{ Ensemble deep Learning }

(Review)

Author : MA Ganaie
Minghui Hu
A.K. Malik
M Tanveer
PN Suganthan

Ensemble learning defined as a combination of several individual model, The combination increase the Robustness, Generalised Performance of the models. Again deep learning techniques are used for getting the better result when we are considering both data centric and model centric approach.

The Deep Ensemble Learning models combines all the advantages of these two models to increase the performance of a model.

Again when we are using supervised learning Approach then we are focused on predictions, Combination of several different prediction from different models has proven as a elegent approach to get final product that is known as ensemble model.

Ensemble Deep Learning

- **Theory**
  - Bias Variance Decomposition
  - Statistical Decomposition & Representational
  - Diversity
- **Ensemble Strategy**
  - Classical method
    - Bagging
    - Boosting
    - Stacking
  - General method
    - Negative Correlation Learning
    - Explicit
    - Homogeneous / Heterogeneous
- **Fusion Strategy**
  - Unweighted Model Averaging
  - Majority Voting
  - Bayes optimal Classifier
  - Super learner
  - Consensus
  - Binary By Committee
- **Applied**
  - Healthcare
  - Speech
  - Image Classification
  - Forecasting
  - Others

THEORY :-

There are several theories are there by using these theories the Ensemble deep Learning is developed.

(i) Bias Variance Decomposition
(ii) Statistical Decomposition & Representational
(iii) Diversity

Bias Variance Decomposition :

We already know the terminology 'Bias-Variance Tradeoff'. It describes that the High Variance returns "overfitting" and High bias returns "underfitting".

The decomposition of the loss into bias and variance helps us understand learning Algorithms. As these things are correlated to overfitting and underfitting.

$$\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

$$\text{Variance}(\hat{\theta}) = E\left[(E[\hat{\theta}] - \hat{\theta})^2\right]$$

Statistical, Computational and Representational Aspects :

Thomas. G. Diettrich, Stanford university provided Statistical, computational and Representational reason. for succeee of ensemble models.

The learning model is viewed as the search of optimal hypothesis h among several hypothesis in the search space.

When the training data is smaller compared to size of hypothethis space then statistical problems arises Due to statistical problem, The learning algorithm identifies different hypothethis which gives same performance on training sample.

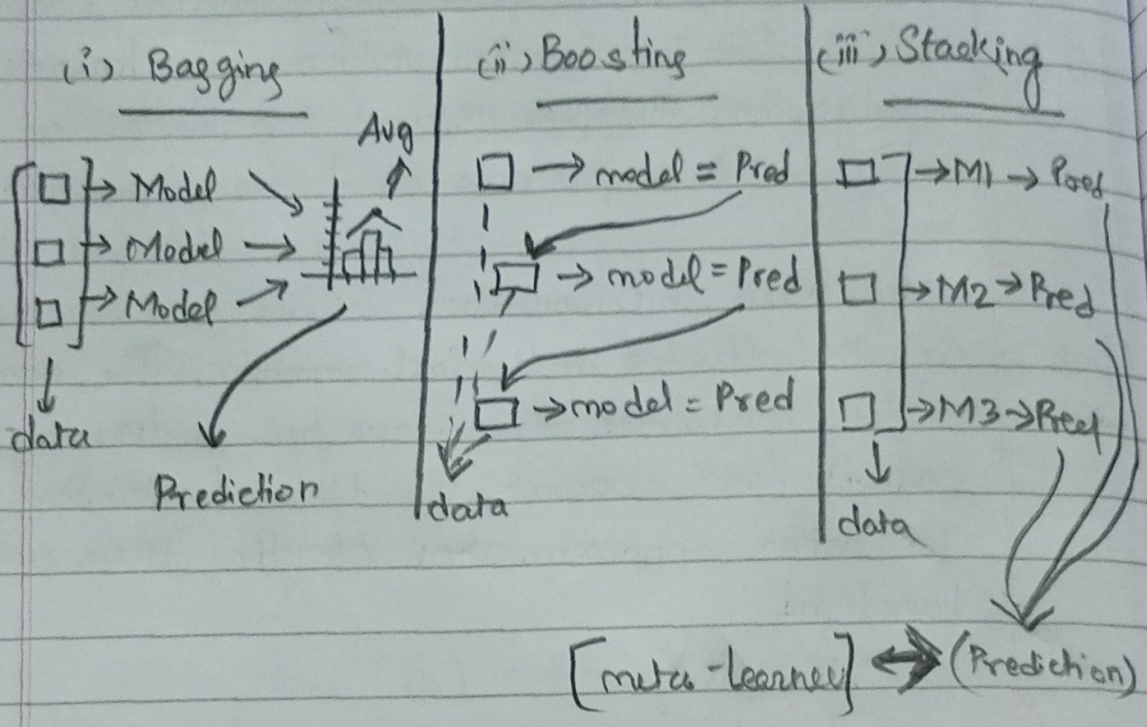
When, in Computation where in a learning algorithm stucks in the local minima, The ensemble model overcomes this issue by performing some form of local search via different starting points for better approximation of the true unknown function.

The third reason is representation where in none of the hypothethis among the set of hypothethis is able to represent the true unknown function. Hence ensembling of these hypothethis via some weight techniques results which expands the representable functional space.

Diversity Theory:

Diversity is defined as a measure of disagreement between the ensemble members - This measure should have a clear relation to the overall ensemble error. The theory should have a clear relation to previously established result and expand our understanding to a wider Range of Learning Scenario.

Ensemble Learning Stratagies :- (Classic Method)



Bagging :-

It's a boostraping Aggregation method designed to increase the accuracy and robustness of a ML model. It follows these methods

Training (Bootstraping)

(i) Data Sampling :- Create N Samples of a data
(ii) Model training :- Raining a separate model for each sample
(iii) Aggregation :- Combine the predictions from all indivisual models.

Usecase:
→ When variance is high
→ Increase the accuracy.

Example: Random forest

Boosting :-

Boosting is another strong technical emtemble technique used to form a strong model using combination of weak models.

→ Sequential Training
→ Weight Adjustment
→ Model Combination

Usecase :-
→ Reduce Bias
→ Produce Strong Predictors

Example: XGBoost, AdaBoost, GBM, LightGBM

Stacking :-