האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

ELSC The Edmond & Lily Safra Center for Brain Sciences

GATSBY

Deep Learning and the Brain

January 20-22, 2019

Becker Auditorium, Goodman Brain Sciences Building

Edmond J. Safra Campus, The Hebrew University, Jerusalem

Abstracts

Broadening and deepening the role of Artificial Intelligence in Computational Neuroscience

Daniel Yamins, Stanford University

Advances combining artificial intelligence techniques with computational neuroscience have shown that time-averaged neural responses in the primate visual and auditory systems can be modeled with reasonable accuracy by task-optimized deep neural networks. I'll discuss our lab's recent work to broaden and deepen these results, using recurrent convolutional networks to capture neural dynamics in the visual system. I'll also talk about attempts to plug the biggest hole in the task-optimized theory --- moving beyond unrealistic labelled supervision by creating self-supervised interactive agents that create powerful sensory representations --- and discuss the connection between these ideas and development. Moving beyond sensory systems, I'll describe models bridging to decision-making and memory, in the context of modular continual learning. Finally, I'll discuss how these results fit into the historical trajectory of AI and computational neuroscience, and discuss future questions of great interest that may benefit from a similar approach

Old new frontiers in visual object recognition using deep learning: curriculum learning

Daphna Weinshall, The Hebrew University

The incredible recent progress in this visual object recognition, and the availability of very effective public domain tools for this task, allows us to reopen old questions and approach them from new directions with new tools. In this talk I will focus on curriculum learning, where a learner is exposed to examples whose difficulty level is gradually increased. This heuristic dominates human learning; empirically, it has also been shown to improve the outcome of learning in various machine learning models. Our main contribution is a theoretical result, showing that learning with a curriculum speeds up the rate of learning in the context of the regression and the hinge loss. Interestingly, we also show how curriculum learning and hard-sample mining, although conflicting at first sight, can coexist harmoniously within the same theoretical model. I will show empirical results using deep CNN models for image classification, where curriculum learning is shown to speed up the rate of learning and improve the final generalization performance. In the context of human cognition, I will show how this empirical approach can be used to investigate related questions in perceptual learning.

The functional neuroanatomy of face perception: from brain measurements to deep neural networks

Kalanit Grill-Spector, Stanford University

A central goal in neuroscience is to understand how processing within the ventral visual stream enables rapid and robust perception and recognition. Recent neuroscientific discoveries have significantly advanced understanding of the function, structure and computations along the ventral visual stream that serve as the infrastructure supporting this behaviour. In parallel, significant advances in computational models, such as hierarchical deep neural networks (DNNs), have brought machine performance to a level that is commensurate with human performance. Here, we propose a new framework using the ventral face network as a model system to illustrate how increasing the neural accuracy of present DNNs may allow researchers to test the computational benefits of the functional architecture of the human brain. Thus, we will (i) consider the specific neural implementational features of the ventral face network, (ii) describe similarities and differences between the functional architecture of the brain and DNNs, and (iii) provide hypotheses for the computational value of implementational features within the brain, which may ultimately improve DNN performance. Importantly, this new framework promotes the incorporation of neuroscientific findings into DNNs in order to test the computational benefits of fundamental organizational features of the visual system.

Work done together with Kevin S. Weiner, Jesse Gomez, Anthony Stigliani, and Vaidehi S. Natu

Perceptual learning in a mouse model: a progress report

Adi Mizrahi, The Hebrew University

Abstract TBA

Decoupling gating from linearity

Shai Shalev-Shwartz, The Hebrew University

The gap between the empirical success of deep learning and the lack of strong theoretical guarantees calls for studying simpler models. By observing that a ReLU neuron is a product of a linear function with a gate (the latter determines whether the neuron is active or not), where both
share a jointly trained weight vector, we propose to decouple the two. We introduce GaLU networks --- networks in which each neuron is a product of a Linear Unit, defined by a weight vector which is being trained, with a Gate, defined by a different weight vector which is not being trained. Generally speaking, given a base model and a simpler version of it, the two parameters that determine the quality of the simpler version are whether its practical performance is close enough to the base model and whether it is easier to analyze it theoretically. We show that GaLU networks perform similarly to ReLU networks on standard datasets and we initiate a study of their theoretical properties, demonstrating that they are

indeed easier to analyze. We believe that further research of GaLU networks may be fruitful for the development of a theory of deep learning.

