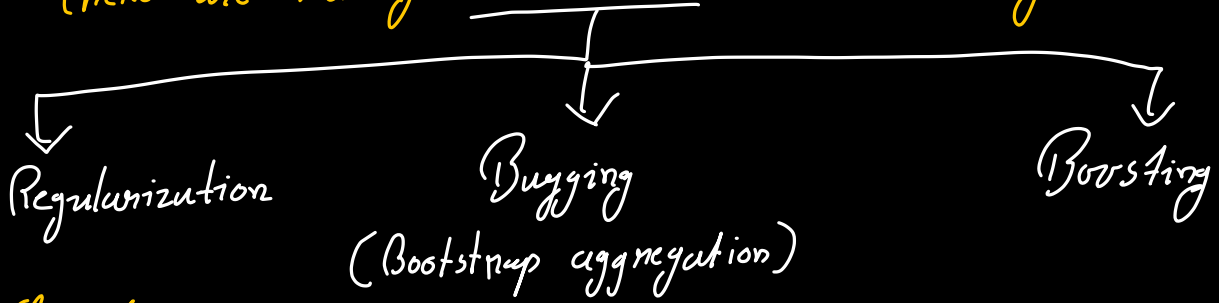
Bagging vs Boosting

→ Model should have Low bias and Low variance.

→ This type of model is called as generalized model.

→ There are mainly 3 methods to achive a generalized model.



Questions

Q1 : Type of model used

Q2 : Sequential vs parallel

Q3 : Weightage of base learners.

Q1 : Type of model used

[a]. Bagging

→ In bagging, we selects such model which has low Bias and High variance (clearly overfitting)

ex:- • fully-grown decision tree
• KNN (with smaller 'k')

[b]. Boosting

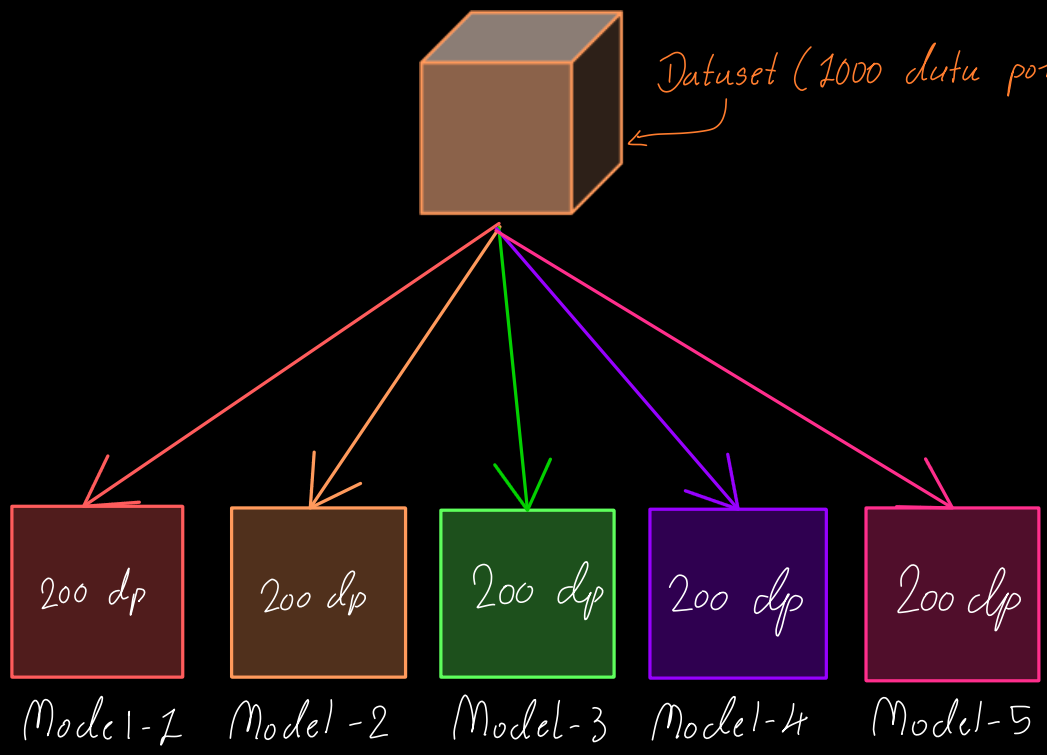
→ In boosting, generally the model which we select has High Bias and Low Variance.

ex. Decision Tree (whose 'max_depth' is 1)

Q2 : Sequential vs parallel

[a] Bagging

-> Bagging follows parallel execution, where each model will be trained in parallel.



