**Inputs:**
1. Training points
2. Prediction points
3. Number of folds (k)
4. Clustering algorithm (hierarchical or k-means)

**Outputs:**
1. Fold indices
2. Nearest neighbour distance functions
3. W statistic
4. Number of intermediate clusters (q)

---

Compute $\hat{G}_j(r)$ and $\hat{G}_{ij}(r)$

KS test for clustering

No clustering

Return a random k-fold CV

Cluster N training points into $q \in Q$ groups

where $Q = [k, \ldots, N]$

For each $q > k$:

Merge $q$ clusters into $k$ folds along the first principal component of the training points’ coordinates

Select $q$ such that $W$ is minimized

For each $q$:

Compute $\hat{G}_j^*(r, L)$ and $W$ statistic between $\hat{G}_j^*(r, L)$ and $\hat{G}_{ij}(r)$

Return a kNNDM k-fold CV

Cluster N training points into $q \in Q$ groups

where $Q = [k, \ldots, N]$

For each $q > k$:

Merge $q$ clusters into $k$ folds along the first principal component of the training points’ coordinates

Select $q$ such that $W$ is minimized

For each $q$:

Compute $\hat{G}_j^*(r, L)$ and $W$ statistic between $\hat{G}_j^*(r, L)$ and $\hat{G}_{ij}(r)$

Return a kNNDM k-fold CV