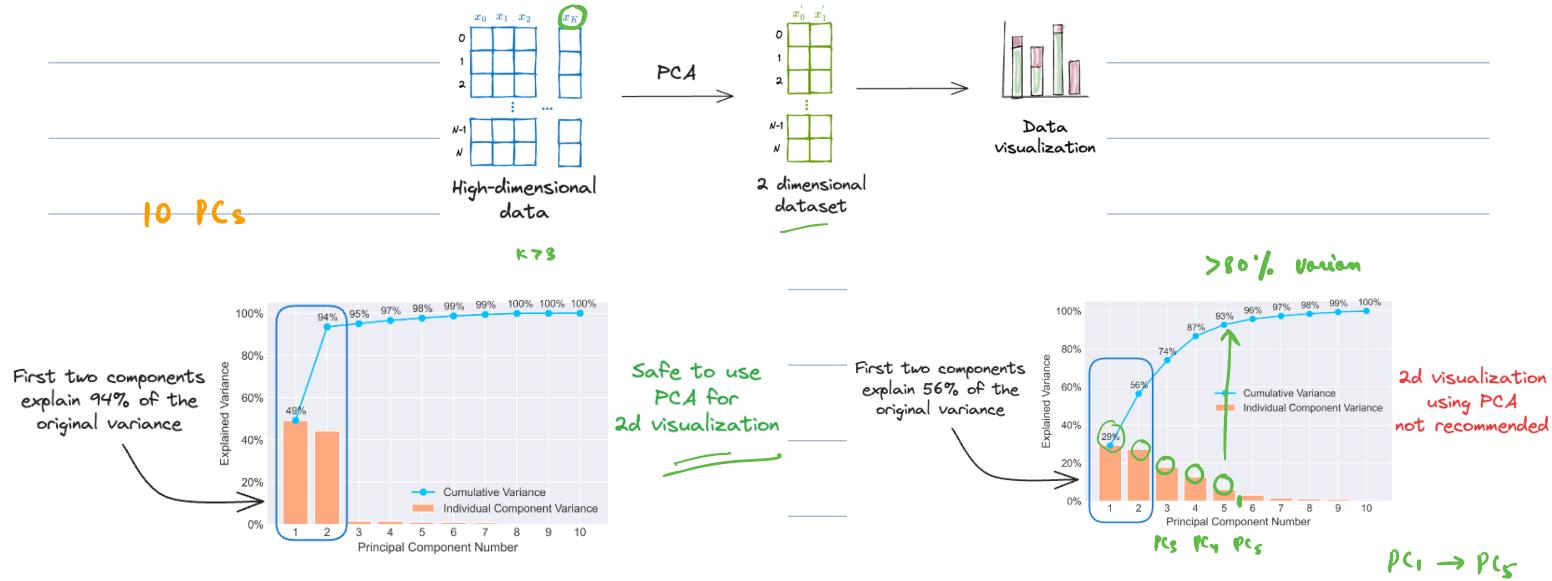
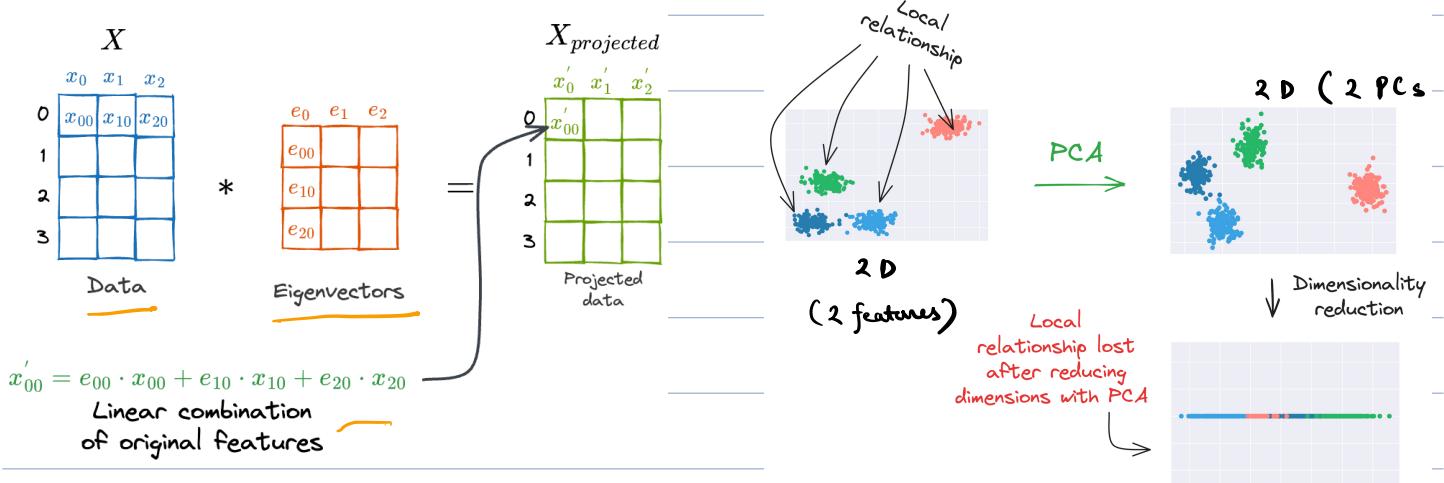


