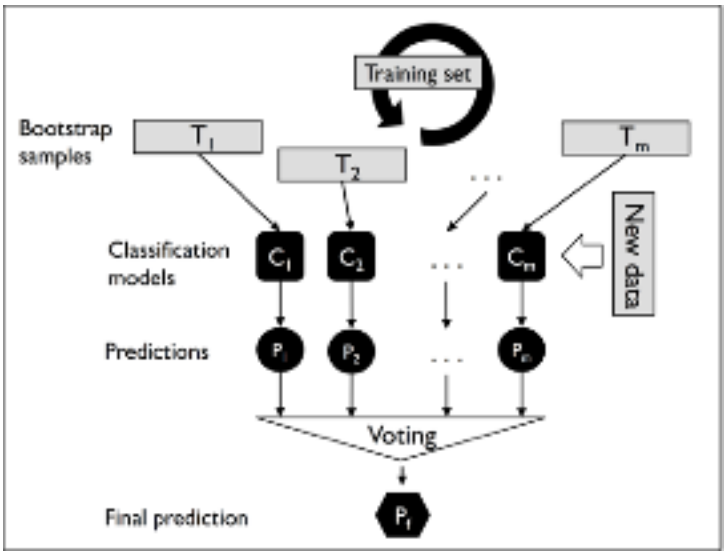
VIRTUAL LAB ENSEMBLE LEARNING

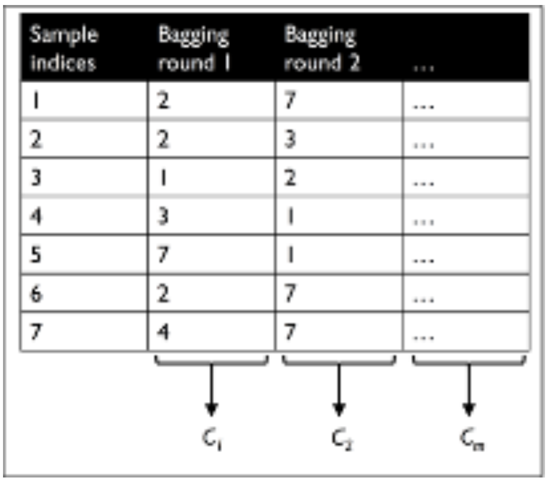
RAJ CHOUDHARY
16BCE1384
SLOT: L39, L40

Q. Find and report why does Bagging or Boosting Work? How much % improvement did you get over SVM/MLP etc.

Bagging:

Bagging is an ensemble learning technique that is closely related to the Majority Vote. However, instead of using the same training set to fit the individual classifiers in the ensemble, we draw bootstrap samples (random samples with replacement) from the initial training set, which is why bagging is also known as bootstrap aggregating.





Sample indices	Bagging round 1	Bagging round 2	...
1	2	7	...
2	2	3	...
3	1	2	...
4	3	1	...
5	7	1	...
6	2	7	...
7	4	7	...

From the above figure, we see that from the sample indices, a subset of the indices is chosen with replacement for training the classifier which is typically an unpruned decision tree. each classifier receives a random subset of samples from the training set. Each subset contains a certain portion of duplicates and some of the original samples don't appear in a resampled dataset at all due to sampling with replacement. Once the individual classifiers are fit to the bootstrap samples, the predictions are combined using majority voting.

Boosting:

In boosting, the ensemble consists of very simple base classifiers, also often referred to as **weak learners**, which often only have a slight performance advantage over random guessing—a typical example of a weak learner is a decision tree stump. The key concept behind boosting is to focus on training samples that are hard to classify, that is, to let the weak learners subsequently learn from misclassified training samples to improve the performance of the ensemble.
In contrast to bagging, the initial formulation of boosting, the algorithm uses random subsets of training samples drawn from the training dataset without replacement; the original boosting procedure is summarized in the following four key steps:

1. Draw a random subset of training samples $d1$ without replacement from training set D to train a weak learner $C1$.
2. Draw a second random training subset $d2$ without replacement from the training set and add 50 percent of the samples that were previously misclassified to train a weak learner $C2$.
3. Find the training samples $d3$ in training set D , which $C1$ and $C2$ disagree upon, to train a third weak learner $C3$.
4. Combine the weak learners $C1$, $C2$, and $C3$ via majority voting.

Boosting can lead to a decrease in bias as well as variance compared to bagging models. In practice, however, boosting algorithms such as AdaBoost are also known for their high variance, that is, the tendency to overfit the training data.

Q. Why does XGBoost help to win competitions?

With the benefits of boosting tree models, i.e. adaptively determined neighbourhoods, XGBoost in general should make a better fit than other methods. They are able to perform automatic feature selection and capture high-order interactions without breaking down.
Hence XGBoost help to win competitions.