

Basic of clustering

- 1) Verify Eq. 14.25 in ESL
- 2) Show that $\bar{d}_j = 2 \cdot \text{var}_j$ in Eq. 14.27 in ESL

k-mean and GMM

- 3) Is the k-mean cost function Eq. 9.1 in PRML the same as those in Eq. 14.31 in ESL? If not, discuss what difference can it make.
- 4) Refer to the list of challenges in clustering discussed in the lecture, compare GMM with k-mean in terms of their ability to handle these challenges.
- 5) Read the paragraph under Eq. 9.14 in PRML for the singularity problem of GMM
- 6) Verify Eq. 9.17-22 in PRML and discuss their meaning
- 7) (Difficult) In PRML, starting from the log-likelihood (Eq. 9.14), show that its maximization in the limit $\epsilon \rightarrow 0$ is the same as the minimization of the k-mean cost function (Eq. 9.1). (You should start from Eq. 9.14 and show that it is related to Eq. 9.1 by a multiplication constant as $\epsilon \rightarrow 0$)

Hierarchical clustering

- 8) In Chapter 14 of ESL, the single and complete linkage were introduced as group-to-group distances for bottom-up hierarchical clustering. Two artifacts, chaining for single linkage and failure in handling clusters with long shape for complete linkage, were mentioned in the class. Provide more examples when the single and complete linkage may fail.
- 9) Consider 3 cases, a) data points generated from uniform distribution, b) data points generated from a single multivariate Gaussian, and c) data points generated from two separated multivariate Gaussian with small overlapping. Draw schematically the dendrograms resulted from applying hierarchical clustering to these 3 cases. What are the most distinct features in the trees for the 3 cases?

PCA and MDS

- 10) Show that $W^\top = I_{p \times D} V^\top$ minimize the reconstruction error Eq. 2.12 in NDR.
- 11) The coding $Y \rightarrow X$ in principal coordinates is given by $\hat{X}_{MDS} = I_{p \times N} \Lambda_{MDS}^{1/2} U^\top$, discuss where and how the information of Y enter this coding transformation
- 12) PCA and classical metric MDS are the same when the Euclidean distances are used. Point out explicitly where in PCA and classical metric MDS the assumption of Euclidean distances is imposed.
- 13) Show that the principal coordinates $\hat{X}_{MDS} = I_{p \times N} \Lambda_{MDS}^{1/2} U^\top$ are centered.

Validation

- 14) Refer to the recommended exercises in the lecture note for the validation lectures.

Graph construction

- 15) Give 2 examples (or more), one for kNN and one for mutual kNN graph constructions, that they may fail and then propose solution for each of them.

Graph Laplacian and related

- 16) In class, we discuss that the unnormalized Laplacian is independent of the weights of the self-edges, i.e., ω_{ii} for $i = 1, \dots, N$. This means that all its eigenvalues and eigenvectors are also independent of ω_{ii} . How about the normalized symmetric Laplacian and the random walk Laplacian, and their eigenvalues and eigenvectors?
- 17) The transition probability of a graph is given by w_{ij}/d_i in the random walk Laplacian, show that the stationary distribution is given by $\pi_i = d_i/\text{vol}(V)$, where d_i and $\text{vol}(V)$ are the degree of the i -th node and the volume of the graph, respectively.

- 18) At stationary state and consider that the vertices of a graph is partitioned in only two non-overlapping subsets, A and $\bar{A}$ (i.e. the complement of A). Show that $P(A|\bar{A})P(\bar{A}) = P(\bar{A}|A)P(A)$ and discuss the meaning of this equality.
- 19) Show that the commute time distance (CTD) is invariant if all the weights of a graph is scaled by the same factor, i.e., $w_{ij} \rightarrow a w_{ij}$ with $a > 0$.
- 20) Consider graphs with a single connected component, the CTD embedding coordinates for L_{sym} is $x_{ai}^{\text{sym}} = v_{ai}^{\text{sym}} \sqrt{\frac{\text{vol}(V)}{\lambda_{\alpha}^{\text{sym}} d_i}}$ as discussed in the class. Show that a) $E(x_{\alpha}^{\text{sym}}) = 0$ for $\alpha > 1$ with the weight of each data point given by $P(i) = d_i/\text{vol}(V)$. b) With the same weights in a, find the covariance matrix $E(x_{\alpha}^{\text{sym}} x_{\alpha'}^{\text{sym}})$ for $\alpha, \alpha' > 1$.

Density-based clustering

- 21) Propose 2 (or more) improvements for DBSCAN. Show the details and justify your suggestions.
- 22) Propose 2 (or more) improvements of the density based clustering method in the 2014 paper by Rodriguez & Laio. Show the details and justify your suggestions.

ISOMAP, LLE

- 23) List 2 (or more) possible problems of using ISOMAP for embedding. Justify your answer.
- 24) In LLE, show that the weights w_{ij}^{min} that minimize the cost function

$$\epsilon(w) = \sum_{i=1}^N \left| \vec{x}_i - \sum_{j \in \text{kNN of } \vec{x}_i} w_{ij} \vec{x}_j \right|^2 \text{ is invariant under translation, rotation and rescaling of } \vec{x}.$$

- 25) In LLE, show that the cost function $\epsilon'(w) = \left| \vec{x} - \sum_{j \in \text{kNN of } \vec{x}} w_j \vec{x}_j \right|^2$ with $\sum_{j \in \text{kNN of } \vec{x}} w_j = 1$ can be written as $\sum_{j,k} w_j C_{jk} w_k$ with $C_{jk} = (\vec{x} - \vec{x}_j) \cdot (\vec{x} - \vec{x}_k)$. Then

$$\text{show that the minimization of } \epsilon'(w) \text{ leads to the solution } w_j^{\text{min}} = \frac{\sum_k C_{jk}^{-1}}{\sum_{l,m} C_{lm}^{-1}}.$$

Discuss in what situation that C^{-1} does not exist (i.e., when C is a singular matrix) and what are the implications in terms of the intrinsic dimension of the data structure locally around $\vec{x}$. Finally propose a simple way to evaluate w_j^{min} numerically to avoid the problem of divergence.

- 26) In LLE, show that the cost function $\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j w_{ij} \vec{Y}_j \right|^2$ can be written as

$$\sum_{i,j} M_{ij} (\vec{Y}_i \cdot \vec{Y}_j) \text{ with } M_{ij} = \delta_{ij} - w_{ij} - w_{ji} + \sum_k w_{ki} w_{kj}.$$

How does the matrix M relate to the random walk Laplacian L_{rw} . Does M and L_{rw} share the same eigenvectors (i.e., if the projection of LLE is the same as the projection using the eigenvectors of L_{rw})?