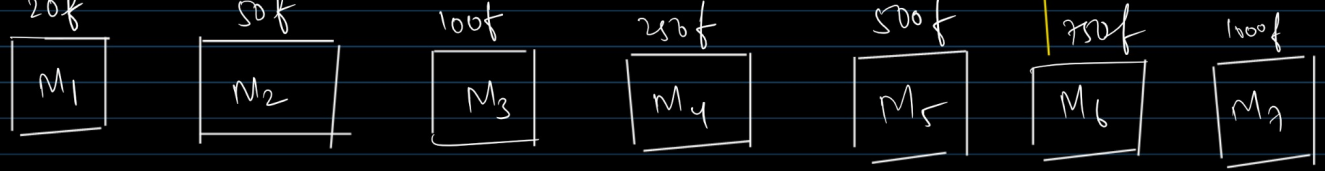
Principle of Component Analysis (PCA)

* Curse of dimensionality.
⇓
dimensions (features)

To predict the price of house

$f_1 f_2 \cdots f_{1000} \mid y$



Acc/Rsquare < Acc/Rsq ↑ < Acc/Rsquare ↑ < Acc ↑ ≃ Rcc/Rsq. ≃ Acc/Rsq ≃ Acc/Rsq.

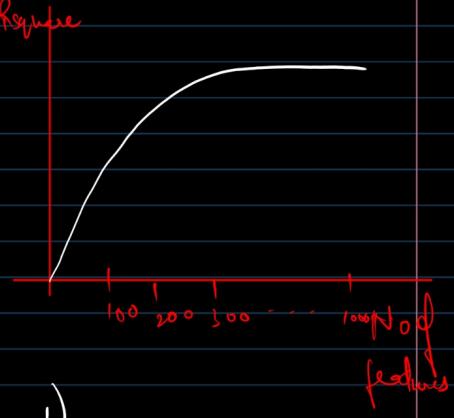
* With increase in no. of features, after one point of time the Acc/Rsquare will not increase in that proportion.

⇓
Why ?

→ Few of features will be multicollinear.
$(f_1 f_2 f_3 \simeq f_4)$

→ few of features might be exactly same.

$f_1 = [1, 1, 1, 1, 1, \ldots]$
$f_2 = [1, 1, 1, 1, \ldots]$



→ No variance | information in feature (Pattern can not be captured)

f_1

1
1.001
1.001
1.002

→ lot of duplicate enteries.

→ No business value.

Curse of dimensionality → With increase in number of features performance of model degrades.

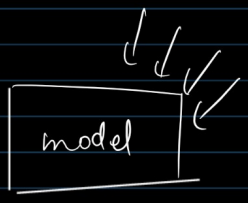
- Execution time
- Evaluation metrics

Analogy

You want to buy a house.

Broker

- 2BHK → 60 Lakhs
- 3BHK → ↑↑
- beach → ↑↑
- Airport → ↑↑
- Grocery shop → ≃
- near to a politician house → ≃
- Animal Shelter → ≃



* Curse of dimensionality → with increase in no of feature the performance of model degrades.

To remove Curse of dimensionality

① Feature Selection

② Feature Extraction

↓

PCA (Dimensionality reduction technique)