Joint work with Jonathan Fiat and Eran Malach.

High-dimensional dynamics of generalization error in neural networks: implications for experience replay
Andrew Saxe, University of Oxford

Learning even a simple task can engage huge numbers of neurons across the cortical hierarchy. How do neuronal networks manage to generalize from a small number of examples, despite having large numbers of tunable synapses? And how does depth—the serial propagation of signals through layered structure—impact a learning system? I will describe results emerging from the analysis of deep linear neural networks. Deep linear networks are a simple model class that retain many features of the full nonlinear setting, including a nonconvex error surface and nonlinear learning trajectories. In this talk I will focus on their generalization error, using random matrix theory to analyze the cognitively-relevant "high-dimensional" regime, where the number of training examples is on the order of or even less than the number of adjustable synapses. Consistent with the striking performance of very large deep network models in practice, I show that good generalization is possible in overcomplete networks due to implicit regularization in the dynamics of gradient descent. Overtraining is worst at intermediate network sizes, when the effective number of free parameters equals the number of samples, and can be reduced by making a network smaller or larger. I identify two novel phenomena underlying this behavior in linear networks: first, there is a frozen subspace of the weights in which no learning occurs under gradient descent; and second, the statistical properties of the high-dimensional regime yield better-conditioned input correlations which protect against overtraining. Turning to the impact of depth, the theory reveals a trade-off between training speed and generalization performance in deep neural networks, and I confirm this speed-accuracy trade-off through simulations. Finally, I will describe an application of these results to experience replay during sleep. The consolidation of learning during sleep is thought to arise from the replay of stored experiences between hippocampus and neocortex. Why is this complex strategy beneficial? As a simple model of this process, we compare the dynamics arising from online learning, in which each example is used once and discarded; and batch learning, in which all examples are stored (for instance, in hippocampus) and replayed repeatedly (for instance, during sleep). While these two strategies yield similar performance when training experience is abundant, we find that replay can be decisively better when training experience is scarce. Our results suggest a normative explanation for a two-stage memory system: replay can enable better generalization from limited training experience.

The computational benefit of the hidden layers in Deep Neural Networks
Naftali Tishby, The Hebrew University

The Information Bottleneck Theory of Deep Learning has three interesting predictions:
1. The layers of Deep Neural Networks (or brains), if achieve good generalization on a specific task by ANY training algorithm, should lie close to the Information Bottleneck limit for this task with effectively compressed representations of the input patterns.
2. Most of the improvement in generalization is achieved through diffusion in the irrelevant directions in the weight space, which amounts to reduction of the signal to noise ration of the irrelevant features of the data and lead to the hierarchically compressed representations the input by the layers.
3. This compression by diffusion leads to dramatic boost of the optimization (training) time with the number of layers: more layers can achieve good generalization much faster.

In this talk I will present new rigorous arguments and experimental results that confirm these predictions. In particular, I will prove that the computational benefit of the hidden layers scales as a power law in the number of layers, with exponent that depends on the Stochastic Gradient Decent (SGD) diffusion exponent and on the efficiency of the different representations of the layers. This power law scaling in the number of layers can become an exponential boost in cases of ultra slow diffusion of the SGD, as reported by others.

Three puzzles in the theory of deep learning
Tomaso Poggio, MIT

In recent years, machine learning researchers has achieved impressive results. Though theory has been lagging behind, some of the main questions about deep learning are now being solved. I will review the state of three main puzzles which include 3 separate branches of mathematics, that is approximation, optimization and machine learning theory:
• Approximation Theory: When and why are deep networks, with many layers of neurons, better than shallow networks which correspond to older machine learning techniques? When can they avoid the curse of dimensionality?
• Optimization: Why is it easy to train a deep network and often achieve global minima of the empirical loss?
• Learning Theory: How can deep learning avoid overfit and predict well for new data despite overparametrization? Do deep networks generalize according to classical theory?
I will also discuss the future of AI. To create artifacts that are as intelligent as we are, we need several additional breakthroughs. A good bet is that several of them will come from interdisciplinary research between the natural science and the engineering of Intelligence. This vision is in fact at the core of the CBMM and of the new MIT Quest for Intelligence, of which I will outline organization and research strategy.

