Function-space Parameterization of Neural Networks for Sequential Learning

Aidan Scannell*, Riccardo Mereu*, Paul Edmund Chang, Ella Tamir, Joni Pajarinen, Arno Solin

**Equal contribution*

Scan

For project website+code

FCAI ICLR

fcai.fi

Motivation: Combine Benefits of Neural Networks and Gaussian Processes

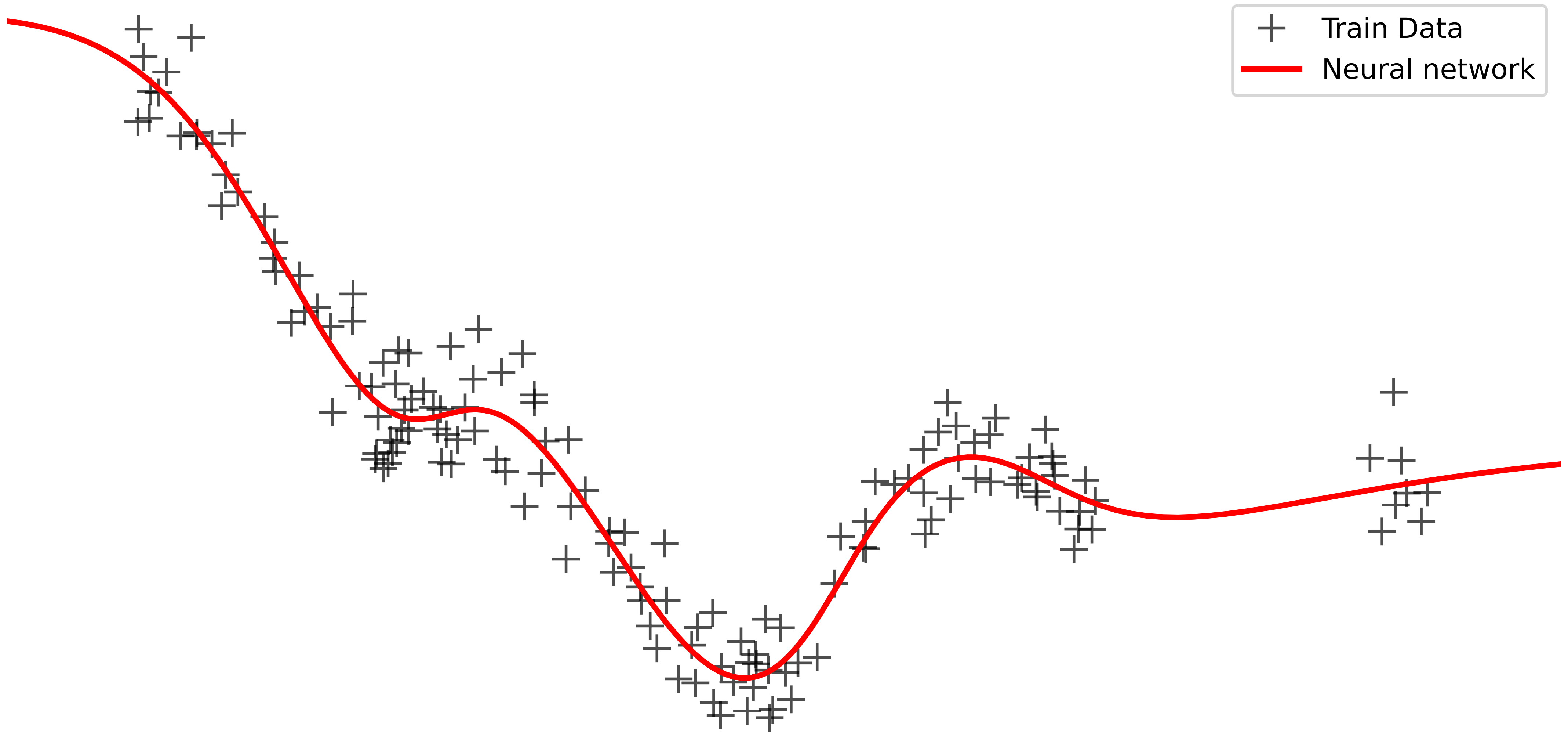
Uncertainty Quantification

Neural networks lack uncertainty estimates

Uncertainty Quantification

