Artificial Intelligence Masterclass

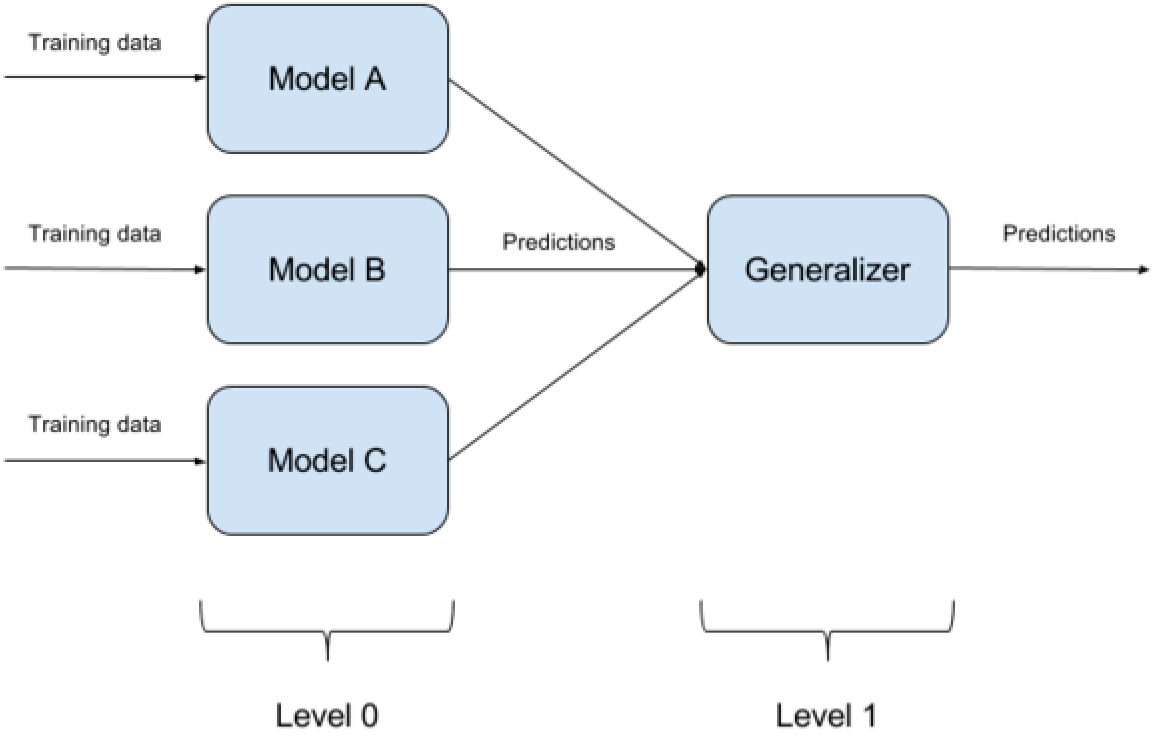
Ensemble Techniques

H.M. Samadhi Chathuranga Rathnayake

M.Sc in CS (SU), PG.Dip in SML (Othm), PG.Dip in HRM (LRN), B.Sc (Hons) in IS (UOC), B.Eng (Hons) in SE (LMU),
P. Dip EP & SBO (ABE), Dip SE, Dip IT, Dip IT & E-Com, Dip B.Mgt, Dip HRM, Dip Eng

Ensemble Techniques

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.

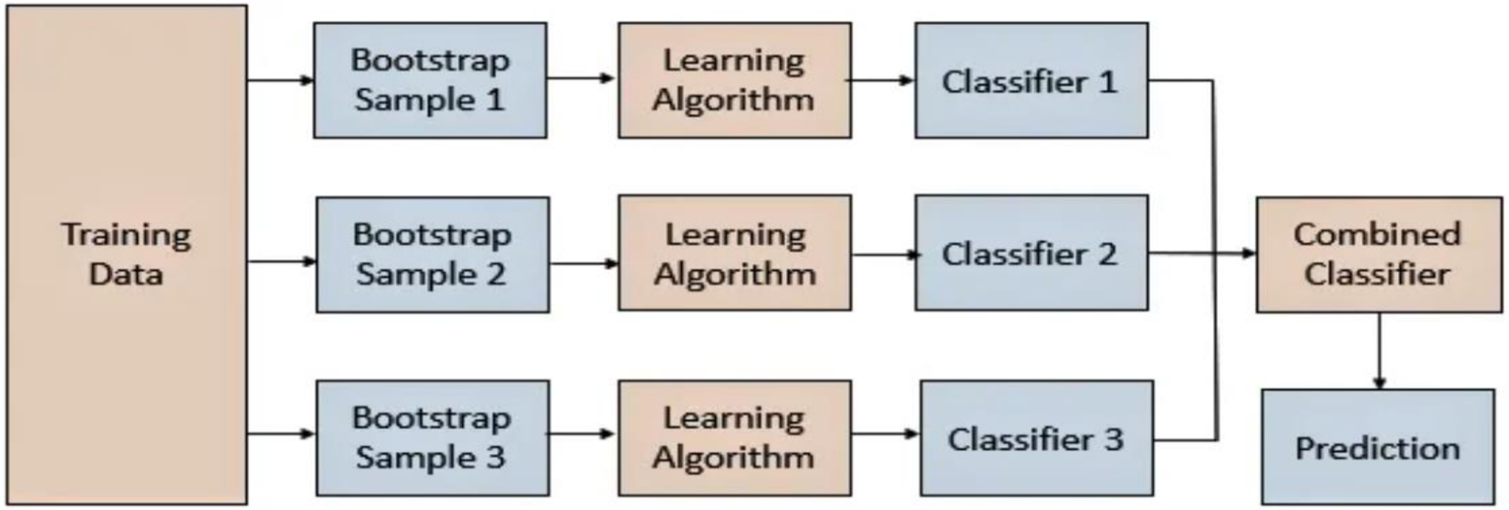


Bootstrapping

This statistical technique consists in generating samples of size B (called bootstrap samples) from an initial dataset of size N by randomly drawing with replacement B observations.

Bagging (Bootstrapping Aggregation)

Bagging in ensemble machine learning takes several weak models, aggregating the predictions to select the best prediction. Here we take many training sets from the population, build a separate prediction model using each training set and average the resulting predictions.



For regression problems, the prediction will be the average prediction. For classification problems, the prediction will be the majority vote.

Random Forest

In bagging, most or all of the trees will use the strong predictor in the top split when there is one very strong predictor along with a number of other moderately strong ones. Consequently all of the bagged trees will look quite similar. Hence the predictions from the bagged trees will be highly correlated.

Bagged trees may be correlated. Random forests improve over bagged trees by decorrelating the trees. As in bagging, we build a number of decision trees on bootstrapped samples. Each time a split is considered, a random sample of m predictors is chosen as candidates from the full set of p predictors. The split is allowed to use only one of those m predictors. Typically,

$$m \approx \sqrt{p}$$

Random Forest

This constructs many decision trees during the training. It predicts the mode of the classes for classification tasks and mean prediction of trees for regression tasks.

It is using random subspace method and bagging during tree construction. It has built-in feature importance.

Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value the more important the feature.