"PCA + tSNE"

① Assumptions of PCA

② Limitations of PCA

③ tSNE

$\rightarrow$ Intuition

$\rightarrow$ Code Implementation
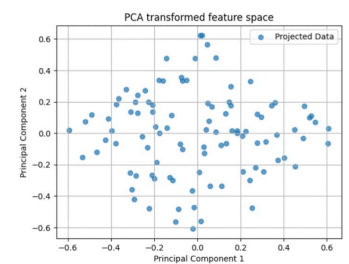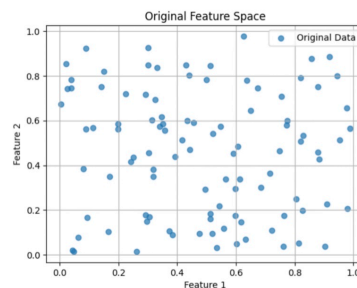
$\rightarrow$ Internals of tSNE

$\rightarrow$ After PCA

• Angle between PCs ? 90°

• Multi-collinearity ? Not present

• Assumptions of PCA

① Correlation between features

② Sensitive to scale of data

Original Feature Space

PCA transformed feature space

- Limitations of PCA :

① Low interpretability of PCs.

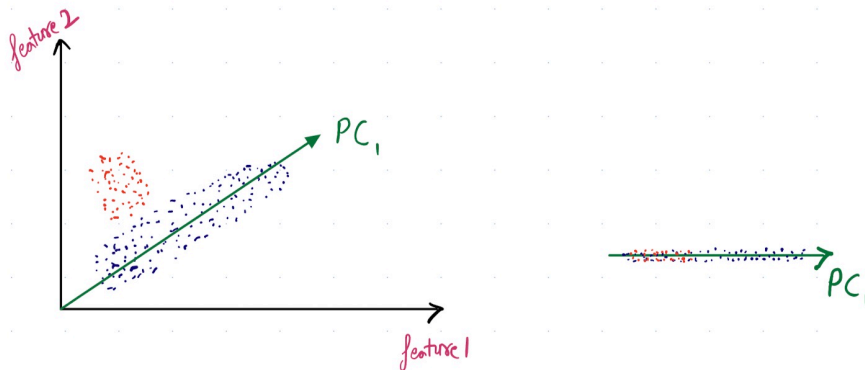② Information Loss (Dim. Reduction)

Robust against Outliers ? NO

→ Tries to find PCs in the direction of high variance
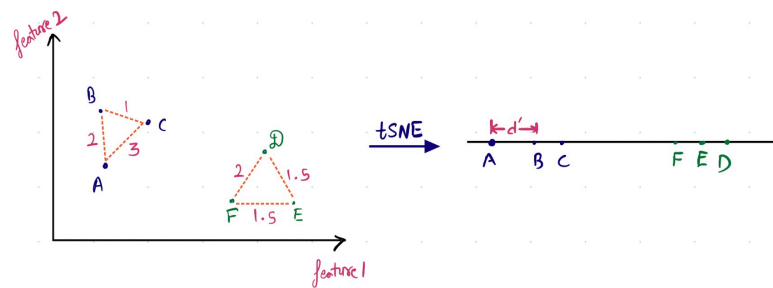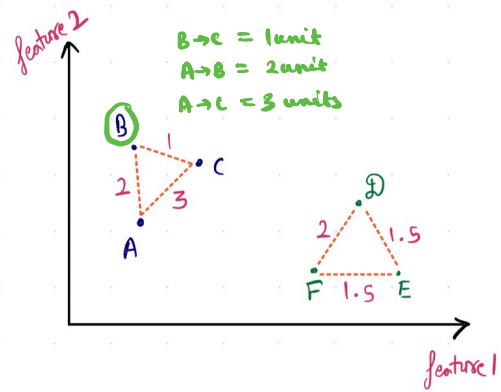
→ If outliers present (variance high)

