UAI 2024 Tutorial

Geometric Probabilistic Models

Viacheslav (Slava) Borovitskiy
@vabor112
ETH zürich

Alexander Terenin
@avt_im
Cornell University

UAI 2024 Tutorial

Geometric Probabilistic Models

Part I

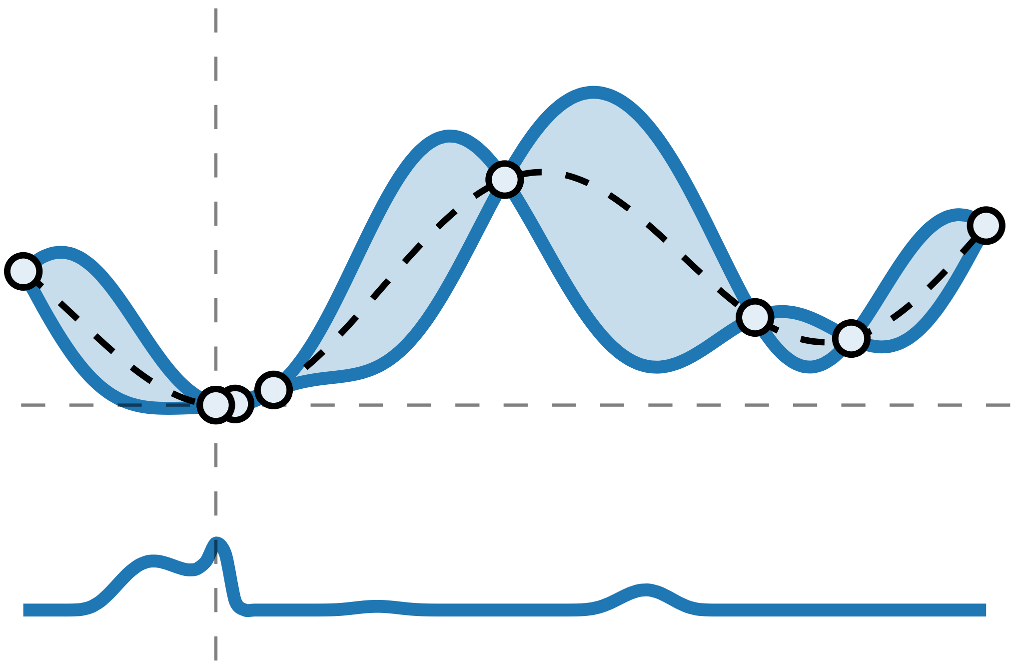
Alexander Terenin

Part I: background, motivation, technical fundamentals

Part II: geometric probabilistic models

Model-based decision-making

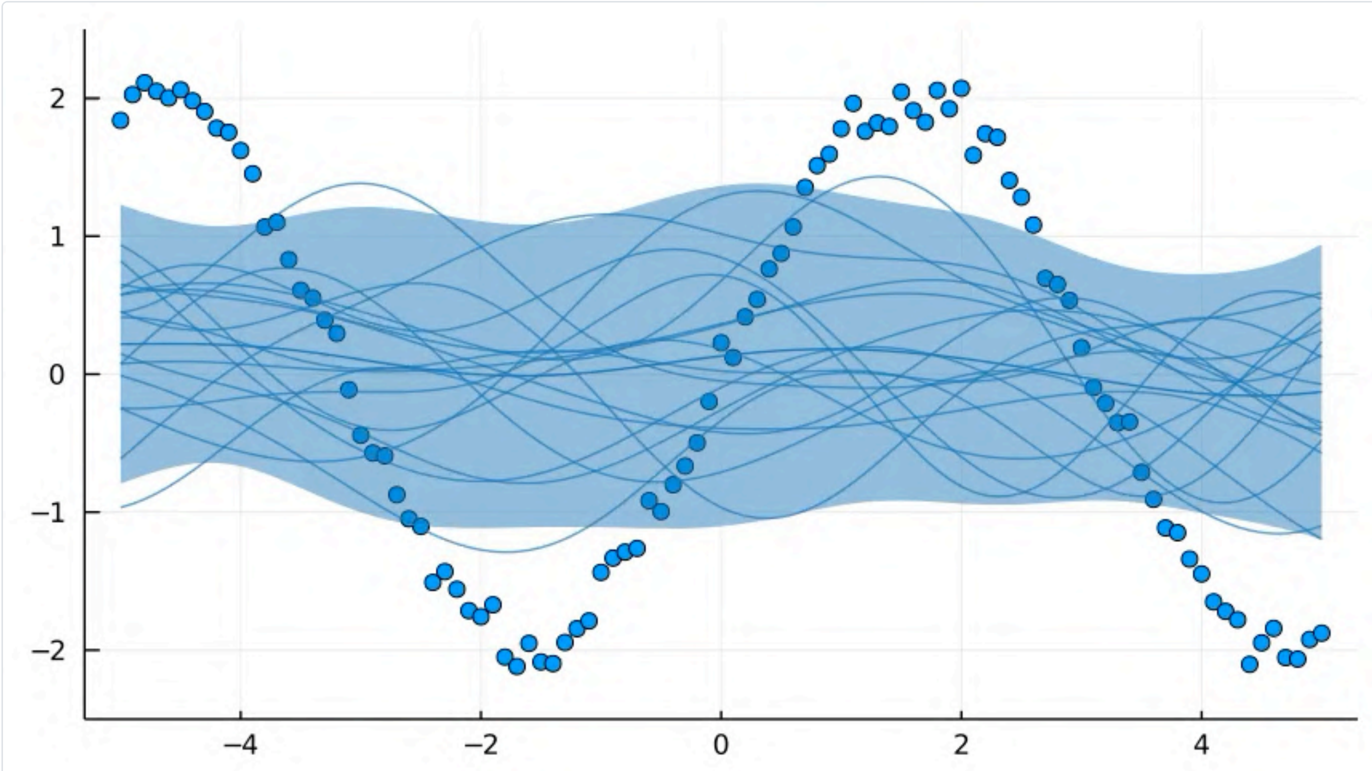
Model-based decision-making: Bayesian optimization



Principle of Separation: prediction and decision-making

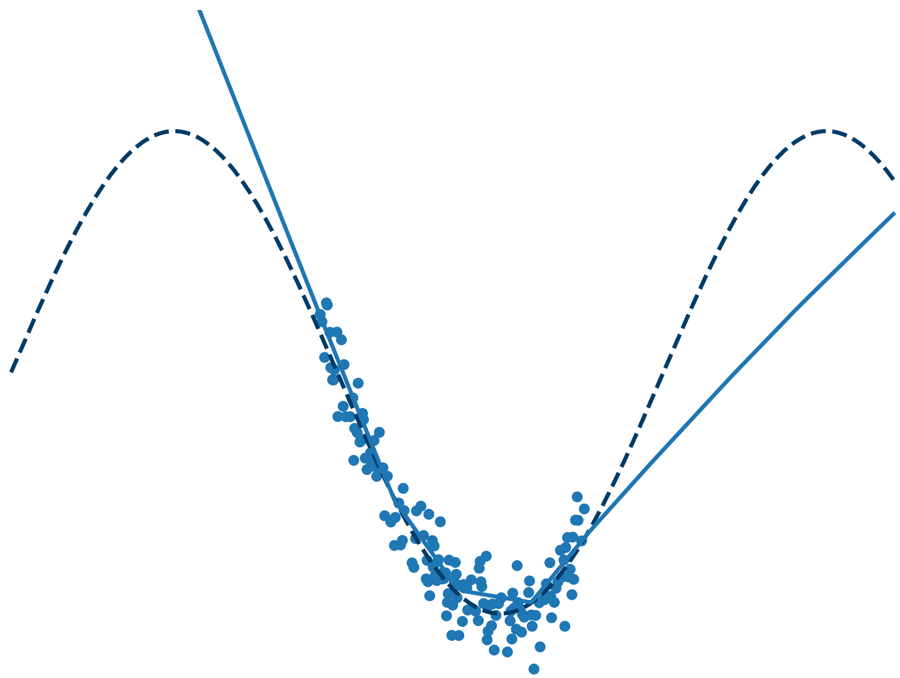
Probabilistic models

Probabilistic models



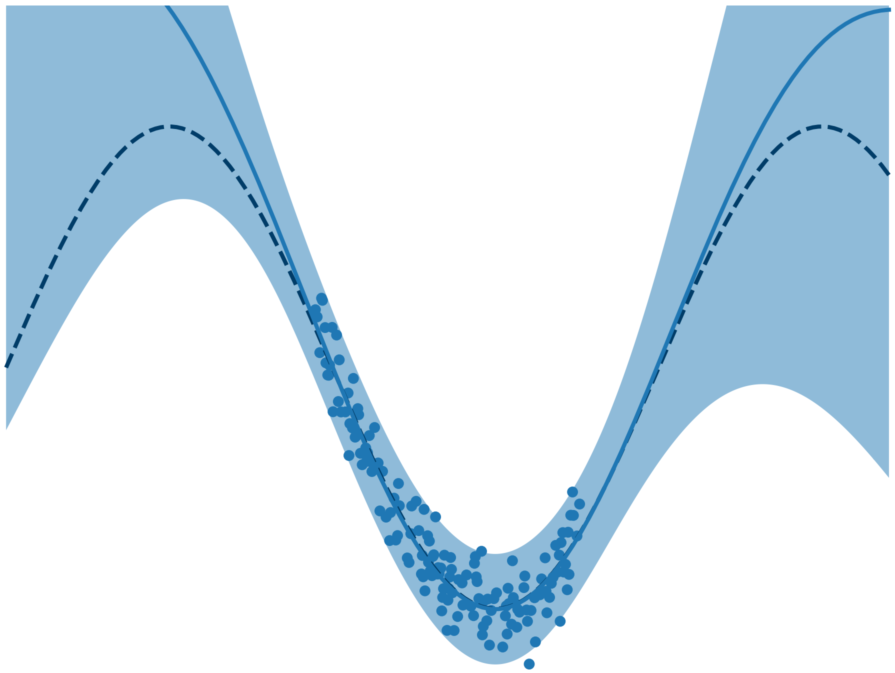
Predictions and uncertainty

Extrapolation and uncertainty



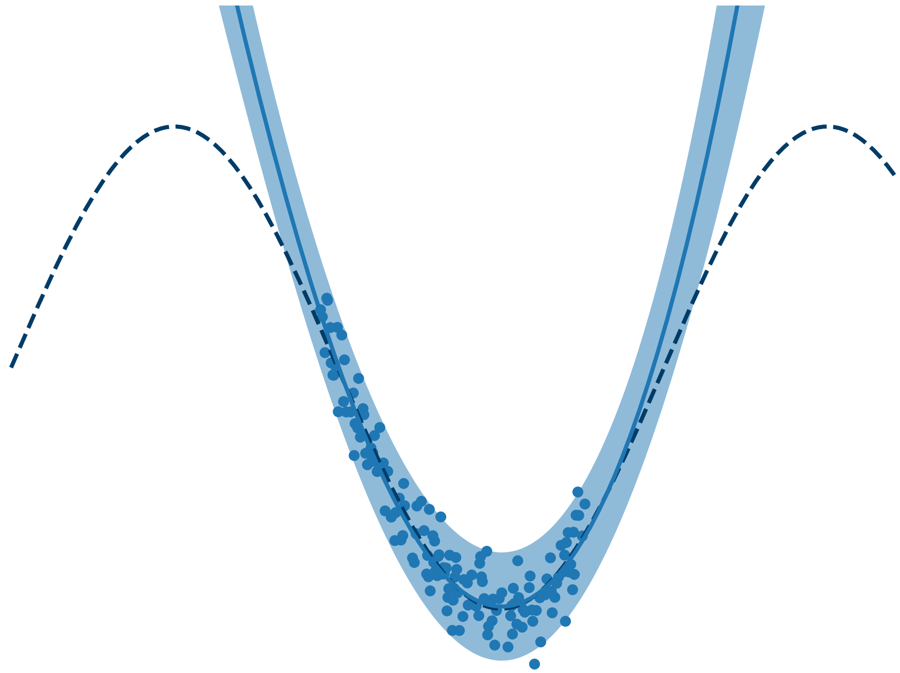
Neural network baseline

Extrapolation and uncertainty



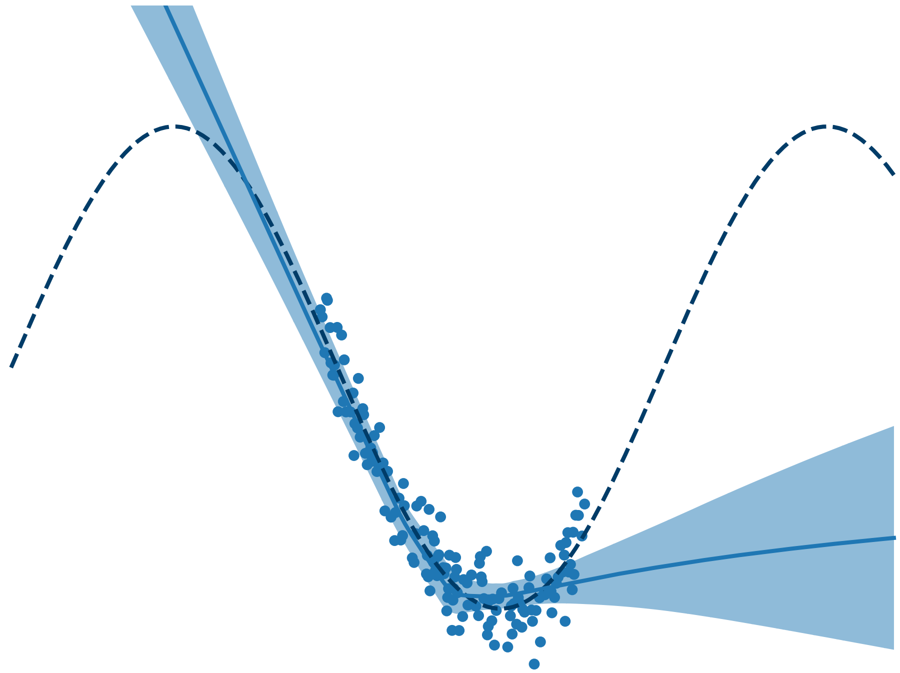
Gaussian process: squared exponential kernel

