Ensemble Methods

Ensemble Methods

- Imagine that you have received an offer letter from an organization but you are still uncertain about the work culture and ethics of the organization.

How will you finalize your decision whether to join the organization or not?

There can be multiple ways to gather the relevant information like:

a) Checking your LinkedIn network to find if you can get such information (This can be a time taking exercise and yet not reliable)

b) Asking your friend who works in the same domain or in that organization (This can be helpful but still there can be bias associated with that single review)

c) Check websites likes glassdoor or ambitionbox and compare the ratings.(If majority of the reviews are positive then join else not)

Ensemble Methods

- The responses, in third case, would be more generalized and diversified since now you have people with different sets of skills and as it turns out – this is a better approach to get honest ratings than the previous cases we saw.
- With these examples, you can infer that a diverse group of people are likely to make better decisions as compared to individuals.
- Similar is true for a diverse set of models in comparison to single models. **This diversification** in Machine Learning is achieved by a technique called **Ensemble Learning**.

Ensemble Learning

- Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model.
- Basic idea is to learn a set of classifiers (experts) and to allow them to vote.
- The biggest advantage of using ensemble machine learning is that it improves the predictive accuracy significantly.

Simple Ensemble Techniques

There are few ensemble techniques which are simple to implement yet those are very useful which are as follows

- Max Voting
- Averaging
- Weighted Averaging

1) Max Voting

- The max voting method is generally used for classification problems.
- In this technique, multiple models are used to make predictions for each data point.
- The predictions by each model are considered as a **'vote'.** The **predictions** which we get **from the majority of the models** are used as the **final prediction**.

For example, when you asked 5 of your colleagues to provide feedback about the organisation which you are willing to join. we'll assume three of them gave positive feedback while two of them gave it negative feedback. Since the majority gave positive feedback , the final decision will be taken as positive.

We can consider this as taking the mode of all the predictions.

Simple Ensemble Techniques

Results of Max Voting

The result of max voting would be something like this:

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
Negative	Positive	Negative	Positive	Positive	Positive

2) Averaging

Similar to the max voting technique, multiple predictions are made for each data point in averaging.

In this method, we take an average of predictions from all the models and use it to make the final prediction. Averaging can be used for **making predictions in regression problems** or while **calculating probabilities for classification problems.**

Suppose in the above example we ask your colleagues to provide ratings out of 5.Then final rating will be average of all the five ratings.

Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
5	4	5	4	4	4.4

Note : In case of averaging in classification problems predicted probabilities are averaged rather than the predictions and accordingly decision is taken.

Simple Ensemble Techniques

3) Weighted Averaging

- This is an extension of the averaging method.
- All models are assigned different weights defining the importance of each model for prediction.
- For instance, if two of your colleagues are in the same domain (Data science), while others have no prior experience in that domain, then the answers by these two colleagues are given more importance as compared to the other people.

	Colleague 1	Colleague 2	Colleague 3	Colleague 4	Colleague 5	Final rating
Ratings	5	4	5	4	4	4.4
Weights	0.35	0.35	0.1	0.1	0.1	
Weighted ratings	1.75	1.4	0.5	0.4	0.4	4.45

Advanced Ensemble Techniques

Previously we have simple but effective ensemble methods and now we will see the advanced ensemble techniques which are

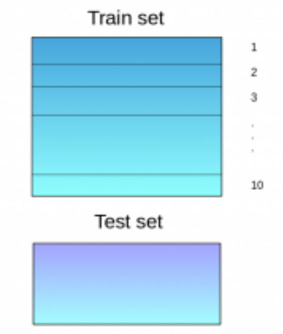
- Stacking
- Blending
- Bagging

Stacking

Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, knn or svm) to build a new model. This model is used for making predictions on the test set.

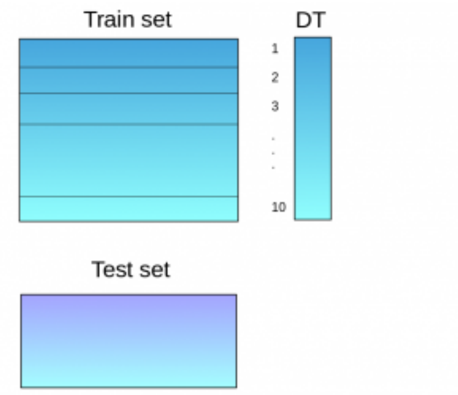
Below is a step-wise explanation for a simple stacked ensemble:

Step 1) The train set is split into 10 parts

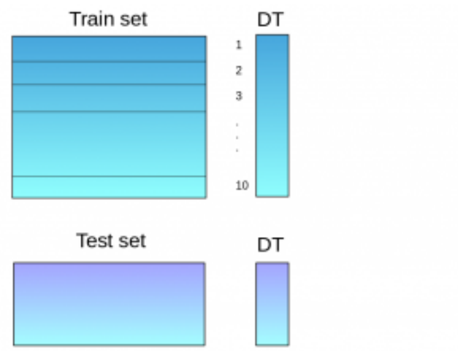


Advanced Ensemble Techniques

Step 2) A base model (suppose a decision tree) is fitted on 9 parts and predictions are made for the 10th part. This is done for each part of the train set.

