data from the training sample, which is chosen randomly with replacement. Now each collection of subset data is used to prepare their decision trees thus, we end up with an ensemble of various models. The average of all the assumptions from numerous tress is used, which is more powerful than a single decision tree.

- **Random Forest** is an expansion over bagging. It takes one additional step to predict a random subset of data. It also makes the random selection of features rather than using all features to develop trees. When we have numerous random trees, it is called the Random Forest.

- These are the following steps which are taken to implement a Random forest:
- Let us consider **X** observations **Y** features in the training data set. First, a model from the training data set is taken randomly with substitution.
- The tree is developed to the largest.
- The given steps are repeated, and prediction is given, which is based on the collection of predictions from n number of trees.

BOOSTING

- Boosting is another ensemble procedure to make a collection of predictors. In other words, we fit consecutive trees, usually random samples, and at each step, the objective is to solve net error from the prior trees.
- f a given input is misclassified by theory, then its weight is increased so that the upcoming hypothesis is more likely to classify it correctly by consolidating the entire set at last converts weak learners into better performing models.
- **Gradient** Boosting is an expansion of the boosting procedure.

1. Gradient Boosting = Gradient Descent + Boosting

- It utilizes a gradient descent algorithm that can optimize any differentiable loss function. An ensemble of trees is constructed individually, and individual trees are summed successively. The next tree tries to restore the loss (It is the difference between actual and predicted values).

- **Advantages of using Gradient Boosting methods:**

1. It supports different loss functions
2. It works well with interactions.

Disadvantages of using a Gradient Boosting methods:

1. It requires cautious tuning of different hyper-parameters

BAGGING VS BOOSTING

agging	Boosting
Various training data subsets are randomly drawn with replacement from the whole training dataset.	Each new subset contains the components that were misclassified by previous models.
Bagging attempts to tackle the over-fitting issue.	Boosting tries to reduce bias.
If the classifier is unstable (high variance), then we need to apply bagging.	If the classifier is steady and straightforward (high bias), then we need to apply boosting.
Every model receives an equal weight.	Models are weighted by their performance.
Objective to decrease variance, not bias.	Objective to decrease bias, not variance.
It is the easiest way of connecting predictions that belong to the same type.	It is a way of connecting predictions that belong to the different types.
Every model is constructed independently.	New models are affected by the performance of the previously developed model.