Neural networks lack uncertainty estimates

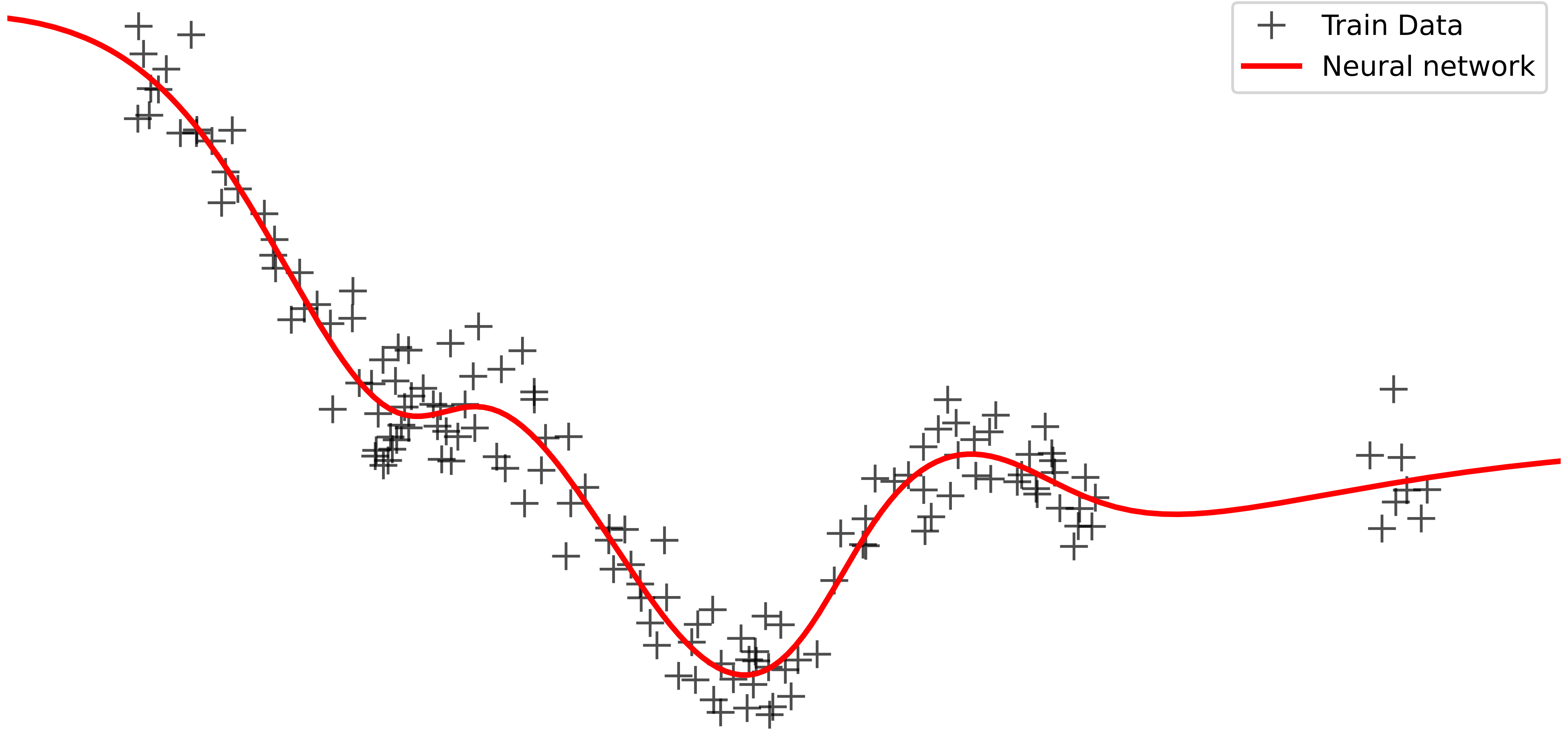
Neural network



Uncertainty Quantification

Neural networks lack uncertainty estimates

Neural network



Gaussian process

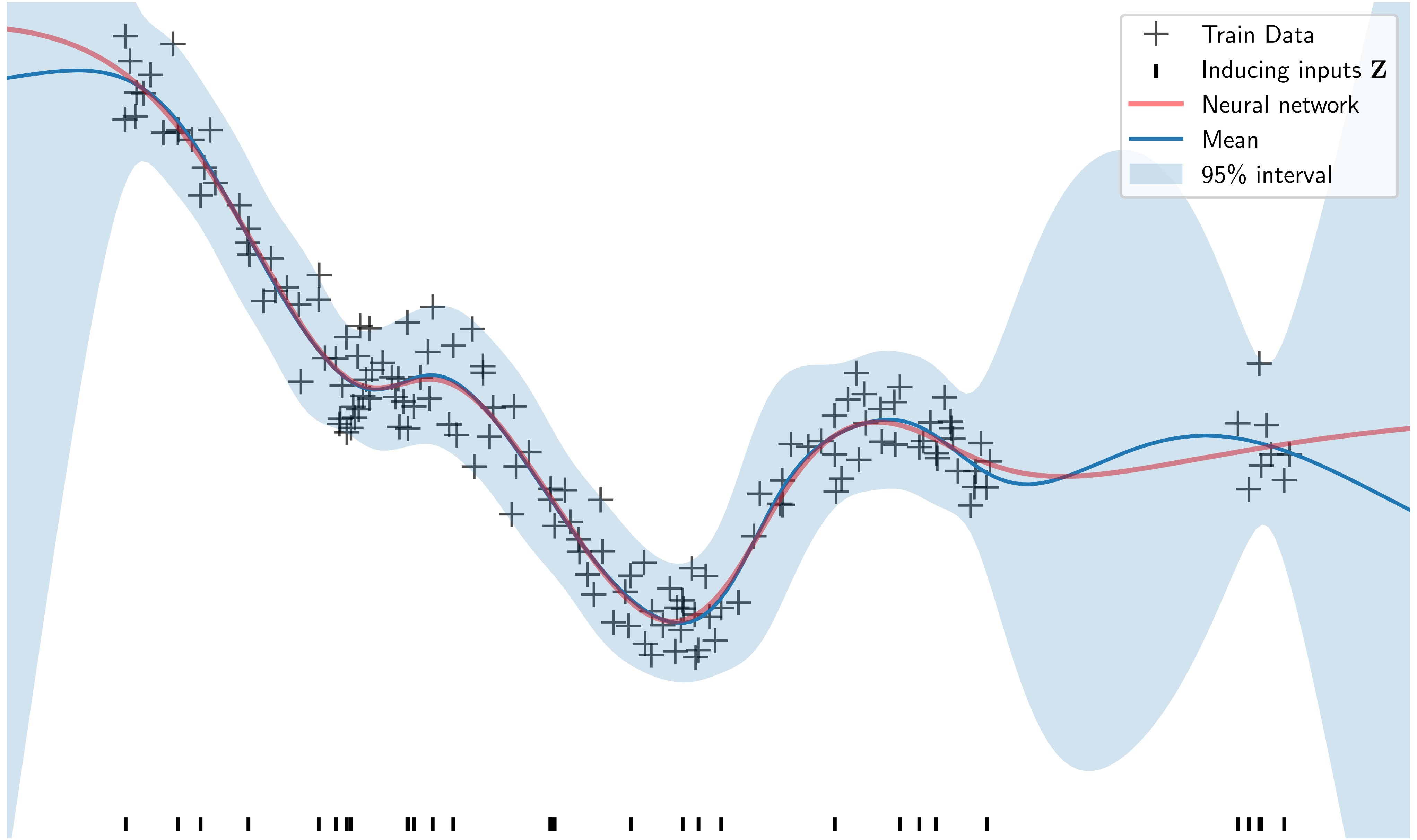


Image Inputs

Neural networks scale to high-dimensional inputs

Image Inputs

Neural networks scale to **high-dimensional inputs**

Neural network

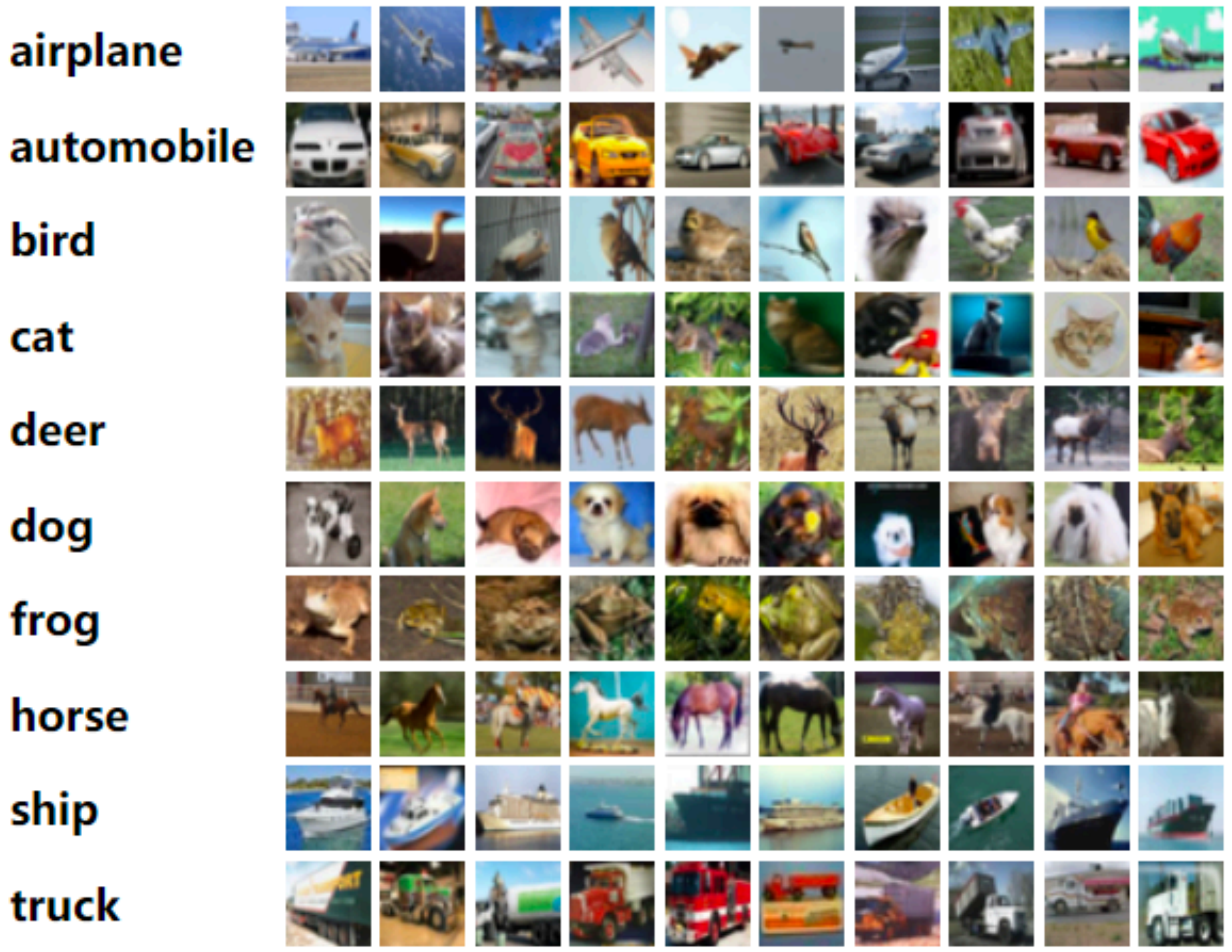
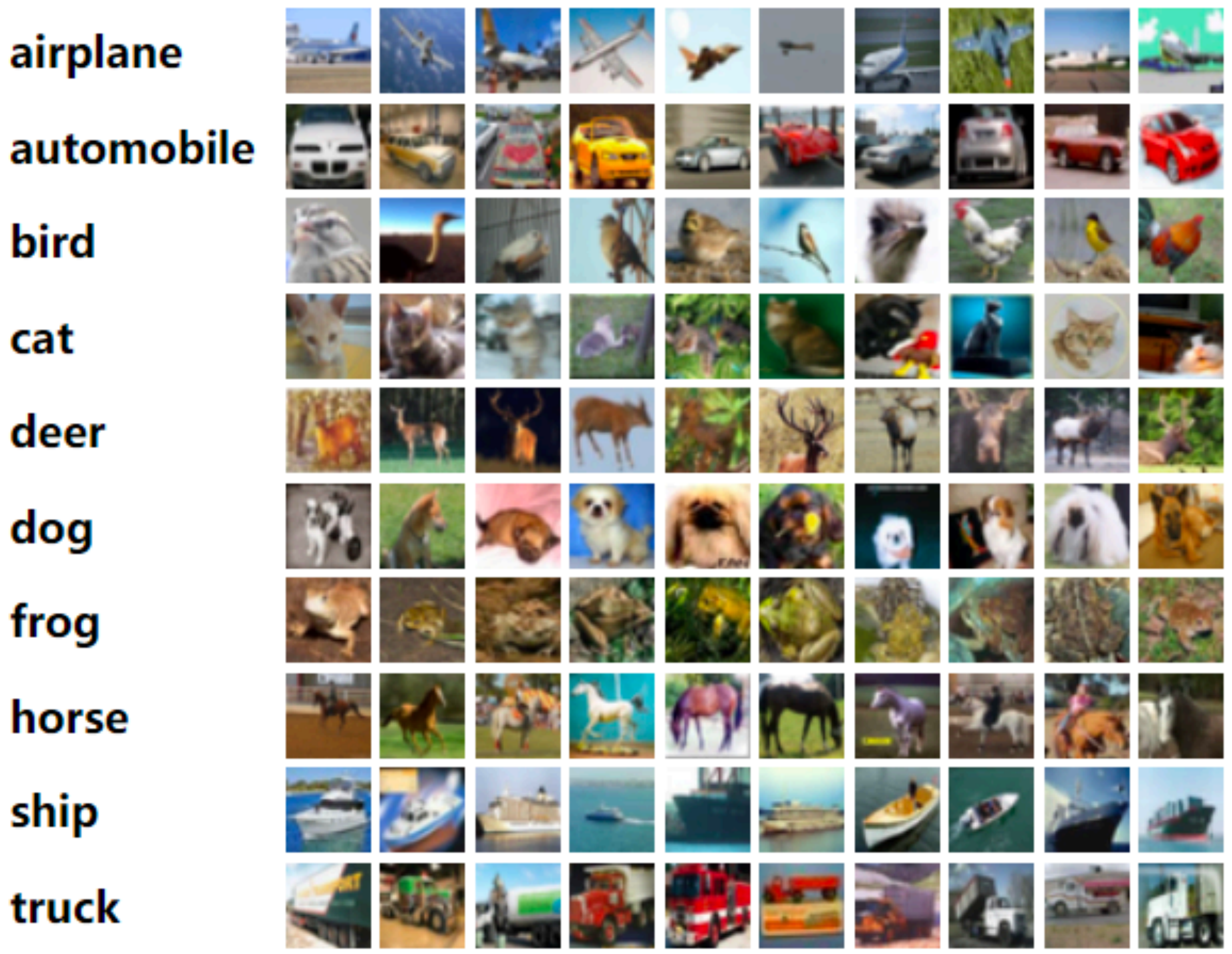


