Symmetries and Neural Tangent Kernels

Jan E. Gerken







in collaboration with



from





Pan Kessel

Symmetries in physics

 $SU(2) \times SU(3) \times U(1)$



Standard Model of Elementary Particles













Equivariance



T.S.COHEN@UVA.NL

M.WELLING@UVA.NL

Group Equivariant Convolutional Networks

Taco S. Cohen University of Amsterdam Max Welling University of Amsterdam

University of Amsterdam University of California Irvine Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeCorrolution layers can be used effectively in a deep network because all the layers in such a network net translative equivariaser: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature mars call leads un to edue-effects. In

T.S.COHEN@UVA.NL

M.WELLING@UVA.NL

Group Equivariant Convolutional Networks

Equivariant Transformer Networks

Kai Sheng Tai	Pater Builly	Greenery Valiant
Provide the second second		Citegory canner

Abstract

How can prior knowledge on the transformation invariances of a domain he incorperated into the architecture of a neural network? We propose Equivariant Transformers (ITs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived canorical coonlinets western IFs1 incorporates functions the scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2016; Marcos et al., 2017; Worrall et al., 2017; Horniques & Vedaldi, 2017; Cohen et al., 2018) has explored new CNN architectures that are survented to the moreoid modified to noteriodite transforma-

Taco S. Cohen University of Amsterdam Max Welling

University of Amsterdam University of California Irvine Canadian Institute for Advanced Research

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network ner translariow equivariant: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the machine feature more, of layers and then shifting

4

Group Equivariant Convolutional Networks

T.S.COHEN@UVA.NL University of Amsterdam M.WELLING@UVA.NL University of Amsterdam

University of California Irvine Canadian Institute for Advanced Research

Taco S. Cohen

Max Welling

Abstract

We introduce Group emissiont Completional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces somely complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are translation equivariant: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature many (at least up to obse-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards. pre-defined continuous transformation groups, Through the use of specially derived canonical cointrough the use of spectally-derived caronical co-onlinate systems. FTs incomorate functions that scaling to each training image). While data augmentation tynically helps reduce the test error of CNN-based models. there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data sugmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016: Dieleman et al., 2016: Marcos et al., 2017: Worrall et al. 2017: Henriques & Vadable 2017: Cohenet al. 2018) has explored new CNN architectures that are

Theory for Equivariant Quantum Neural Networks

Ouvnh T. Nemen 3.2 Louis Schatzki 3.4 Paolo Braeria 3.5 Michael Barone 3.6 Patrick J. Coles,¹ Frédéric Sauvage,¹ Martin Larocca,^{1,7} and M. Cerezo³ ¹Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87515, USA ²School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. Information Sciences, Les Alemas National Laboratory, Les Alemas, New Merico 87515, USA University of Blinois at Urbana-Chempaign, Urbana, Elinois 61801, USA ^bDipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy

⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action com-

Group Equivariant Convolutional Networks

Tara S Calan T.S. COREN @ UNA.NL University of Annolam M. WELLING @ UNA.NL University of Collifornia Freier M. WELLING @ UNA.NL University of Collifornia Freier M. WELLING @ UNA.NL

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeCorrelution layers can be used effectively in a deep network because all the layers in such a network are translation equivariant: shifting the image and then feeding in through a variable of layers is the same as feeding the original image through the same layers and then shifting the resulting feature mans (a layer and then shifting

Equivariant Transformer Networks

Kai Sheng Tai⁺ Peter Bailis⁺ Gregory Valiant⁺

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivarian Transformers (IFUs), a famity of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived carotical coonlinate washime. FC incorporate formition the data scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation imariance will be enforced for data not seen during training.

