The Alan Turing Institute

Mathematics and data

Abstracts

Alexander Gorban, University of Leicester

The simplicity revolutions in the era of complexity and the unreasonable effectiveness of small neural ensembles in high-dimensional brain

We review and analyse biological, physical, and mathematical problems at the core of the fundamental question: how can high-dimensional brain organise reliable and fast learning in high-dimensional world of data by simple tools?

Two critical applications are reviewed: one-shot correction of errors in artificial intellectual systems and emergence of static and associative memories in ensembles of single neurons. Error correctors should be simple; not damage the existing skills of the system; allow fast non-iterative learning and correction of new mistakes without destroying the previous fixes. All these demands can be satisfied by new tools based on the concentration of measure phenomena and stochastic separation theory.

In several words, the stochastic separation theorems state that for an essentially high-dimensional distributions a random point can be separated from a random set by Fisher's linear discriminant with high probability. The number of points in this set can grow exponentially with dimension. Different versions of stochastic separation theorems use different definitions of `random sets' and `essentially high-dimensional distributions' but the essence of these definitions is simple: sets with very small (vanishing) volume should not have high probability even for large dimension.

The talk is based on the work: A.N. Gorban, V.A. Makarov, I.Y. Tyukin, The unreasonable effectiveness of small neural ensembles in high-dimensional brain. Physics of Life Reviews, 2019, https://doi.org/10.1016/j.plrev.2018.09.005

Chaim Even-Zohar, University of California

Local view of combinatorial structures and applications to data analysis

Consider a large combinatorial object such as a permutation or a graph, and look at all its restrictions to constant size substructures. As simple examples, consider the occurrences of triangles in a graph, or unsorted pairs in a permutation. Local properties of this kind often show up in extremal questions in Combinatorics, in the study of random models, pseudo-randomness, constructions of limiting objects, and more. Often, such large graphs or permutations are used to model various data sets, and then the local picture provides efficient ways to analyze the data.

This talk will discuss some basic questions on constant size permutation patterns – these are the k! possible ordering types obtained by restricting a large permutation to k entries. How are the various pattern occurrences distributed in a random permutation? What do they tell us about the global properties of our data? How efficiently can we count the occurrences of such patterns? We will survey these questions, and present some recent progress on them. One tool that we use is the representation theory of the symmetric group.

The Alan Turing Institute

Felipe Rincon, QMUL

An introduction to tropical geometry and connections to deep learning

This talk will give a basic introduction to tropical geometry and show how it serves as a very good language for studying some aspects of deep neural networks. In particular, I will try to give an idea about the result by Zhang, Naitzat, and Lim that the family of feedforward neural networks with ReLU activation is equivalent to the family of tropical rational maps

Ginestra Bianconi, QMUL and The Alan Turing Institute

Classical information theory of networks

Heterogeneity is among the most important features characterizing real-world networks. Empirical evidence in support of this fact is unquestionable. Existing theoretical frameworks justify heterogeneity in networks as a convenient way to enhance desirable systemic features, such as robustness, synchronizability and navigability. However, a unifying information theory able to explain the natural emergence of heterogeneity in complex networks does not yet exist. Here, we fill this gap of knowledge by developing a classical information theoretical framework for networks. We show that among all degree distributions that can be used to generate random networks, the one emerging from the principle of maximum entropy is a power law. We also study spatially embedded networks finding that the interactions between nodes naturally lead to nonuniform distributions of points in the space. The pertinent features of real-world air transportation networks are well described by the proposed framework.

John Oprea, Cleveland State University

Topological complexity and its applications

The motion planning problem asks for an algorithm that carries a state of a system to any other state. This is the same as asking for an algorithmic way to choose a path between any two points of the configuration space associated to the system. Topological complexity is a measure of how hard it is to solve this problem. In some sense then, topological complexity is another instance of Poincare’s thesis that topology constrains solutions to applied problems in Science and Mathematics. This talk will describe topological complexity and some of its properties together with some results about it. Finally, a (surprising?) link between topological complexity and social choice theory will be described.

Nati Linial, Hebrew University of Jerusalem

Expander graphs

Consider a large network of communication. These may be computers passing messages among themselves or neurons in our brain that communicate with each other. You may also think of some large social network or a web of companies that do business together. Let us consider what may happen when some of the lines connecting these entities cease to exist. Perhaps a storm destroys some physical lines in the communication networks, or some illness affects the synapses through which neurons communicate.

You can easily imagine what failure looks like in the other examples. In many such cases it is desirable for the network to “have no bottlenecks”. Namely, while we may not mind it if a tiny part of the network gets disconnected from the rest of it, following such failure, it is highly undesirable that the whole network breaks into two or more parts of substantial size when only few interconnecting lines fail. In mathematical terms a network having this property is called an expander graph. The study of expander graphs is an exciting area of research that spans from some of the deepest parts of pure mathematics to very concrete down-to-earth applications. In this lecture I will try to give you a taste of these fascinating subjects.

Shay Moran, University of Princeton

On the expressiveness of comparison queries

Comparisons are a classical and well studied algorithmic tool that is used in a variety of contexts and applications. We will discuss two manifestations of the expressiveness of these queries in machine learning and algorithms (a more detailed overview is given below). Both manifestations are based on the notion of "inference dimension” that can be viewed as another instance of the fruitful link between machine learning and discrete mathematics - a link dating back to the discovery of the VC dimension.

1. Active classification with comparison queries [Kane, Lovett, Moran, Zhang, FOCS ’17] Active learning is a model for semi-supervised learning that captures situations in which unlabeled data is abundant and manually labelling it is expensive. We consider an extension of active learning in which the learning algorithm may ask the annotator to compare the distances of two examples from the boundary of their label-class. For example, in a recommendation system application (say for restaurants), the annotator may be asked whether she liked or disliked a specific restaurant (a label query); or which one of two restaurants did she like more (a comparison query). We prove that the usage of comparison queries leads to an exponential improvement in the query complexity of some well studied problems. Specifically, for the class of half-spaces, we show that under natural assumptions, such as large margin or bounded bit-description of the input examples, it is possible to reveal all the labels of a sample of size n using approximately O(logn) queries.
2. Nearly optimal linear decision trees for k-SUM and related problems [Kane, Lovett, Moran, JACM ‘19] We use the above framework to construct linear decision trees for a variety of decision problems in combinatorics and discrete geometry. For example, for any k, we construct linear decision trees that solve the k-SUM problem on n elements using O(nk log n) linear queries. Moreover, the queries we use are comparison queries, which compare the sums of two k-subsets; when viewed as linear queries, comparison queries are 2k-sparse and have only {−1,+1} coefficients. We give similar constructions for sorting sumsets A+B and for solving the SUBSET-SUM problem, both with optimal number of queries, up to poly-logarithmic terms.