22/3/24

Q] Why not use multiple strong classifiers?

→ Strong classifiers might have strong biases

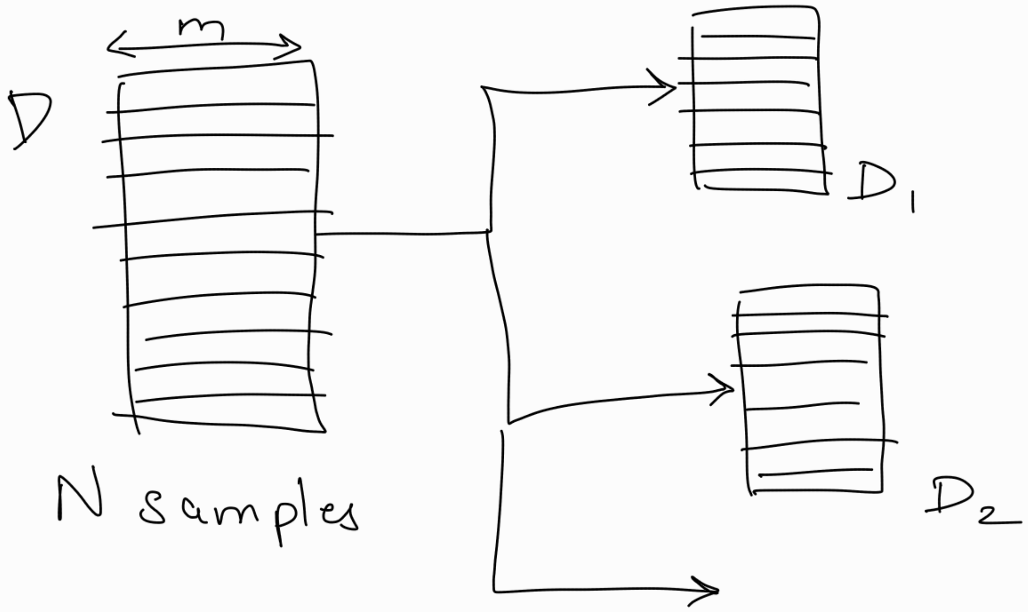
→ High compute

Ensemble consists of multiple weak classifiers

BAGGING (Bootstrap aggregation)

Random Forest Classifier

1] Sampling with replacement



2 approaches for rule building:

1] Sample k features from m and pick best out of these k

2] Pick 1 feature randomly

Build a tree for every $D_1, D_2, \ldots$

Decide no. of decision trees using validation set.

Class is decided by majority voting.

Every tree has equal vote.

Validation set uses out of bag error

Q] Why can't we use SVM for every $D_1, D_2, \ldots$?

→ The support vectors are key to the svm.
There is high probability of all the

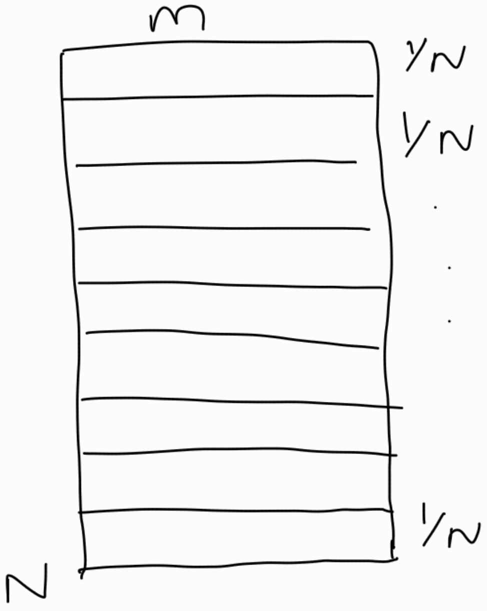
support vectors being picked every time.
This means that the SVM margin is likely
to be same every time.

BOOSTING (Ada Boost algo)

→ Creation of classifiers is sequential as opposed to bagging, where it can be parallel.

→ The vote of each classifier is not equally weighted (amount of say)

→ Most used weak classifier is a single rule decision tree (also called stump)



- Classify using stump
- Increase weight of wrongly classified samples
- Decrease weight of correctly classified samples.

- Now when bagging again, you are more likely to pick wrongly classified samples.

- This means that in the new sampling, the wrongly classified samples occur multiple times. So, when building the stump, you are more likely to classify these correctly.

- $$\varepsilon_t = \frac{\text{No. of wrongly classified}}{\text{Total samples}}$$

$$\alpha \text{ (amount of say)} = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

→ The new dataset helps to find new stump, but amount of say is calculated on the original dataset. Sampling and weight updation is also done on original dataset