samsi
NSF·Duke·NCSU·UNC

Deep Learning Program
Transition Workshop
March 12-13, 2020

SPEAKER TITLES/ABSTRACTS

David Banks
SAMSI

"Teaching Deep Learning to Statisticians"

Most courses on deep learning are taught by computer scientists to computer scientists. Statisticians look at things a little differently. This talk reviews some lessons learned in a first attempt by a statistician to teach a course on deep learning.

Anindya Bhadra
Purdue University

"Deep Neural Network Emulators beyond Gaussian Process Models"

Following Radford Neal's work in the 1990s, Gaussian process (GP) models have been prevalent in building emulators for deep neural networks. Suppose a neural network has one hidden layer, the hidden-to-output weights are independent with bounded variance and the hidden-to-output transfer function is linear. Then an application of the classical central limit theorem ensures convergence to a multivariate normal distribution, so that the limiting process is a GP as the width of the hidden layer approaches infinity. Neal's classical result has also been extended to multilayer feedforward neural networks and recurrent neural networks in recent years to build GP emulators. But what if the hidden-to-output weights are such that the classical CLT does not apply and a convergence to a GP does not occur? A simple example would be any weights in the \alpha stable family with \alpha <2, so that the variances are unbounded. This talk will touch upon some preliminary ideas on building DNN emulators based on more general L\'evy processes. They are attractive because they can model some behaviors that can never arise under GPs, e.g., jumps in the sample path.

Wyatt Bridgman and **Sorin Mitran**
University of North Carolina at Chapel Hill

"Deep Neural Networks as a Coarse-Graining Procedure for Stochastic Microdynamics"

Numerous coarse-graining approaches reduce the dimension of dynamical systems via linear projection onto an orthogonal subspace. These methods can fail to capture how the nonlinearity at the microscale manifests itself in the coarse-grained dynamics. Here we present a novel coarse-graining approach for stochastic dynamical systems that instead uses nonlinear transformations for dimension reduction. The approach is based on constructing stochastic processes to represent the time evolution of weights from deep neural networks (DNN) trained to advance the system forward

in time. These stochastic processes are built from simulation data and used to define a notion of nonlinear modes which can be seen as generalizations of the linear modes common to model reduction methods like proper orthogonal decomposition. As a test bed for this model reduction approach, we consider stochastic molecular dynamics (MD) simulations of a simple protein undergoing alpha helix nucleation. Analysis of the weight sequences indicates non-Gaussian behavior requiring special treatment to achieve dimension reduction to dominant modes. An optimization problem, which can be viewed as a nonlinear generalization of the SVD, is introduced to achieve this reduction.

Guang Cheng
Purdue University

"Sharp Rate of Convergence for Deep Neural Network Classifiers under the Teacher-Student Setting"

Classifiers built with neural networks handle large-scale high dimensional data, such as facial images from computer vision, extremely well while traditional statistical methods often fail miserably. In this paper, we attempt to understand this empirical success in high dimensional classification by deriving the convergence rates of excess risk. In particular, a teacher-student framework is proposed that assumes the Bayes classifier to be expressed as ReLU neural networks. In this setup, we obtain a dimension-independent and sharp rate of convergence, i.e., O(n^{−2/3}), for classifiers trained based on either 0-1 loss or hinge loss. This rate can be further improved to O(n^{-1}) when data is separable. Here, n represents the sample size. Such a dimension independent result provides a theoretical explanation for the empirical successes of deep neural network classifiers, in particular for high dimensional data.

Jhuma Das
University of North Carolina

Adrian Green
North Carolina State University

"Leveraging High-Throughput Screening Data and Conditional Generative Adversarial Networks to Advance Predictive Toxicology"

There are currently 85,000 chemicals registered with the EPA under the Toxic Substances Control Act, but only a small fraction (~5%) have any measured toxicological data. To address this data gap, the National Toxicology Program and National Academy of Sciences published visions calling for more efficient approaches in predicting how substances impact human health. Advancements in high-throughput screening (HTS) methods are central to realizing this vision. As part of one such HTS effort, nearly 200,000 embryonic zebrafish were used to examine 22 morphological and mortality endpoints from 1003 unique chemicals at six concentrations from the ToxCast chemical set. We hypothesized that by using a Generative Adversarial Network (GAN) and leveraging this large set of toxicity data, plus chemical structure information, we could efficiently predict toxic outcomes of unknown/untested chemicals. The ToxCast chemical dataset was divided into training (n = 802) and validation (n = 201) sets. Using the CAS numbers for each chemical, we generated textual files containing three-dimensional structural information for each chemical in the Protein Data Bank (PDB) file format. Utilizing a novel method, we converted the 3D structural information into a weighted set of points while retaining all information about the structure. Missing morphology values in the toxicity data, due to mortality, were imputed. The toxicity data were then transformed to improve learning, consolidated into endpoint x concentration matrices (22 x 6), and normalized to the number of fish exposed to each chemical. Next, considering that each matrix is associated with a chemical, we needed our network to generate an accurate matrix, given a

particular chemical structure. Therefore, we switched from a GAN to a conditional GAN (cGAN). Currently, we are in the process of training a generator, using both cGAN architecture and regression, and exploring alternative data transformations to reduce noise, increase meaning, and improve learning. The overall goal is to build a model that accurately predicts observed toxicity of known chemicals and provides insight into as-yet untested areas of chemical space.

