Final Exam
Mathematical Foundations of DS

Date: 96.11.2

Time: Start (9am 96.11.2) – Stop (13pm 96.11.4)

***You can consult anyone or anything except your classmates.**

!Unreadable documents will not go through the grading procedure

1. Let $X = [a, b]$ be a closed interval in the real line with its natural metric and the Lebesgue measure, and also let $Y = P_n$ be a path on n vertices with weight one for all edges and vertices. Moreover, $Z = Pet$ is the Petersen graph with vertex-weights equal to one in which all edges of the outer and inner 5-cycles have weight one but the edges of the matching in between have weights equal to 2. Compare the connectivity of these three spaces using the following parameters (note: weights represent similarity).
   a) Best constant for a Poincare inequality.
   b) Relaxation time of the natural random walk. (Ignore X if you do not know about Brownian motions.)
   c) How can you compare the connectivity of X with Y. Try to propose a method so that you can deduce which one is more connected. (discuss)
   d) Someone asks about discretizing continuous spaces and how this may affect their connectedness. What can you say? (discuss)
   e) Using your answer, does it make sense to consider a continuous manifold and a smoothness procedure for a given (discrete) data-set and then use the smoothness to analyze the data in a better way using differentiability? (discuss in detail)
   f) Use spectral clustering to bisect Z into two parts. Write down all the details of your computations. Does your answer make sense intuitively? Try this again for three parts and justify your answer. How good is this method?

2. Let X and Y be nonempty closed and bounded subsets in $\mathbb{R}^3$ and let

$$X + Y = \{x + y \mid x \in X,\ y \in Y\}.$$

   a) Prove that $Vol(X + Y)^{1/3} \geq Vol(x)^{1/3} + Vol(Y)^{1/3}$. What does this say about subsets of the space? (discuss)
   b) Does the real space satisfy some sort of concentration of measure? What about the real sphere? (discuss/prove)

c) Does the real space satisfy some sort of isoperimetric inequality? (discuss/prove)

d) Does the property of part (a) has anything to do with a Johnson-Lindenstrauss property for the points on the sphere? why?

3. Let $Cost_G(A_1, A_2, ...A_k)$ be a cost function where minimizers are best subpartitions of a graph G in some sense. Define the new cost function

$$Cost_{G,\lambda}(A_1, A_2, ...A_k) - \lambda w(A^*)$$

in which $A^* = V(G) - \cup_i A_i$ is the residual of the subpartition, λ is a real constant and $w(A^*)$ is the weight (i.e. measure) of the residual.

a) Discuss the nature of the optimizers of $Cost_{G,\lambda}$.

b) Choose one of the well-known cost functions and explain the new optimization problem.

c) Can you express the mean isoperimetry cost function in this form? what is λ in this case?

d) Discuss the computational complexity of the new optimization problem? Is it easier or harder than the original problem? why?

4. Someone is asked to classify a given dataset in the plane by a line in two parts. Intuitively, it is clear that the farther the two clusters of data are the clearer is the bisection. Provide a concrete proof of this intuitive fact using VC-dimention and the fundamental theorem of statistical learning.

5. Explain how matrix-tree theorem can be interpreted to give an evidence for the fact that the number of spanning trees of a graph G, denoted by $\tau(G)$, is a measure of connectedness.

a) Can you estimate $\tau(G)$ by other connectivity parameters?

b) How hard is to find $\tau(G)$ computationally?

c) Does it make sense to propose a sparsification method in which one tries to extract a couple of *carefully chosen* spanning trees of a graph and define the sparsifier as their union? (discuss)

d) Try to prove facts about correctness or inapplicability of the above idea. (Better results/counterexamples and deeper discussions get better marks.)

6. Is it correct to say that the theory of finite Markov chains is equivalent to the theory of finite electrical networks? why? (prove or disprove)

7. Develop a kernel-based PCA algorithm that can also be applied to data with arbitrary large dimension. Explain in detail how you use RKHS-theory to design your algorithm and provide a concrete numerical example to show the usefulness of your algorithm.

8. Consider the Petersen graph Z of Problem 1 with edge distances equal to the inverse of similarities. Compute the distortion of embedding this Petersen graph with the corresponding shortest path metric into the Euclidean space $\mathbb{R}^n$ (i.e. with the ℓ^2 norm).
   Note: Approximations also achieve partial credits!