AMS Seminar review 1

Saturday September 14, 2019, 2:00 p.m.-5:00 p.m.

Time table

1. 2:00 p.m. Average-Case Reductions Between Statistics Problems: From GOE to Wishart and Other Precise Distributional Maps. Matthew Brennan, Massachusetts Institute of Technology Guy Bresler*, Massachusetts Institute of Technology (1150-62-182)
2. 2:30 p.m. Scalable and Model-free Methods for Multiclass Probability Estimation. Helen Zhang*, University of Arizona (1150-62-157)
3. 3:00 p.m. Unbiased estimators for random design regression. Michal Derezinski*, UC Berkeley Manfred K. Warmuth, UC Santa Cruz and Google Inc. Daniel Hsu, Columbia University (1150-62-222)
4. 3:30 p.m. Fractal Dimension Estimation with Persistent Homology. Benjamin Schweinhart*, The Ohio State University (1150-60-616)
5. 4:00 p.m. Statistical learning with evolutionary-related correlated random variables. Vu C Dinh*, University of Delaware (1150-00-491)
6. 4:30 p.m. Reducing AI Bias using Truncated Statistics. Constantinos Daskalakis*, EECS and CSAIL, MIT (1150-62-651)

I would like to summarize 2 talks I felt interesting the most

1. 2:30 p.m. Scalable and Model-free Methods for Multiclass Probability Estimation. Helen Zhang*, University of Arizona

   (a) **Summary** There are classical ways to classify data: 1. Model based classifier 2. Using Machine learning. For the first way, it has limitation that we have to assume specific model and need parameters which makes this method not being robust and computationally expensive. For the second method, there is no formal way to find out convergence properties and it's too heuristic.
   The presenter suggests two main way to overcome above classical methods.
   (1).Generalizing weighted SVM for binary problem using truncated loss: But it also has it's limitation that problem set becomes non convex and can not be handled if classes are large
   (2). Divide and Conquer technique : It achieves convex problem and consistency.

   (b) **Qualitative Assessment**
   The presenter gave natural way to solve a given problem. She generalized and extended SVM to apply multiclasses case and did truncation to overcome Fisher consistency part. Also for divide and Conquer technique, she simply divided into many tests and just used existing theorem. I think there needs discussion about multiple comparison like problems especially, she only compared two classes at a time and repeated many times. It seems like results could be different if classes are like $P_1 < P_2$, $P_2 < P_3$ but $P_3 < P_1$

2. 4:30 p.m. Reducing AI Bias using Truncated Statistics. Constantinos Daskalakis*, EECS and CSAIL, MIT (1150-62-651)

   (a) **Summary** Prediction bias happens when distribution of training data set is different from that of test data set. The presenter gave really good examples how wrong prediction is made: (1) NBA data (2) Supervised learning in gender recognition.
   (1) is truncated by y axis and this data is only data we can get(we can't get data about ordinary people) (2) is truncated by x axis. Training data is biased (there are a lot data which have bright color) and this kind of truncation can not be detected until the results came to be different from real observation) How can he recover θ from truncated statistics? He used stochastic gradient descent method and got unbiased sample for loglikelihood function.

   (b) **Qualitative Assessment** He avoids the difficulty that stems from unknown distribution D over covariate x, unknown link function and unknown noise distribution by taking derivative and finding gradient based on stochastic gradient descent method. I think it's clever way to delete unknown parameter and recover unbiased sample. My main questions are how much good θ is through this method and how much bad it is when data is not truncated compared to ordinary estimator of θ

AMS Seminar review 2

Sunday September 15, 2019, 8:30 a.m.-10:50 a.m.

Time table

1. 8:30 a.m. Clustering small datasets in high-dimension using random projection. Mireille Boutin*, Purdue University (1150-60-120)
2. 9:00 a.m. Learning nearest neighbor graphs from noisy distance samples. Blake J Mason*, University of Wisconsin - Madison Ardhendu Tripathy, University of Wisconsin - Madison Robert Nowak, University of Wisconsin - Madison (1150-62-477)
3. 9:30 a.m. Multi-Level Graph Spanners. Reyan Ahmed, University of Arizona Faryad Darabi Sahneh, University of Arizona Keaton Hamm*, University of Arizona Stephen Kobourov, University of Arizona Mohammad Latifi Jebelli, University of Arizona Richard Spence, University of Arizona (1150-05-172)
4. 10:00 a.m. Is Manifold Learning for toy data only? Marina Meila*, University of Washington Samson Koelle, University of Washington Dominique Perrault-Joncas, Google James McQueen, Amazon.com Yu-chia Chen, University of Washington Hanyu Zhang, University of Washington Jacob VanderPlas, Google Zhongyue Zhang, Seattle, WA (1150-53-663)