Extrapolation and uncertainty



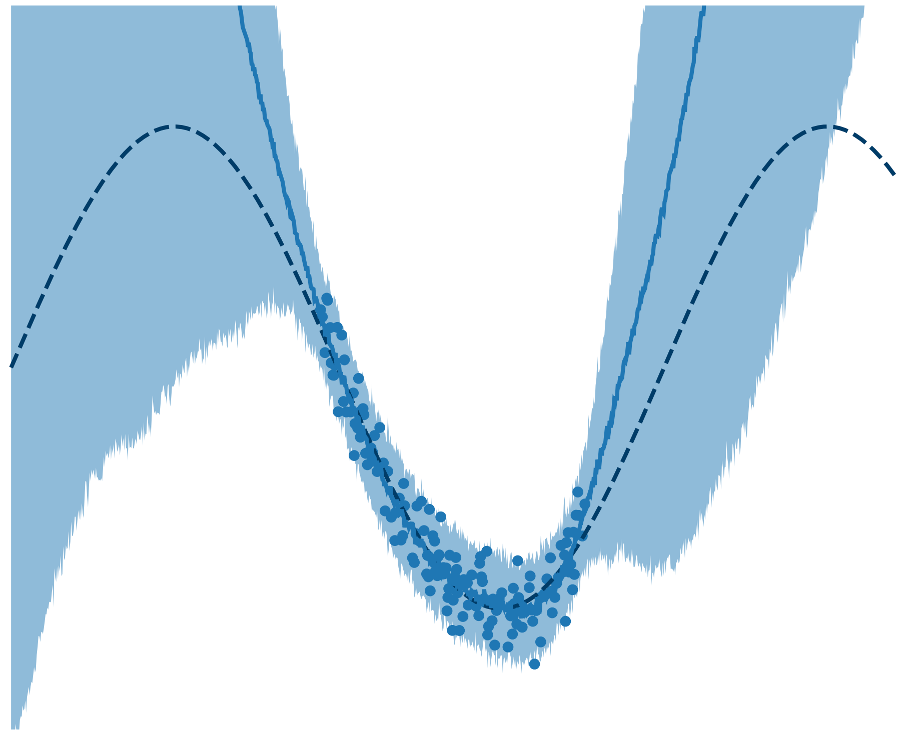
Gaussian process: polynomial kernel

Extrapolation and uncertainty



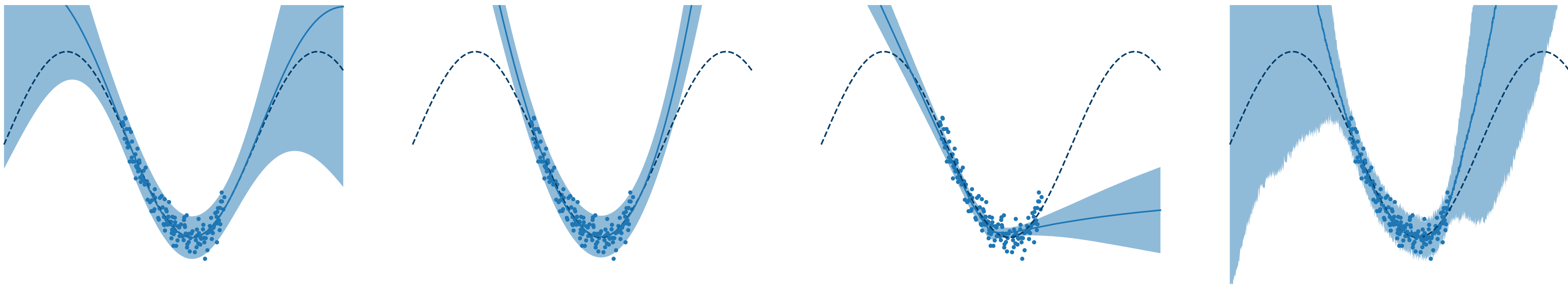
Neural network ensemble

Extrapolation and uncertainty



Bayesian neural network using MCMC

Extrapolation and uncertainty



Gaussian process: stationary kernel

Gaussian process: polynomial kernel

Ensemble

Bayesian neural network

Goal: reasonable control of model's uncertainty behavior

Vague idea: regulate uncertainty using symmetries

Geometric Learning

What are we trying to do?

Non-linear regression and classification:

$$\min_{f \in \mathcal{F}} \sum_{i=1}^{N} L(y_i, f(x_i))$$

where

- (x_i, y_i): training data
- L: loss function, such as square loss $\|y_i - f(x_i)\|^2$
- $\mathcal{F}$: some space of functions $f : X \to Y$

Spaces X *and/or* Y *carry geometric structure*

What are we trying to do?

Probabilistic non-linear regression and classification:

$$y_i \sim p_{y|f}(\cdot \mid f(x_i))$$

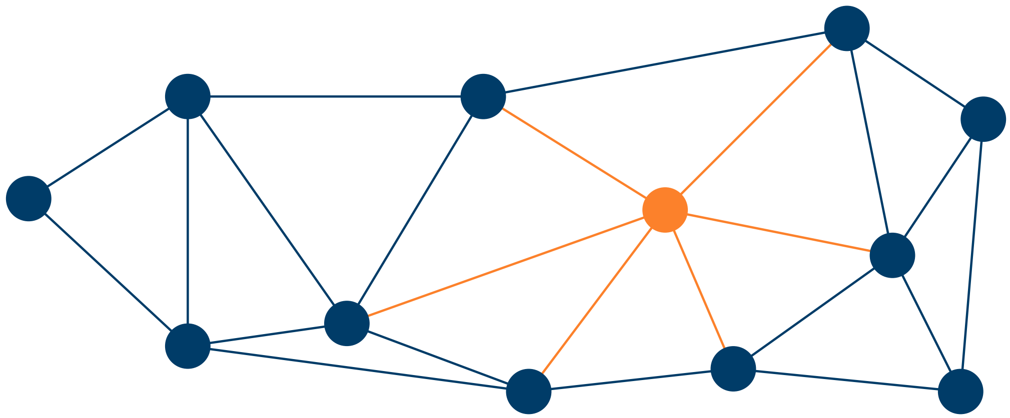
where

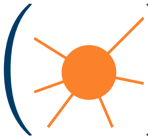
- (x_i, y_i): training data
- $p_{y|f}$: likelihood, such as Gaussian $y_i \sim \mathrm{N}(f(x_i), \sigma^2)$
- Bayesian approach: place a prior p_f over functions $f : X \to Y$

Most models can be formulated in both ways

Spaces X and Y carry geometric structure

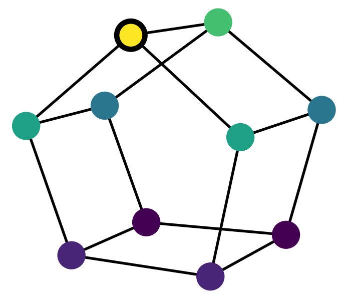
What are we trying to do? An example: graphs



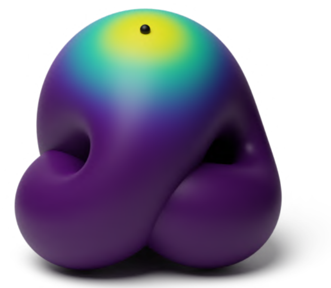
$f\Big($  $\Big) \to \mathbb{R}$

Geometric learning: settings

Meshes



Graphs



Manifolds

Vector Fields

And many more!

Challenge: how do we represent data?

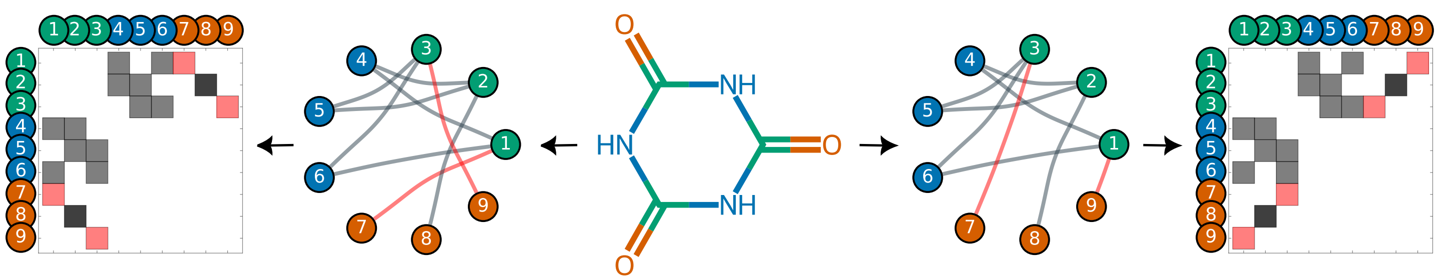
Graph: $G = (V, E)$ where

- V: set of vertices
- $E \subseteq V \times V$: edges

Numerical representation:

- $\boldsymbol{V}$: integers
- $\mathbf{E}$: binary matrix

Different orderings represent same graph $\rightsquigarrow$ *permutation invariance*



$\boldsymbol{V} \leftrightsquigarrow \mathbf{P}\boldsymbol{V}$ $\quad$ $\mathbf{E} \leftrightsquigarrow \mathbf{P}\mathbf{E}\mathbf{P}^T$ $\quad$ $\mathbf{P}$: permutation matrix

Challenge: how do we represent data?

