OM-S20-20: Optimization for Manifold Learning (Cont.)

C. V. Jawahar

IIIT Hyderabad

http://preon.iiit.ac.in/om_quiz

31 Mar 2020

Announcements

- We had our earlier evaluation scheme
  - Exams: Mid: 20 and Final: 30
  - HW: 25, Assignments: 15 and Reviews: 10
- We need to revisit this in the new situation
  1. All online Modules (eg. manifold learning) including a set of questions: (15)
  2. Participation in the online discussions in the team channels (10)
  3. Increase the weight of reviews by 5% with possibly more questions and some redundancy (take best 85% questions).

Review

We will start with the review today. Please downlaod questions and and answer in 15 minutes.

We have now a set of questions to initiate discussions in "team channel". Please discuss in "Manifold Learning Channel". We will spend 5 mins in discussing here. But continue for more than a day (until tomorrow late evening).

Q1

Q1: We know K Nearest neighbour graph.
S1: "If $K = N$, this graph is fully connected"
S2: "If $K = 1$, this graph can never be a single connected component"
A reverse NN Graph can be constructed by connecting all samples to other sample whose K nearest neighbour contains this one.
S3: "Both K NN Graph and Rerverse K NN Graphs are the same"
Discuss True/False with examples, where ever appropriate.

Q2

Q2: How is the Laplacian eigenmap (Refer 39:17, Ali Ghodsi, Lec 6) related to spectral clustering? Analytically show how objectives are different or related.

Q3

Q3: We have a set of 1000 (N) points in 20 (d) dimensional space in a line. Samples are equidistant. (Think of equidistant points on a line in a high dimensional space). We are interested in reducing the dimension to 1 (p). d is our original dimension and p is the new dimension.
"In this case, PCA and LLE will yield the same answer".

- Agree/Disagree
- Experimentally validate
- Analytically/Formally Prove
- Rewrite into a more formal and right statement. (A lemma in your name!)

Q4

Q4: We have a set of 1000 (N) points in 20 (d) dimensional space in a line. Samples are equidistant. (Think of equidistant points on a line). We are interested in reducing the dimension to 1 (p).
"In this case, the Y computed out of ISOMAP are independent of K".
(i.e., whether $K = 2, 5$ or 10, we will get the same Y)

- Agree/Disagree
- Experimentally validate
- Analytically/Formally Prove
- Rewrite into a more formal and right statement. (A lemma in your name!)

Q5

Q5: Consider 10 points in 2D (two well seperated clusters of 5 points each)

$$(1,1),(0,0),(0.5,0.5),(0,1)(1,0)$$

$$(11,11),(10,10),(10.5,10.5),(10,11)(11,10)$$

Will K Means and Ratio Cut yield same answer for these?
Discuss

Q6

Q6: With an example set of > 5 points and $K = 2$, show that K means can oscillate.
Suggest a heuristic that can avoid oscillation in this case.

Q7

Q7: With an example set of > 5 points and $K = 2$, show that K means lead to a local optima that is suboptimal.
(i.e., show a superior solution (better objective) to this problem exist, but K means did not give us in our attempt, may be due to bad initialization.)

Q8

Q8: We know that MDS and ISOMAP(?) optimize $J = ||D^X - D^Y||$.
Where D^Y is the pairwise distance matrix in p dimension.

Statement: "If p increases J decreases."
Always True/Never True/Sometime True/There are Exceptions/Analytically Provable.(create a lemma in your name!)

Q9

Q9: List some (say 10) advances to LLE and ISOMAP that are popular. Mention algorithm name, reference (say paper/project page/code). (Lecture mentions some. use google beyond!)

Q10

Q10: We know how to find W in LLE as an optimization problem. Write the objective. Write the solution as steps (clearly define variables, matrix dimensions, and matrix operations)

Q11

Q11: We know how to find Y in LLE as an optimization problem. Write the objective. Write the solution as steps (clearly define variables, matrix dimensions and matrix operations)