Collaborators: Martin Mohlenkamp, Jhuma Das, Meenal Chaudhari, Lisa Truong, Robyn L. Tanguay, and David Reif

Bianca Dumitrascu
SAMSI

"Optimal Nonlinear Marker Selection for Cell Type Discrimination in Single Cell Analyses"

Single-cell technologies characterize complex cell populations across multiple data modalities at un-precedented scale and resolution. Multi-omic data for single cell gene expression, *in situ* hybridization, or single cell chromatin states are increasingly available across diverse tissue types. When isolating specific cell types from a sample of disassociated cells or performing *in situ* sequencing in collections of heterogeneous cells, one challenging task is to select a small set of informative markers to identify and differentiate specific cell types or cell states as precisely as possible. Given single cell RNA-seq data and a set of cellular labels to discriminate, scGene-Fit selects gene transcript markers that jointly optimize cell label recovery using label-aware compressive classification methods, resulting in a substantially more robust and less redundant set of markers than existing methods. In this talk we discuss work in progress regarding nonlinear dimensionality reduction and connections with Learn-to-explain models that generalize the scGene-Fit framework to a nonlinear marker selection problem.

David Dunson & **Edric Tam**
Duke University

"Graph-structured Inference using Neural Nets"

We will first provide a brief summary of the Bayesian deep learning group activities. We then focus the majority of the talk on new approaches for (a) inferring low-dimensional embeddings of brain structural connectivity networks using graph auto encoders; and (b) regularizing deep neural nets incorporating and respecting the graph structure - in particular, using the Fiedler value of the neural network's underlying graph as a tool for regularization. For (b): we provide theoretical support for this approach via spectral graph theory. We demonstrate the convexity of this penalty and provide an approximate, variational approach for fast computation in practical training of neural networks. We provide bounds on such approximations. We provide an alternative but equivalent formulation of this framework in the form of a structurally weighted L1 penalty, thus linking our approach to sparsity induction. We performed experiments on datasets that compare Fiedler regularization with traditional regularization methods such as dropout and weight decay. Results demonstrate the efficacy of Fiedler regularization.

Haiyang Huang
Duke University

"Dimension Reduction and Manifold Learning: a brief survey"

Dimensionality reduction (DR) techniques that utilize pairwise constraints such as t-SNE, LargeVis, and UMAP have demonstrated impressive visualization performance on high dimensional real world data. In the SAMSI interpretable deep learning reading group in Fall 2019, we aimed to better understand the important elements in these algorithms. In this talk we will discuss what we have found to be the most significant factors leading to high visualization quality.

Pulong Ma
SAMSI and Duke University

"Kriging: Beyond Matérn"

The Matérn covariance function is a popular choice for prediction in spatial statistics and uncertainty quantification literature. A key benefit of the Matérn class is that it is possible to get precise control over the degree of differentiability of the process realizations. However, the Matérn class possesses exponentially decaying tails, and thus may not be suitable for modeling long range dependence. This problem can be remedied using polynomial covariances; however one loses control over the degree of differentiability of the process realizations, in that the realizations using polynomial covariances are either infinitely differentiable or not differentiable at all. We construct a new family of covariance functions using a scale mixture representation of the Matérn class where one obtains the benefits of both Matérn and polynomial covariances. The resultant covariance contains two parameters: one controls the degree of differentiability near the origin and the other controls the tail heaviness, independently of each other. Using a spectral representation, we derive theoretical properties of this new covariance including equivalence measures and asymptotic behavior of the maximum likelihood estimators under infill asymptotics. The improved theoretical properties in predictive performance of this new covariance class are verified via extensive simulations. Application using NASA's Orbiting Carbon Observatory-2 satellite data confirms the advantage of this new covariance class over the Matérn class, especially in extrapolative settings.

Matthew Phillips
LifeOmic

"Landmark Priors for Biomedical Image Segmentation"

Objects of interest in biomedical images typically bear consistent spatial relationships to other objects or to global position. Current approaches make use of this spatial context either through multi-resolution pyramid approaches or by the creation of attentional maps. Here we present a novel, alternative way of using contextual information to inform biomedical image segmentation, based on CoordConv [Liu et al. 2018]. We segment an image in two fundamental stages: In the first, easily-identifiable 'landmarks' are identified, such as the optic disk in the retina fundus. In the second, gradients representing the locations of these objects are added as separate input channels to a second segmentation model. Thus the model, even when dealing with only a small piece of the total image (or volume) at a time, is able to incorporate spatial context without sacrificing resolution. We report the effect this has on segmentation of retinal vessels as well as on other segmentation tasks.