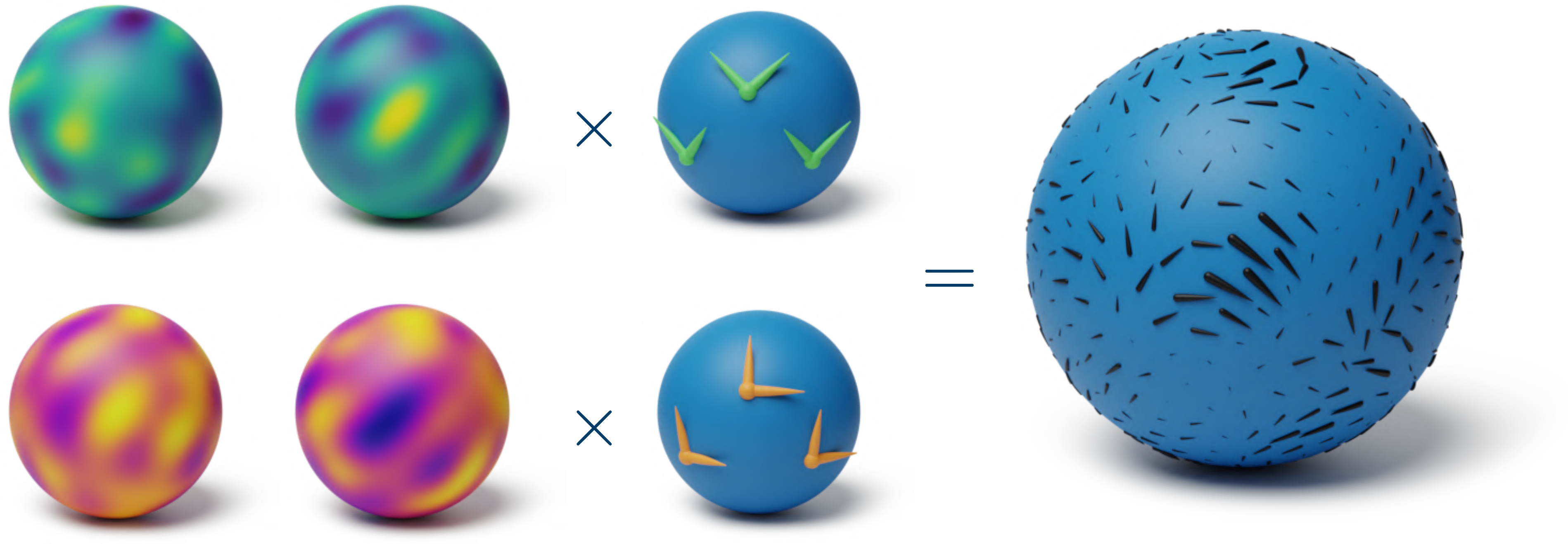
Two ways of representing points on a sphere:

- Points in $\mathbb{R}^3$: $(x, y, z) \in \mathbb{R}^3$ with $x^2 + y^2 + z^2 = 1$
- Angles: (θ, ϕ) representing angle relative to a chosen point

Different numbers: *same point on sphere*

Predictions should depend *on our data,*
not on how we store it numerically

Challenge: how do we represent data?



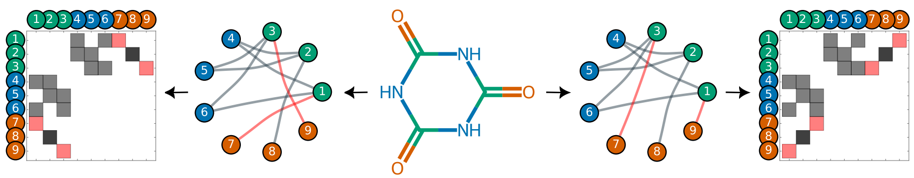
Same vector field: represented differently in different *frames*

Predictions should depend *on our data,* not on the choice of frame

Symmetries: predictions that depend only on data

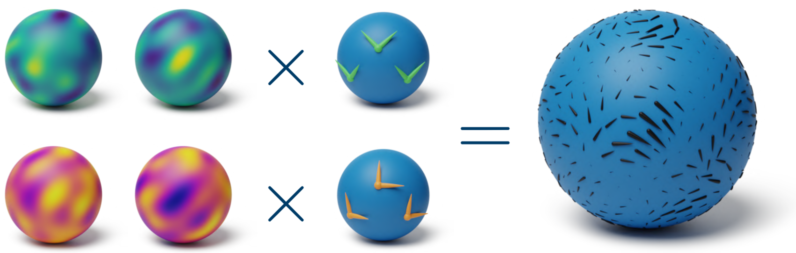
Need models that respect representation's symmetries

Invariance



Transformations of representation
don't affect output

Equivariance



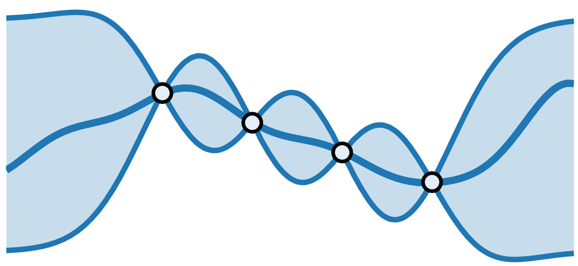
Transformations correctly carry over
from input to output

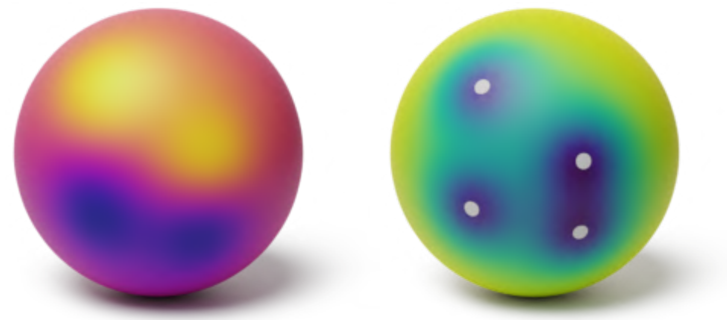
Challenge: model design

Symmetries: extrapolation and uncertainty

Design principle: models which respect the *space's symmetries*

- Symmetry-based extrapolation
- Symmetry-based uncertainty





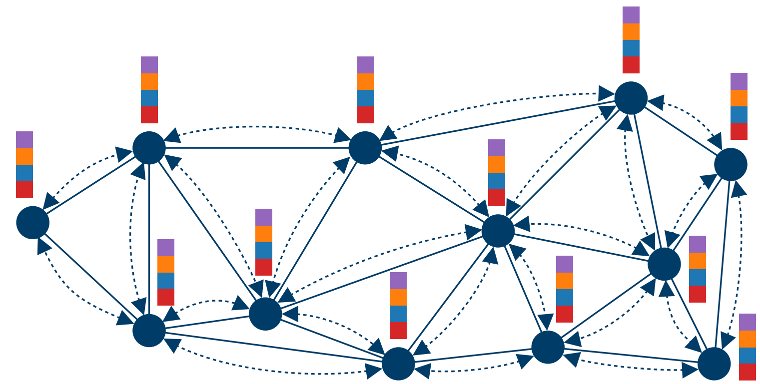
..unless data indicates otherwise?

Just as relevant in classical Euclidean setting

Geometric learning: right technical language for studying this

Geometric deep learning: models

Graph neural networks



- Geometric analogs of ConvNets
- Most-common model by far

Idea:

1. Pixel-space convolution: discretization of Euclidean convolution
2. Replace Euclidean convolution with non-Euclidean analog
3. Discretize: obtain graph neural network architecture

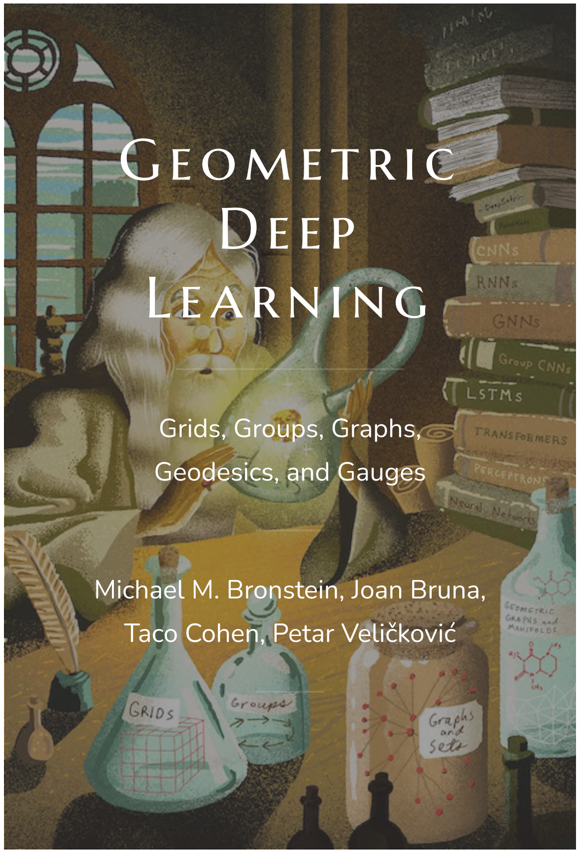
Challenge: what is the *right* notion of convolution?

Geometric deep learning: models

Other models:

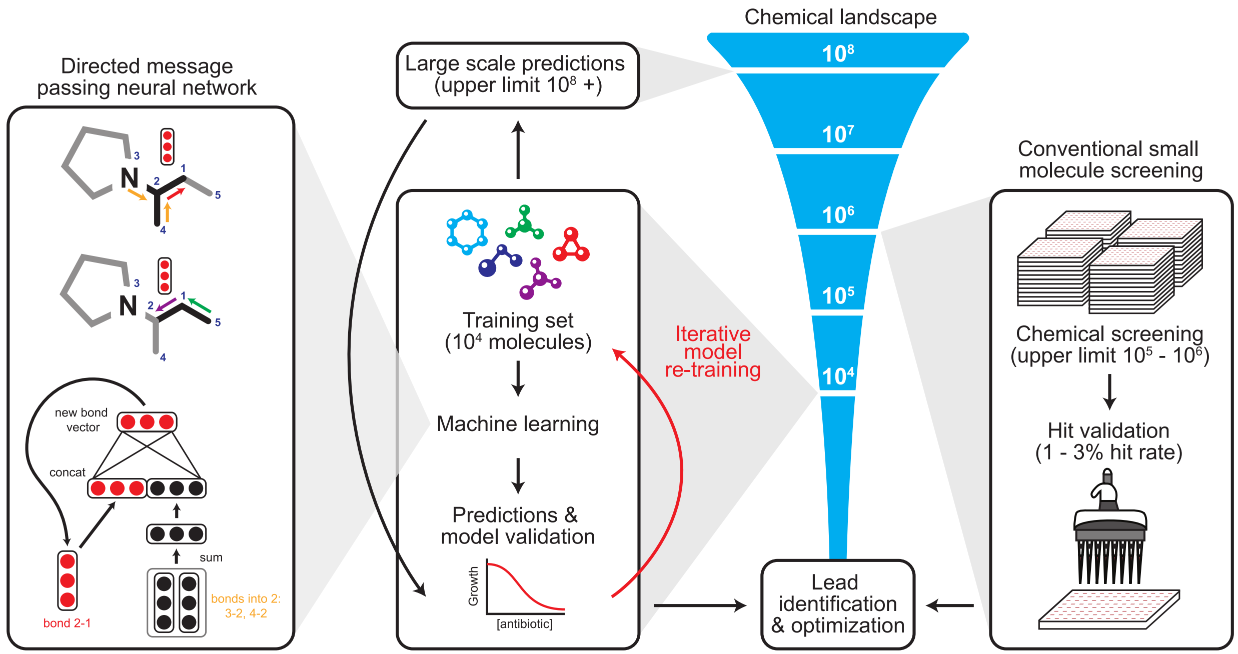
- Attention and graph transformers
- Invariant and equivariant networks
- Lots of other ideas

Check out the recent book, available with slides and other content at HTTPS://GEOMETRICDEEPLEARNING.COM



Applications of Geometric Learning

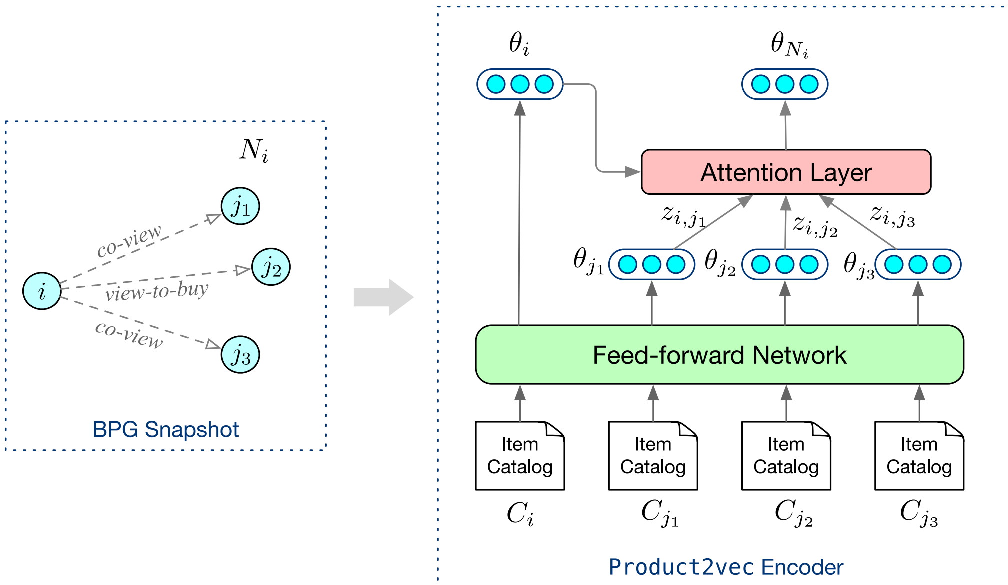
Applications: drug discovery



Graph neural network: predicted antibiotic activity in halicin, prev. studied for diabetes
Shown in mice to have broad-spectrum antibiotic activity