• PCA v/s t-SNE



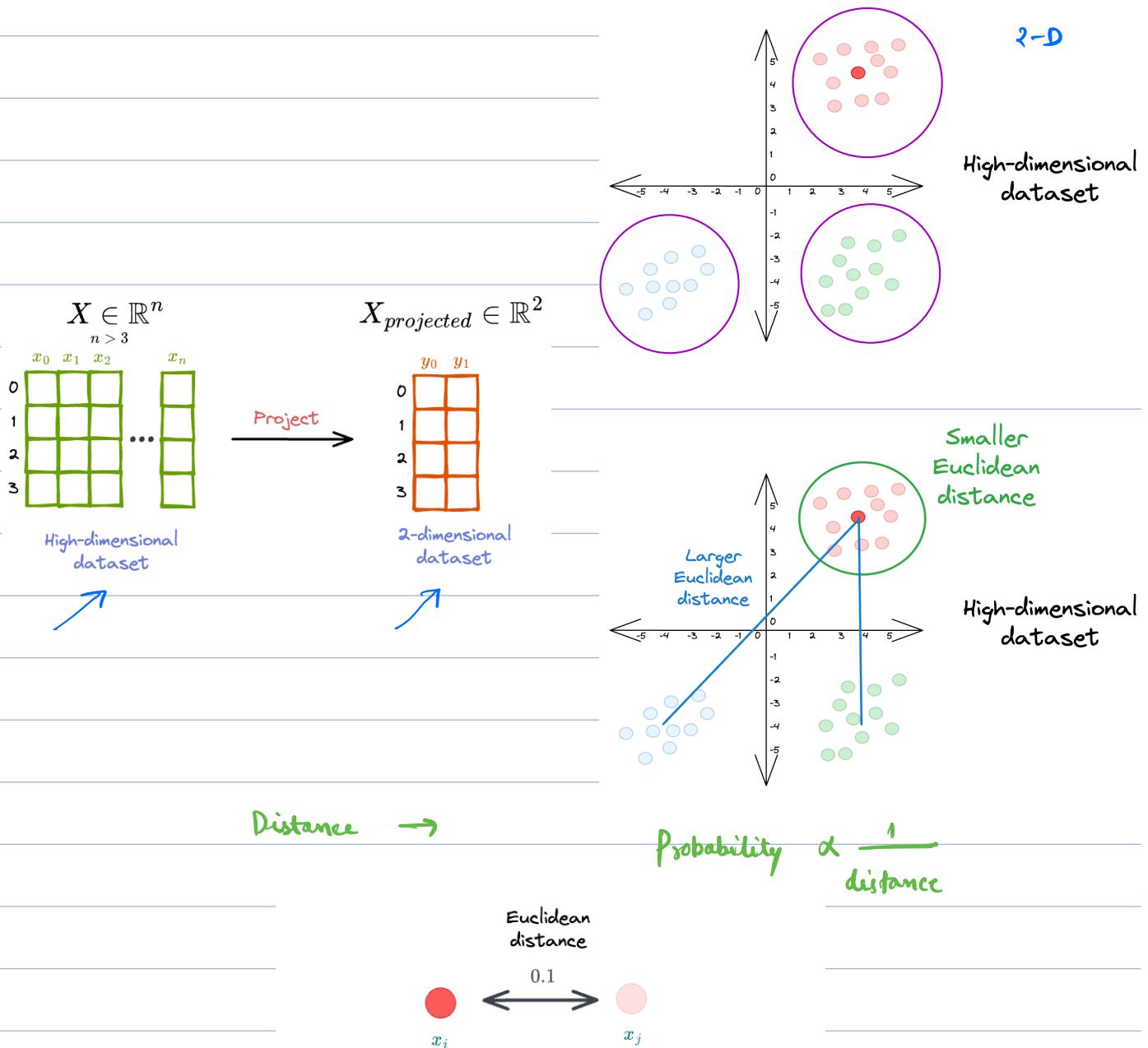
PCA process



PCA		t-SNE	
Purpose	Dimensionality reduction	Data visualisation in low dimensions (typically 2D)	
Type of algorithm	Deterministic	Stochastic	
Unique solution?	Yes	No	
Projection type	Linear	Non-linear	
Type of approach	Global	Local/Global*	
Interpretation	PCA is just a rotation of axes	Subjective interpretation	