I would like to summarize 2 talks I felt interesting the most

1. 8:30 a.m. Clustering small datasets in high-dimension using random projection. Mireille Boutin*, Purdue University (1150-60-120)

   (a) **Summary** The presenter used projection to cluster data into several groups. The most interesting thing was she uses random projectors to make separation and go down to lower dimension. She randomly projects data for several times and find the best one to cluster. Though she uses very simple approach for clustering data, the results are modest and sometimes even better than other complex methods.

   (b) **Qualitative Assessment** The main weak points of this research are she can't explain why it works well and have no gaurantee that it will work well later. Even though she justified why it worth to do using simulations, I think she needs to justify why this method would work mathematically. Considering all benefits it has(like easy to implement, simple algorithm and modest performance), however, it really worth to do further research.

2. 10:00 a.m. Is Manifold Learning for toy data only? Marina Meila*, University of Washington Samson Koelle, University of Washington Dominique Perrault-Joncas, Google James McQueen, Amazon.com Yu-chia Chen, University of Washington Hanyu Zhang, University of Washington Jacob VanderPlas, Google Zhongyue Zhang, Seattle, WA (1150-53-663)

   (a) **Summary** To be honest, I couldn't understand her research fully because it has many unfamiliar concepts of topology. Based on what I understood, she uses diffeomorphism which preserves both topological structure and geometry to make it easier to analyize data. Finding out what transformation I will use preserving both topological and geometrical structure is one of the hardest problems but she adds Riemannian metric g to data, to preserve geometry and successfully found $\phi(g)$ using mathematical theorems

   (b) **Qualitative Assessment** The impressive technique she uses was adding Riemannian metric and finding relationship between diffeomorphism and target function by taking derivatives on both side. I could learn one good process of how to find good transformation which loses data less but preserve many mathematical structures.

SILO Seminar (Christina Yu)

Wednesday September 18, 2019, 12:30

Iterative Collaborative Filtering for Sparse Noisy Tensor Estimation

Abstract: We present a generalization of the collaborative filtering algorithm for the task of tensor estimation, i.e. estimating a low-rank 3-order n-by-n-by-n tensor from noisy observations of randomly chosen entries in the sparse regime. Not only does the algorithm have desirable computational properties, it also provably achieves sample complexity that (nearly) matches the conjectured lower bound on the sample complexity. Furthermore, our analysis results in high probability bounds on the infinity norm of the error, as opposed to the weaker MSE bounds achieved by previous approaches. Our proposed algorithm uses the matrix obtained from the "flattened" tensor to compute similarity with respect to a corresponding observation graph. The algorithm recovers the tensor with max entrywise error decaying to 0 with high probability as long as the entries are sampled uniformly at a density of..$\Omega(n^{-3/2+\epsilon})$ for any arbitrarily small $\epsilon > 0$. This sample complexity threshold (nearly) matches a conjectured lower bound as well as the "connectivity threshold" of the corresponding observation graph used in our algorithm, providing a different angle to explain the conjectured lower bound.

1. **Summary** One goal of tensor analysis is to estimate underline structure based on sparsely observed data, it's not only fill missing entries but also to estimate noisily observed entries. In an e-commerce platform, datapoint collected from an interaction may be associated to a user, product, and date or time.
   Her problem setting is Matrix estimation with latent varianle model as follows. $M_{uvw} = f(\alpha_u, \beta_v, \gamma_w) + \epsilon_{uvw}$
   She uses Nearest Neighbors similarity based collaborative filtering(ex.Movie rating estimation) to estimate above matrix. Bottleneck is to how to measure similarities: key idea was an expand neighborhood in data graph and check if they overlap. Overallalgorithm steps is step1. Cosntruct bipartite graph from tensor
   stpe2. Construc breadth first search trees
   step3. compare depth l bfs trees
   step4. Average similar datapoints
   tensor PCA, learning latent similarities for online/sequence decision making.

2. **Qualitative Assessment** I felt two things from this talk were impressive. First of all, she makes matricization to transform a tensor into a matrix and analyze in matrix setting and get it back into a tensor again. This makes me want to find out a method that uses matricization, calculates in matrix setting and transforms back to a tensor. The other is she uses a kind of network and stretch out this network to define similarities which are hard to do. I think these two things are good example how can they find a way to solve hard problem making different setting.