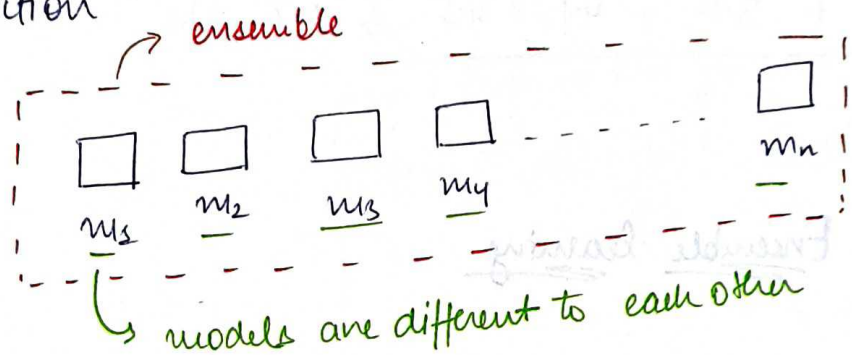
Ensemble Learning

Core Idea

→ training

→ prediction



- All model are different like m_1 → linear reg. m_2 → Logistic ...

OR

- All model are same but trained with different data like all models are linear regression and $m_1, m_2 \dots m_n$ trained with different different data, m_1 trained with dif data, m_2 trained with different data and so on.

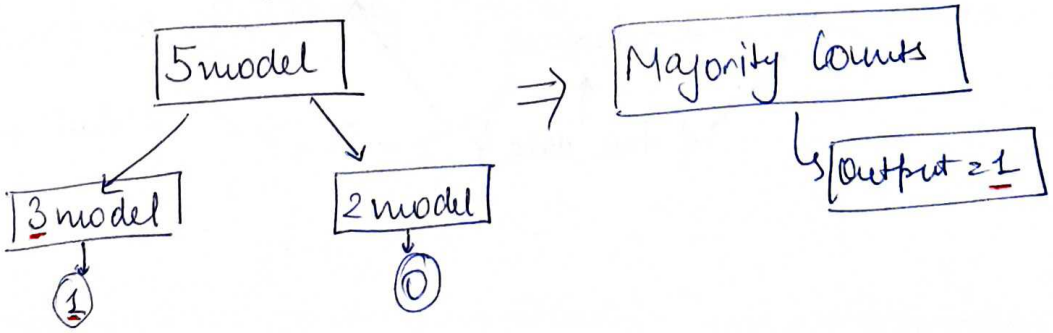
OR

- All model are different and trained with different data.

Classification

cgpa	ia	placement
-	-	1
-	-	0

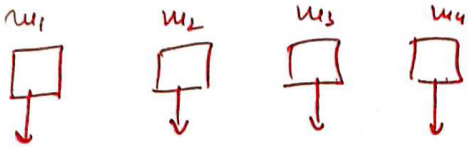
{8.1, 85} → 1/0?