Empirical and neural network modeling approaches to understanding human memory and consolidation

Anna Schapiro, Harvard Medical School

There is a fundamental tension between storing discrete traces of individual experiences, which allows recall of particular moments in our past without interference, and extracting regularities across these experiences, which supports generalization and prediction in similar situations in the future. This tension is resolved in classic memory systems theories by separating these processes anatomically: the hippocampus rapidly encodes individual episodes, while the cortex slowly extracts regularities over days, months, and years. This framework fails, however, to account for the full range of human learning and memory behavior, including: (1) how we often learn regularities quite quickly—within a few minutes or hours, and (2) how these memories transform over time and as a result of sleep. I will present evidence that the hippocampus, in addition to its well-established role in episodic memory, is in fact also responsible for our ability to rapidly extract regularities. I will show a neural network model of the hippocampus that demonstrates how these two competing learning processes can coexist in one brain structure. Finally, I will present empirical and simulation work showing how these initial hippocampal memories are replayed during offline periods to help stabilize and integrate them into cortical networks.

Representation learning in rats and men

Yael Niv, Princeton University

On the face of it, most real-world world tasks are hopelessly complex from the point of view of reinforcement learning mechanisms. In particular, due to the ”curse of dimensionality”, even the simple task of crossing the street should, in principle, take thousands of trials to learn to master. But we are better than that.. How does our brain do it? In this talk, I will argue that the hardest part of learning is not assigning values or learning policies, but rather deciding on the boundaries of similarity between experiences, which define the ”states” that we learn about. I will show behavioral evidence that humans and animals are constantly engaged in this representation learning process, and suggest that in a not too far future, we may be able to read out these representations from the brain, and therefore find out how the brain has mastered this complex problem. I will formalize the problem of learning a state representation in terms of Bayesian inference with infinite capacity models, and suggest that an understanding of the computational problem of representation learning can lead to insights into the machine learning problem of transfer learning, and psychological/neuroscientific questions about the interplay between memory and learning.

The Star Cells of Learning: Astrocytes modulate local neuronal activity to affect global behavior.

Inbal Goshen, The Hebrew University

Neurons in the hippocampus perform complicated computations, but not all of them participate in those tasks all the time. How are the active populations selected? And can the selection process be modulated by other cells?
Astrocyte can sense both pre synaptic and post synaptic activity, and modulate synaptic communication with precision. However, whereas the supportive roles of astrocytes, such as glucose metabolism maintenance, glutamate levels monitoring and neurotrophic factors secretion, are well recognized, their direct effects on neuronal activity remains elusive.
We chose to target astrocytic activity as a way to modulate synaptic plasticity and cognitive performance. In parallel, we image the activity and structure of neurons and astrocytes (separately and simultaneously) in the hippocampus of behaving mice, to study the interaction between these populations. To directly and specifically modulate astrocytic activity we employed excitatory and inhibitory chemogenetic tools.
We discovered that astrocytic activation is not only necessary for synaptic plasticity, but also sufficient to induce NMDA-dependent de-novo long term potentiation in the hippocampus, which persisted after astrocytic activation ceased. In-vivo, astrocytic activation enhanced memory allocation, i.e. it increased neuronal activity in a task-specific way, only when coupled with learning but not in home-caged mice. Furthermore, astrocytic activation using either chemogenetic or optogenetic tools during acquisition resulted in memory recall enhancement on the following day. Conversely, directly increasing neuronal activity resulted in dramatic memory impairment.
Astrocytic inhibition during memory acquisition impairs remote, but not recent, recall. We show that this effect is mediated by a specific disrupting the projection from the hippocampus to the anterior cingulate cortex by astrocytes.
Finally, we imaged neuronal activity and astrocytic morphology in mice navigating a virtual reality, and tested the significance of the affiliation to a certain astrocytic domain to the activity of the neurons within this discrete domain.

