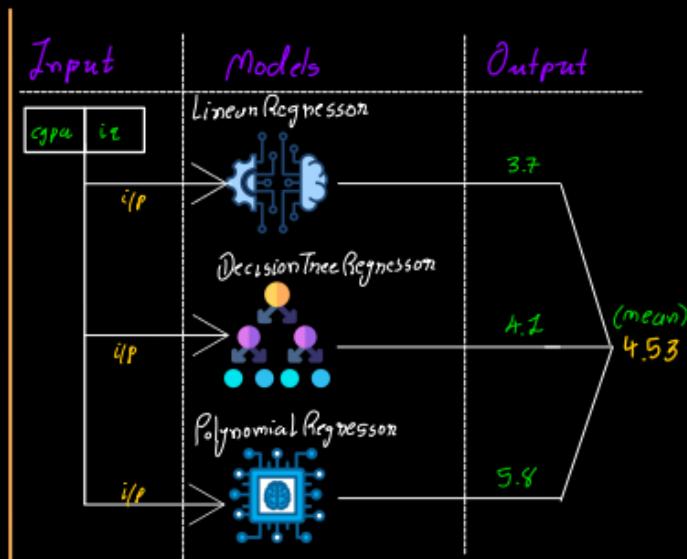


Stacking and Blending

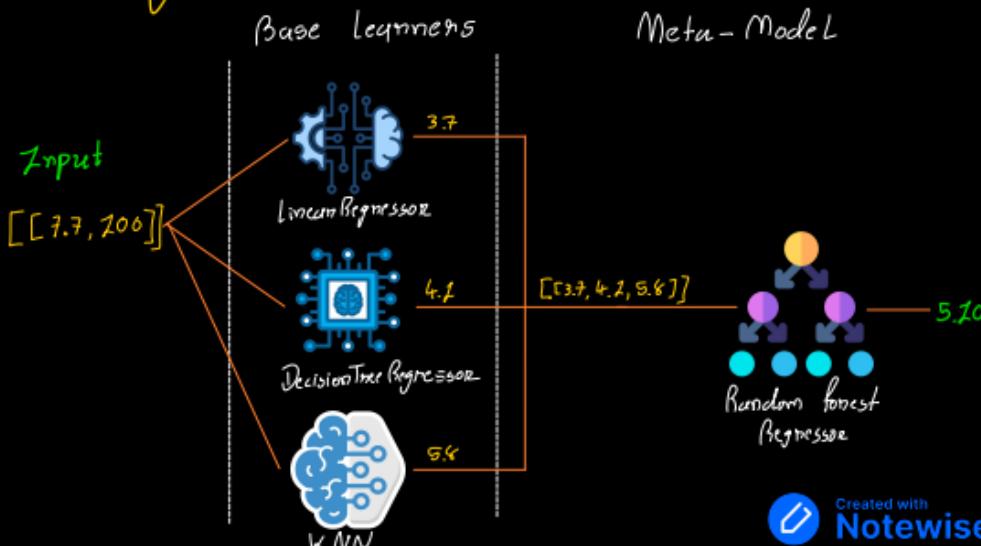
- Stacking extends voting ensemble technique
- Suppose we're working on regression problem....

Voting Ensemble

cgpa	iq	package
7.8	210	4.5
8.5	220	5.2
6.9	205	6.0



Stacking Ensemble



Advantages

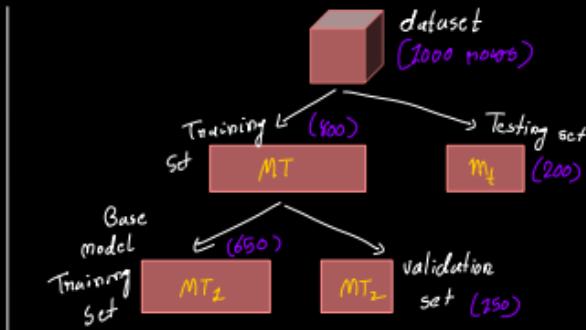
- We can use different types of base models (Not allowed in bagging / boosting)
- We can directly use the o/p of own base model, we can use o/p of base model as training data for meta model (Not in bagging / boosting)

Overfitting

- There is high possibility for overfitting because we're using same data for training & testing.

Solution

2. Hold out method (Blending)



→ How these methods works:

[a] Base models will be trained on MT_1 dataset.