Cynthia Rudin
Duke University

"Two Projects on Interpretable Deep Learning: case-based reasoning and concept whitening"

I will discuss two projects on interpretable deep neural networks. In both cases, the networks have interpretability constraints imposed, but do not lose accuracy over their black-box counterparts.

Project 1: Deep Learning for Case-Based Reasoning (joint work with Chaofan Chen, Oscar Li, Alina Barnett, Daniel Tao, and Jonathan Su)

When we are faced with challenging image classification tasks, we often explain our reasoning by dissecting the image, and pointing out prototypical aspects of one class or another. The mounting evidence for each of the classes helps us make our final decision. In this work, we introduce a deep network architecture – prototypical part network (ProtoPNet), that reasons in a similar way: the network dissects the image by finding prototypical parts, and combines evidence from the prototypes to make a final classification. The model thus reasons in a way that is qualitatively similar to the way ornithologists, physicians, and others would explain to people on how to solve challenging image classification tasks.

This Looks Like That: Deep Learning for Interpretable Image Recognition. NeurIPS (spotlight), 2019 Chaofan Chen, Oscar Li, Alina Barnett, Jonathan Su, Cynthia Rudin https://arxiv.org/abs/1806.10574

Project 2: Concept Whitening for Interpretable Deep Learning (joint work with Zhi Chen and Yijie Bei)

We introduce a mechanism, called concept whitening (CW), to alter a given layer of the network to allow us to better understand the computation leading up to that layer. When a concept whitening module is added to a CNN, the axes of the latent space can be aligned with concepts of interest. By experiment, we show that CW can provide us a much clearer understanding for how the network gradually learns concepts over layers without hurting predictive performance.

Concept Whitening for Interpretable Image Recognition, in progress, 2020 Zhi Chen, Yijie Bei, Cynthia Rudin https://arxiv.org/abs/2002.01650

Deborshee Sen
Duke University

"Bayesian Dimension Reduction using Neural Networks"

In conducting non-linear dimensionality reduction and feature learning, it is common to suppose that the data lie near a lower-dimensional manifold. One class of model-based approaches for such problems includes latent variables in an unknown non-linear regression function; this includes Gaussian process latent variable models (GP-LVMs) and variational auto-encoders (VAEs) as special cases. VAEs use neural networks and additionally employs approximations to make the computation tractable; however, current implementations lack adequate uncertainty quantification in estimating the unknown density and the lower-dimensional subspace, and can be unstable and lack reproducibility in practice. We attempt to solve this problem by designing Markov chain Monte Carlo (MCMC) sampling algorithms for fully Bayesian inferences in neural network models with latent variables. We address issues of identifiability by imposing constraints on the neural network parameters as well as by using anchor points.

Shan Shan
Duke University

"Dimension Reduction with Fiber Bundles"

Dimension reduction is one of the fundamental techniques in unsupervised data analysis. In this talk, we will present how fiber bundles can be used in dimension reduction and probabilistic modeling. We will begin with viewing the data space as a manifold and review Diffusion maps (DM). We present a new parameter tuning test for DM that utilizes the semigroup properties of the diffusion kernel. We will then move on to think data as lying on or close to a fiber bundle. We review Horizontal Diffusion maps (HDM) and introduce a probabilistic model for learning the generative process of data. At the end of the talk, we will discuss our working plan of using deep neural networks to help with the computation of dimension reduction techniques on fiber bundles.

Quoc Tran-Dinh
University of North Carolina

"Shuffling and Sampled-Based Schemes for Non-convex Optimization: Convergence Guarantees"

In the first part of this talk, we propose a unified convergence analysis for a class of shuffling gradient algorithms for minimizing a finite-sum non-convex problem. In the second part, we propose a stochastic Gauss-Newton method for solving non-convex compositional optimization problems. We establish convergence for two difference choices of estimators and its oracle complexity.

Linjun Zhang
Rutgers University

"Exploring Model Sensitivity via Adversarial Influence Functions"

While robust optimization has been widely used in nowadays data science, such as adversarial training in machine learning, nothing much has been said about how it affects the optimization over parameters and the predictions on natural data theoretically. In this paper, we introduce the adversarial influence function (AIF) as a tool to provide insights, which enjoys a closed-form and can be calculated efficiently. To illustrate the power of AIF, we apply it to study model sensitivity ---- a quantity defined to capture the change of prediction loss of natural data after implementing robust optimization. Specifically, we use AIF to analyze how model complexity and randomized smoothing affect model sensitivity with respect to a broad class of functions. We further derive AIF for kernel regressions and experimentally demonstrate the effectiveness of characterizing via AIF on how robust optimization affects the optimizer of kernel regressions with neural tan- gent kernels. Lastly, we extend our theory for AIF from robust optimization to distributional robust optimization.