Learning and generalization in visual question answering

Aaron Courville, University of Montreal

Numerous models for grounded language understanding and visual reasoning have been recently proposed, including (i) generic modules that can be used easily adapted to any given task with little adaptation and (ii) intuitively appealing modular models that require background knowledge to be instantiated. In this talk, I will briefly review some representative models and compare them in how much they lend themselves to a particular form of systematic generalization. Using a synthetic VQA test, we show that the generalization of modular models is much more systematic and that it is highly sensitive to the module layout, i.e. to how exactly the modules are connected. We furthermore investigate if modular models that generalize well

could be made more end-to-end by learning their layout and parametrization. Our results suggest that, in addition to modularity, systematic generalization in language understanding may require explicit regularizers or priors.

Vector-based navigation using grid-like representations in artificial agents
Andrea Banino, DeepMind

Deep neural networks have achieved impressive successes in fields ranging from object recognition to complex games such as Go. Navigation, however, remains a substantial challenge for artificial agents, with deep neural networks trained by reinforcement learning failing to rival the proficiency of mammalian spatial behaviour, which is underpinned by grid cells in the entorhinal cortex. Grid cells are thought to provide a multi-scale periodic representation that functions as a metric for coding space and is critical for integrating self-motion (path integration) and planning direct trajectories to goals (vector-based navigation). Here we set out to leverage the computational functions of grid cells to develop a deep reinforcement learning agent with mammal-like navigational abilities. We first trained a recurrent network to perform path integration, leading to the emergence of representations resembling grid cells, as well as other entorhinal cell types. We then showed that this representation provided an effective basis for an agent to locate goals in challenging, unfamiliar, and changeable environments—optimizing the primary objective of navigation through deep reinforcement learning. The performance of agents endowed with grid-like representations surpassed that of an expert human and comparison agents, with the metric quantities necessary for vector-based navigation derived from grid-like units within the network. Furthermore, grid-like representations enabled agents to conduct shortcut behaviours reminiscent of those performed by mammals.
Our findings show that emergent grid-like representations furnish agents with a Euclidean spatial metric and associated vector operations, providing a foundation for proficient navigation. As such, our results support neuroscientific theories that see grid cells as critical for vector-based navigation, demonstrating that the latter can be combined with path-based strategies to support navigation in challenging environments.

Theoretical and empirical investigation of several common practices in Deep Learning
Daniel Soudry, Technion

We examine several empirical and theoretical results on the training of deep networks. For example,

- Why are common "over-fitting" indicators (e.g., very low training error, high validation loss) misleading?
- Why, sometimes, early-stopping time never arrives?
- Why can adaptive rate methods (e.g., adam) degrade generalization?
- Why commonly used loss functions exhibit better generalization than others?

- Can we train with large batch sizes, without hurting generalization?
- Why use weight decay before batch-norm?
- When can we use low numerical precision, and how low can we get?

and discuss the practical implications of these results to data parallelism and resource efficiency in deep networks.

Why do deep convolutional networks generalize so poorly to small image transformations?
Yair Weiss, The Hebrew University

Deep convolutional network architectures are often assumed to guarantee generalization for small image translations and deformations. In this paper we show that modern CNNs (VGG16, ResNet50, and InceptionResNetV2) can drastically change their output when an image is translated in the image plane by a few pixels, and that this failure of generalization also happens with other realistic small image transformations. Furthermore, the deeper the network the more we see these failures to generalize. We show that these failures are related to the fact that the architecture of modern CNNs ignores the classical sampling theorem so that generalization is not guaranteed. We also show that biases in the statistics of commonly used image datasets makes it unlikely that CNNs will learn to be invariant to these transformations. Taken together our results suggest that the performance of CNNs in object recognition falls far short of the generalization capabilities of humans.

Unsupervised learning via video prediction
Rob Fergus, New York University

One approach to unsupervised learning is the prediction of future elements in a sequence, given previous ones. We explore this paradigm in the context of video, showing how robust and stable representations can be learned. Two key issues are (i) what is the right representation space in which to perform the prediction task? and (ii) how to address and model the inherent uncertainty in video sequences? We introduce models that address both challenges and are able to generate realistic samples many frames into the future.

Joint work with Emily Denton.