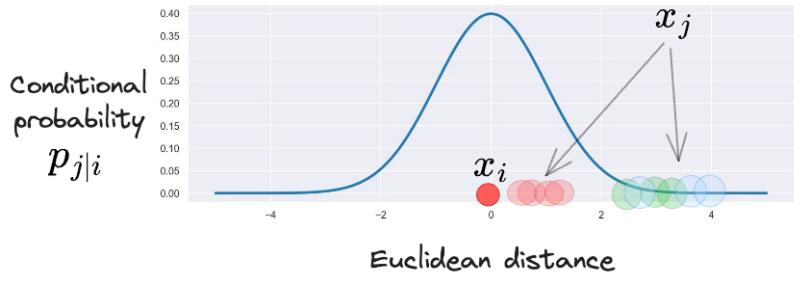
*We do not explicitly specify global structure preservation in tSNE.

Because of
Probabilities

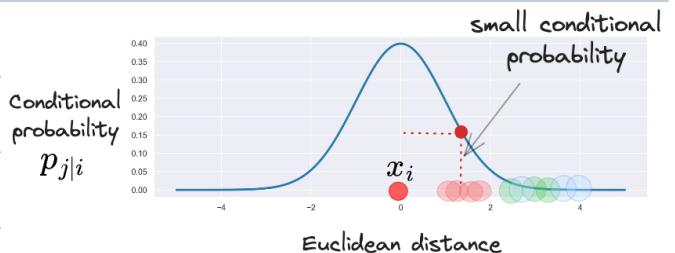
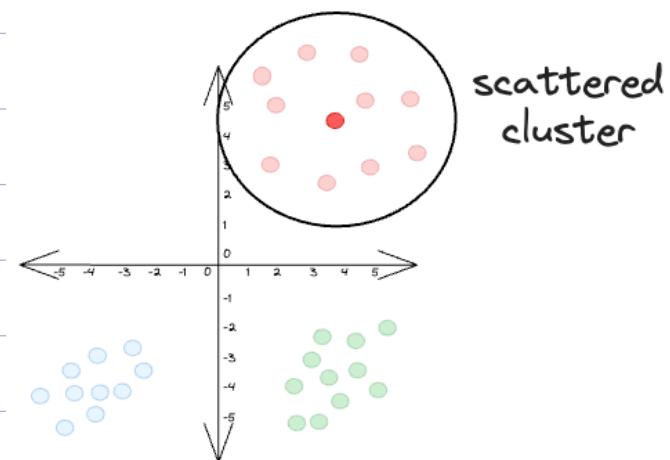
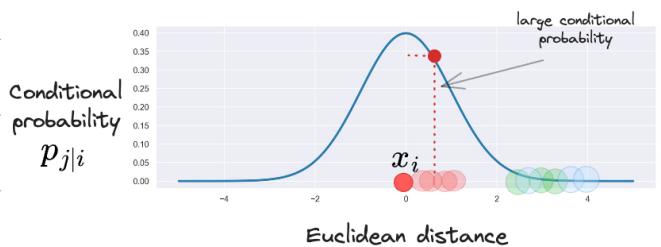
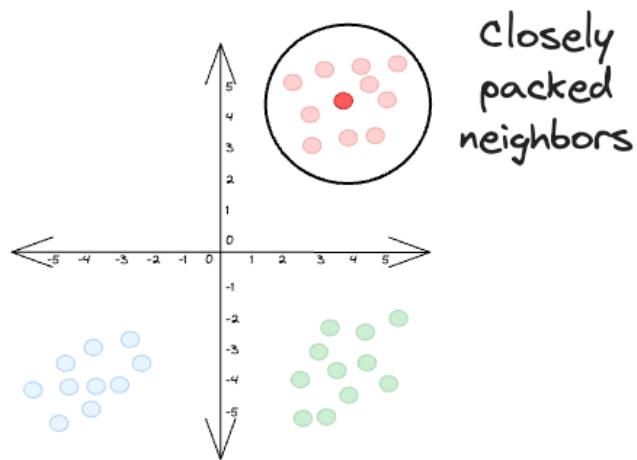


$p_{j|i} \rightarrow$ point i will select point j as its neighbor.