Figure and results: Stokes et al. (Cell, 2020)

Applications: recommendation systems

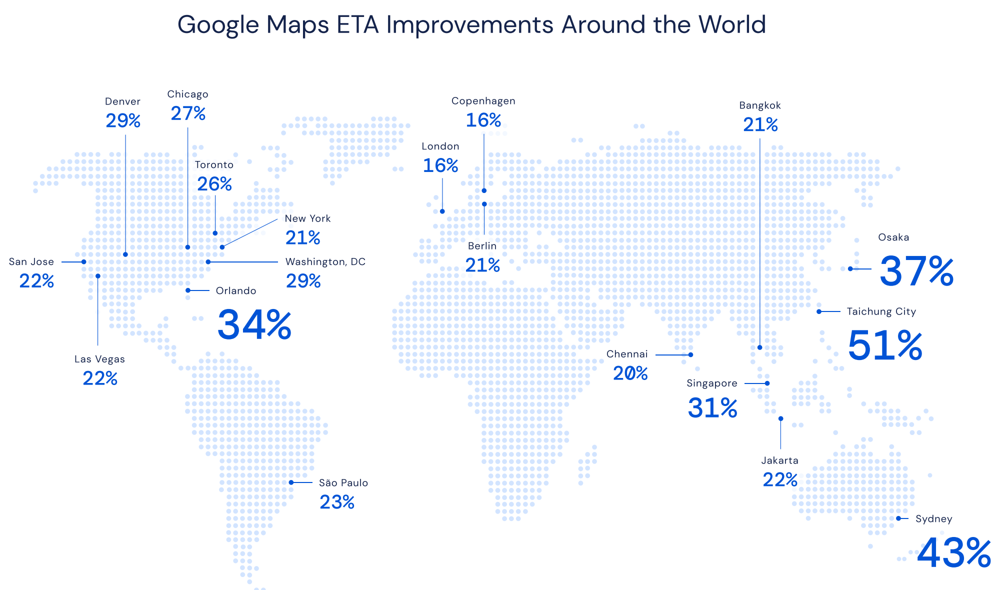


Graph neural networks: used to improve complementary product recommendation

Improved performance in production compared to Amazon's prior approach

Figure and results: Hao et al. (CIKM 2020)

Applications: traffic prediction



Graph neural networks: Google Maps traffic ETA prediction
Improved accuracy in production compared to prior approach

Figure and results: Derrow-Pinion et al. (CIKM 2021)

UAI 2024 Tutorial

Geometric Probabilistic Models

Viacheslav (Slava) Borovitskiy
@VABOR112

ETH zürich

Alexander Terenin
@AVT_IM

Cornell University

UAI 2024 Tutorial

Geometric Probabilistic Models

Part II

Viacheslav (Slava) Borovitskiy

Part I: background, motivation, technical fundamentals

1. Model-based decision-making
2. Probabilistic modeling and uncertainty
3. Geometric learning: working with structured data

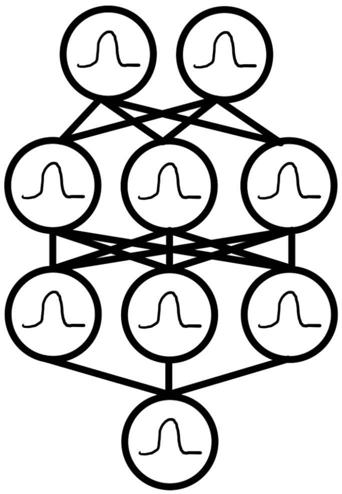
Part II: geometric probabilistic models

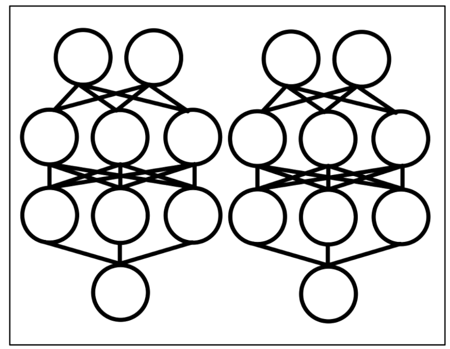
1. Three types of probabilistic models: an example problem
2. Deep dive: geometric Gaussian processes
3. Emerging applications
4. Software
5. Summary

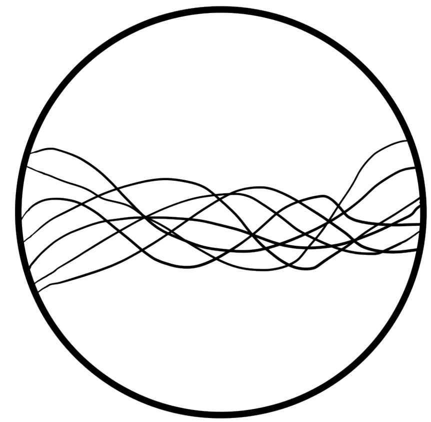
References

Basic types of probabilistic models

Basic types of probabilistic models







Bayesian Neural Networks

Deep Ensembles

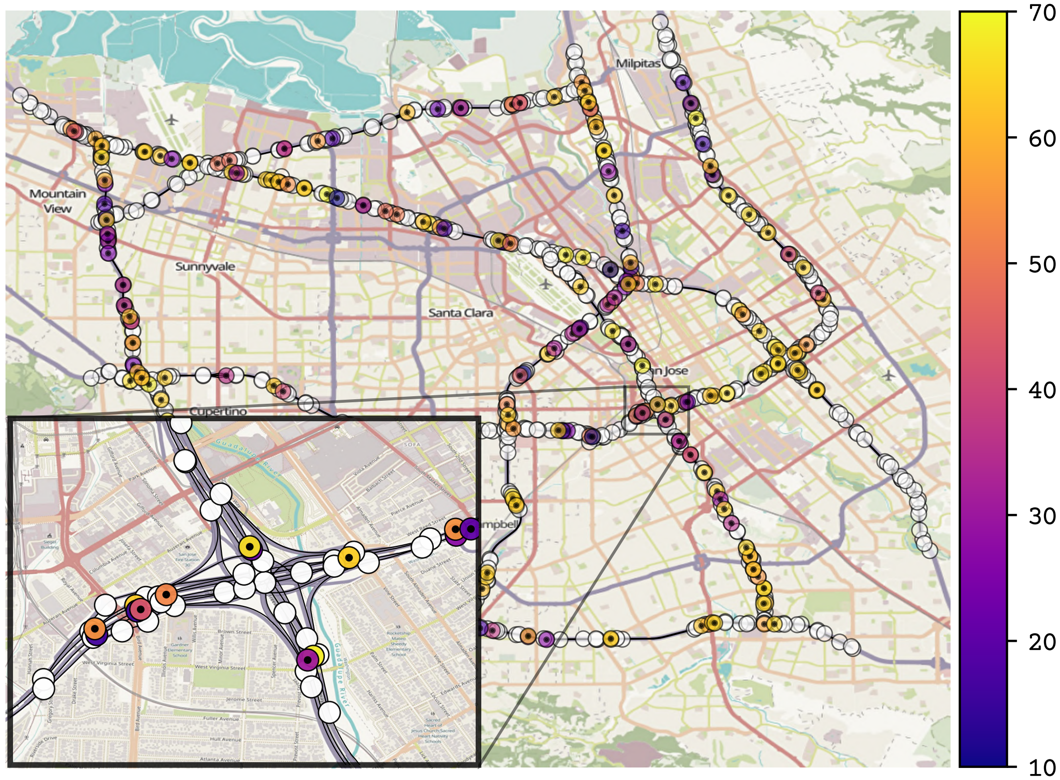
Gaussian Processes

require specialized architectures

require specialized kernels

An example problem

An example problem

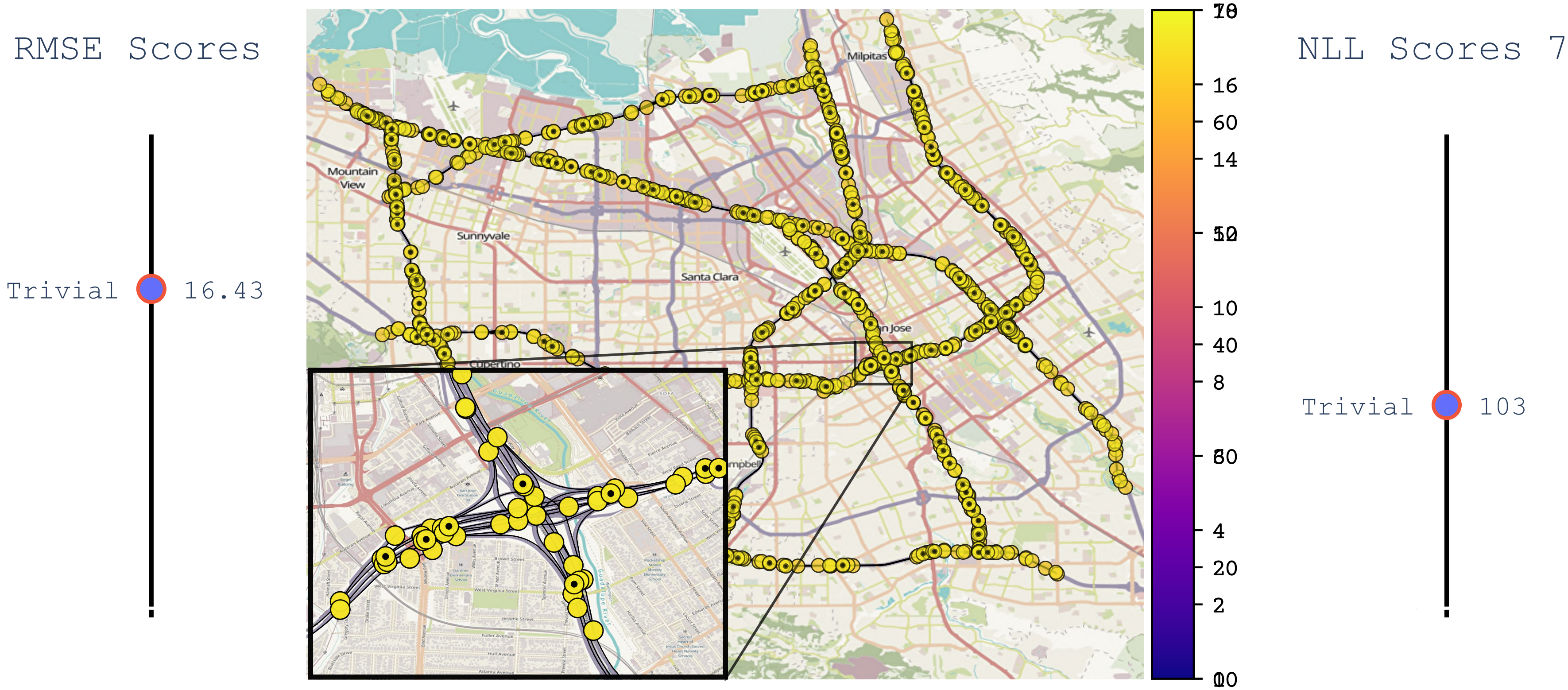


Use 250 labeled nodes for training data and 75 for test data

Dataset details: Borovitskiy et al. (AISTATS 2021)

