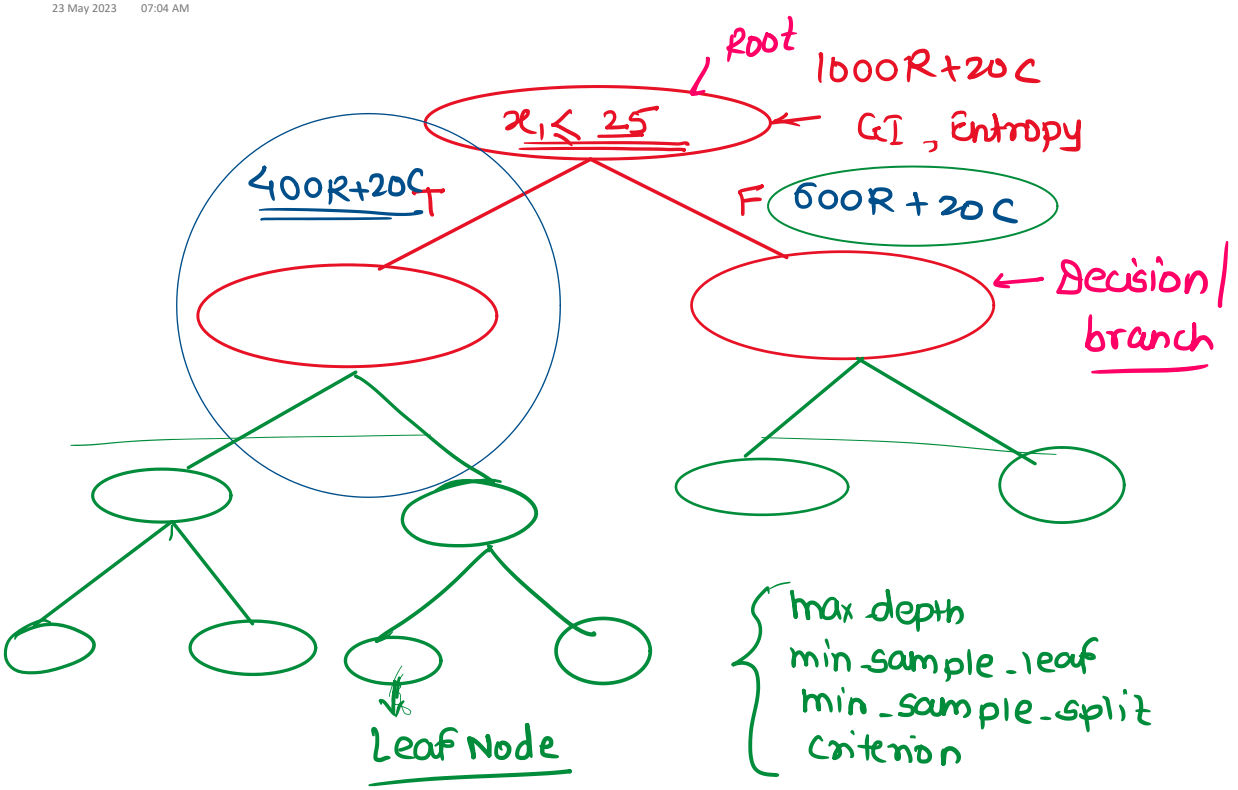
Ensemble Techniques

23 May 2023 07:04 AM

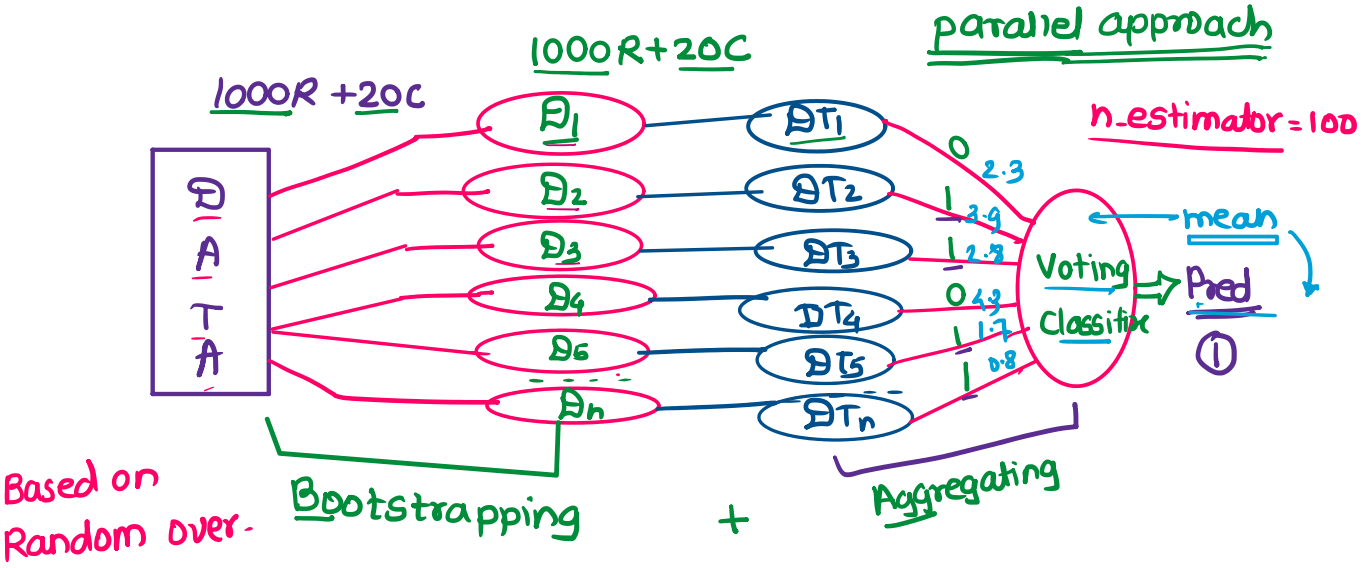


Ensemble Methods: Combining of no. of decision trees

1. Bagging
2. Boosting
3. Stacking

1. Bagging → Bootstrapping + aggregating

parallel approach



