

Ensembling Models

Collection

/

Bagging

Boosting

Random Forest

1 Model \rightarrow 60% Acc.

10 Models \rightarrow 70% Acc.

Bagging (Bootstrapping Aggregation)

Sampling
with
Replacement

Model 1 (logistic)

Model 2 (logistic)

Model 3 (logistic)

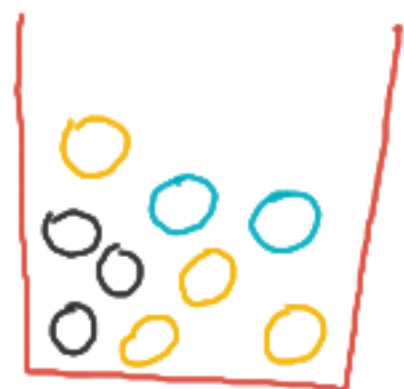
dataset

Pred 1

Pred 2

Pred 3

} Mean
or
Mode



$$P(B) \rightarrow 3/9$$

$$P(\text{Blue}) \rightarrow 2/8, 1/7$$

$$P(\text{Yellow}) \rightarrow 4/7$$

w/o Replacement

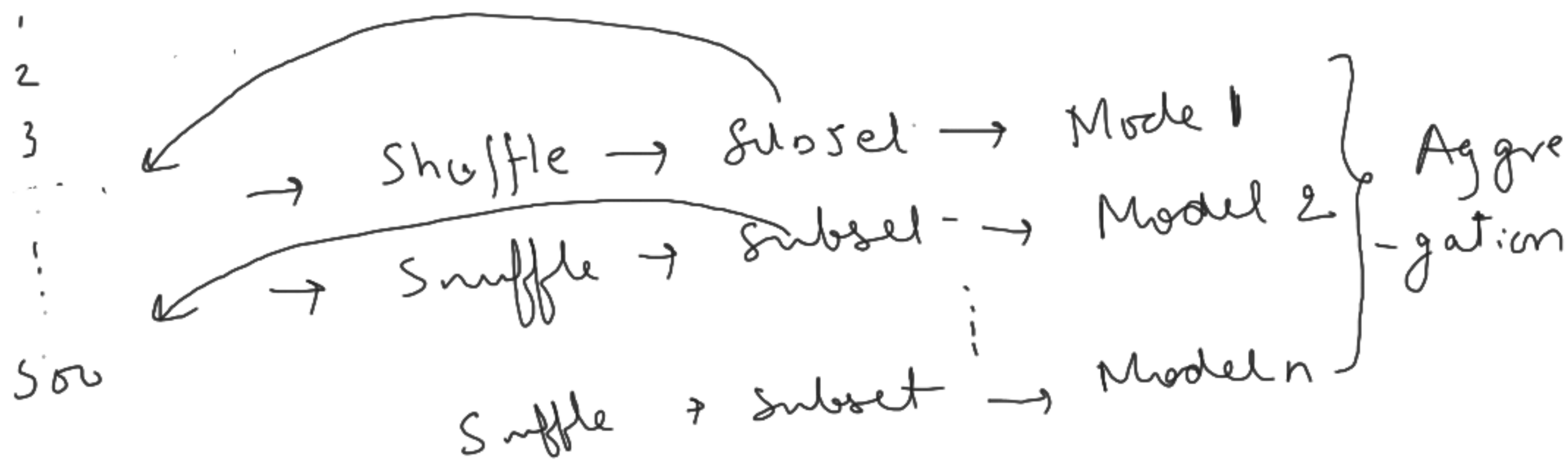
$$P(B) \rightarrow 3/9$$

$$P(\text{Blue}) \rightarrow 2/9, 2/9$$

$$P(\text{Yellow}) \rightarrow 4/9$$

w/ Replacement

○ → Black
○ → Blue
○ → Blue



Random Forest-

Sampling w/ Replacement

(Not Only Rows take sample of Columns)

Subset 1 (p rows, q columns) \rightarrow model 1
Subset 2 (m rows, n columns) \rightarrow model 2
...

} Aggregation.

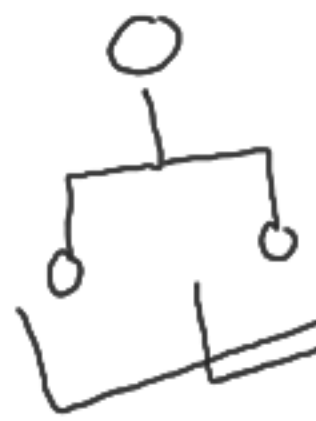
log reg, DT, NB..... \Rightarrow Bagging / Random Forest.

Boosting (D.T)



X GB Boost
Light GBM

(Predictions - Originals)


$$(\hat{y} - 1) \quad 1.550$$

• • •



Value

Class

final
Predictions

false predictions
↓
features
more weighting

Tree1 → Gender (Root) → Age (False Node)
Age (Root) → BMI (FP)

→ BMI (F-P)

$\rightarrow (\uparrow \text{wt to Age})$ Tree 2 \rightarrow 14
 \rightarrow Tree 3