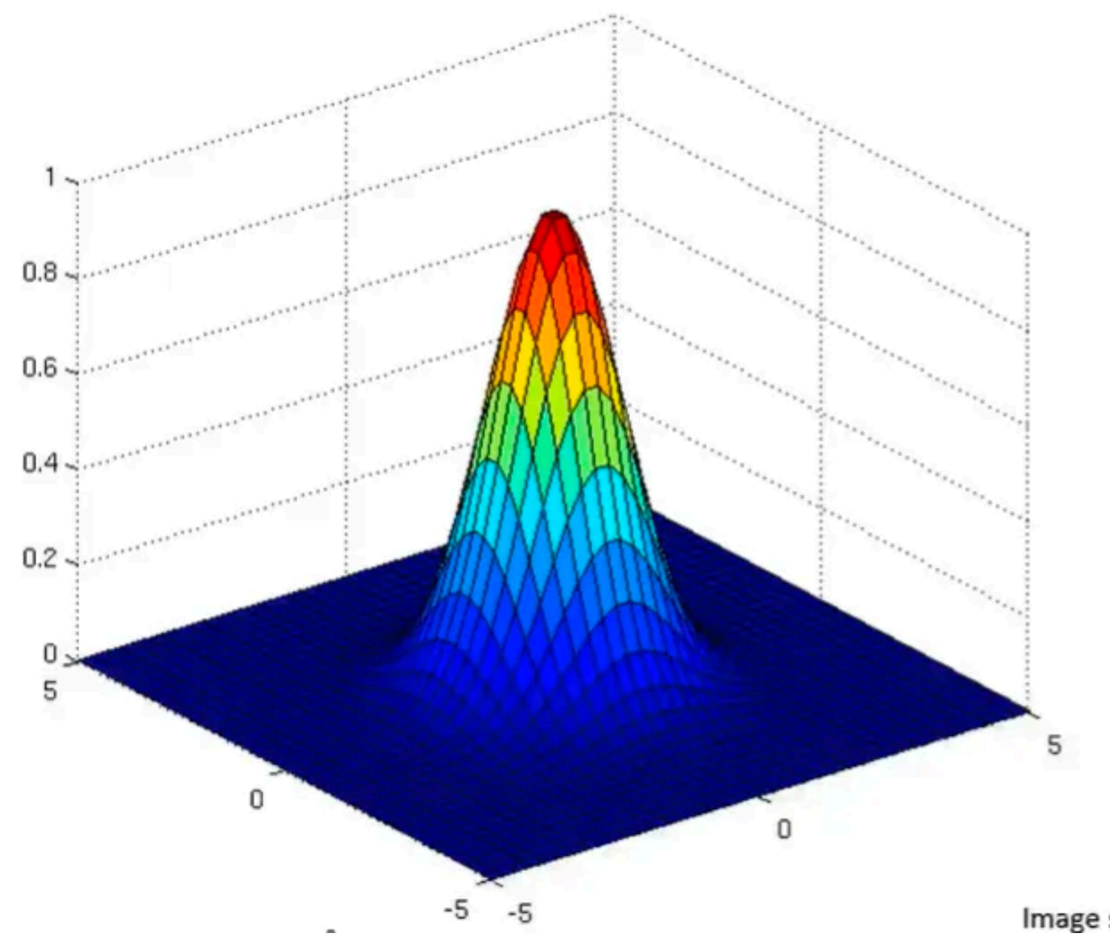
Image Inputs

Neural networks scale to high-dimensional inputs

Neural network



Gaussian process



$$K(\vec{x}, \vec{l}^{i}) = e^{-\frac{\|\vec{x}-\vec{l}^{i}\|^{2}}{2\sigma^{2}}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