Based on Random over-sampling

Bootstrapping + Aggregating

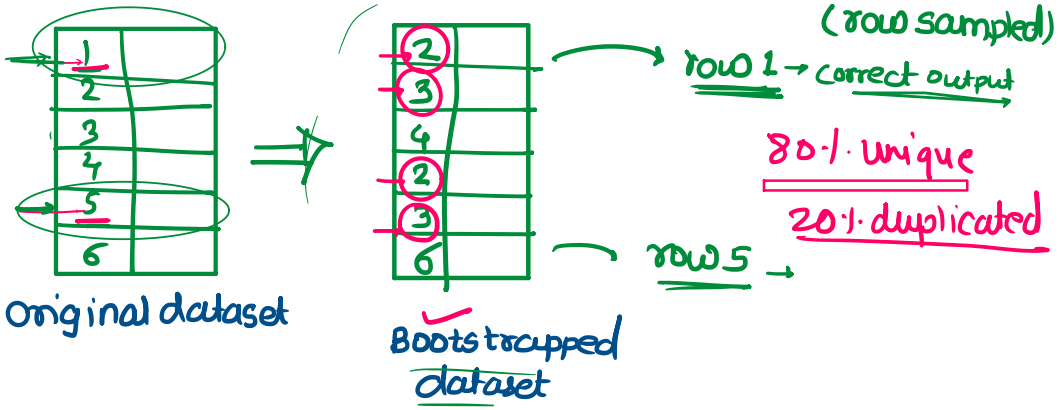
→ D_1 → 750 Unique

250 Duplicated

→ D_1 → 750 Unique

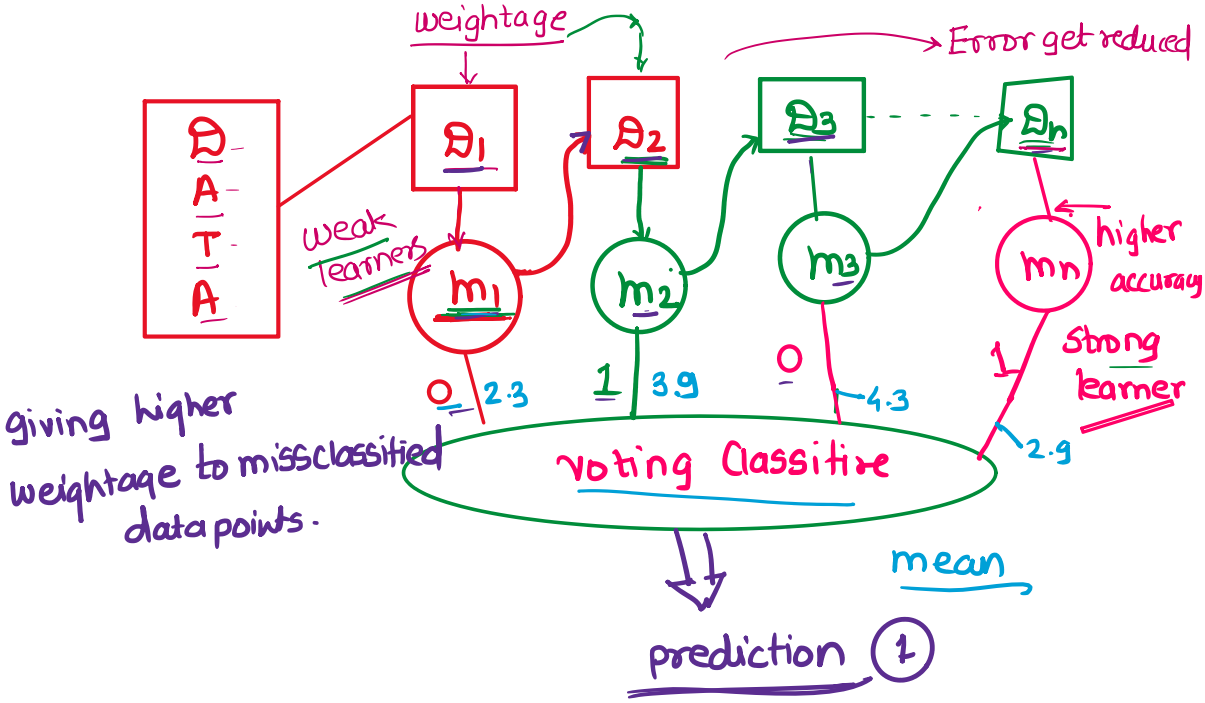
250 Duplicated

Bootstrapped dataset → 80% unique + 20% oversampled dataset (row sampled)