Regression

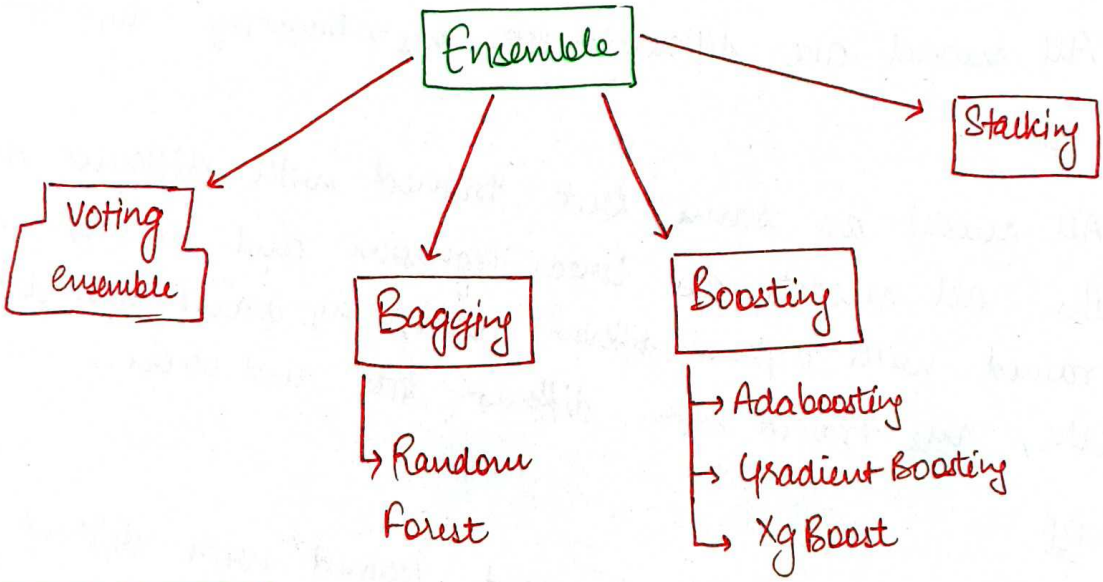
cgpa	iq	lpa placement
-	-	-

{8.1, 8.5} → lpa?