Large Data Sets

Neural networks scale to large data sets

Large Data Sets

Neural networks scale to **large data sets**

Neural network

✅

Large Data Sets

Neural networks scale to large data sets

Neural network

✅

Gaussian process

❌

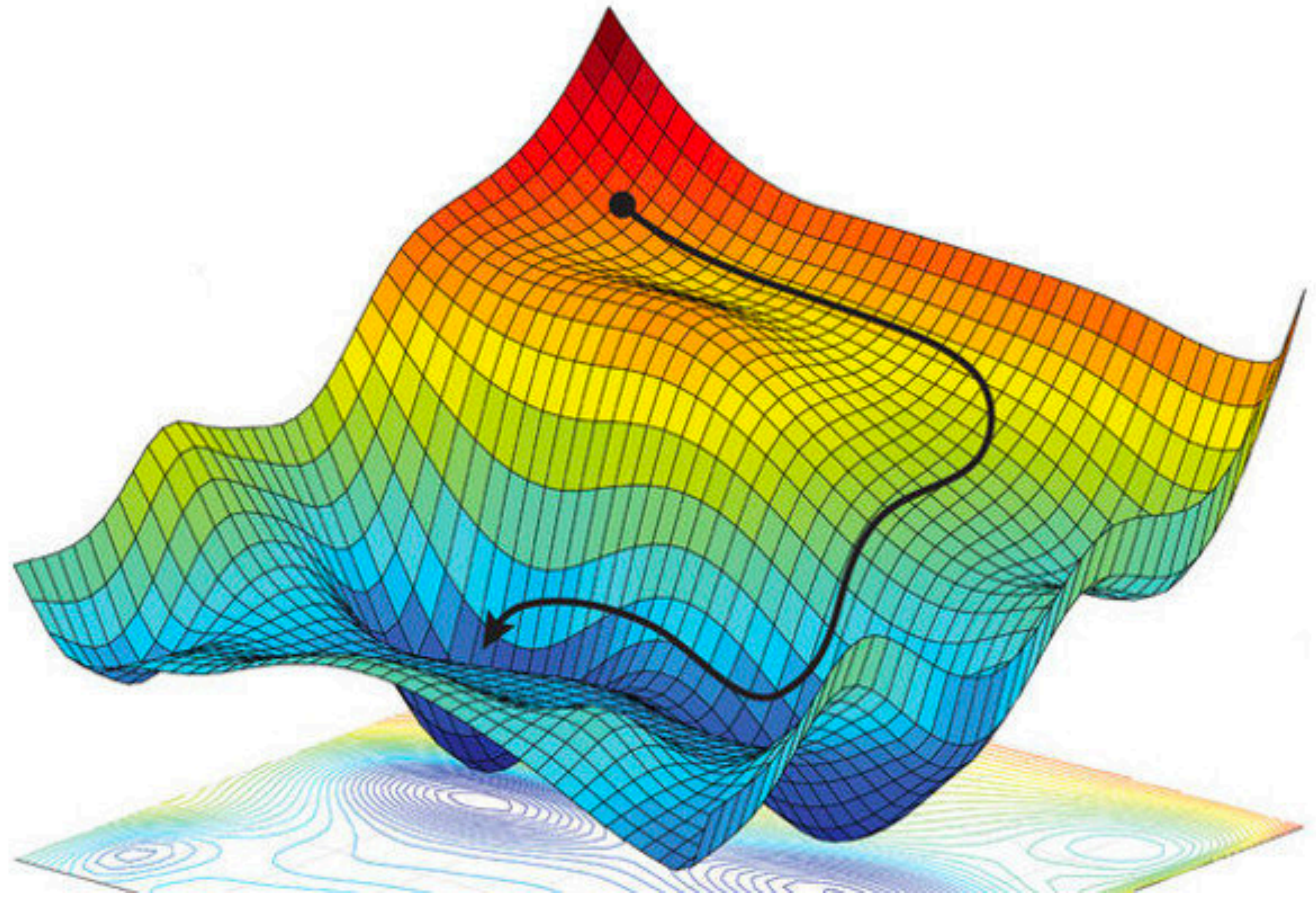
Incorporating New Data

Neural networks require costly retraining

Incorporating New Data