In contrast to training time approaches like data sugmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dieleman et al., 2017; Heuriques et al., 2017; Wernal et al., 2017; Heuriques & Vedalds, 2017; Cohen et al., 2018) has explored new CNN architectures that are established to nurveed unell-tothe to satisfast transformasarranted to nurveed unell-tothe to satisfast transforma-

Theory for Equivariant Quantum Neural Networks

Quynh T. Negrus,^{1,2} Louis Schattal,^{1,4} Pinalo Buscela,^{1,3} Michael Ragnen,^{1,4} Patrick, J. Chel,^{1,4} Piofiti Schurger,¹ Muth. Lacces,¹ and M. Crewpl³ ¹Thematical Division. Les Alarone Neural Liberatory. Les Alarone, New Maxime 8744, USA ²Theorem and Applet Science, Revered University, Contrologi, Estavathanelli 1923, USA ³Departments Neural Control (1998), 2008. Control (1998), 2008. ⁴Departments Neural (1998), 2009. ⁴Department (1998), 2018. ⁴Departments (1998), 2019. ⁴Department (1998), 2019. ⁴Department (1998), 2019. ⁴Department of Microsof Microsoft (1998), 2019. ⁴Department (1998), 2019. ⁴Department (1998), 2019. ⁴Department (1998), 2019. ⁴Department (1998), ²Department (1998), ⁴Department (1998), ⁴Departmen

^bDipartimoda i Puise Astronomia, Università di Primar, State Forentino (PJ), 60019, Raly Dipartimoda di Puise Astronomia, Università di Primar, State Forentino (PJ), 60019, Raly Department of Mathematica, Università qui Primar, State Storentino (PJ), 60019, Raly Center for Nouliceor Statian, Les Altamas Mathematica, Las Alaxana, New Mexico ST454, USA

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, necessi trackthroughs in machino learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized largely the usage of equivariant neural networks whose action comneurs with their of the symmetry. In this ways of the innerest them ishness to the maximum eachs har

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,c,1} Qi Meng^b Jue Zhang^b Huilin Qu^c Congqiao Li^d Sitian Qian^d Weitao Du^c Zhi-Ming Ma^a Tie-Yan Liu^b

*Andring of Mathematics and Systems Science, Charse Andrewy of Sciences, Damogeneous Death Rood, Beijng 100180, China *Mercung Research Asia, Daming Street, Breijing 100080, China *CERN, RP Dopartment, CH1111 Grossov, B. SosterAnda *School of Physics, Pring Starrenths Charnets Rood, Brains 100807, China

Group Equivariant Convolutional Networks

Taro S. Calan TA COREN & UNA SAL University of Annatian MARLENG UNA SAL University of Calorian University MARLENG UNA SAL

Abstract

We introduce Group equivariant Convolutional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that reduces sample complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are trauslation equivariant: shifting the image and then feeding in through a number of layers is the same as feeding the original image through the same layers and then shifting the provider feature same, the lowar on to observe (fresh).

Equivariant Transformer Networks

Kai Sheng Tai⁺ Peter Bailis⁺ Gregory Valiant⁺

Abstract

How can prior knowledge on the transformation imrainaces of a domain be incorporated into the architecture of a neural network? We propose Equivatian Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards pre-defined continuous transformation groups. Through the use of specially-derived carotical coonlines weatmen. FTs incorporates functions that scaling to each training image). While data augmentation typically helps reduce the test error of CNN-based models, there is no guarantee that transformation imariance will be enforced for data not seen during training.

In contrast to training time approaches like data augmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016; Dietensnet 41a, 2017; Woerall et al., 2017; Henriques & Vedaldi, 2017; Cohen et al., 2013) has explored new CNN architectures that are surrestated to recorded meet CNN architectures that are

Theory for Equivariant Quantum Neural Networks

Quyah T. Negrus,^{1,4} Louis Schatzlah,^{1,4} Mealo Buecch,^{1,5} Michael Ragona,^{1,4} Patrick J. Code,¹ Priotic Scarroger,¹ Matth. Corec.³ and M. Creent,³ ¹Theorem Code,¹ Territor Scarroger, Martin Lances, Cree Merice 3744, USA School of Department and Applied Sciences, However, Lion Almon, New Mexice 3744, USA ¹Supervision Sciences, Los Manues Mained Laboretory, Lio Almons, New Mexice 3744, USA ¹Supervision Sciences, Los Manues Mained Laboretory, Lio Almons, New Mexice 3744, USA ¹Supervision Sciences, Conf. Proc. Network, Sciences, Conf. 2018, 2019,