Simple baseline

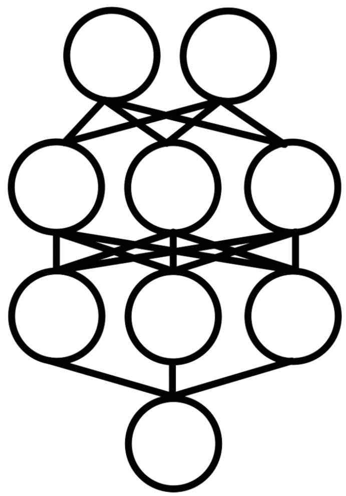
Simple baseline: uncertainty



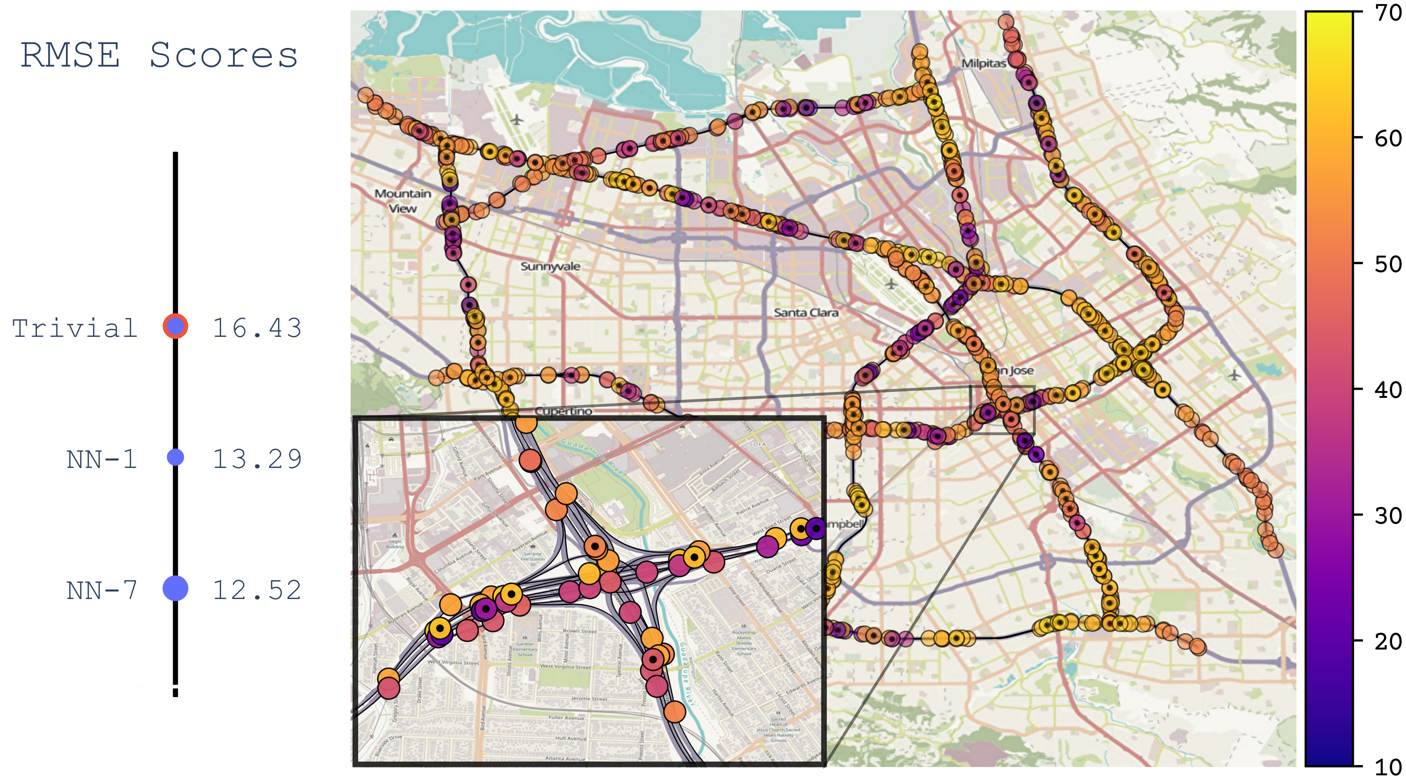
Trying out different probabilistic models

Link to code at the end of the tutorial

Sanity Check:
Deterministic Graph CNN

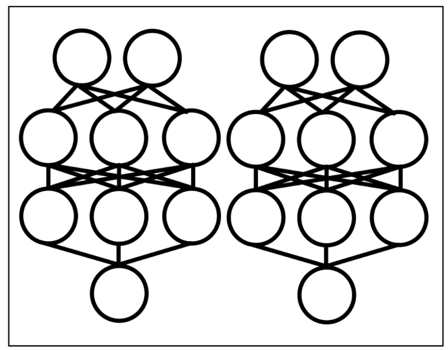


Deterministic Graph CNN: prediction (7 layers)