\rightarrow Conditional prob. is assumed to be proportional to probability density of "Gaussian" centered at x_i



$$p_{j|i} \propto \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)$$



$x_a \xrightarrow{1} x_b$
 $2.2 \xrightarrow{1.4} x_c$

Probability → A lot of variance in probability

Normalization of Probability

$$p_{j|i} = \frac{p_{j|i}}{\sum_{k \neq i} p_{k|i}} \quad \text{Sum of } p_{k|i}$$

$x_a, x_b \rightarrow 0.6$

$x_b, x_c \rightarrow 0.3$

$x_c, x_a \rightarrow 0.1$

$$= \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)}$$

High Dimension $\rightarrow p_{ij}$

Low Dimension $\rightarrow q_{ij}$ Steps Iteration

$\xrightarrow{\text{Project}}$

x_i
High-dimensional data point

y_i

Low-dimensional data point

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)}$$

After 'n' iterations

\hookrightarrow differentiated points

$$p_{j|i} \longleftrightarrow \text{minimum difference}$$

"KL-Divergence"

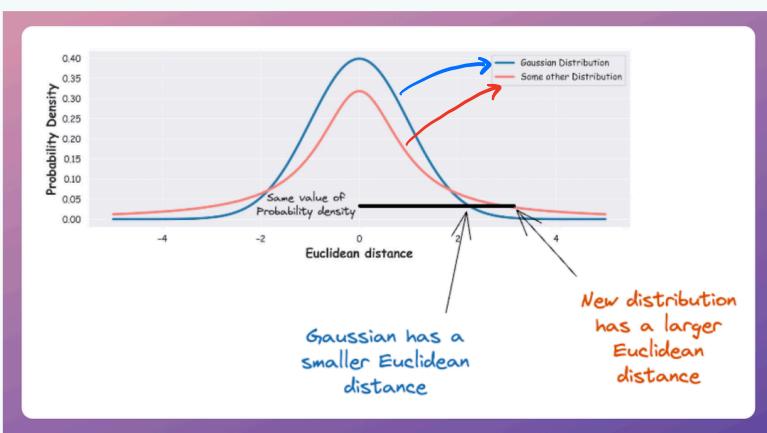
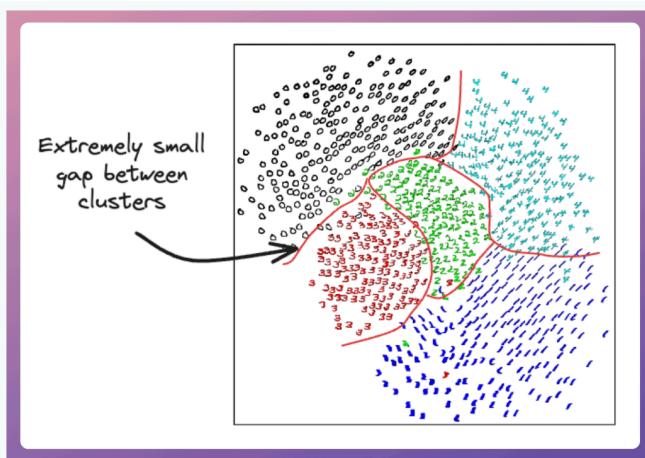
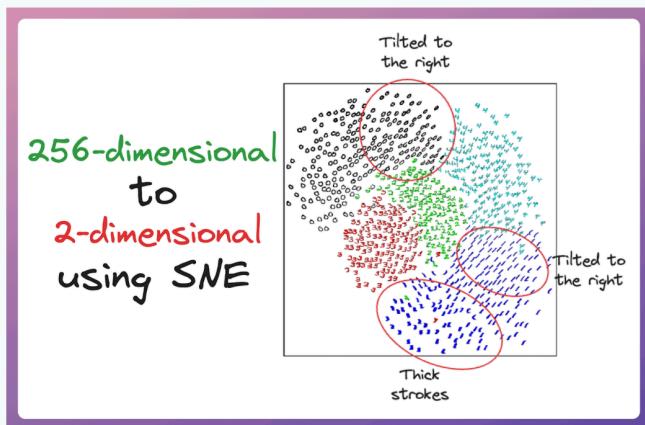
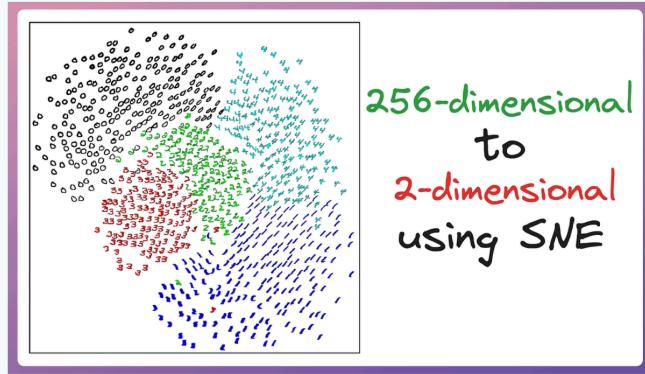
$$D_{\text{KL}}(P \parallel Q) = \sum_x P(x) \cdot \log\left(\frac{P(x)}{Q(x)}\right)$$

