

# ## UMAP ##

It is a dimensionality reduction technique, takes high dimensional data and projects it to a lower-dim space.

To perform UMAP:

1) Calculate probabilities of the points, i.e. one point  $j$  being a meaningful neighbour of  $i$ . Steps to do that:

i) Calculate pairwise Euclidean distances

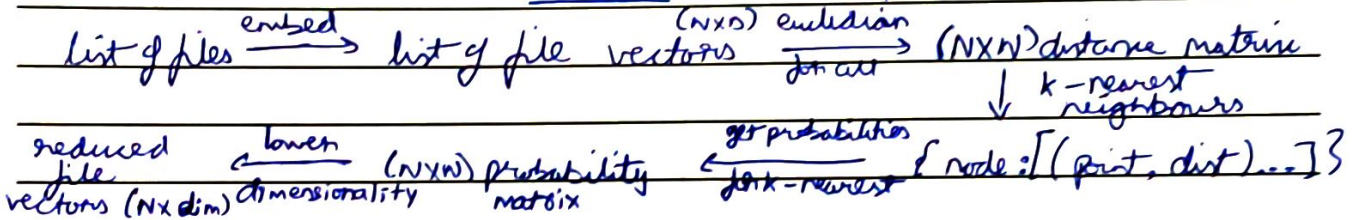
ii) Select nearest neighbours.

iii) Calculate probabilities

iv) Symmetrize probabilities.

WORKFLOW

$$\sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$



2) Perform dimensionality reduction using obtained probability matrix.

$$\text{Loss} = \sum_{i,j} [p'_{ij} \log q_{ij} + (1 - p'_{ij}) \log (1 - q_{ij})]$$

where  $q_{ij} = \frac{1}{1 + d_{ij}^2}$  (reduced dim distance)

high dim symmetric probabilities  $\rightarrow$  low dim probability

$$\therefore \frac{\partial \text{loss}}{\partial \text{embedding of point } i} = \sum_j 2 \cdot (p'_{ij} - q_{ij}) \cdot \frac{x_i - x_j}{(1 + d_{ij}^2)}$$

Update embedding using sgd and above gradient.