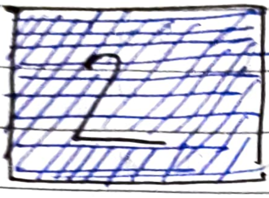
CURSE OF DIMENSIONALITY

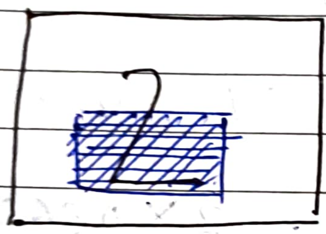
→ MNIST Data

① with all set of features



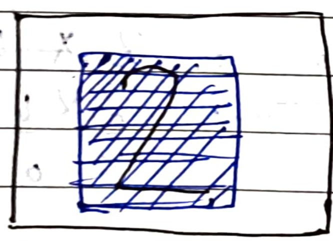
- Curse of dimensionality
- High computation cost

② with very less no features



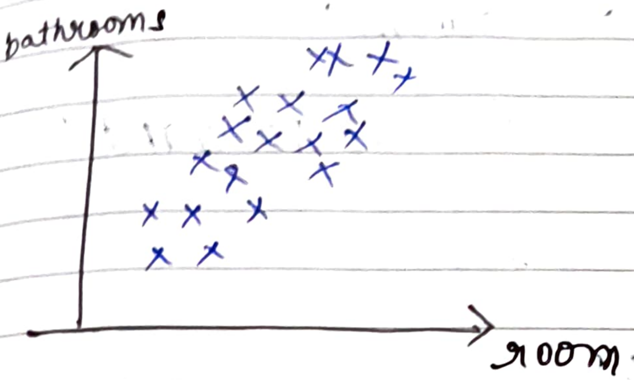
- low accuracy

③ With optimum no. features



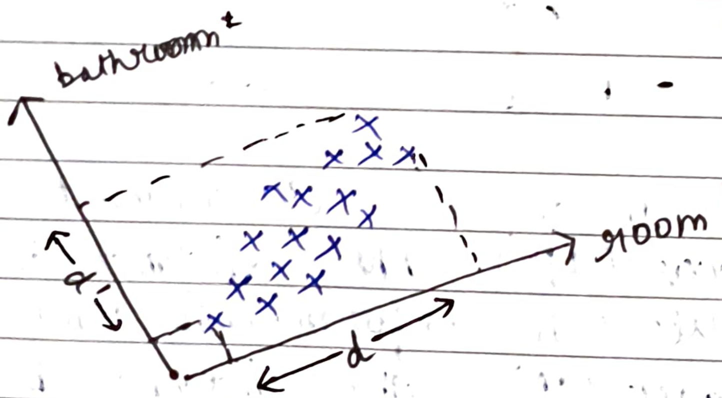
- No curse of dimensionality
- No high computation cost
- High accuracy

PCA



- Apply PCA

  - It will rotate the coordinate axis which will create new form of axis.



- variance of data on room' axis = d
- variance of data on bathroom' axis d'
- Clearly, d > d'
- Hence, we will keep room' as its variance (d) is more than bathroom.

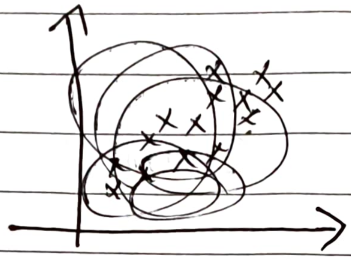
Note: When you shifted the coordinate, the new axis are called ~~PCA~~ Principle components

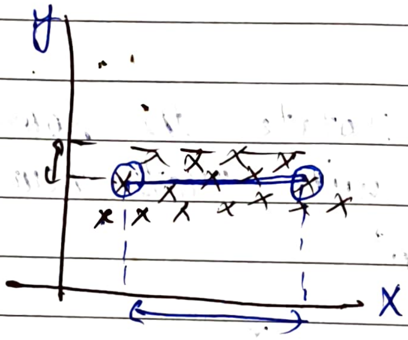
- Room' → PC 1
  Bathroom' → PC 2

• and then transform 'data', acc to your selected principle component ie PC1 → 'room'

Why Variance is important ?

• data in 2D





• The actual distance in 2D cordinates is given

in PCA

• ~~Bu~~ On x axis (1D), the distance is still comparable (ie information isn't loose much)
• But on y axis (1D), the distance becomes too small (Hence information is loose)
• That's in P.C.A, you select a component with high variance
• because when you bring data to lower dimension, you don't want to degrade the relationship between data points, that's why we maximize the variance