Neural networks require costly retraining

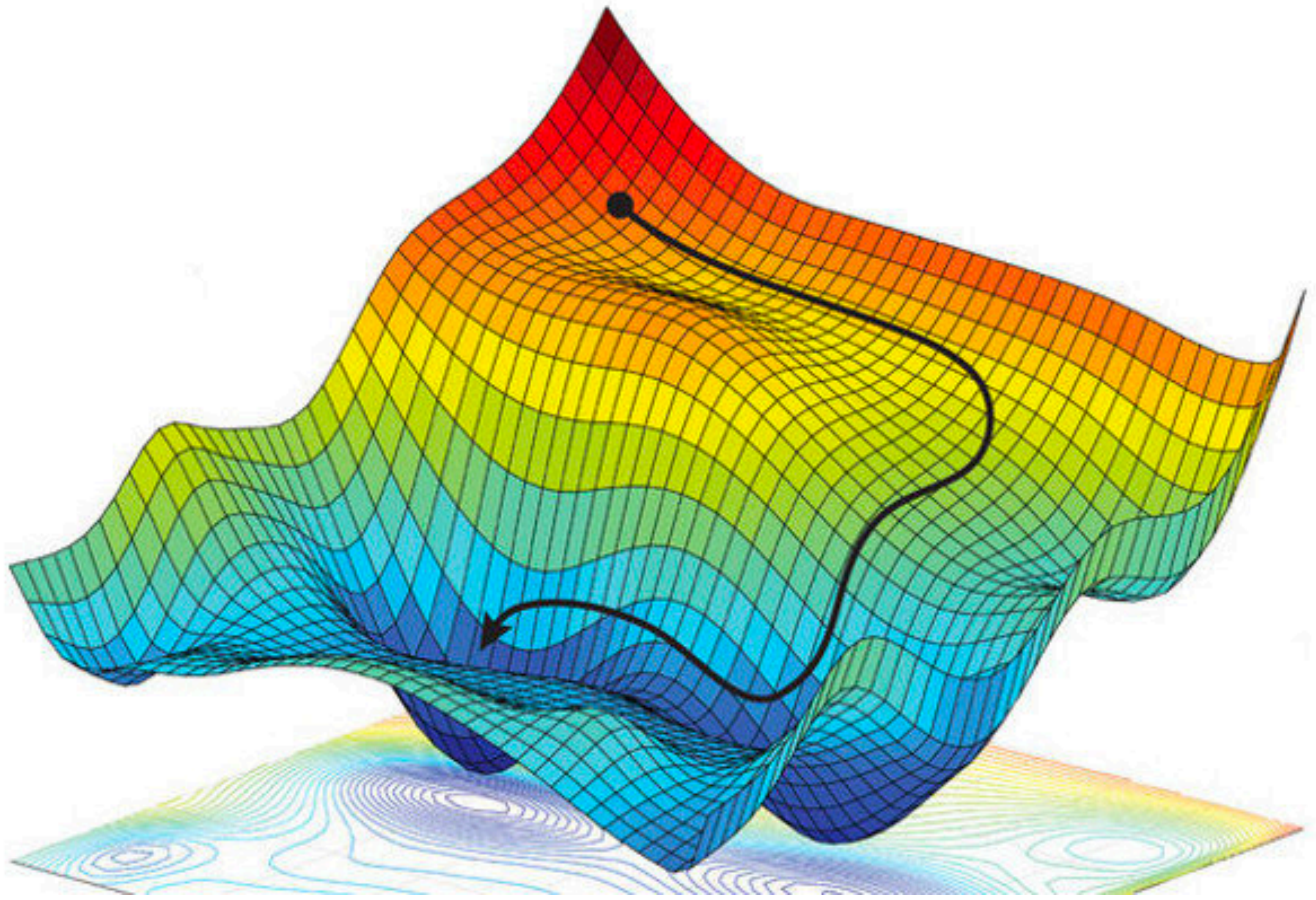
Neural network



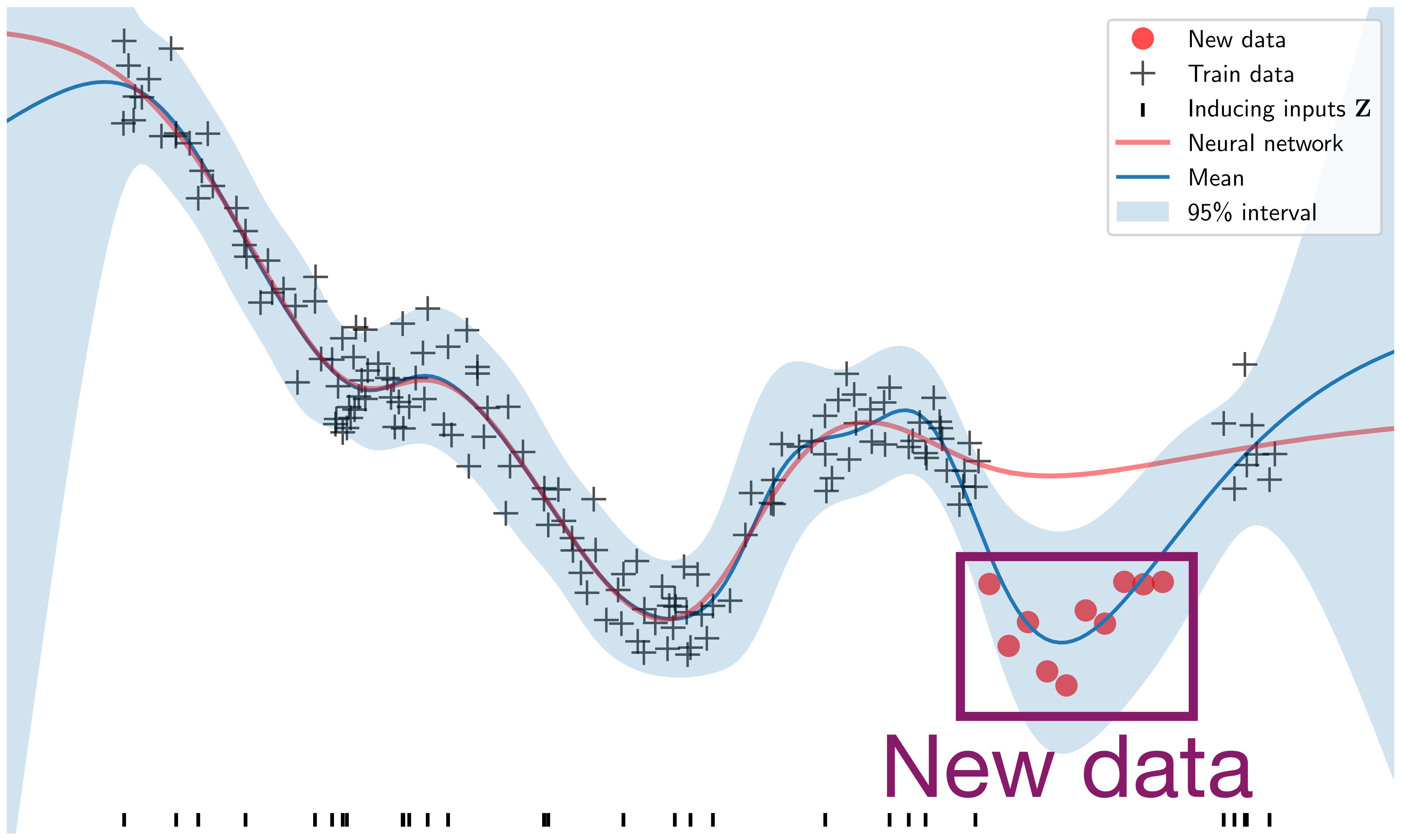
Incorporating New Data

Neural networks require costly retraining

Neural network



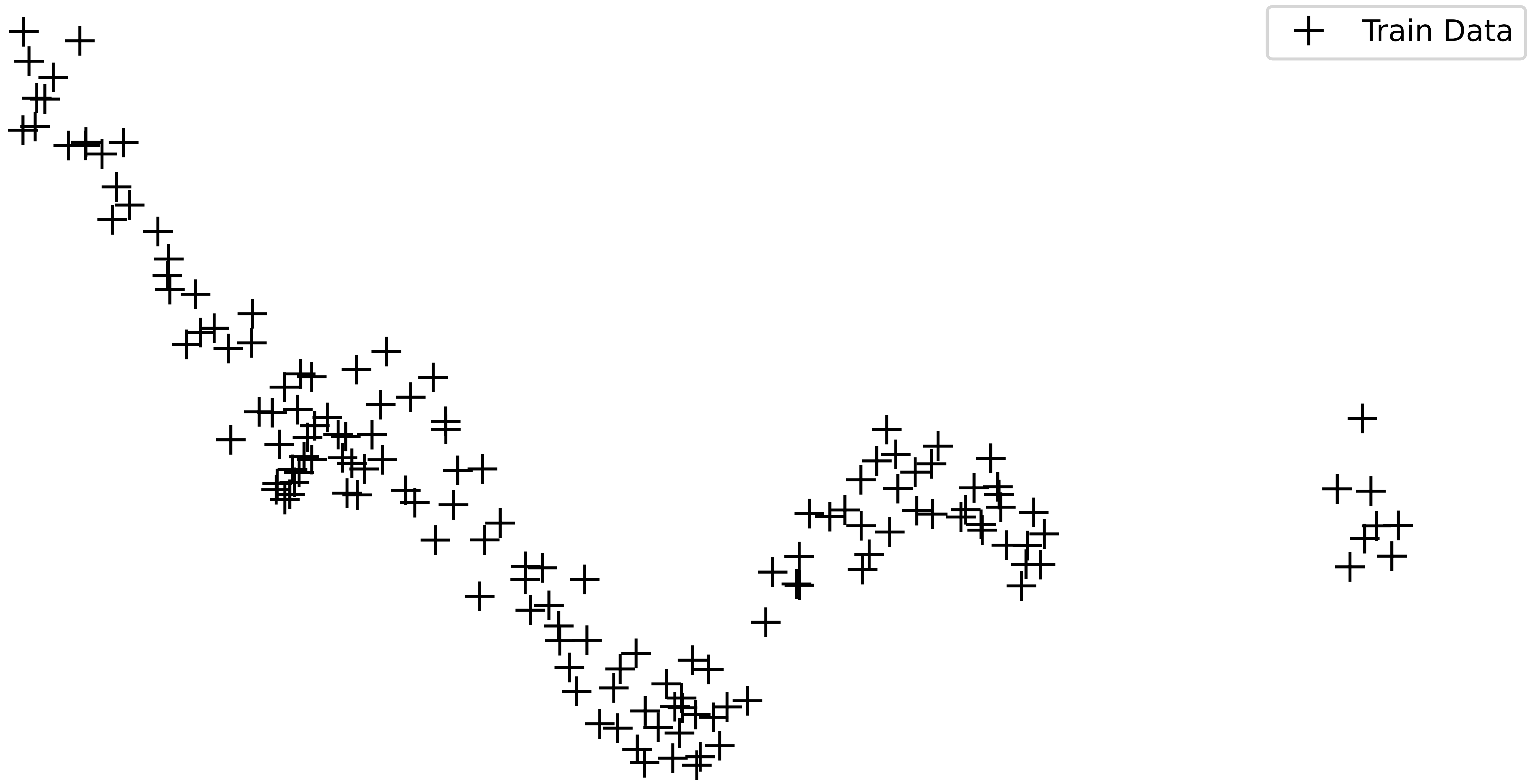
Gaussian process



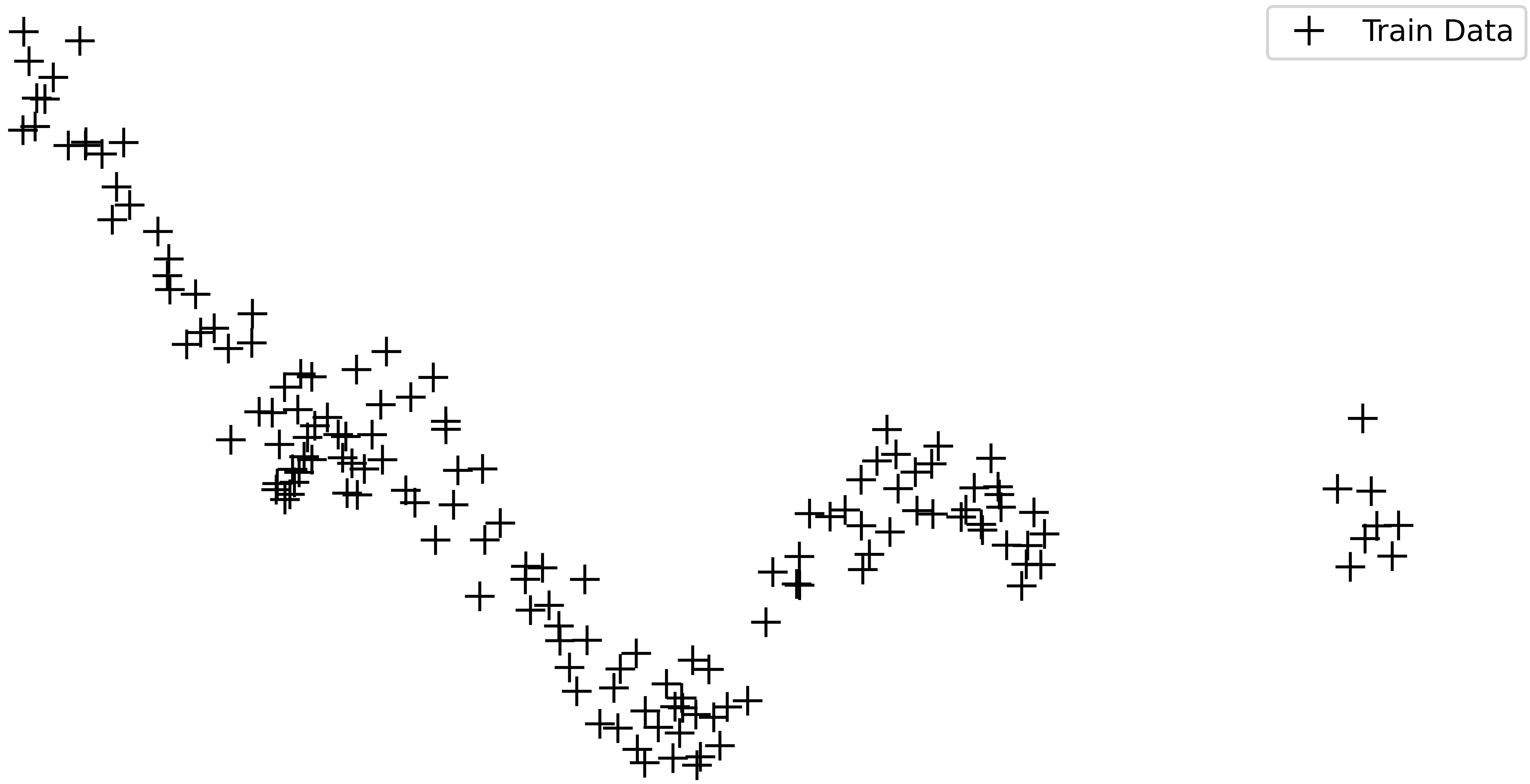
SFR: Sparse Function-space Representation