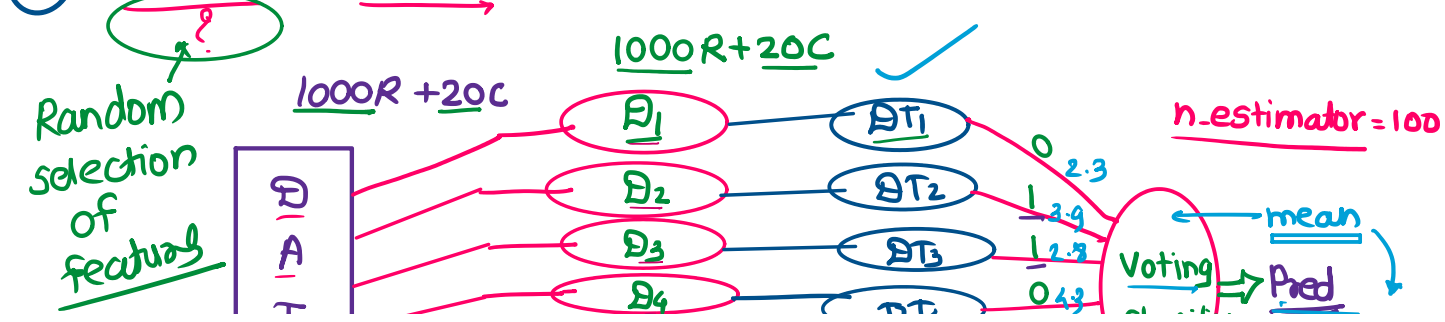


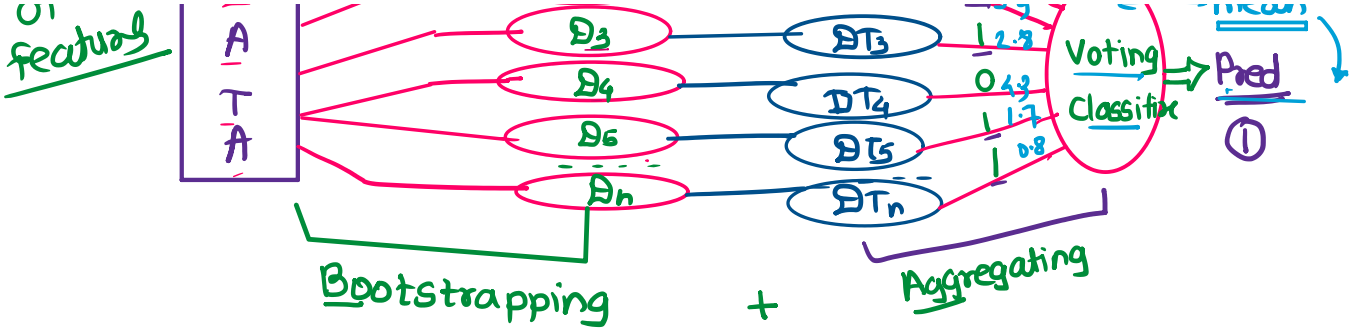
② Boosting:- Sequential approach

n_estimator = 50



① Random Forest → Forest of decision trees





1000 R + 20C

Bootstrapping

D_1 → 780 uniques + 20C
220 – duplicates

D_2 → 850 → unique + 20C
→ 150 → duplicate

D_3 → 730 – unique + 20C
150 → duplicate

All datasets are unique

oob → out of bag

present in original dataset but not present in bootstrapped dataset

