Curse of Dimensionality (COD)

columns
↓
features
↓
dimensions

When feature are increasing, it can improve model performance upto a limit.
If feature number → saturate.
↓
not any significant improve in model performance after adding any features.
↓
Model performance can decrease or model or computation become complex.

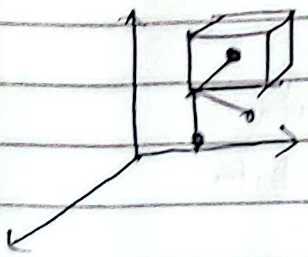
i.e.
fn < f(m)
↓ 5
↙
max no of features where model is working best if added can be in loss.

SOL. Curse of dimensionality says
when features add → model performance ↑
↓ but to
upto optimal point.
↓
after that
features add no improvement

Cause
↓
Higher Dimension
↳ problems.
↓
Sparsity
↳ Data sparse हुन्छ
means separating far from each other.

→ Reason: जति जति Higher Dimension हुन्छ त्यति त्यति sparsity हुन्छ ie than Data point खोज्न गाह्रो हुन्छ।

So, Higher Dimension मा ML algorithms will fail because of sparsity.
e.g.: KNN



So, affect of COD:
1. Performance decreases
2. Computation ↑

↳ Solving Technique
↓
Dimensionality reduction (reduce the no. of column)

Feature Selection
- Forward Selection
- Backward elimination

Feature Extraction
- PCA (Principle Component Analysis)
- LDA (Linear Discriminant Analysis)
- TSNE

- Principal Component Analysis (PCA)

  First paper of PCA → 1950

  So, old & Reliable technique

- Introduction

  PCA → Feature Extraction Technique

  Main ! → COD ↓ → features ↓ → model improved & fast ↑

  goal

- PCA ले Higher Dimension Data लाई lower Dimension Data मा लयाउछ ।
- PCA ले Best possible lower Dimension Data मा लेउछ So, Maximum property of HD Data will also be captured in LD.

  ↳ Conserves the behaviour of Higher Dimension data so that ML algorithm will perform good on LD Data also.

- Benefits of using PCA

1. faster execution of algorithms
2. Visualization (100 लाई · 2D मा · Convert गरेर graph plot गर्न सकिन्छ ।)