sum of KL divergences over all data points

$$C = \sum_i D_{\text{KL}}(P_i \parallel Q_i) = \sum_i \sum_j p_{j|i} \log\left(\frac{p_{j|i}}{q_{j|i}}\right)$$

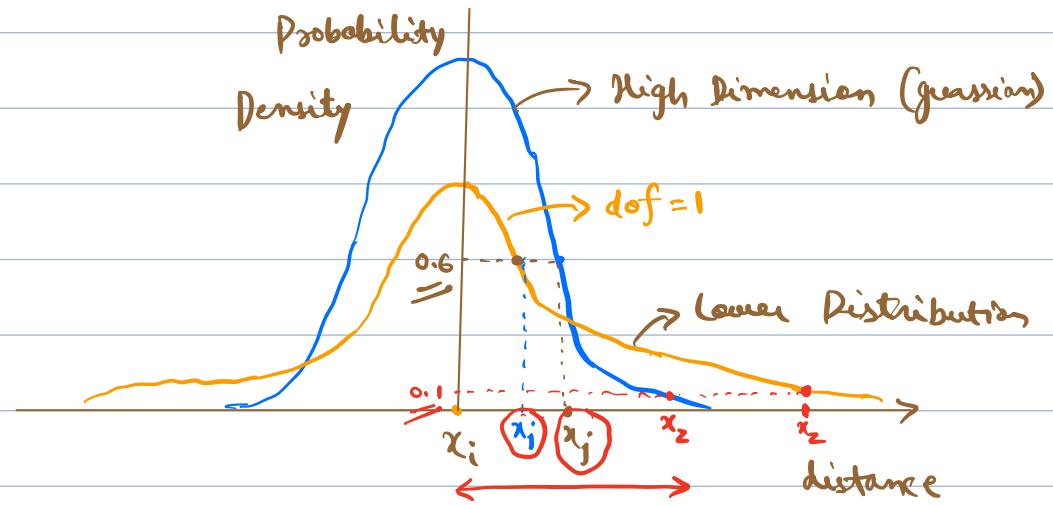
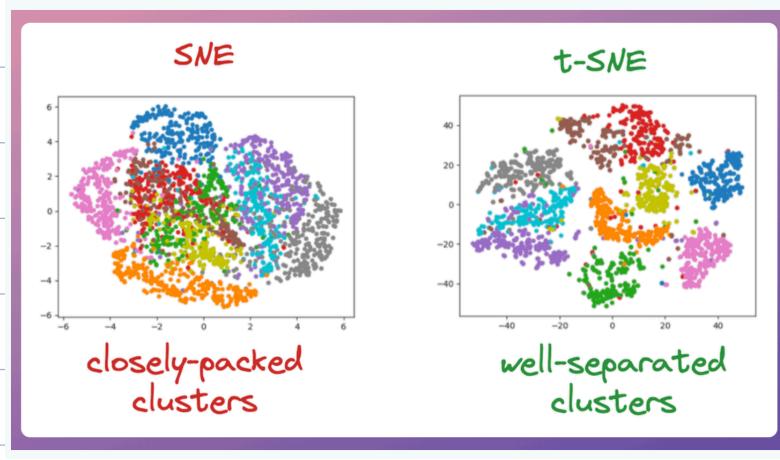
Cost function

Ideal situation,
 $P_{j|i} = Q_{j|i}$



Higher Dimension \rightarrow Gaussian

Lower Dimension \rightarrow t-distribution



• DOF (Degree of Freedom)

\hookrightarrow dof = 1 \rightarrow t

dof = 10 \rightarrow t

dof = 25 \rightarrow t

dof = 30 \longrightarrow Gaussian
(Standard Normal)

• Hyperparameters

\rightarrow Perplexity \rightarrow Range [0 - 50]

\rightarrow n-iter (steps) typically around 30

Limitations of tSNE:

① Algorithm is slow

② Hyper-parameters tuning