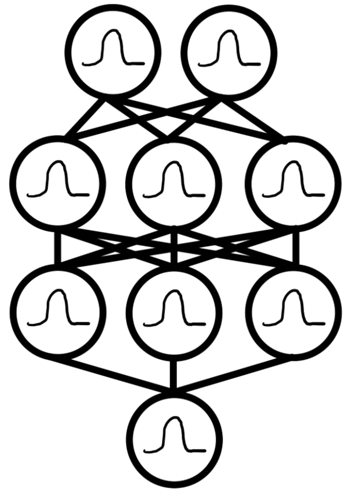
Probabilistic models

Deep Ensemble

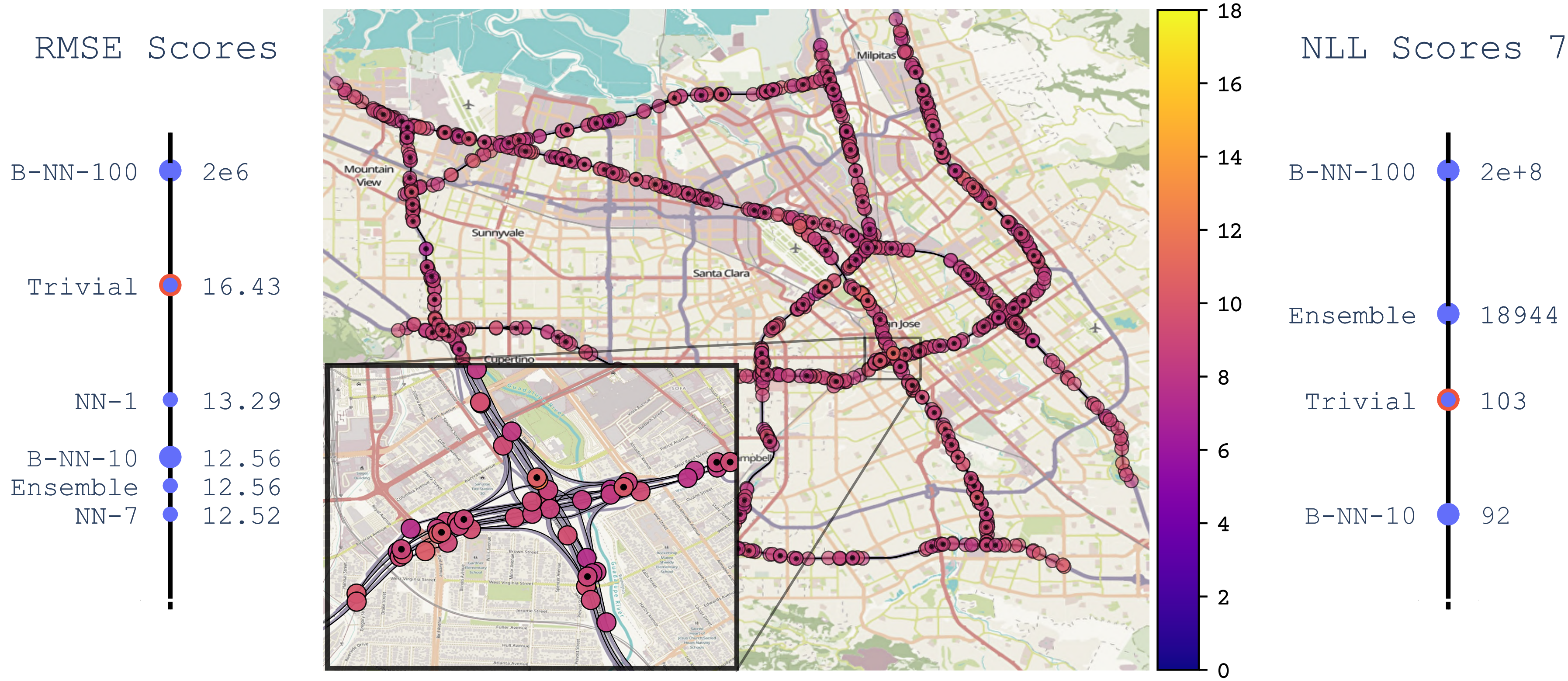


Deep Ensemble: uncertainty

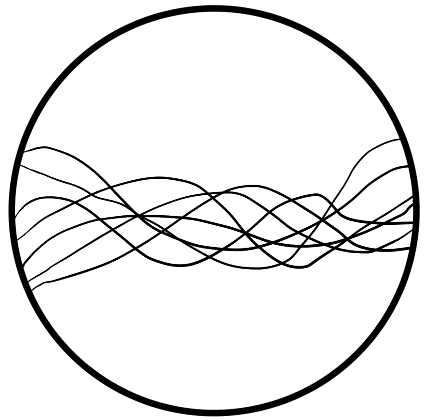
Bayesian Graph CNN



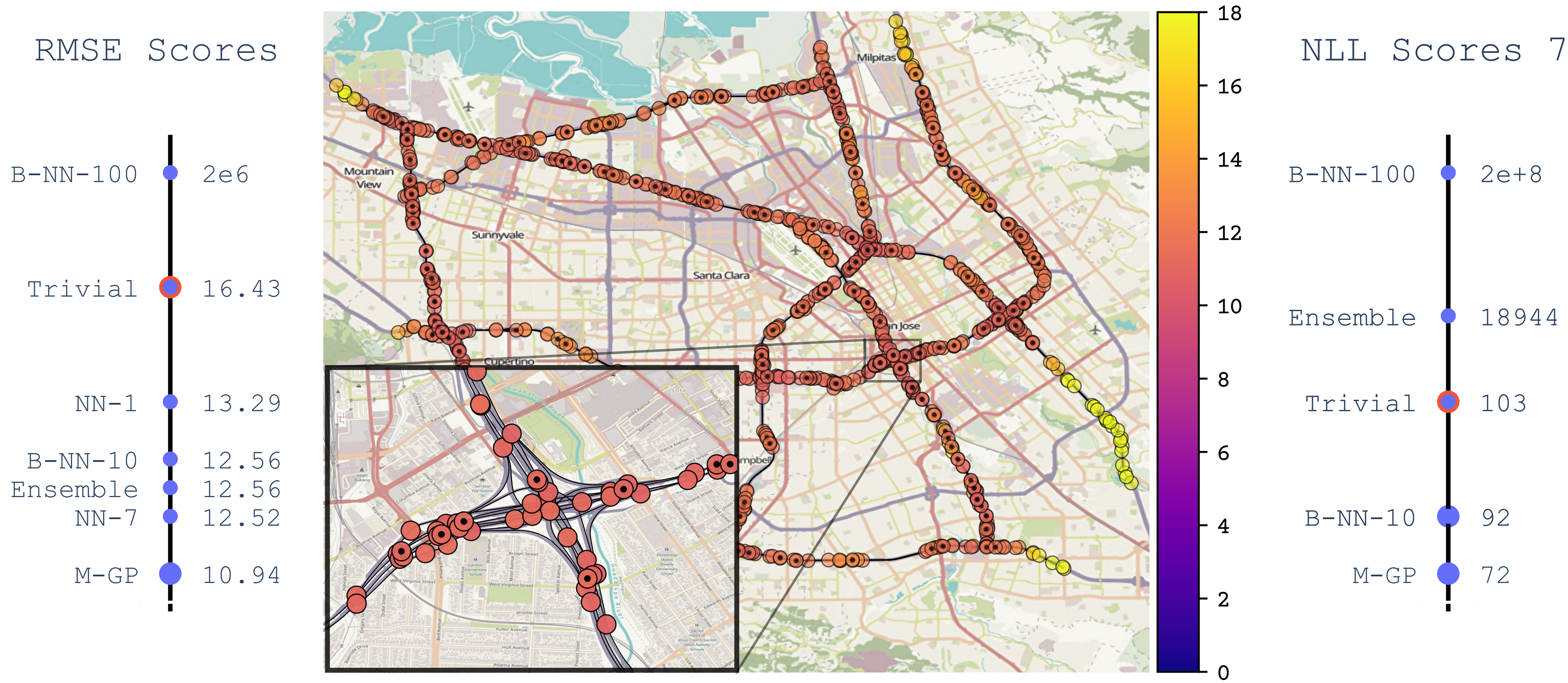
Bayesian Graph CNN



Gaussian Process

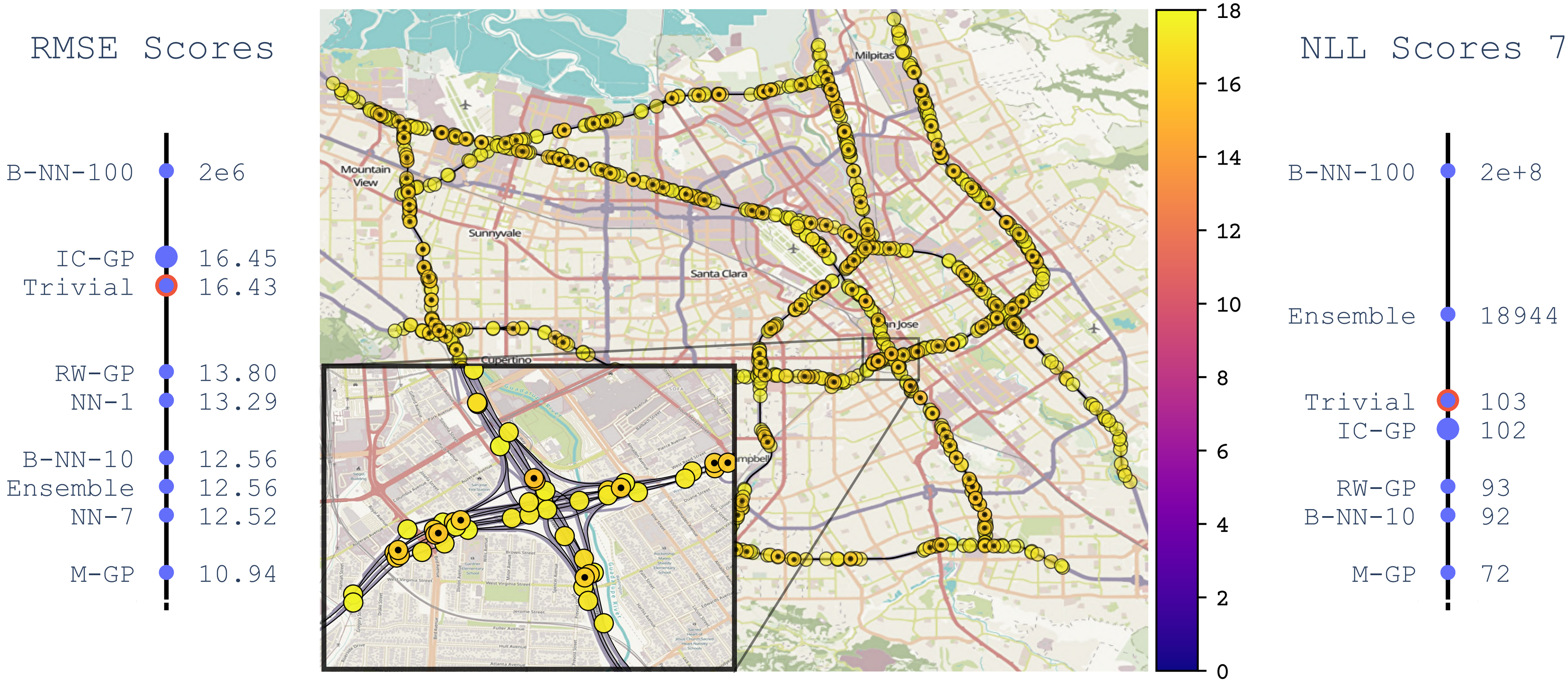


Gaussian Process: geometric Matérn kernel – uncertainty



Not all Gaussian processes work well

Gaussian Process: inverse cosine kernel, uncertainty



A way to think about this

A way to think about this

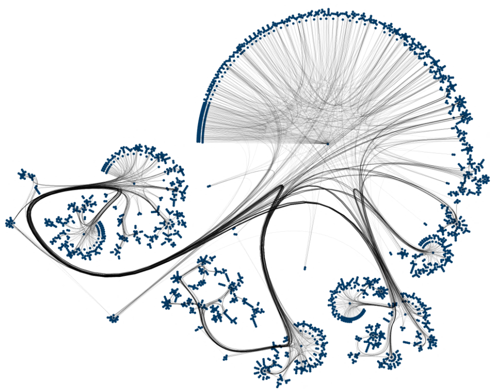
This example: easy-to-use and *visualize* benchmark for regression on geometric data

Geometric Gaussian processes:

- Reliable off-the-shelf probabilistic models
- Solid baselines to benchmark other models against

Deep Dive:
Geometric Gaussian Processes

Geometric (non-Euclidean) domains

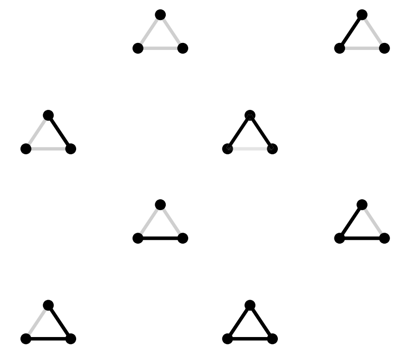


Graphs

road networks

Manifolds

physics, robotics



Spaces of graphs

drug design

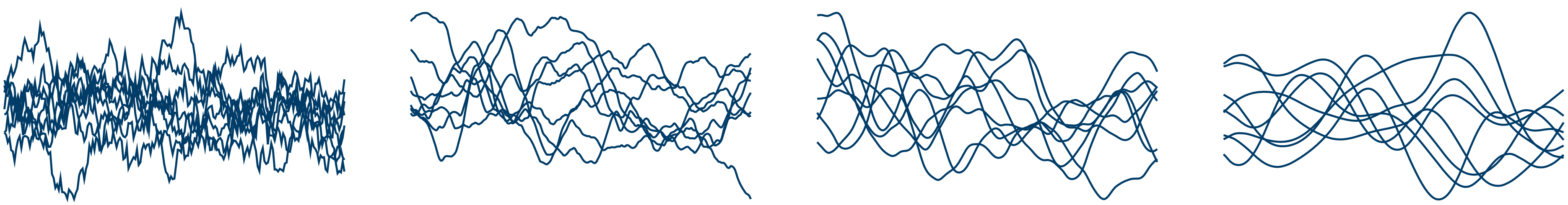
Standard kernels in $\mathbb{R}^n$

Matérn kernels

$$k_{\nu,\kappa,\sigma^2}(x,x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|x - x'|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{|x - x'|}{\kappa} \right)$$

σ^2: variance κ: length scale ν: smoothness

$\nu \to \infty$: RBF kernel (Gaussian, heat, diffusion)



$\nu = 1/2$ $\nu = 3/2$ $\nu = 5/2$ $\nu = \infty$

How to generalize these to non-Euclidean settings?

Distance-based approach

$$k^{(d)}_{\infty,\kappa,\sigma^2}(x, x') = \sigma^2 \exp\left(-\frac{d(x, x')^2}{2\kappa^2}\right)$$

Geometry-aware, but...

Manifolds: not well-defined unless the manifold is isometric to a Euclidean space
Feragen et al. (CVPR 2015)

Graphs: not well-defined unless nodes can be isometrically embedded into a Hilbert space
Schonberg (TAMS 1938)

Spaces of graphs: what is a *space* of graphs?

Need a different generalization

Another approach: stochastic partial differential equations

$$\underbrace{\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{\nu}{2}+\frac{d}{4}} f = \mathcal{W}}_{\text{Matérn}}$$

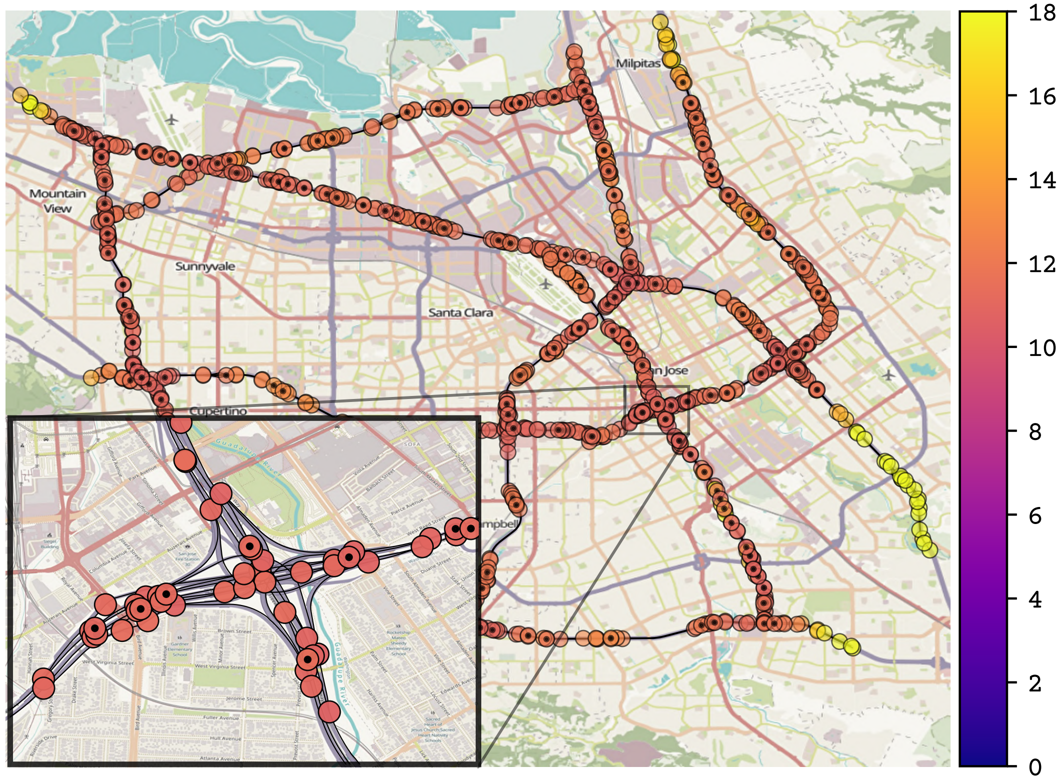
Δ: Laplacian $\quad \mathcal{W}$: Gaussian white noise $\quad d$: dimension

Whittle (ISI 1963)

Lindgren et al. (JRSSB 2011)

Making it explicit: node sets of graphs

Motivating example: traffic speed modeling



Matérn kernels on weighted undirected graphs

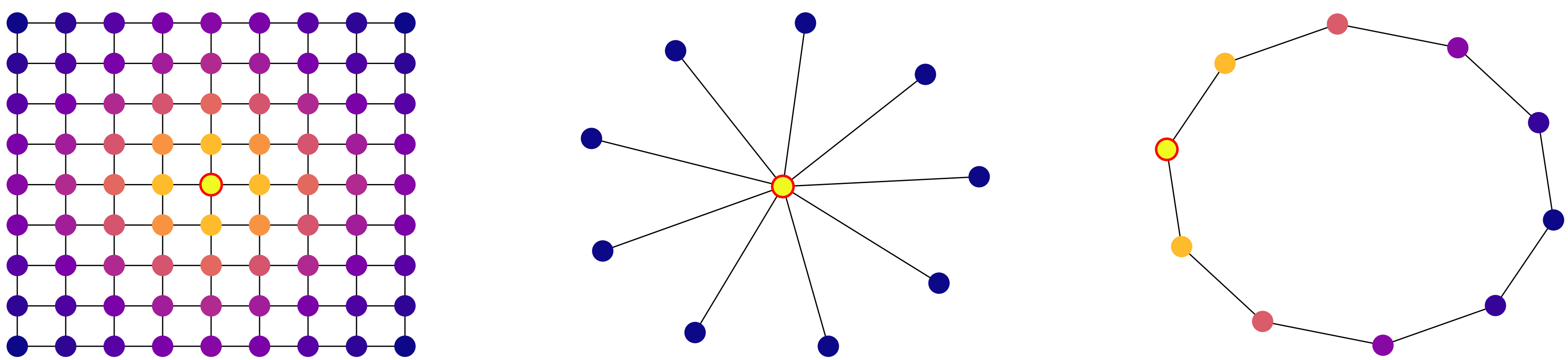
SPDE turns into a stochastic linear system: solution has kernel

$$k_{\nu,\kappa,\sigma^2}(i,j) = \frac{\sigma^2}{C_{\nu,\kappa}} \sum_{n=0}^{|V|-1} \Phi_{\nu,\kappa}(\lambda_n) \boldsymbol{f_n}(i) \boldsymbol{f_n}(j)$$

λ_n, $\boldsymbol{f_n}$: eigenvalues and eigenvectors of the Laplacian matrix

$$\Phi_{\nu,\kappa}(\lambda) = \begin{cases} \left(\frac{2\nu}{\kappa^2} - \lambda\right)^{-\nu} & \nu < \infty : \text{ Matérn} \\ e^{-\frac{\kappa^2}{2}\lambda} & \nu = \infty : \text{ heat (sq. exp., RBF)} \end{cases}$$

