

Curse of Dimensionality

Columns \rightarrow features (dimension)

High dimensional data:
Image
Text

Higher dimension

\rightarrow Sparsity

If we have more features than observations than we run the risk of massively overfitting our model - this would generally result in terrible out of sample performance.

Performance decreases
Computation Increases

Solution : Dimensionality Reduction

feature selection

1) Forward Selection

2) Backward Elimination

feature extraction

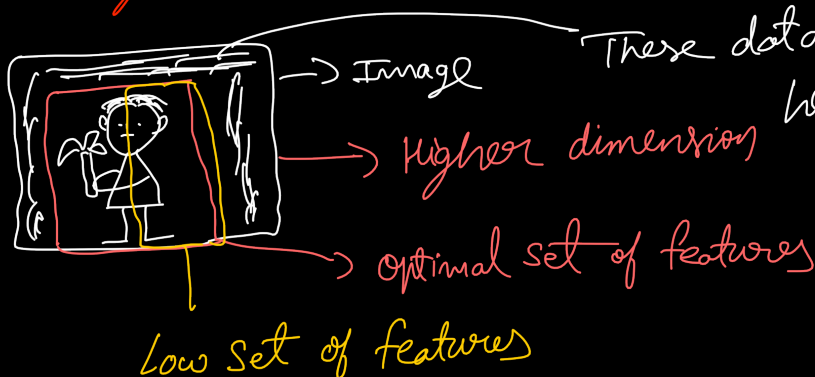
1) PCA

(from given set of features, creating new set of features)

2) LDA (Linear discriminant analysis)

3) tSNE

Higher dimensional data is dangerous for ML, because after a certain point, it is difficult to work in higher dimension.



These data points will not help in classifying given image.