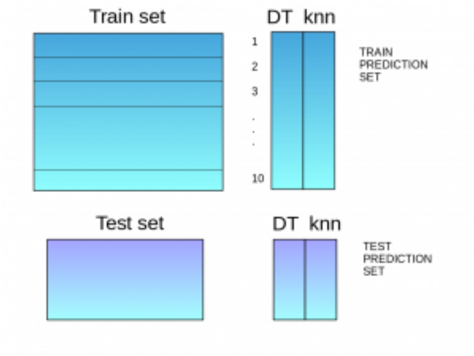


Step 3) The base model (in this case, decision tree) is then fitted on the whole train dataset.
Step 4) Using this model, predictions are made on the test set.

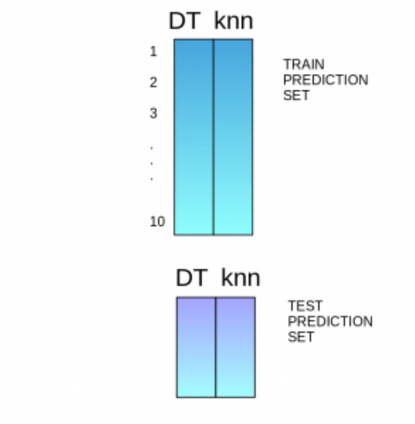


Advanced Ensemble Techniques

Step 5) Steps 2 to 4 are repeated for another base model (say knn) resulting in another set of predictions for the train set and test set.



Step 6) The predictions from the train set are used as features to build a new model.



Step 7) This model is used to make final predictions on the test prediction set.

Advanced Ensemble Techniques

Blending

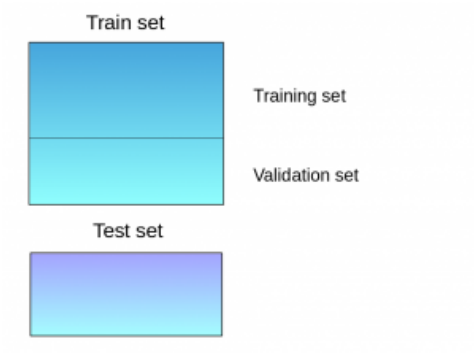
Blending follows the same approach as stacking but uses only a **holdout (validation)** set from the train set to make predictions.

In other words, **unlike stacking, the predictions are made on the holdout set only.**

The holdout set and the predictions are used to build a model which is run on the test set.

Steps of the blending process:

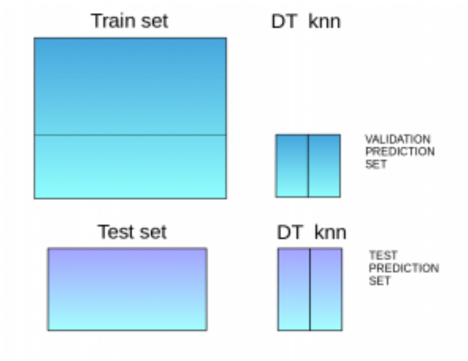
Step 1) The train set is split into training and validation sets



Step 2) Model(s) are fitted on the training set

Advanced Ensemble Techniques

Step 3) The predictions are made on the validation set and the test set.



Step 4) The validation set and its predictions are used as features to build a new model.

Step 5) This model is used to make final predictions on the test and meta-features.

Advanced Ensemble Techniques

Bagging

- The idea behind bagging is combining the results of multiple models (for instance, all decision trees) to get a generalized result.
- But If you create all the models on the same set of data and combine it, will it be useful?
- There is a high chance that these models will give the same result since they are getting the same input.
- So how can we solve this problem?
- One of the techniques is **bootstrapping.**

Bootstrapping

- Bootstrapping is a sampling technique in which we create subsets of observations from the original dataset, **with replacement**.
- The size of columns of the subsets is the same as the size of the original set.

Bagging **(or Bootstrap Aggregating)** technique uses these **subsets (bags) to get a fair idea of the distribution** (complete set).

The size of rows of subsets created for bagging may be less than the original set.

Advanced Ensemble Techniques

Steps involved in Bagging

1) Multiple subsets are created from the original dataset, selecting observations with replacement.

2) A base model (weak model) is created on each of these subsets.

3) The models run in parallel and are independent of each other.

4) The final predictions are determined by combining the predictions from all the models.