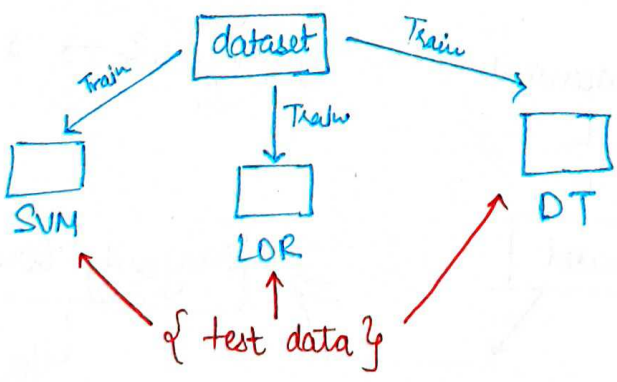


$$\frac{8.1 + 3.2 + 4.1 + 7.5}{4} = 6.6 \text{ lpa}$$

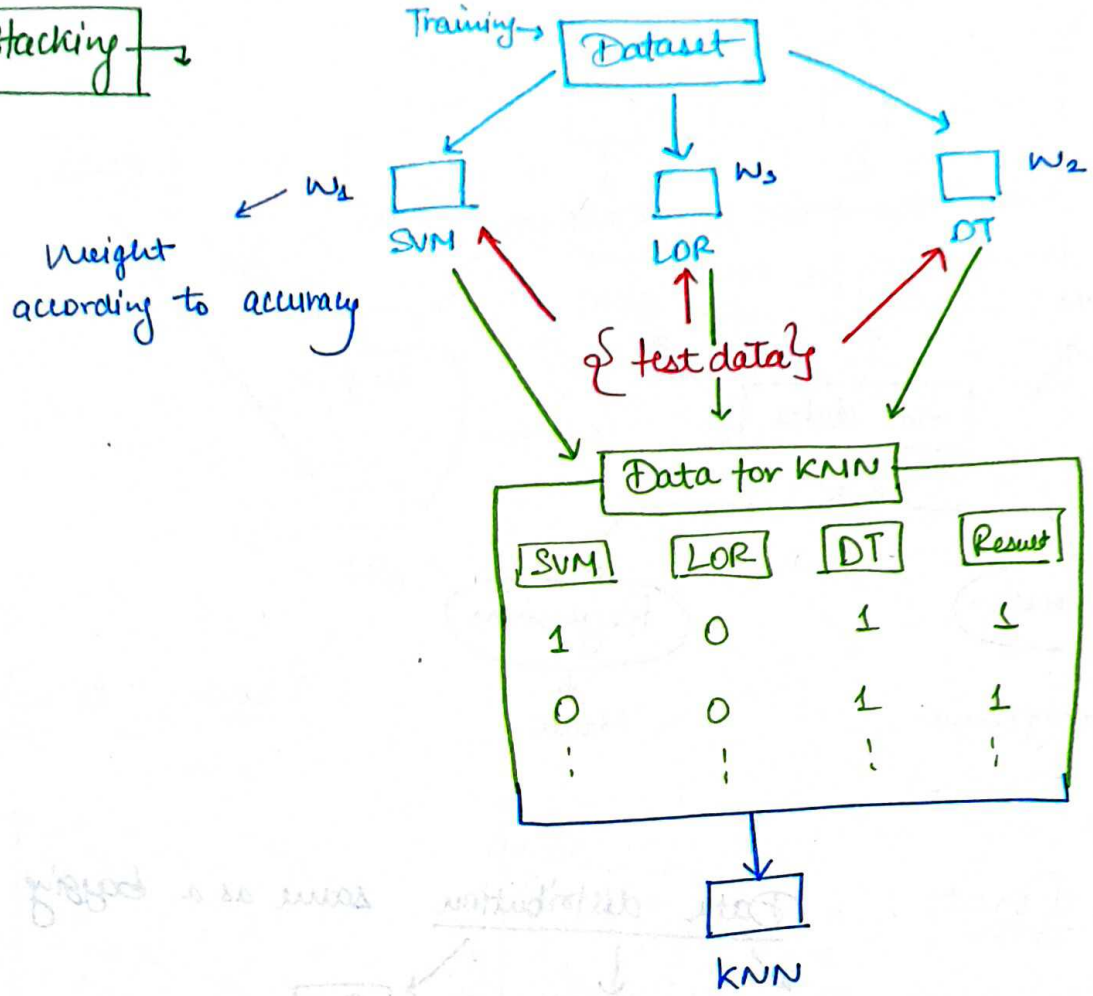
Type of Ensemble learning



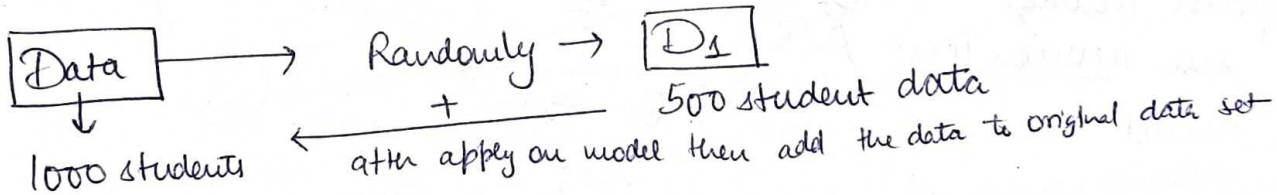
Voting Esemble → Base model
↳ diff algo

