Ensemble Learning

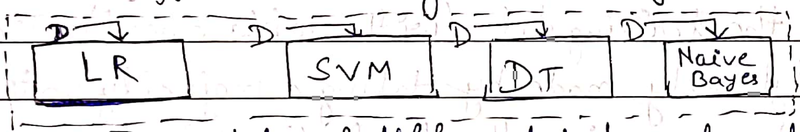
Ensemble Learning is a machine learning technique where multiple models (often referred to as "weak learners" or "base models") are combined to solve a particular problem and improve the overall performance compared to individual models.

Key concepts of ensemble learning:

1. **Weak / Base Learners:** These are individual models that may perform slightly better than random guessing. By themselves, they may not be very accurate, but when combined, their strengths can be aggregated.

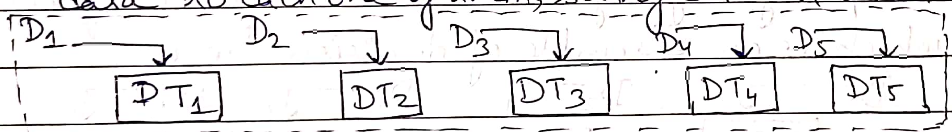
2. **Diversity:** The base models used for ensembling should be ideally diverse, meaning they should make different errors. The diversity helps in cancelling out the individual weaknesses of each base model.
The base models can be made diverse in three ways—

(a) We use different kind of models / algorithms with same data.



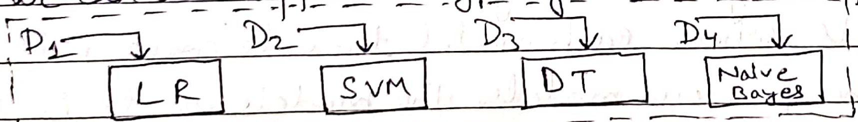
Ensemble of different types of models

(b) We use same model for all but we provide different data to each one of them, so they can be trained differently.



Ensemble of same type of model but different dataset.

(c) We can use different types of models with different data.

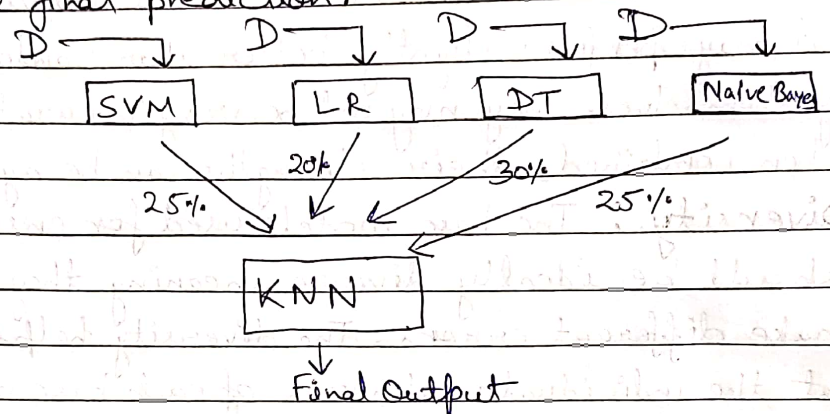


Ensemble of different models with different data.

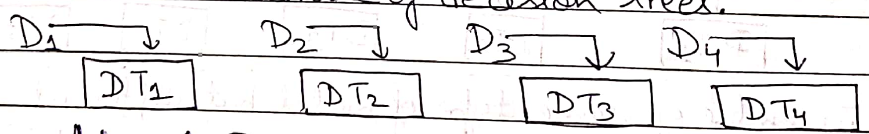
3. **Types of ensemble learning:**

(a) **Voting:** In this method, multiple different base models are trained, and then their predictions are combined by a voting mechanism. In case of a classification problem, the final class is the one that gets the majority of votes, while in regression, predictions are averaged.

(b) **Stacking:** In this method, **different** models are trained on the **same dataset**, and a new model 'meta-learner' is trained to combine their predictions. The meta-learner typically uses the outputs of the base models and learns how to combine them i.e., which base learner should have how much say in the final prediction.



(c) **Bagging:** In this method, multiple instances of the **same** model are trained on **different subsets** of the training data (with replacement). The final prediction is typically made by taking a majority vote in classification or by averaging the predictions in regression. It is also known as '**Bootstrap Aggregation**. A popular example of Bagging is the Random Forest algorithm, which is an ensemble of decision trees.



(d) **Boosting:** Boosting involves sequential training of models, where each model tries to correct the errors of the previous models. The models are trained in sequence and each subsequent model gives more weight to data points that were misclassified by previous models. Examples are – Ada boost, Gradient Boost, XG Boost.

4. **Advantages:** Few advantages of ensemble modelling are–

(a) Improved accuracy: By combining models, ensemble methods can significantly improve prediction accuracy.

(b) They promote low bias and low variance thus help in making better generalized models that perform better on both training and testing data.

(c) Robustness: These models are more robust to outliers and noisy data.