A less artificial Intelligence
Andreas Tolias, Baylor College of Medicine

Despite major advances in artificial intelligence through deep learning methods, computer algorithms remain vastly inferior to mammalian brains, and lack a fundamental feature of animal intelligence: they generalize poorly outside the domain of the data they have been trained on. This results in brittleness (*e.g.* adversarial attacks) and poor performance in transfer

learning, few-shot learning, causal reasoning, and scene understanding, as well as difficultly with lifelong and unsupervised learning — all important hallmarks of human intelligence. We conjecture that this gap is caused by the fact that current deep learning architectures are severely under-constrained, lacking key model biases found in the brain that are instantiated by the multitude of cell types, pervasive feedback, innately structured connectivity, specific non-linearities, and local learning rules. There is ample behavioral evidence that the brain performs approximate Bayesian inference under a generative model of the world (also known as inverse graphics or analysis by synthesis), so the brain must have evolved a strong and useful model bias that allows it to efficiently learn such a generative model. Therefore, our goal is to learn the brain's model bias in order to engineer less artificial, and more intelligent, neural networks. Experimental neuroscience now has technologies that enable us to analyze how brain circuits work in great detail and with impressive breadth. Using tour-de-force experimental methods we have been collecting an unprecedented amount of neural responses (*e.g.* more than 1.5 million neuron-hours) from the visual cortex, and developed computational models that we use to extract principles of functional organization of the brain and learn the brain's model biases.

Less-artificial vision with artificial neural networks
Matthias Bethge, University of Tübingen

Deep neural networks have become an ubiquitous tool in a broad range of AI applications. Resembling important aspects of rapid feed-forward visual processing in the ventral stream they can be trained to match human behavior on standardized pattern recognition tasks. Outside the training distribution, however, decision making of artificial neural networks exhibits large discrepancies to biological vision systems. I will present recent results of my lab to quantify and overcome these discrepancies in the context of domain adaptation, few-shot learning, task transfer and adversarial robustness. More generally, I will discuss the importance of generative modeling and causal representations for the design of more data-efficient, interpretable and robust learning machines.

Connecting the structure and function of neural circuits
Srinivas Turaga, HHMI Janelia Research Campus

In this talk, I will describe how we developed deep learning based computational tools to solve two problems in neuroscience: inferring the activity of a neural network from measurements of its structural connectivity, and inferring the connectivity of a network of neurons from measurements and perturbation of neural activity.

1. Can we infer neural connectivity from noisy measurement and perturbation of neural activity? Population neural activity measurement by calcium imaging can be combined with cellular resolution optogenetic activity perturbations to enable the mapping of neural connectivity in vivo. This requires accurate inference of perturbed and unperturbed neural

activity from calcium imaging measurements, which are noisy and indirect. We built on recent advances in variational autoencoderes to develop a new fully Bayesian approach to jointly inferring spiking activity and neural connectivity from in vivo all-optical perturbation experiments. Our model produces excellent spike inferences at 20K times real-time, and predicts connectivity for mouse primary visual cortex which is consistent with known measurements.

2. Are measurements of the structural connectivity of a biological neural network sufficient to predict its function? We constructed a simplified model of the first two stages of the fruit fly visual system, the lamina and medulla. The result is a deep hexagonal lattice convolutional neural network which discovered well-known orientation and direction selectivity properties in T4 neurons and their inputs. Our work demonstrates how knowledge of precise neural connectivity, combined with knowledge of the function of the circuit, can enable in silico predictions of the functional properties of individual neurons in a circuit, leading to an understanding of circuit function from structure.

Neural networks and the brain: from the retina to semantic cognition, and beyond
Surya Ganguli, Stanford University

A synthesis of machine learning, neuroscience and psychology has the potential to elucidate how striking computations emerge from the interactions of neurons and synapses, with applications to biological and artificial neural networks alike. We discuss two vignettes along these lines. First we demonstrate that modern deep learning methods yield state-of-the-art models of the retina that predict the retinal response to natural scenes with high precision, recapitulate the functional properties of the retinal interior, and generalize to simultaneously account for decades of physiological studies with artificial stimuli. Second, we review work on how deep neural networks can describe a wide array of psychology experiments about the developmental time course of infant semantic cognition, including the hierarchical differentiation of concepts as infants get older as well as the notion of coherent versus incoherent categories. We describe a mathematical analysis that not only provides a natural explanation for the dynamics of human semantic development and category processing, but also leads to better algorithms for speeding up learning in artificial neural networks.