^b Dipartimento di Fusica e Artoneovis, Università di Pirma, Sette Formatine (PI), 50019, Italy Departmento di Fusica e Artoneovis, Università di Pirma, Sette Formatine (PI), 50019, Italy Department of Mathematica, Università di California Davis, Davis, California 95616, USA Center for Neulineovistatica, Leo Altennos Melanza Laboratory, Las Adavaco, New Mesico ST452, USA

Quantum neural network architectures that have little-to-no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent trackthroughs in machino learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the mags of equivariant neural networks whose action comneurs with their of the symmetry. In b this work more more three inducts to the surrounder modes of the symmetry.

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shiqi Gong^{a,c,1} Qi Meng^b Jue Zhang^b Huilin Qu^c Congqiao Li^d Sitian Qian^d Weitao Du^c Zhi-Ming Ma^a Tie-Yan Liu^b

*Androng of Mathematics and Systems Science, Chance Andremy of Sciences, Damogeneous Deat Road, Beijan (1019), China *Mercang Research Asia, Daming Street, Breijan (1008), China *CERN, RP Doputment, CH1111 Graves B., SosterAnda *School of Physics, Poling University, Chaneta Beaul, Brains (1008); China

E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials

Simon Batzner⁴,¹ Albert Musaelian,¹ Lixin Sun,¹ Mario Geiger,² Jonathan P. Mailon,³ Mordechai Kornbluth,³ Nicola Molinari,¹ Tess E. Smidt,^{4,5} and Boris Kozinsky^{4,1,3}

¹John A., Fundom Rohn of Digmorray and Applied Sciences, Branned University, Candrolph, M. 20155, USA Basel Polycologue Foldenk et Lansauxe, 2015 Lansauxe, Distributi Rohn Polycologue Foldenk et Lansauxe, 2015 Lansauxe, 2016 Control Computational Research Distance and Contexp For Manned Mathematics for Europy Research Applications, Lawrence Revisite National Laboratory, Berkels, CA 3720, USA Engineering and Computer Science, Carding Mathematics, Experiment and Computer Science, Carding Mathematics, 2018.

This work presents Neural Equivalent Interatoric Potentials (RequiP), an E(3)-equivalent energy large the large difference in the energy simulation for molecular dynamics inimitations. While most contemporary symmetry-source models use intraviate convolutions and only act on source and Neurope energy E(3)-equivalent action of generative transverse models are intraviated convolutions of interactions of generative transvers, resulting in a more information-rich and faithful representation of atomic environments. The method advises state-of-the-anti-order and faithful representation of a distribution of the models of the other state of the state-of-the-anti-order and environments.

Group Equivariant Convolutional Networks

Taco S. Cohen University of Amsterdam	T.S.COHEN@UVA.NL
Max Welling	M.WELLING@UVA.NL
University of Amsterdam	
University of California Irvine	
Canadian Institute for Advanced Research	

Abstract

We introduce Group emissiont Completional Neural Networks (G-CNNs), a natural generalization of convolutional neural networks that redaces sounds complexity by exploiting symmeConvolution layers can be used effectively in a deep network because all the layers in such a network are translation equivariant: shifting the image and then feeding it through a number of layers is the same as feeding the original image through the same layers and then shifting the resulting feature mans (at least up to obse-effects). In

Equivariant Transformer Networks

Kai Sheng Tai 1 Peter Bailis 1 Gregory Valiant

Abstract

How can prior knowledge on the transformation invariances of a domain be incorporated into the architecture of a neural network? We propose Equivariant Transformers (ETs), a family of differentiable image-to-image mappings that improve the robustness of models towards. pre-defined continuous transformation groups, Through the use of specially-derived caponical coonlinate systems. FTs incomprate functions that scaling to each training image). While data augmentation tynically helps reduce the test error of CNN-based models. there is no guarantee that transformation invariance will be enforced for data not seen during training.