Examples: $k_{5/2}(\bullet, \cdot)$ on node sets of different graphs



Making it explicit: other settings

Making it explicit: other settings

Compact manifolds:

$$
\begin{aligned}
k_{\nu,\kappa,\sigma^2}(x,x') &= \frac{\sigma^2}{C_{\nu,\kappa}} \sum_{n=0}^{\infty} \Phi_{\nu,\kappa}(\lambda_n) f_n(x) f_n(x') \\
&= \frac{\sigma^2}{C_{\nu,\kappa}} \sum_{l=0}^{\infty} \Phi_{\nu,\kappa}(\lambda_l) \underbrace{\sum_{s=1}^{d_s} f_{l,s}(x) f_{l,s}(x')}_{G_l(x,x')}
\end{aligned}
$$

Infinite summation where λ_n, f_n are Laplace–Beltrami eigenpairs

Addition theorems: more efficient $G_l(x, x')$-based expressions

Making it explicit: other settings

Non-compact manifolds:

- No summation: instead, integral approximated via Monte Carlo

Spaces of graphs:

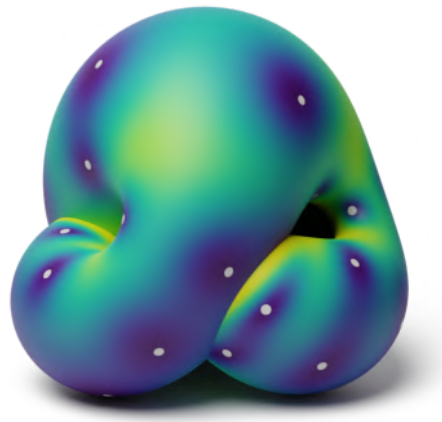
- Geometry: make a set of graphs to be a node set of a larger graph
- Then, heat and Matérn kernels are automatically defined
- Larger graph is huge: need clever computational techniques

Gaussian processes: other geometric settings

Gaussian processes: other geometric settings

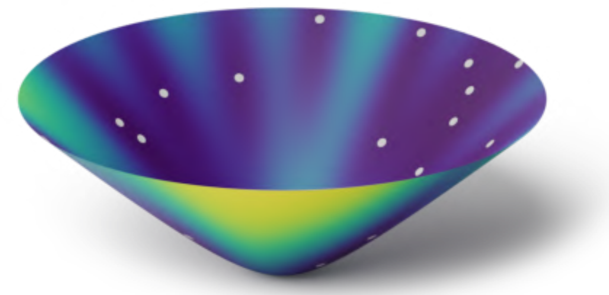
Simple manifolds: meshes, spheres, tori

Borovitskiy et al. (AISTATS 2020)



Compact Lie groups and homogeneous spaces: $\mathrm{SO}(n)$, resp. $\mathrm{Gr}(n, r)$

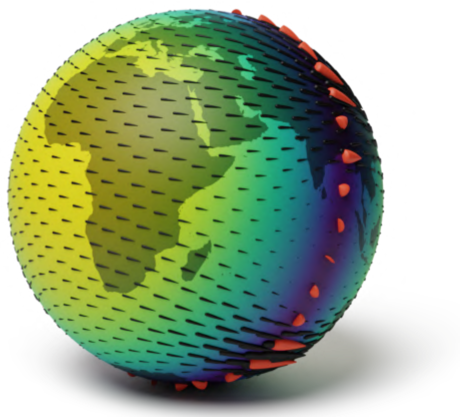
Azangulov et al. (JMLR 2024)



Non-compact symmetric spaces: $\mathbb{H}_n$ and $\mathrm{SPD}(n)$

Azangulov et al. (JMLR 2024)

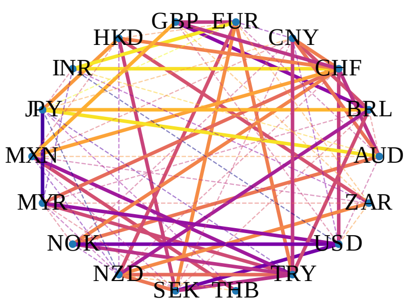
Gaussian processes: other geometric settings



Vector fields on simple manifolds

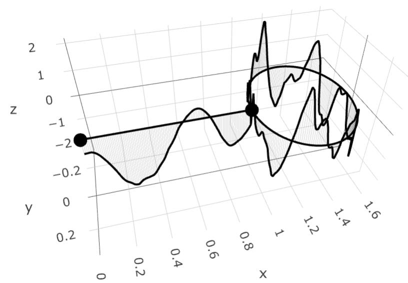
Robert-Nicoud et al. (AISTATS 2024)

Gaussian processes: other geometric settings



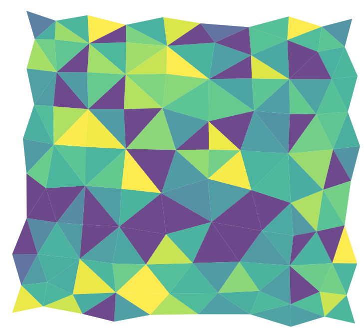
Edges of graphs

Yang et al. (AISTATS 2024)



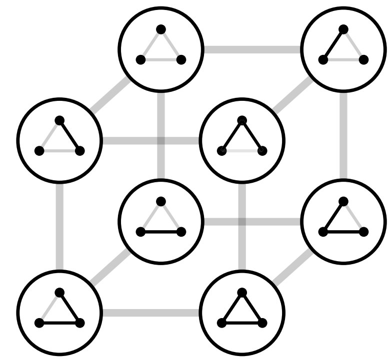
Metric graphs

Bolin et al. (Bernoulli 2024)



Cellular Complexes

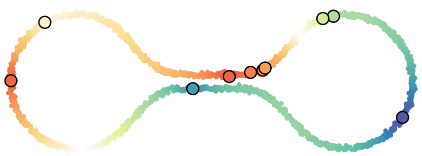
Alain et al. (ICML 2024)



Spaces of Graphs

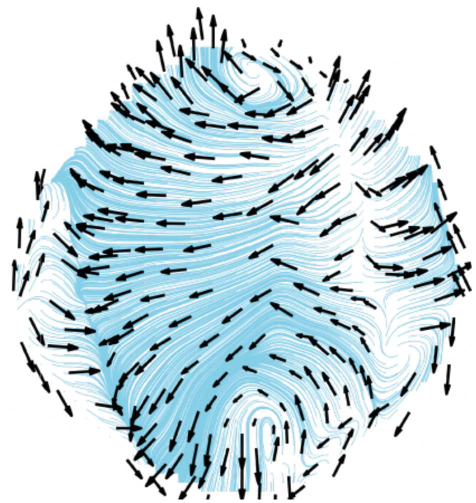
Borovitskiy et al. (AISTATS 2023)

Implicit geometry



Semi-supervised learning:
scalar functions

Fichera et al. (NeurIPS 2023)



Semi-supervised learning:
vector fields

Peach et al. (ICLR 2024)

Theoretical results

Theoretical results

(A) $\quad d \leq 3$

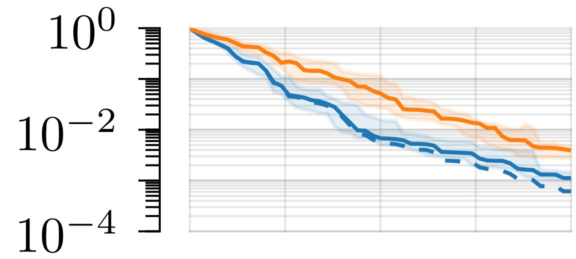
$\sigma_1^2 / C_{\nu_1,\alpha_1} = \sigma_2^2 / C_{\nu_2,\alpha_2},\ \nu_1 = \nu_2$

(B) $\quad d \geq 4$

$\sigma_1^2 = \sigma_2^2 \ \textit{and}\ \alpha_1 = \alpha_2,\ \nu_1 = \nu_2$

Hyperparameter inference

Li et al. (JMLR 2024)

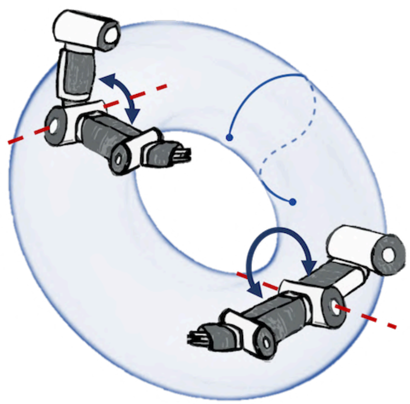


Minimax errors

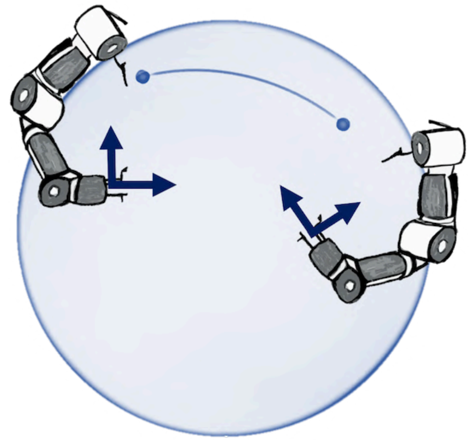
Rosa et al. (NeurIPS 2023)

Emerging real-world applications

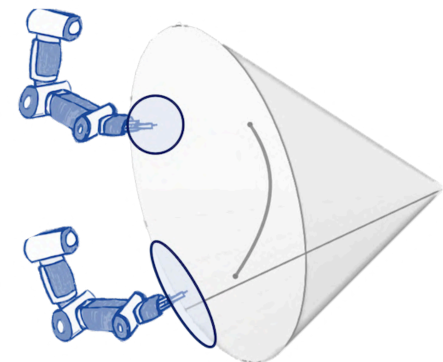
Robotics: Bayesian optimization of control policies



Joint postures: $\mathbb{T}^d$



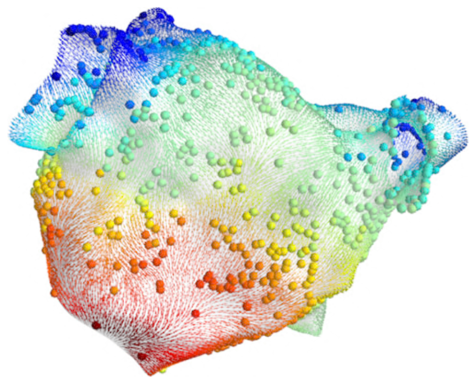
Rotations: $\mathbb{S}_3$, $\mathbf{SO(3)}$



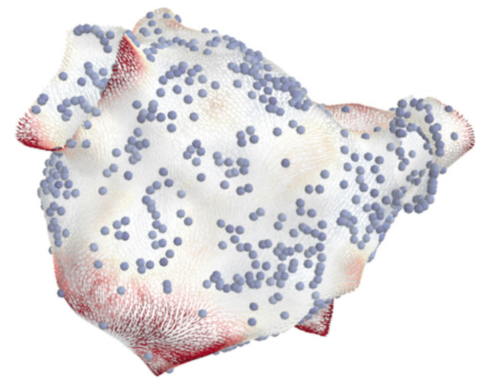
Manipulability: $\mathbf{SPD(3)}$

Jaquier et al. (CoRL 2019, CoRL 2021), figures by the authors

Medicine: probabilistic atrial activation maps



Prediction

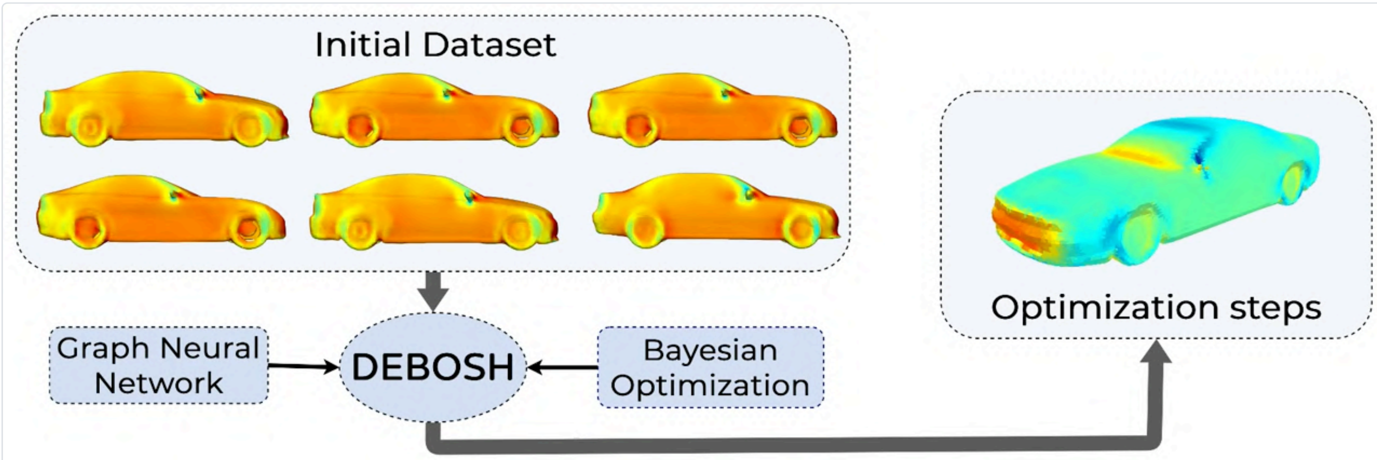


Uncertainty

Matérn Gaussian processes on meshes (independently proposed)

Coveney et al. (TBE 2019, PTRSA 2020), figures by the authors

Engineering: Bayesian optimization of shapes



Based on ensembles of geodesic CNNs

Neural Concept, video from neuralconcept.com

Join us!