5. **Disadvantages:** Few disadvantages of ensemble method–

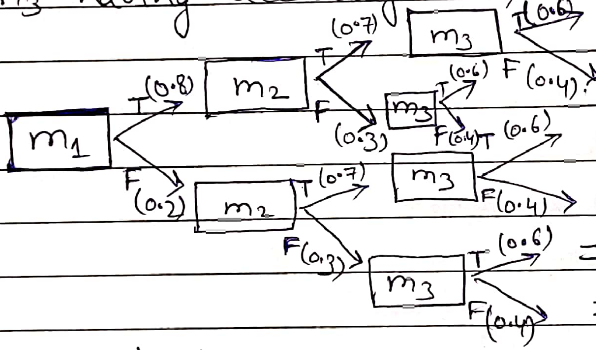
(a) Increased complexity: The ensemble models which involve many base models can be hard to interpret compared to simpler models like linear regression or decision trees. This can make it difficult to understand why a particular prediction was made, which is a drawback in situations where model transperancy is important.

(b) Higher computational cost.

Voting Ensembles

Voting ensembles combine the prediction of multiple base models (trained using different algorithms) to produce a more accurate and robust prediction. The underlying principle is that by combining diverse models, the ensembles can reduce the impact of individual model errors and improve overall performance.

The ensemble is most effective when the base models have similar performance and when they make different errors (i.e., the errors are not perfectly correlated). If the base models are not perfectly correlated (i.e., they make different errors), the ensemble's accuracy will be higher than the average of the individual models. Suppose, we have three base models m_1, m_2 and m_3 having accuracy 0.8, 0.7, 0.6 respectively then-



= 0.336 ← TTT
= 0.224 ← TTF
= 0.144 ← TFT
= 0.096 ← TFF
= 0.084 ← FTT
= 0.056 ← FTF
= 0.036 ← FFT
= 0.024 ← FFF

Correct prediction will be made by ensemble model when at least two predictions are true i.e.,

$P(T) = TTT + TTF + TFT + FTT$

$P(T) = 0.336 + 0.224 + 0.144 + 0.084$

$P(T) = 0.788$ is the probability of making correct prediction by the ensemble model.

The average accuracy of all the base models $= \frac{0.8 + 0.7 + 0.6}{3} = 0.7$

Hence, we can say that the ensemble's accuracy can be potentially higher than the average of individual models.

* When the individual base models have poor accuracy suppose 0.2, 0.3, 0.4, then ensemble accuracy will be lower than mean of base models i.e., $\underset{\text{(mean)}}{0.3} > \underset{\text{(ensemble model)}}{0.212}$

Voting Ensemble Classifiers Types

There are two primary types of voting classifiers: **hard voting** and **soft voting**.

1. **Hard voting:** In hard voting each base model makes a prediction (votes for a particular class), and the class with most votes becomes the final prediction of ensemble. If $m_1, m_2, m_3, \ldots, m_n$ are individual base models and $\hat{y}_i$ is the prediction from the i^{th} model, the final prediction $\hat{y}$ in hard voting is:

$$\boxed{\hat{y} = \text{mode}(\hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots, \hat{y}_n)}$$

This means the class that appears the most frequently among the models' prediction is chosen as the final output.

2. **Soft Voting:** In soft voting, the classifier considers the predicted probability for each class from each base model rather than just the final class prediction. The predicted probabilities are averaged, and the class with the highest average probability is chosen as the final prediction. For Ex - Suppose, we have three classifiers predicting an image is a "cat" or a "dog" and the predicted probabilities are as follows-

model 1 : 70% cat, 30% dog
model 2 : 40% cat, 60% dog
model 3 : 20% cat, 80% dog

The soft voting classifier will create the average probability for each class: cat $= \frac{70+40+20}{3} = 43.3\%$
dog $= \frac{30+60+80}{3} = 56.7\%$, the soft voting classifier would predict "dog" as the final output since "dog" has the higher output probability.

Suppose for different models, accuracies again are -

model 4: 45% cat, 55% dog
model 5: 48% cat, 52% dog
model 6: 70% cat, 30% dog

as per hard voting dog will be final output.

But, Mean $P_{cat} = \frac{45+48+70}{3} = 54.3\%$ and $M p_{dog} = \frac{55+52+30}{3} = 45.7$
thus in this case soft margin classifier predicts "cat" as the final output.

Voting Ensemble Regressor

$D \rightarrow \boxed{m_1} \rightarrow x_1$

$D \rightarrow \boxed{m_2} \rightarrow x_2$

$D \rightarrow \boxed{m_3} \rightarrow x_3$

simple averagaing | weighted averaging

$\frac{x_1+x_2+x_3}{3}$ | $\frac{w_1x_1+w_2x_2+w_3x_3}{3}$

A voting ensemble regressor is used in regression tasks, where the prediction of multiple regression models (base models) are combined to produce a final prediction. It combines the prediction by averaging the values from multiple models.
We can do both normal averaging and weighted averaging as per our requirement.