STOCKHOLM UNIVERSITY
DEPT. OF MATHEMATICS
Div. of Mathematical statistics

MT 7039
EXAMINATION
26 Apr 2021

Exam in Unsupervised Learning
26 Apr 2021, time 8:00-13:30

Examinator: Chun-Biu Li, cbli@math.su.se.
Permitted aids: When writing the home exam, you may use any literature.
Return of the exam: Results of the exam will be announced via the course page.

NOTE: The exam consists of 3 problems with 100 points in total. Logical explanation and steps leading to the final solution must be clearly shown in order to receive full marks. Minimum points to receive a given grade are as follows:

A	B	C	D	E
90	80	70	60	50

NOTE: If you studied the recommended exercises together with your classmates, please **write with your own words** in answering the questions in this exam.

NOTE: For those parts require explanations, your writing must be to the point, **redundant writing irrelevant to the questions will result in point deduction**.

Problem 1 (Basics of unsupervised learning, total 36p)

Part a and b of this problem refer to the course book "Pattern Recognition and Machine Learning".

a) Given the log likelihood, $\ln p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma})$, of Gaussian Mixture Model (GMM) in Eq. 9.14, show that $\partial \ln p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma})/\partial \mu_{\mathbf{k}} = 0$ gives Eq. 9.17 and Eq. 9.18. **(6p)**

b) Show that $\partial \ln p(\mathbf{X}|\pi, \mu, \mathbf{\Sigma})/\partial \pi_{\mathbf{k}} = 0$ constrained by $\sum_{k=1}^{K} \pi_k = 1$ gives Eq. 9.22 **(10p)**.

Part c and d refer to the course book "Nonlinear Dimensionality Reduction".

c) Show that PCA minimizes the reconstruction error given in Eq. 2.12. **(12p)**

d) We learned that PCA and classical MDS are equivalent when the feature space is Euclidean. Explain concisely in which parts of the PCA construction, classical MDS construction and in the proof of their equivalence that this Euclidean assumption are imposed. You can refer to the equations in the course book "Nonlinear Dimensionality Reduction". **(8p)**

Problem 2 (Graph based methods, total 38p)

a) Consider the kNN method to construct a graph with a **single connected component** from data, give one example that this method may end up with a very big value of k resulting in a lot of short circuits in the graph. **(5p)** Then propose a solution for this problem. **(5p)**

b) Consider the **mutual** kNN method to construct a graph with a **single connected component** from data, give one example that this method may end up with a very big value of k resulting in a lot of short circuits in the graph. **(5p)** Then propose a solution for this problem. **(5p)**

c) Do the eigenvalues of the unnormalized Laplacian L depend on the self weight of a graph, i.e., w_{ii}, with $i = 1, 2, \cdots, N$? **(3p)** How about in the case of the normalized Laplacians L_{sym} and L_{rw}? **(5p)** Please justify your answers.

d) The commute time distances (CTDs) c_{ij} between the i-th and j-th data points can be expressed in terms of the eigenvalues and eigenvectors of both the unnormalized Laplacian and the symmetric Laplacian as

$$c_{ij} = vol(G) \sum_{\alpha=2}^{N} \frac{1}{\lambda_\alpha}(v_{\alpha i} - v_{\alpha j})^2 = vol(G) \sum_{\alpha=2}^{N} \frac{1}{\lambda_\alpha^{sym}}(\frac{v_{\alpha i}^{sym}}{\sqrt{d_i}} - \frac{v_{\alpha j}^{sym}}{\sqrt{d_j}})^2,$$

where $vol(G)$ is the graph volume, d_i is the node degree, λ_α and v_α are eigenvalues and eigenvectors of the unnormalized Laplacian, and λ_α^{sym} and v_α^{sym} are eigenvalues and eigenvectors of the symmetric Laplacian.

When comparing with the results in part c, argue **in words** if the CTDs depend on the self weights w_{ii} of the graph from both the viewpoints of the unnormalized Laplacian **(5p)** and the symmetric Laplacian **(5p)**.

Problem 3 (Validation Methods, total 26p)

This problem refers to the lecture note on validation methods (See "Lecture note" on Lecture 6: Feb 4 in the course page)

a) Referring to the 10-cluster example in P.13 of the lecture note, explain in terms of the concepts of cohesion and separation why there is an initial rise and then a subsequent drop of the Silhouette index before and after the cluster number 10. **(10p)**

b) Name one limitation of the Silhouette plot and index to validate clustering results and propose a solution for it. **(6p)**

c) Referring to P.21 of the lecture note on co-rank matrix, show that $Q_{NX}(K) = \frac{1}{Kn} \sum_{i=1}^{n} |\Psi_K(i) \bigcap \Psi'_K(i)|$. **(10p)**

Good Luck!