Combines benefits of neural networks and Gaussian processes

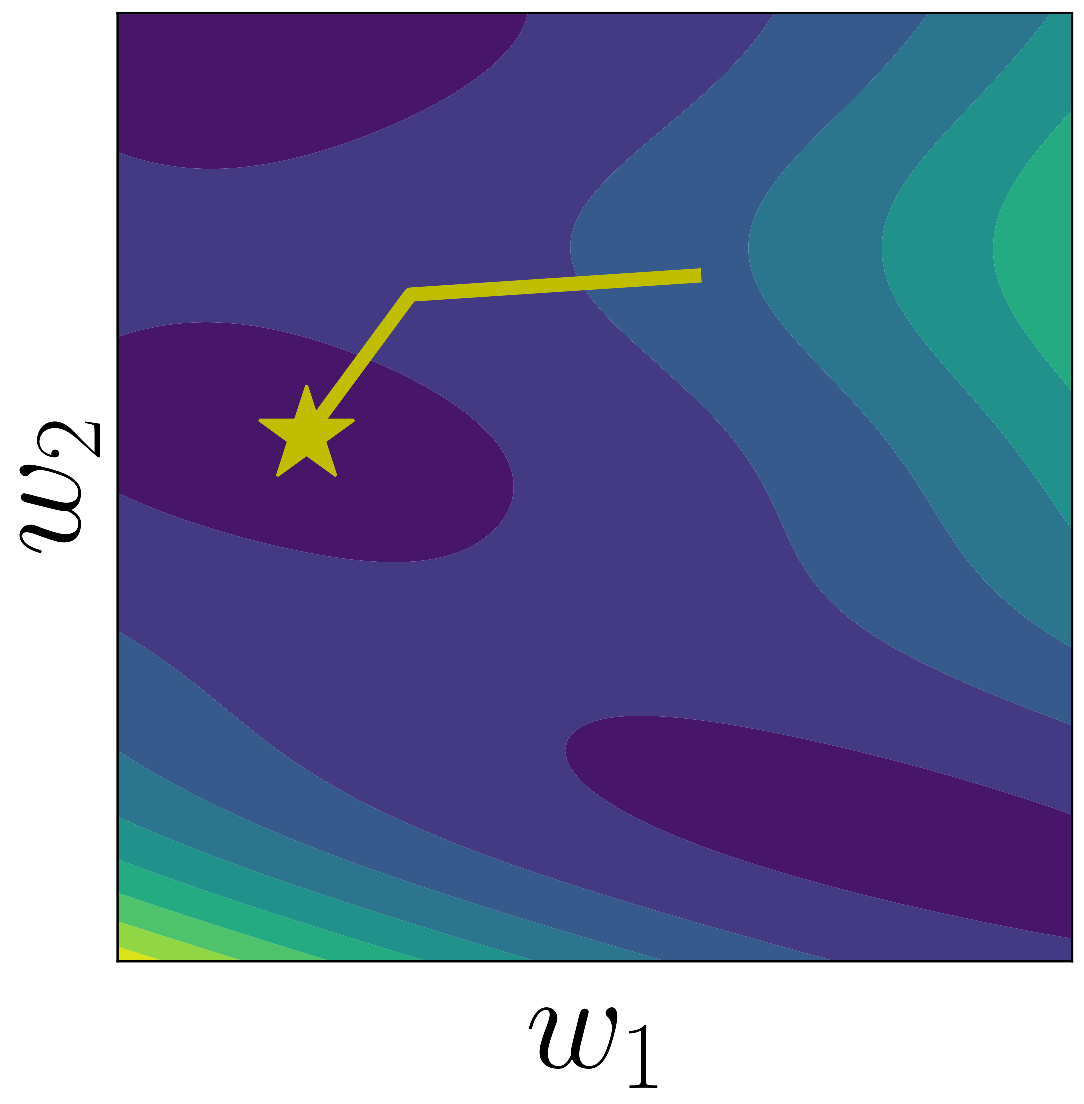
SFR: Sparse Function-space Representation