Assessing the scalability of biologically-motivated deep learning algorithms and architectures
Timothy Lillicrap, DeepMind

The backpropagation of error algorithm (BP) is impossible to implement in a real brain. The recent success of deep networks in machine learning and AI, however, has inspired proposals for understanding how the brain might learn across multiple layers, and hence how it might approximate BP. As of yet, none of these proposals have been rigorously evaluated on tasks where BP-guided deep learning has proved critical, or in architectures more structured than

simple fully-connected networks. Here we present results on scaling up biologically motivated models of deep learning on datasets which need deep networks with appropriate architectures to achieve good performance. We present results on the MNIST, CIFAR-10, and ImageNet datasets, explore variants of target-propagation (TP) and feedback alignment (FA) algorithms, and examine performance in both fully- and locally-connected architectures. Many of these algorithms perform well for MNIST, but for CIFAR and ImageNet we find that TP and FA variants perform significantly worse than BP, especially for networks composed of locally connected units, opening questions about whether new architectures and algorithms are required to scale these approaches.

Bounded learning - biological constraints in cortical learning
Yonatan Loewenstein, The Hebrew University

I will discuss several biological constraints that should be considered when discussing learning in cortical networks. First, I will present evidence for substantial volatility of excitatory synapses in the living cortex and discuss its implications regarding learning and memory. Second, I will explain why excitatory and inhibitory synapses differ in their ability to underlie memory storage – because the average firing rates of excitatory and inhibitory neurons substantially differ. Finally, I will discuss how the stability of the dendritic and axonal arborizations affect the capacity of the network to learn new memories.

Neural constraints on learning
Byron Yu, Carnegie Mellon University

Learning has been studied at multiple levels, including behavior, brain regions, individual neurons, and synapses. However, little is known about how populations of neurons change their activity in concert during learning. Are there network constraints on the types of new neural activity patterns that can be achieved? We studied this question using a brain-computer interface (BCI), which allows us to specify which population activity patterns lead to task success. I will address why learning some tasks is easier than others, as well as how populations of neurons change their activity in concert during learning.

The Brain as a hierarchical adaptive learner
Sophie Denève, École Normale Supérieure

We combined mathematical results from non-linear adaptive control theory and the principle of efficient coding in order to derive how the brain may optimally learn to perform complex tasks, while using the minimum number of spikes, and with only local synaptic plasticity rules. When applied to unsupervised learning of naturalistic stimuli, the model accounts for receptive fields properties and spiking statistics in sensory cortices. When applied to the learning of

dynamical systems (e.g. sensorimotor system), we demonstrate that recurrent spiking networks can learn to predict and control temporal variables evolving with arbitrary dynamics, using a minimal number of spikes. This is achieved by local, biophysically plausible "hebbian" learning rules based on pre-synaptic activity and feedback error signals (in contrast to FORCE of backprop in RNN, for example, who do not have local learning rules). When applied to supervised learning based on input/output exemples (e.g. a classification task) and multi-layer networks, this model provides an alternative to backpropagation, e.g. the learning rules are local including between hidden layers. Albeit such hierarchical spiking networks have been tested so far only in toy example, our preliminary results suggest that the network may at least as performant as backprop while requiring less training examples to achieve high generalization performance. We are thus gearing towards a general theory of learning in brain networks, where membrane potential represent prediction errors, spikes signal temporal updates in an internal representation, and error feedbacks must both drive downstream neurons and modulate synaptic plasticity. Interestingly, a tight balance between excitatory and inhibitory is at the heart of this framework. E/I balance must be maintained in each unit, and in fact achieving the tightest E/I balance becomes equivalent to learning the tasks.

Building a state space for song learning
Michale Fee, MIT

Research on the avian song system has shed light on how the brain produces precise sequences that control behavior, and how the brain implements reinforcement learning (RL) of a complex behavior. While RL is a powerful strategy for learning, it depends critically on having an appropriate representation of the state space of the task. In the songbird, RL is thought to operate on a representation of song timing, but this representation is not present in young birds. I will describe a model for how the songbird brain could construct timing sequences to support RL, and will offer a hypothesis for how the auditory system could shape these sequences to align with a memory of the tutor song, thus facilitating song evaluation.