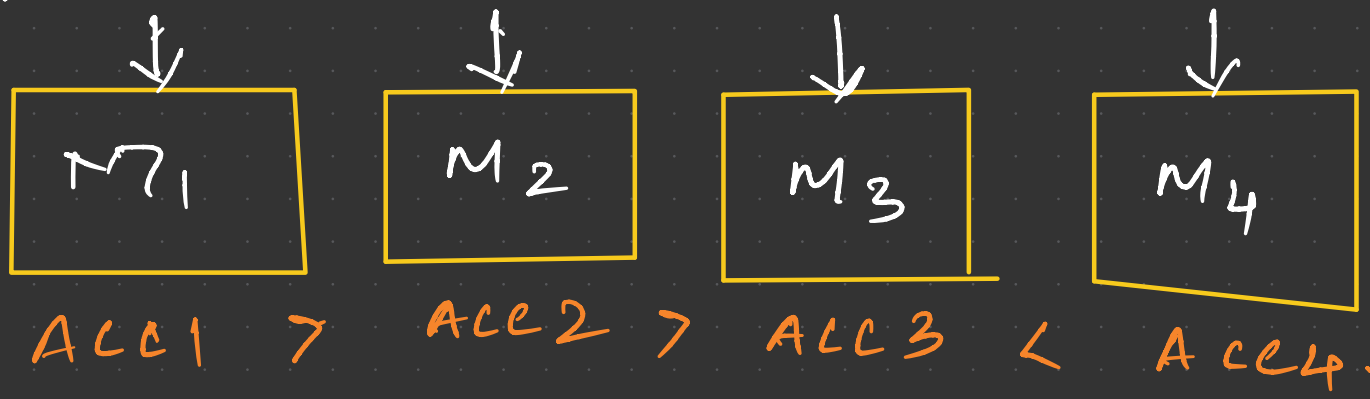
Principle component Analysis (PCA) [Dimensionality Reduction]

① Curse of Dimensionality

Dimensions →

Dataset = 500 features

3 features | 6 features | 15 features | 50 features



Price of the house

House size

no. of bedrooms

no. of washrooms

100 features →

← 500 features

For each model i will be getting an accuracy score.

From the above diagram, if you observe when the model was trained with 15 features, we were able to get the good accuracy score. However, when the model trained with higher number of features, the accuracy score decreases. If you look at the model M_4 which is trained with 50 features but the resultant accuracy score is low, that is because, the model is trained with the feature that are not relevent. this is called as model is overfeeded.

② Model performance Degrade

→ As the number of dimension increases model performance degrades.

Two ways to remove the curse of Dimensionality

① Feature selection

② Dimensionality reduction (PCA)

⇑

we take important features and train our model

⇑

Feature extraction