In contrast to training time approaches like data sugmentation, recent work on group equivariant CNNs (Cohen & Welling, 2016: Dieleman et al., 2016: Marcos et al., 2017: Worrall et al. 2017: Henriques & Vadaldi. 2017: Cohenet al. 2018) has explored new CNN architectures that are

Theory for Equivariant Quantum Neural Networks

Ouvnh T. Nemen 3.2 Louis Schatzki 3.4 Paolo Braeria 3.5 Michael Barone 3.6 Patrick J. Coles,¹ Frédéric Sauvage,¹ Martín Larocea,^{1,7} and M. Cerezo³ ¹Theoretical Division, Los Alamos National Laboratoru, Los Alamos, New Mexico 87515, USA ²School of Engineering and Applied Sciences. Harvard University, Cambridge, Mussachusetts 02138, USA. Information Sciences, Les Alemas National Laboratory, Les Alemas, New Merico 87515, USA ⁴Department of Electrical and Computer Engineering University of Blinois at Urbana-Chempaign, Urbana, Elinois 61801, USA ^bDipartimento di Fisica e Astronomia, Università di Firenze, Sesto Fiorentino (FI), 50019, Italy

⁶Department of Mathematics, University of California Davis, Davis, California 95616, USA

Quantum neural network architectures that have little to no inductive biases are known to face trainability and generalization issues. Inspired by a similar problem, recent breakthroughs in machine learning address this challenge by creating models encoding the symmetries of the learning task. This is materialized through the usage of equivariant neural networks whose action com-

An Efficient Lorentz Equivariant Graph Neural Network for Jet Tagging

Shini Gong^{4,c,1} Qi Meng⁶ Jue Zhang⁶ Huilin Qu^c Congniao Li^d Sitian Qian^d Weitao Du^o Zhi-Ming Ma^o Tio-Yan Liu^b

* Academy of Mathematics and Sostems Science. Chinese Academy of Sciences. Thomassensore Fost Road Reiling 100190 China ^bMicrosoft Research Asia. Danling Street, Briting 100089, China CERN, EP Department, CH-1211 Geneva 23. Switzerland distant of Dission Dahima University Chengla Road. Beijing 100871. China

E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials

Simon Batzner^{* 1} Albert Musaelian ¹ Livin Sun ¹ Mario Geiger ² Jonathan P. Mallon ³ Mordechai Kornbluth.³ Nicola Molinari.¹ Tess E. Smidt.^{4,5} and Boris Kozinsky^{*1,3}

John A. Postson School of Engineering and Ambiol Sciences. Harvard University, Cambridae, MA 02138, USA ³Robert Basch Research and Technology Center, Cambridge, MA 02139, USA 4 Computational Research Division and Center for Advanced Mathematics for Energy Research Applications. Engineering and Commuter Science, Cambridge, MA 02119, USA

This work networks Neural Equivariant Interatomic Potentials (NegaIP), an E(3)-equivariant dynamics simulations. While most contemporary symmetry-aware models use invariant convolutions and only act on scalars. NanuIP condens E(3) conjunctant consolutions for interactions of reconctrice tensors, resulting in a more information-rich and faithful representation of atomic environments. The method achieves state-of-the-art accuracy on a challenging and diverse set of molecules and

HIERARCHICAL, ROTATION-EQUIVARIANT NEURAL NETWORKS TO SELECT STRUCTURAL MODELS OF PROTEIN COMPLEXES

Stephan Eismann*	Raphael J.L. Townshend*	Nathaniel Thomas"
Department of Applied Physics	Department of Computer Science	Department of Physics
Stanfeed University	Stanford University	Stanford University
seismann@stanford.edu	raphael@cs.stanford.edu	nthomas1030gmail.com
Millind Jagota	Bowen Jing	Ron O. Dror
lepertment of Electrical Engineering	Department of Computer Science	Department of Computer Science
Stanford University	Stanford University	Stanford University
mjagota@atanford.edu	bjing@atanford.edu	rondror@cs.stanford.edu

ABSTRACT

Predicting the structure of multi-protein complexes is a grand challenge in biochemistry, with make implications for basic science and drug discovery. Computational structure prediction methods presently investigation of the structural features to distinguish accurate structural markels from less accurate ones. This raises the operation of whether it is rowaible to learn characteristics of accurate models directly from atomic coordinates of protein complexes, with no prior sequentions. Here we introduce a machine learning method that learns directly from the 3D positions of all atoms to