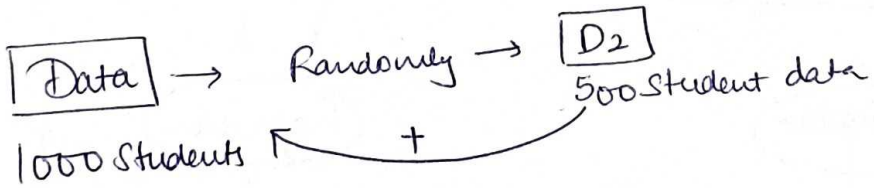


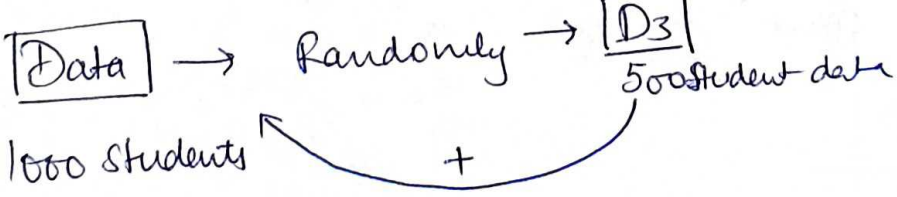
Stacking



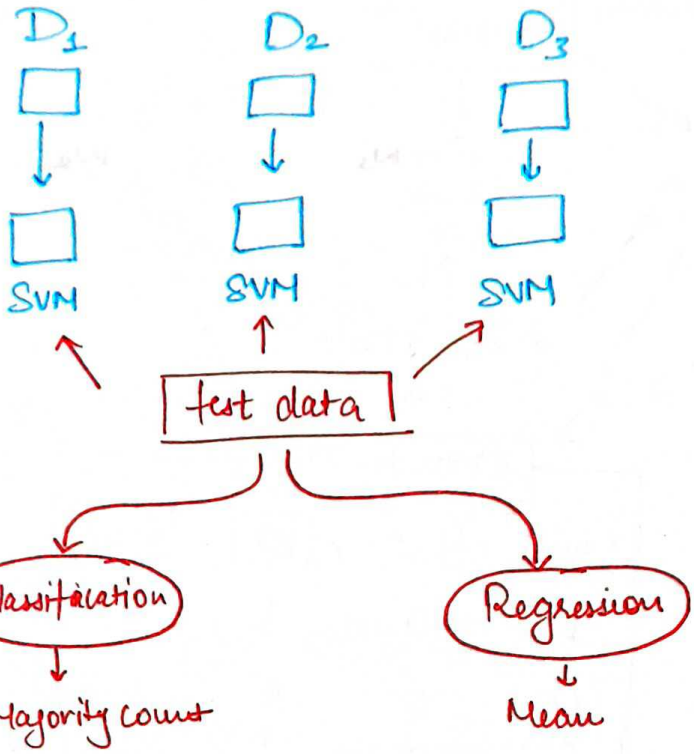
Bagging : Bootstrap Aggregation



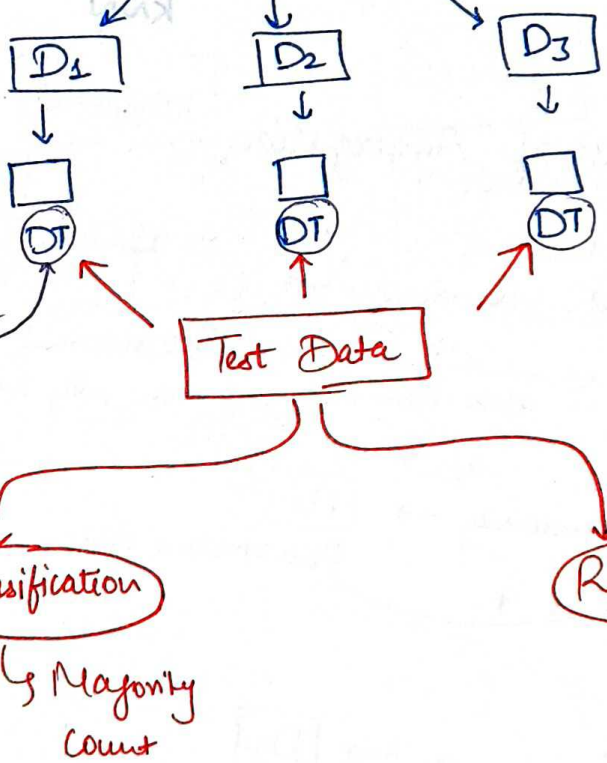




* High prob that D_1, D_2 and D_3 are different from each other.