      ↳ find PCs in that direction
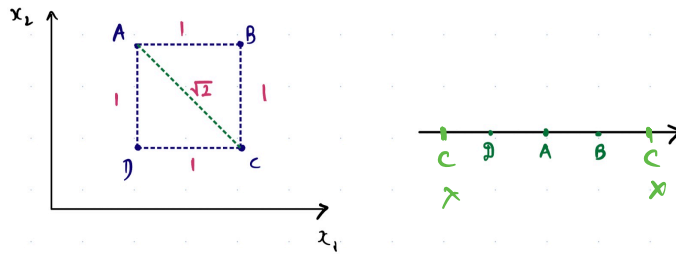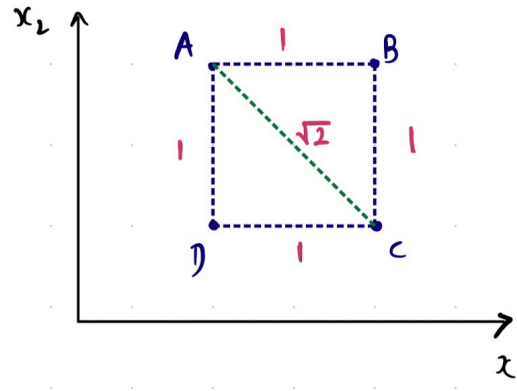
Does PCA capture local structure ? NO

• tSNE (t-distributed Stochastic Neighbor Embedding)

feature 2

B→C = 1 unit
A→B = 2 unit
A→C = 3 units

B
2   1
A   3   C

D
2   1.5
F 1.5 E

feature 1

A   B C   F E D

feature 2

B   1   C
2   3
A

D
2   1.5
F 1.5 E

tSNE

|←d'→|
A   B C   F E D

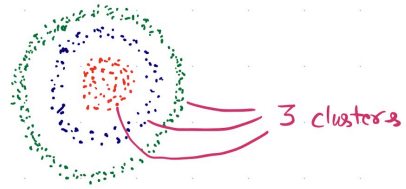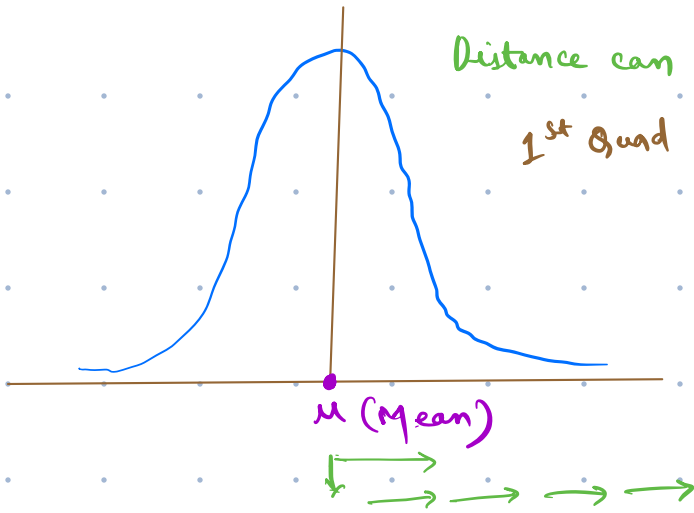feature 1

→ To summarize,

No matter where we place C,

↳ we will not be able to satisfy the condition

to maintain the distance proportion in lower dimension
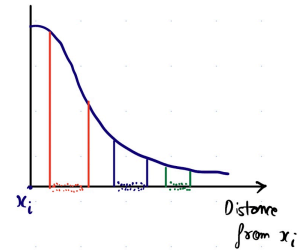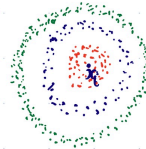
"CROWDING PROBLEM"

- Gaussian Distribution

Distance can never be '-ve'

1st Quad

μ (Mean)

3 clusters

$$P \propto \frac{1}{dist(a_i, a_j)}$$

$$P \propto \frac{1}{\sqrt{a_i - a_j}}$$

x_i

Distance
from x_i

Student's
→ t-distribution

- Points closer will have 'high' probability

- Points are far → Probability will be 'low'

(close to zero)