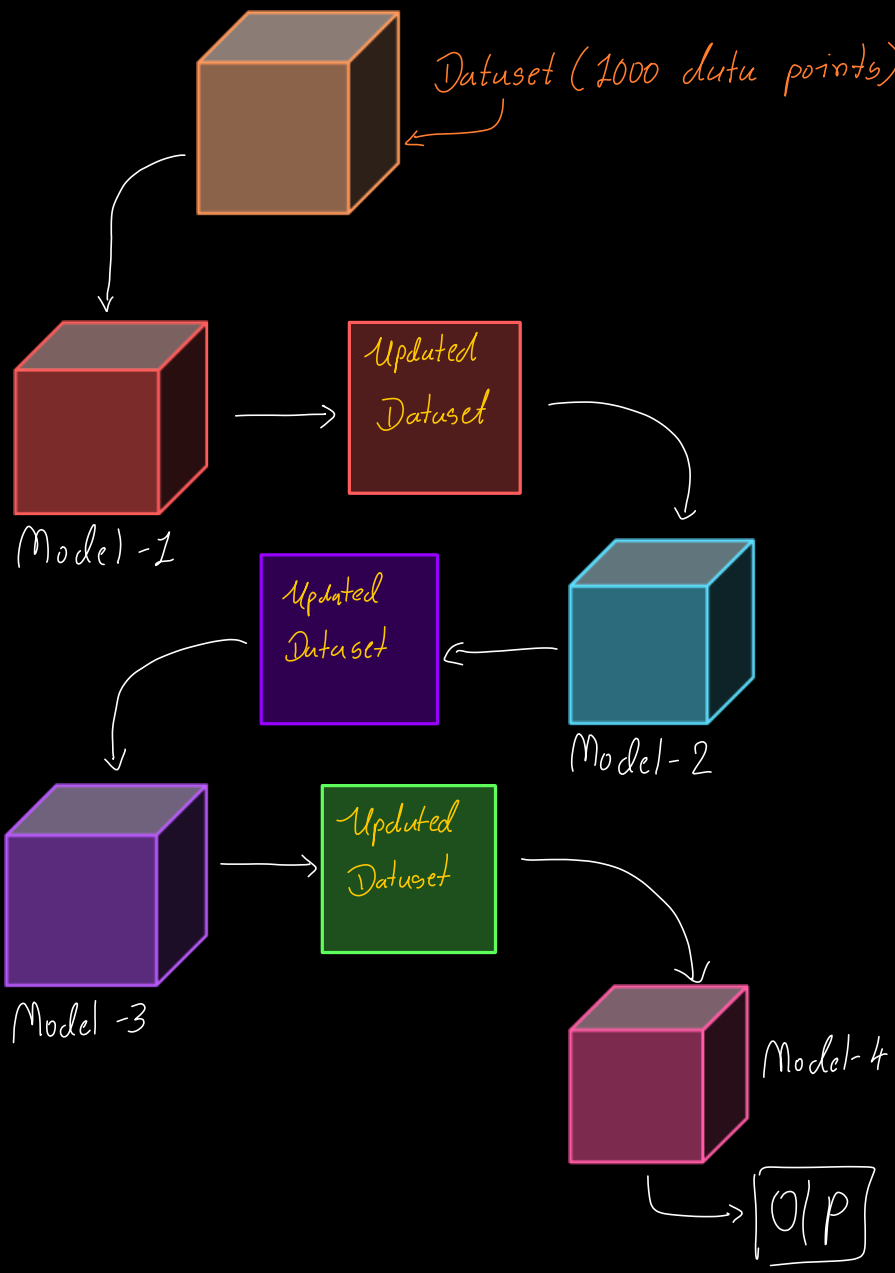
[b]. Boosting

-> Process goes sequentially.

-> First model will train on given data, based on the performance of first model, Next model will get data (Highting error data points) and so on....

-> Based on the performance of current model, importance of each data-points will be updated.

-> Mis-classified data-points will be highlighted for next model so that next model will focus on highlighted data points.



Q3 : Weightage of base learners.

[a]. Bagging

-> Each model has equal amount of weight like democracy.

[b]. Boosting

-> Last model has higher priority than other models.