SFR: Sparse Function-space Representation

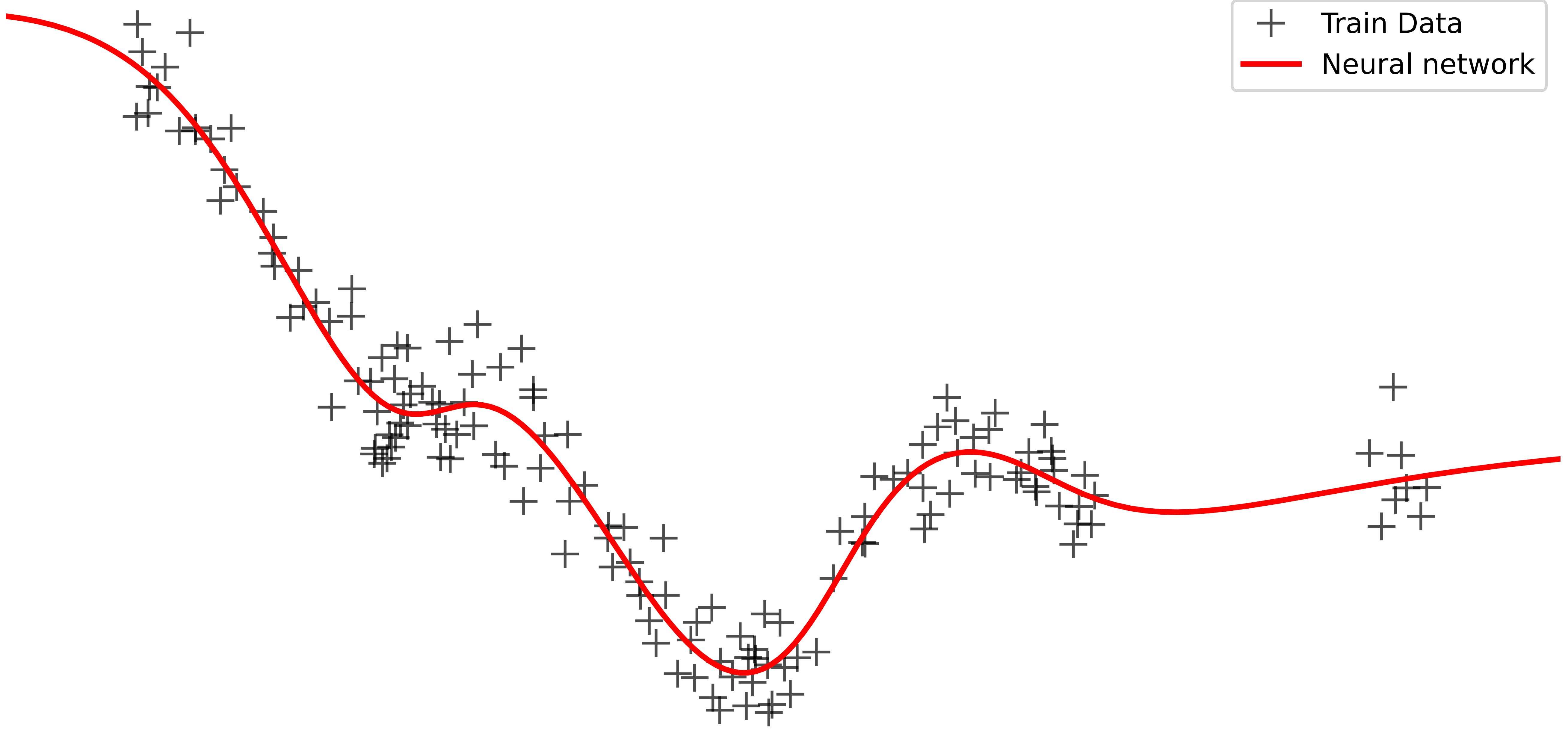


Weight-space NN training

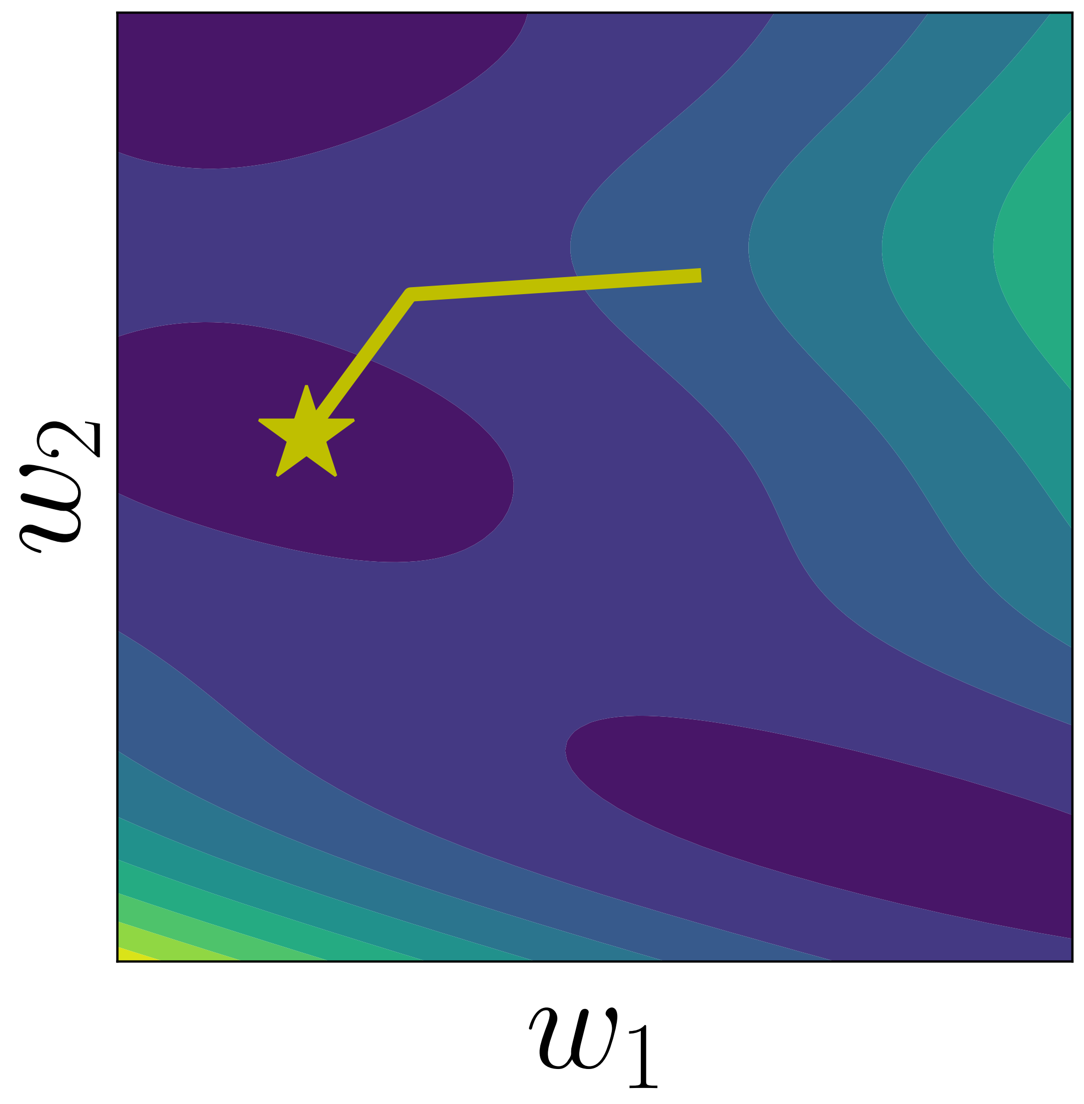


$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i|f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathcal{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

SFR: Sparse Function-space Representation



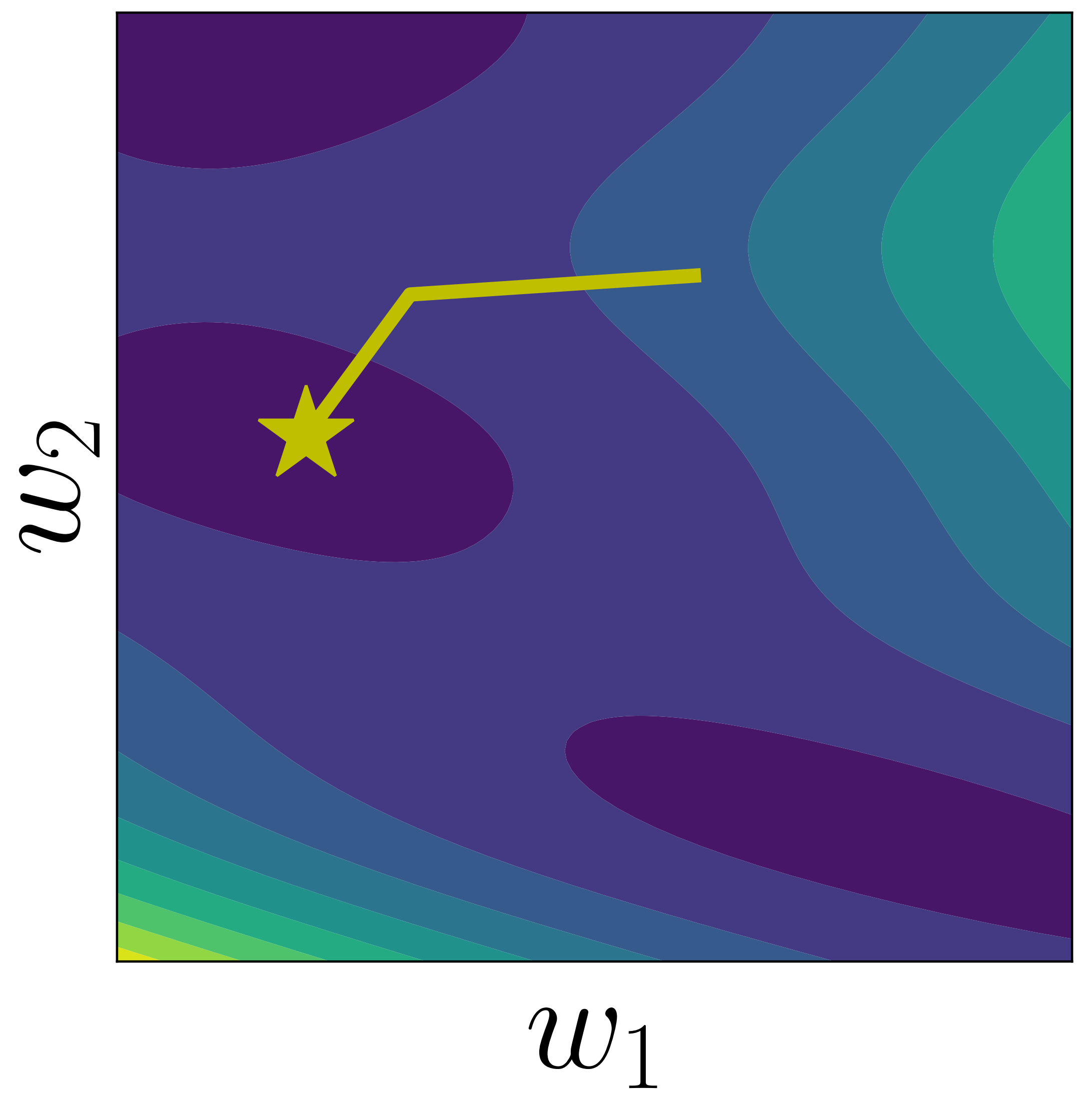
Weight-space NN training



$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i | f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathscr{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