CH-1211 Geneva 23. Switzerland

distant of Dission Dahima University

Chengla Road. Beijing 100871. China

Group Equivariant Convolutional Netw Equivariant Quantum Neural Networks ² Louis Schatzki 3.4 Paolo Braeria 1.5 Michael Ramme 1.6 Geometric Deep Learning and Equivariant Neural Networks Frédéric Sauvage,¹ Martín Larocca,^{1,7} and M. Cerezo³ Taco S. Cohen Alamos National Laboratoru, Los Alamos, New Mexico 87515, USA University of Amsterdam Autom Parallelia Educatory, Edi Antonio, New Interior 97340, USA died Sciences, Harvard University, Cambridge, Massachusetts 02138, USA Alemas National Laboratory, Las Alemas, New Merico 87515, USA Max Welling rtment of Electrical and Computer Engineering University of Amsterdam linois at Urbana-Chempaien, Urbana, Illinois 61801, USA University of California Irvine stronomia, Università di Firenze, Sesto Fiorentino (FI), 50619 , Italy Canadian Institute for Advanced Research dica University of California Davis, Davis, Colifornio 95616, USA JAN E. GERKEN¹, JIMMY ARONSSON^{1*}, OSCAR CARLSSON^{1*}, HAMPUS LINANDER², Los Alamas National Laboratory, Los Alamos, New Mexico 87545, USA architectures that have little-to-no inductive biases are known to face Fredrik Ohlsson³, Christoffer Petersson^{1,4} and Daniel Persson¹ Abstract on issues. Inspired by a similar problem, recent breakthroughs in ma-Convolution layers car challenge by creating models encoding the symmetries of the learning work because all the I through the usage of equivariant neural networks whose action com-We introduce Group amiustiant Completional lation conjugations, sh rough the usage or experiment neuron networks whose accord com-Neural Networks (G-CNNs), a natural generalit through a number of ization of convolutional neural networks that reoriginal image through ¹ Chalmers University of Technology, Department of Mathematical Sciences duces somple complexity by exploiting symmethe resulting feature ma SE-412.96 Gothenburg. Sweden ² Gothenburg University, Department of Physics SE-41296. Gothenburg. Sweden TATION-EOUIVARIANT NEURAL NETWORKS An Efficient Lorentz Equivariant Graph ³ Umeå University, Department of Mathematics and Mathematical Statistics TURAL MODELS OF PROTEIN COMPLEXES Network for Jet Tagging SE-901.87. Umeå. Sweden Rephael J.L. Townshend Nathanial Thomas" Department of Physics 4 Zenseact Shini Gong*.e.1 Oi Meng^b Jue Zhang^b Huilin Qu^c Congniao Li^d ranhanifers stanford edu nthomas 1038email.com Due Zhi-Ming Mat Tie-Yan Link SE-41756. Gothenburg. Sweden Basson Day *Academy of Mathematics and Sustems Science. Chinese Academy of St Department of Computer Science Department of Computer Science Thomassumerous Foot Board Beijing 100190, China Stanford University Microsoft Research Asia. * equal contribution hiingPatanford.edu rondror@cs.stanford.edu Danling Street, Briting 100089, China CERN. EP Department. ABSTRACT

sulti-motois complexes is a grand challenge in biochemistry, with major tence and drug discovery. Computational structure prediction methods penerally leverate needefined structural features to distinguish accurate structural models from less accurate ones. This raises the operation of whether it is rowaible to learn characteristics of accurate models directly from stornic coordinates of protein complexes, with no prior securetions. Here we introduce a machine learning method that learns directly from the 3D positions of all atoms to

Ron O. Drog





🖒 Easy to implement

- 🖒 Easy to implement
- 凸 No specialized architecture necessary
- ர No exact equivariance

🖒 Easy to implement

凸 No specialized architecture necessary

ர No exact equivariance

Can we understand data augmentation theoretically?