Lots of worthwhile problems to study!

Software

Example Problem: Code

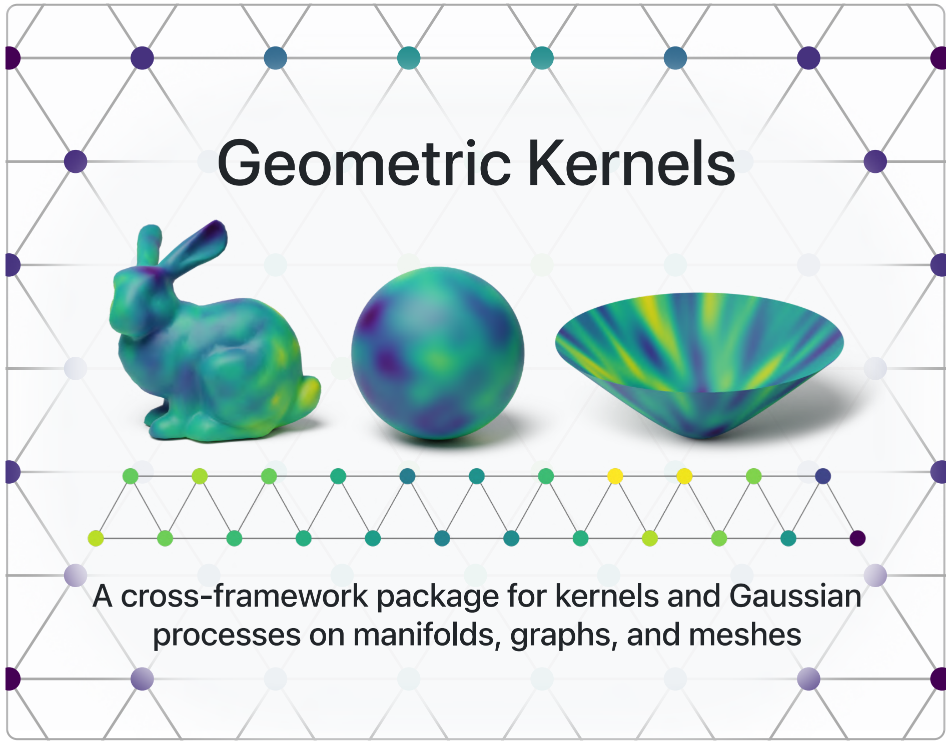
Stack:

- PyTorch Geometric: graph neural networks
- Pyro: probabilistic programming
- GPyTorch: Gaussian processes
- GeometricKernels: geometric Matérn kernels

Implementation:

Colab notebook

GeometricKernels: Matérn and heat kernels on geometric domains



HTTPS://GEOMETRIC-KERNELS.GITHUB.IO/

Matérn and Heat Kernels on Geometric Domains

Example code:

```
>>> # Define a space (2-dim sphere).
>>> hypersphere = Hypersphere(dim=2)

>>> # Initialize kernel.
>>> kernel = MaternGeometricKernel(hypersphere)
>>> params = kernel.init_params()

>>> # Compute and print out the 3x3 kernel matrix.
>>> xs = np.array([[0., 0., 1.], [0., 1., 0.], [1., 0., 0.]])
>>> print(kernel.K(params, xs))

[[1.    0.356     0.356]
  [0.356 1.        0.356]
  [0.356 0.356 1.       ]]
```

Multi-backend: *Numpy, JAX, PyTorch, TensorFlow*

Software and frameworks

	PyTorch	**JAX**	**TensorFlow**
Graph Neural Networks	PyTorch Geometric	Jraph	TensorFlow GNN
Probabilistic Programming	Pyro	NumPyro	TensorFlow Probability
Gaussian Processes	GPyTorch	GPJax	GPflow
Geometric Kernels	GeometricKernels		

Neural networks beyond the graph setting: various small libraries

Summary

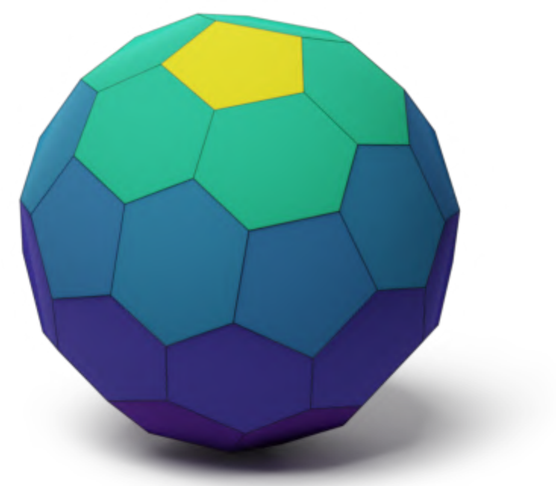
Summary

Geometric machine learning: increasingly popular

- Waves: kernels on structured data, geometric deep learning

Geometric probabilistic models: emerging research area

- Geometric Gaussian processes: solid off-the-shelf baselines
- Increasingly available but actively developing: join us!



Thank you!

@vabor112 · @avt_im

Slides at:

ETH zürich

Cornell University

References

References: geometric deep learning and applications

M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. Available at: https://geometricdeeplearning.com, 2021.

J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, and J. J. Collins. A deep learning approach to antibiotic discovery. *Cell*, 2020.

J. Hao, T. Zhao, J. Li, X. L. Dong, C. Faloutsos, Y. Sun, and W. Wang. P-companion: A principled framework for diversified complementary product recommendation. *Conference on Information and Knowledge Management*, 2020.

A. Derrow-Pinion, J. She, D. Wong, O. Lange, T. Hester, L. Perez, M. Nunkesser, S. Lee, X. Guo, B. Wiltshire, P. W. Battaglia, V. Gupta, A. Li, Z. Xu, A. Sanchez-Gonzalez, Y. Li, and P. Veličković. ETA prediction with graph neural networks in Google Maps. *Conference on Information and Knowledge Management*, 2021.

References: probabilistic neural networks in geometic settings

Y. Zhang, S. Pal, M. Coates, and D. Ustebay. Bayesian graph convolutional neural networks for semi-supervised classification. *AAAI Conference on Artificial Intelligence*, 2019.

M. Stadler, B. Charpentier, S. Geisler, D. Zügner, and S. Günnemann. Graph posterior network: Bayesian predictive uncertainty for node classification. *Advances in Neural Information Processing Systems*, 2021.

H. Shi, X. Zhang, S. Sun, L. Liu, and L. Tang. A survey on Bayesian graph neural networks. *Intelligent Human-Machine Systems and Cybernetics*, 2021.

A. Hasanzadeh, E. Hajiramezanali, S. Boluki, M. Zhou, N. Duffield, K. Narayanan, and X. Qian. Bayesian graph neural networks with adaptive connection sampling. *International Conference on Machine Learning*, 2020.

N. Durasov, A. Lukoyanov, J. Donier, and P. Fua. DEBOSH: Deep bayesian shape optimization. *Technical Report: Neural Concept*, 2021.

Q. Wang, S. Wang, D. Zhuang, H. Koutsopoulos, and J. Zhao. Uncertainty quantification of spatiotemporal travel demand with probabilistic graph neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 2024.

L. Kahle and F. Zipoli. Quality of uncertainty estimates from neural network potential ensembles. *Physical Review E*, 2022.

References: geometric Gaussian processes

V. Borovitskiy, I. Azangulov, A. Terenin, P. Mostowsky, M. P. Deisenroth, and N. Durrande. Matérn Gaussian Processes on Graphs. *Artificial Intelligence and Statistics*, 2021.

A. Feragen, F. Lauze, and S. Hauberg. Geodesic exponential kernels: When curvature and linearity conflict. *Computer Vision and Pattern Recognition*, 2015.

I. J. Schoenberg. Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 1938.

P. Whittle. Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 1963.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 2011.

V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth. Matérn Gaussian Processes on Riemannian Manifolds. *Advances in Neural Information Processing Systems*, 2020.

I. Azangulov, A. Smolensky, A. Terenin, and V. Borovitskiy. Stationary Kernels and Gaussian Processes on Lie Groups and their Homogeneous Spaces I: the compact case. *Journal of Machine Learning Research*, 2024.

I. Azangulov, A. Smolensky, A. Terenin, and V. Borovitskiy. Stationary Kernels and Gaussian Processes on Lie Groups and their Homogeneous Spaces II: non-compact symmetric spaces, 2024.

D. Robert-Nicoud, A. Krause, and V. Borovitskiy. Intrinsic Gaussian Vector Fields on Manifolds. *Artificial Intelligence and Statistics*, 2024.

M. Yang, V. Borovitskiy, and E. Isufi. Hodge-Compositional Edge Gaussian Processes. *Artificial Intelligence and Statistics*, 2024.

D. Bolin, A. B. Simas, and J. Wallin. Gaussian Whittle-Matérn fields on metric graphs. *Bernoulli*, 2024.

M. Alain, S. Takao, B. Paige, and M. P. Deisenroth. Gaussian Processes on Cellular Complexes. *International Conference on Machine Learning*, 2024.

References: geometric Gaussian processes

V. Borovitskiy, M. R. Karimi, V. R. Somnath, and A. Krause. Isotropic Gaussian Processes on Finite Spaces of Graphs. *Artificial Intelligence and Statistics*, 2023.

B. Fichera, V. Borovitskiy, A. Krause, and A. Billard. Implicit Manifold Gaussian Process Regression. *Advances in Neural Information Processing Systems*, 2023.

R. Peach, M. Vinao-Carl, N. Grossman, and M. David. Implicit Gaussian process representation of vector fields over arbitrary latent manifolds. *International Conference on Learning Representations*, 2024.

D. Li, W. Tang, and S. Banerjee. Inference for Gaussian Processes with Matern Covariogram on Compact Riemannian Manifolds. *Journal of Machine Learning Research*, 2023.

P. Rosa, V. Borovitskiy, A. Terenin, and J. Rousseau. Posterior Contraction Rates for Matérn Gaussian Processes on Riemannian Manifolds. *Advances in Neural Information Processing Systems*, 2023.

N. Jaquier, V. Borovitskiy, A. Smolensky, A. Terenin, T. Asfour, and L. Rozo. Geometry-aware Bayesian Optimization in Robotics using Riemannian Matérn Kernels. *Conference on Robot Learning*, 2021.

N. Jaquier, L. Rozo, S. Calinon, and M. Bürger. Bayesian optimization meets Riemannian manifolds in robot learning. *Conference on Robot Learning*, 2020.

S. Coveney, C. Corrado, C. H. Roney, R. D. Wilkinson, J. E. Oakley, F. Lindgren, S. E. Williams, M. D. O'Neill, S. A. Niederer, and R. H. Clayton. Probabilistic interpolation of uncertain local activation times on human atrial manifolds. *IEEE Transactions on Biomedical Engineering*, 2019.

S. Coveney, C. Corrado, C. H. Roney, D. OHare, S. E. Williams, M. D. O'Neill, S. A. Niederer, R. H. Clayton, J. E. Oakley, and R. D. Wilkinson. Gaussian process manifold interpolation for probabilistic atrial activation maps and uncertain conduction velocity. *Philosophical Transactions of the Royal Society A*, 2020.