

Curse of Dimensionality

Features are also called dimension, so we can also call this curse of features.

As soon as number of features increases there is only a optimal number of features that can be utilized.

Even if you add more features they won't increase the accuracy whereas as they only increase the computational power.

$F_n \rightarrow$ optional features

$F_n \rightarrow$ features

$$F(\alpha) < F(\gamma)$$

\hookrightarrow Any number of features above F_n are just waste of computation power

Ex MNIST digit classifier

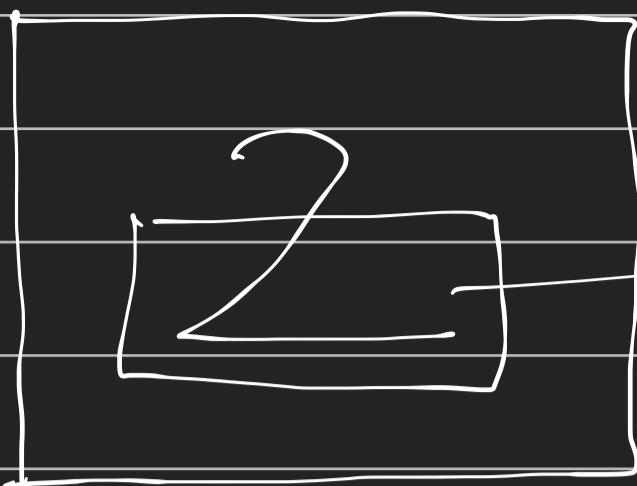
28×28

784



So we converted each pixel to a feature so

that means we had 784 features



Best if model
works on this area
the accuracy won't

be good best what if
model worked on this
area then we would had
the best accuracy possible



Need understand giving rest of this
area in green would be just waste
of computational power as it won't
help in classification.

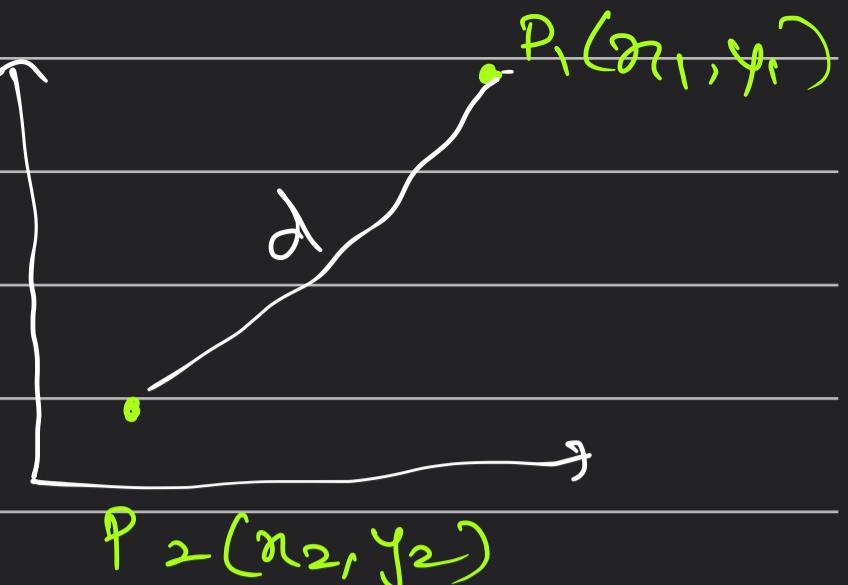
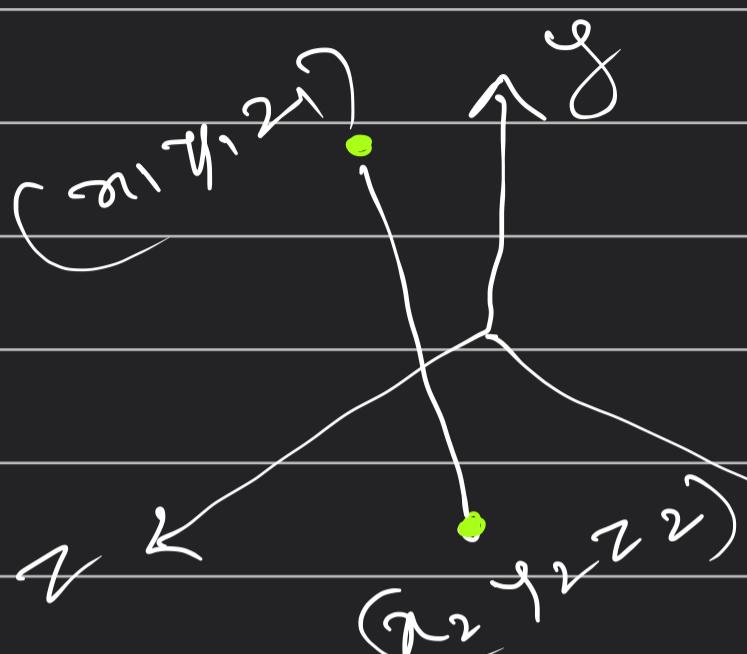
So curse of dimensionality states
that as soon as you start adding
feature your accuracy will increase
but tillle will be an optimal number
beyond which adding feature would
have any impact on the accuracy

☞ Best what exactly is the problem
in higher dimension space.

Ans ⇒ Sparsity

As soon as you increase the
dimension the data present will
become far away from each

other.

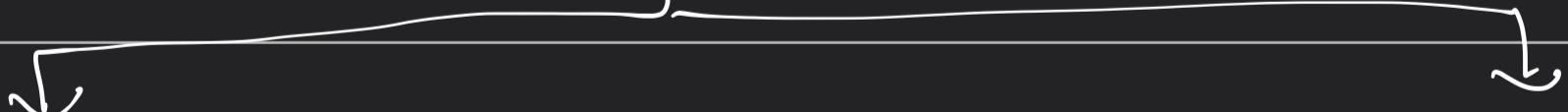


So as soon as dimension increases distance between points increases.

most of ML Algo are statistics based for ex KNN, neighbours gets determined on distance, so any neighbour won't be considered due to distance so ML Algo performance decreases and computational overhead increases.

Solutions :

Dimensionality Reduction



Feature Selection

- 1) forward Selection
- 2) Backward Elimination

Feature Extraction

- 1) PCA
- 2) Linear Discriminant Analysis
- 3) T-SNE

Conclusion : Higher Dimension data is dangerous.