Neural tangent kernel

Empirical NTK

Training dynamics under continuous gradient descent:



Empirical NTK

Training dynamics under continuous gradient descent:

$$\frac{d\mathcal{N}_{\theta}(x)}{dt} = -\frac{\eta}{N} \sum_{i=1}^{N} \Theta_{\theta}(x, x_i) \frac{\partial L}{\partial \mathcal{N}(x_i)}$$
training sample

with the empirical neural tangent kernel (NTK)

$$\Theta_{\theta}(x,x') = \sum_{\mu} \frac{\partial \mathcal{N}(x)}{\partial \theta_{\mu}} \frac{\partial \mathcal{N}(x')}{\partial \theta_{\mu}}$$

[Jacot et al. 2018]



[Jacot et al. 2018]



[Jacot et al. 2018]



🖒 NTK becomes independent of initialization

[Jacot et al. 2018]



凸 NTK becomes independent of initialization

Ճ NTK becomes constant in training



凸 NTK becomes independent of initialization

凸 NTK becomes constant in training

凸 NTK can be computed for most networks



- 凸 NTK becomes independent of initialization
- 凸 NTK becomes constant in training
- 凸 NTK can be computed for most networks
- ✓ Training dynamics can be solved

[Jacot et al. 2018]

$$\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X)t}) Y$$

[Jacot et al. 2018]

$$\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X)t}) Y$$

[Jacot et al. 2018]

$$\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X)t}) Y$$
train data

[Jacot et al. 2018]



[Jacot et al. 2018]



$$\mu_t(x) = \Theta(x, X) \Theta(X, X)^{-1} (\mathbb{I} - e^{-\eta \Theta(X, X)t}) Y$$









$$\mu_t(\rho(g)x) = \Theta(x,X)\Theta(X,X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})\underbrace{\rho(g)Y}_{=Y}$$
for invariance

$$\mu_t(\rho(g)x) = \Theta(x,X)\Theta(X,X)^{-1}(\mathbb{I} - e^{-\eta\Theta(X,X)t})\underbrace{\rho(g)Y}_{=Y}$$

$$= \mu_t(x)$$
for invariance

 $\mu_t(x)$

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}}[\mathcal{N}_{\theta_t}(x)]$$

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}} [\mathcal{N}_{\theta_t}(x)] = \lim_{n \to \infty} \frac{1}{n} \sum_{\theta_0 = \text{init}_1}^{\text{init}_n} \mathcal{N}_{\theta_t}(x)$$

$$\mu_t(x) = \mathbb{E}_{\theta_0 \sim \text{initializations}} [\mathcal{N}_{\theta_t}(x)] = \lim_{n \to \infty} \underbrace{\frac{1}{n} \sum_{\theta_0 = \text{init}_1}^{\text{init}_n} \mathcal{N}_{\theta_t}(x)}_{\text{mean prediction of deep ensemble}}$$

- ✓ Proof of exact equivariance for
 - full data augmentation
 - infinite ensembles
 - at infinite width

- ✓ Proof of exact equivariance for
 - full data augmentation
 - infinite ensembles
 - at infinite width
- ✓ Equivariance holds for all training times

- Proof of exact equivariance for
 - full data augmentation
 - infinite ensembles
 - at infinite width
- ✓ Equivariance holds for all training times
- ✓ Equivariance holds away from the training data

Experiments









Relative Standard Deviation







Key takeaways

Key takeaways

- பsing physics approaches in deep learning can be very fruitful

Key takeaways

- பsing physics approaches in deep learning can be very fruitful
- பி Neural tangent kernels provide a powerful theoretical handle

Papers

• Emergent Equivariance in Deep Ensembles Jan E. Gerken*, Pan Kessel* ICML 2024 (Oral)

* Equal contribution

Papers

- Emergent Equivariance in Deep Ensembles Jan E. Gerken*, Pan Kessel* ICML 2024 (Oral)
 - * Equal contribution
- Equivariant Neural Tangent Kernels Philipp Misof, Pan Kessel, Jan E. Gerken Preprint

Papers

- Emergent Equivariance in Deep Ensembles Jan E. Gerken*, Pan Kessel* ICML 2024 (Oral)
 - * Equal contribution
- Equivariant Neural Tangent Kernels Philipp Misof, Pan Kessel, Jan E. Gerken Preprint



Group Website