oob_score = no. of correctly predicted rows from out-of-bag sample

oob_error = no. of wrongly predicted rows from oob-sample

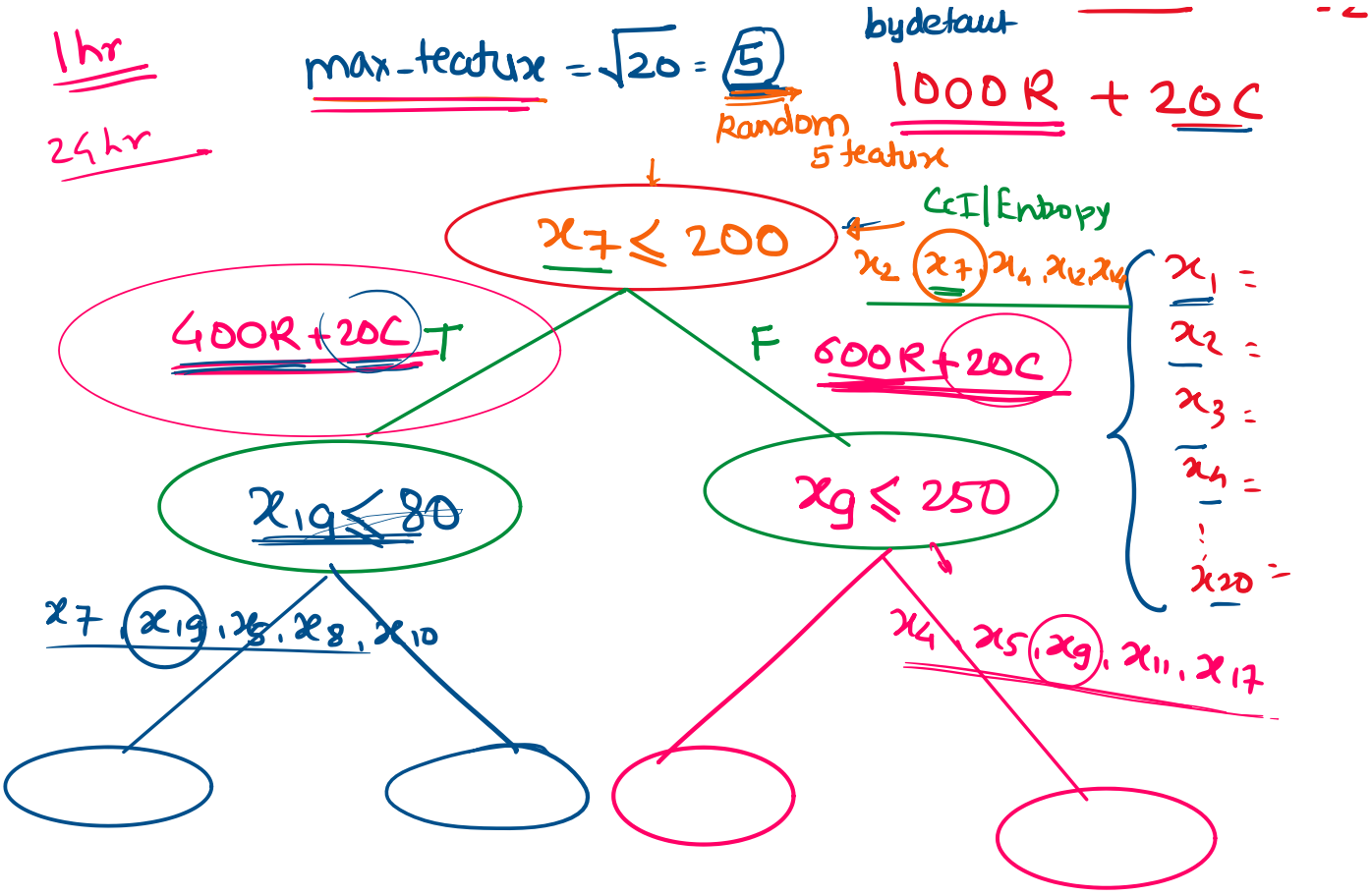
max_features :- max_feature in DT → none → All the feature

RF → max_feature → 'sqrt' (by default) | 'Auto' | $\log_2$

1 hr

max_feature = $\sqrt{20}$ = 5

1000 R + 20C



To reduce time & model complexity

max_feature → less execution time

→ Hyper Parameter tunned

1 DT —— 100 DT RF

High variance

prediction

less variance :-