[b] Test base models on MT_2 dataset & make predictions

the prediction of these base models will create another dataset

pred - LR	pred - DTR	pred - KNN	pred - package
-	-	-	-

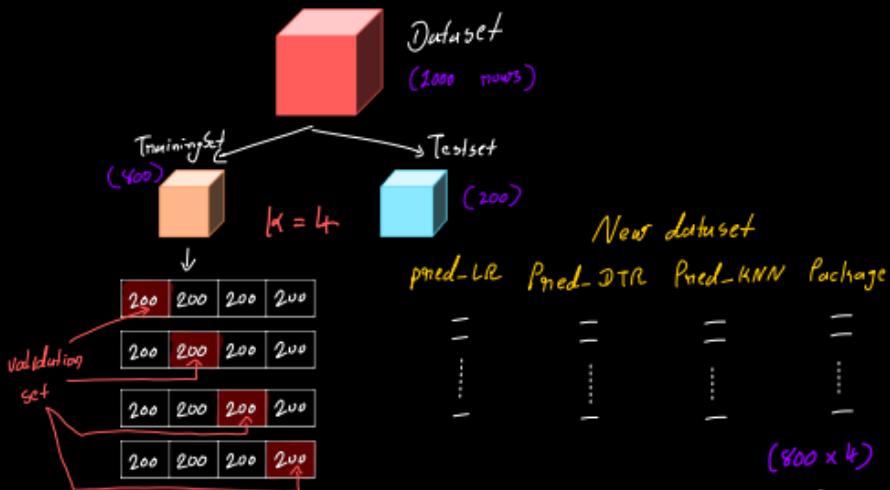
(150×4)

[c]. Train meta model on above dataset.

[d]. Use Mt dataset for test entire chain of models.

2. k-fold method (Stacking)

↳ Sampling through k-fold cross validation.



→ Now we will train a linearModel on first Split, 600 news for training & 200 for making prediction. for another 3 splits, we will train 3 more LR models.

→ at the end, now we have total 12 base models and nearly generated dataset based on predictions

4 LR
4 DTR
4 KNN

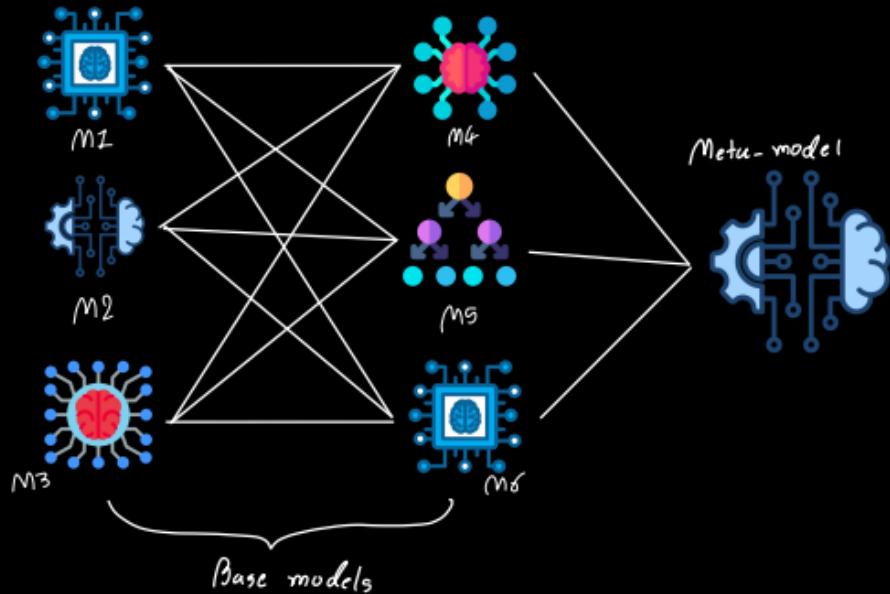
→ Now train Meta-model on nearly generated (800, 4) dataset where last column is actual output column.

→ At this stage, we have 12 trained base models, but we won't use it.

→ We will train our 3 base model on entire training dataset

→ The unique thing is that, Meta-model is trained before base-models.

Multi-layer Stacking



Training Process

- Models M1, M2, M3 will be trained on Td1 dataset
- These 3 models M1, M2, M3 will make prediction on Td1 dataset & will generate a new dataset ND1 (including OLP column) of (300, 4)
- Now model M4, M5, M6 will be trained over ND1 dataset and will make predictions over Td2 dataset.
- New dataset ND2 of size (300, 4) will be generated.
- Final meta-model will be trained over ND2 dataset.
- Mostly these kinds of architecture can be used in Kaggle competitions.