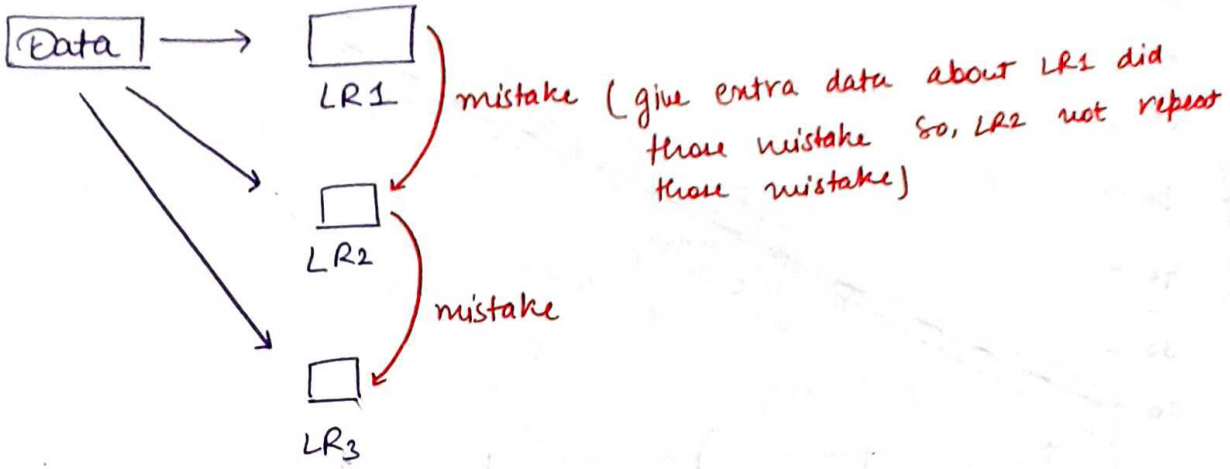


Random Forest : Data distribution same as a bagging

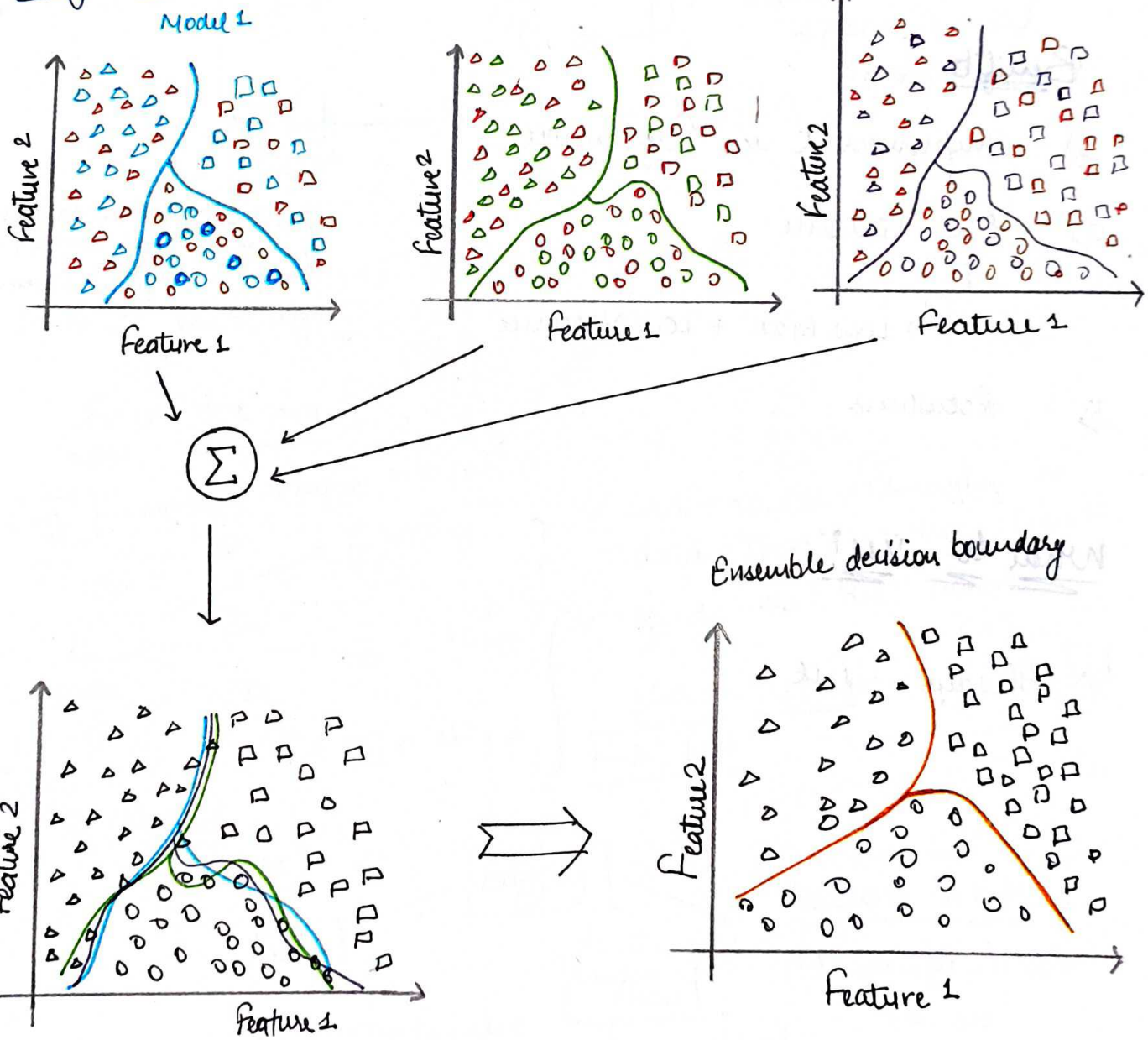


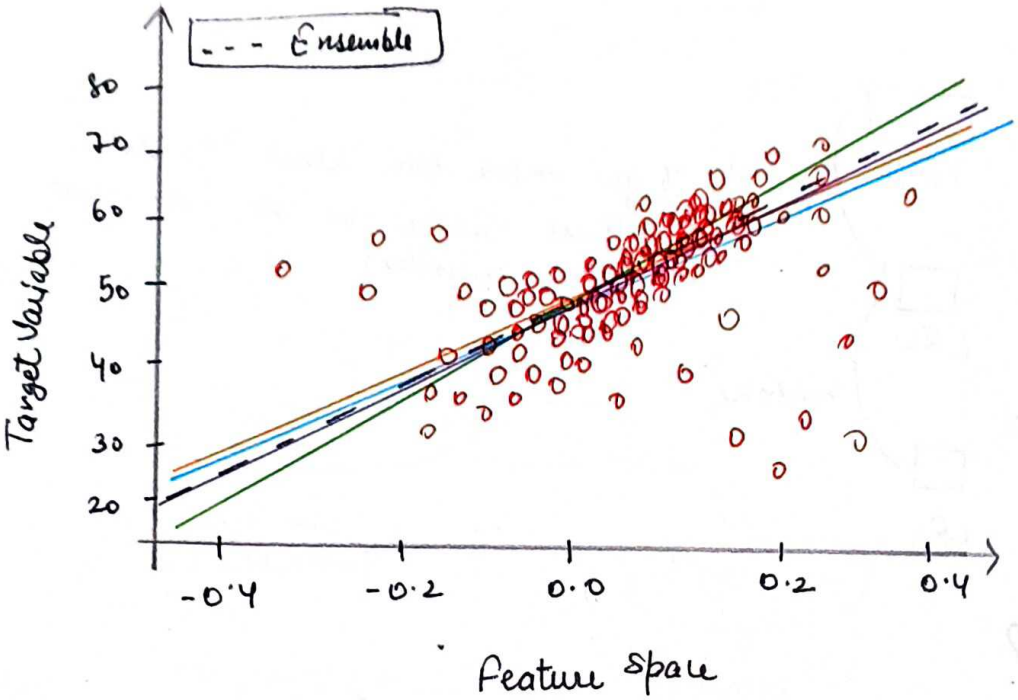
* In Random forest always use decision Tree model

Boosting :



Why it works ?





Benifits

1) Improvement in Performance

2) Bias Variance
   ↳ Low Bias + Low Variance

3) Robustness

Low Bias High Variance
↳ (DT) → ensemble

When to use?

↳ Always use