SFR: Sparse Function-space Representation

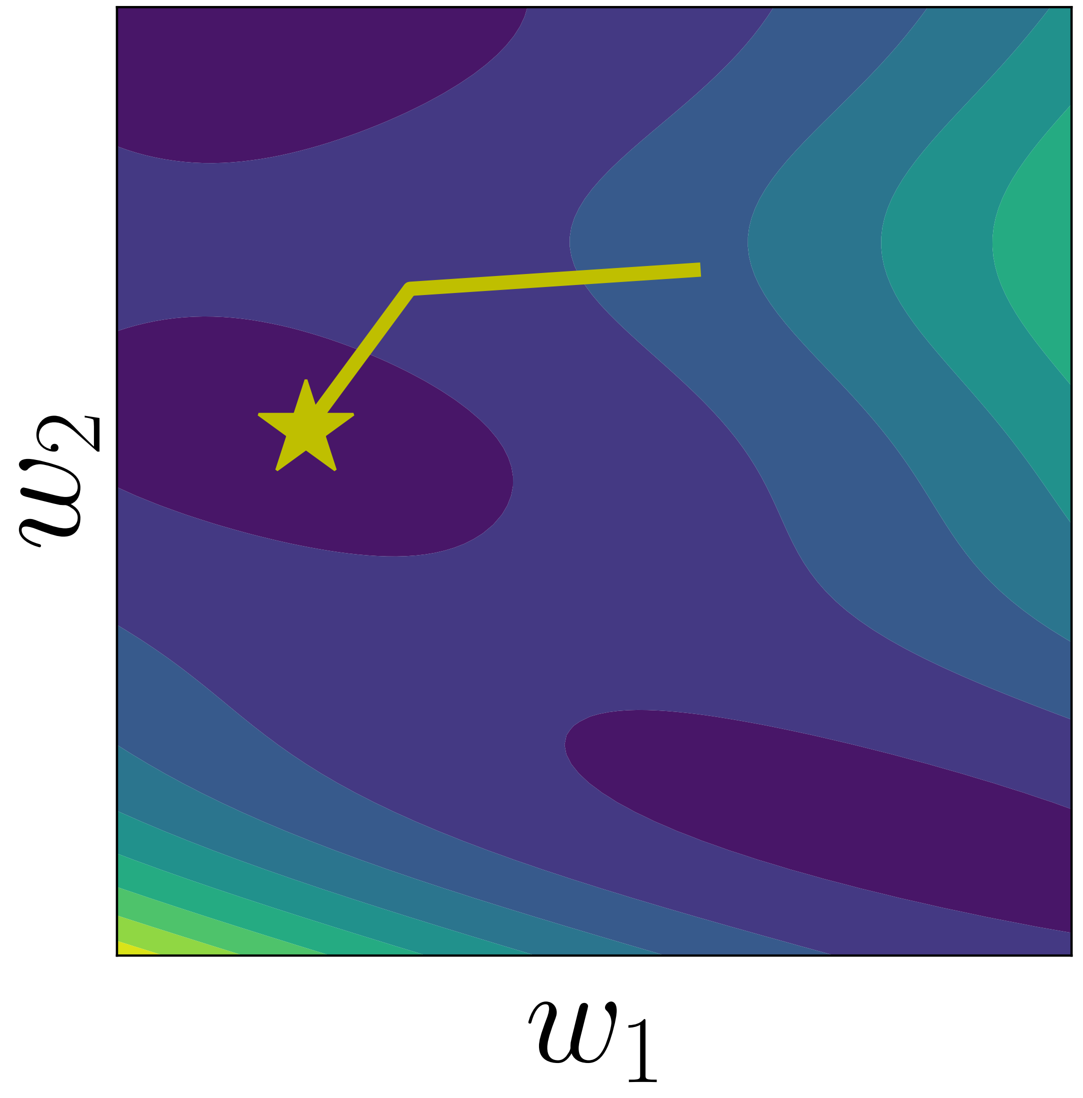
Weight-space NN training



$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i | f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathscr{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

SFR: Sparse Function-space Representation

Weight-space NN training

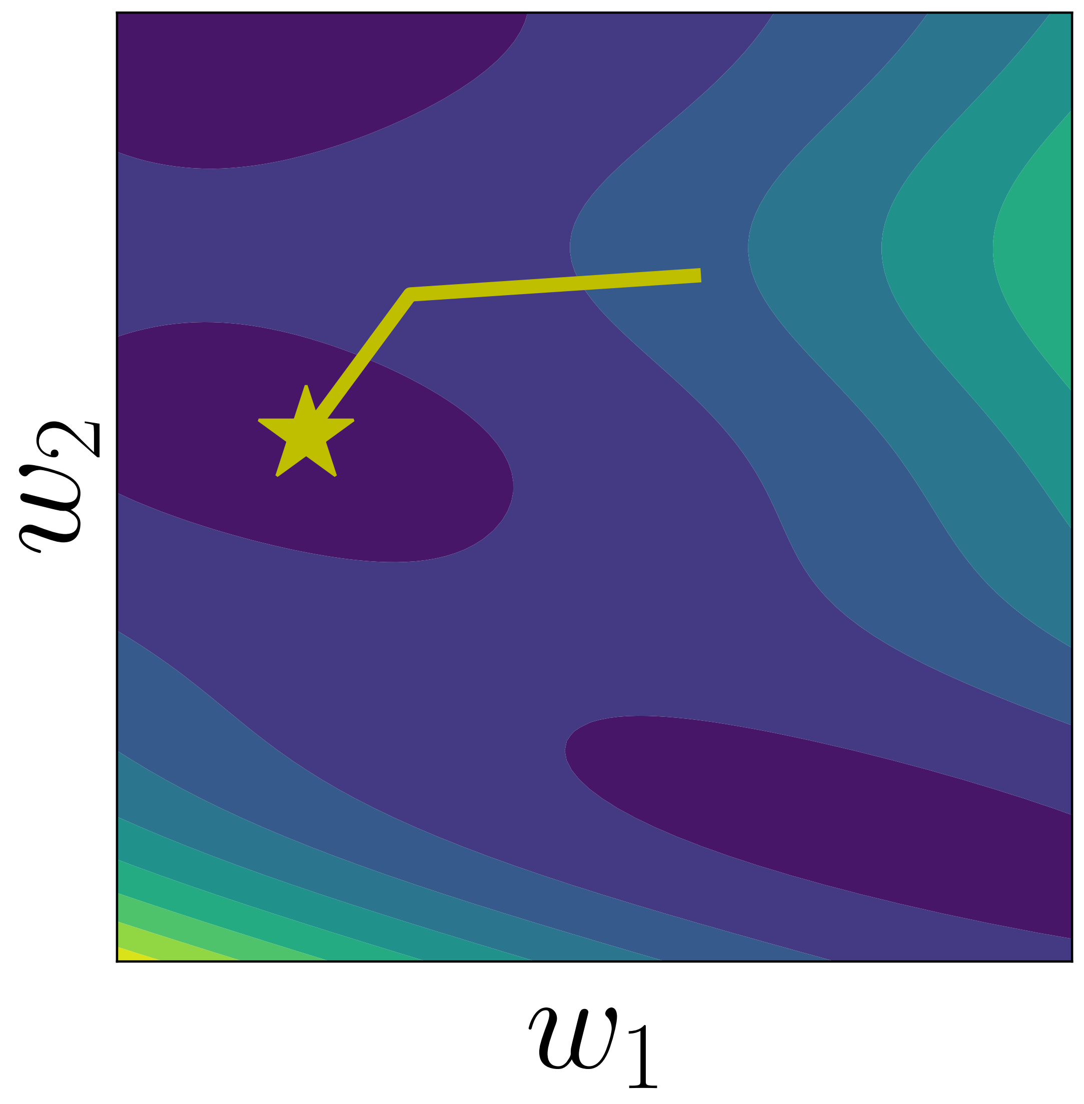


Linearise to get kernel formulation

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i|f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathscr{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

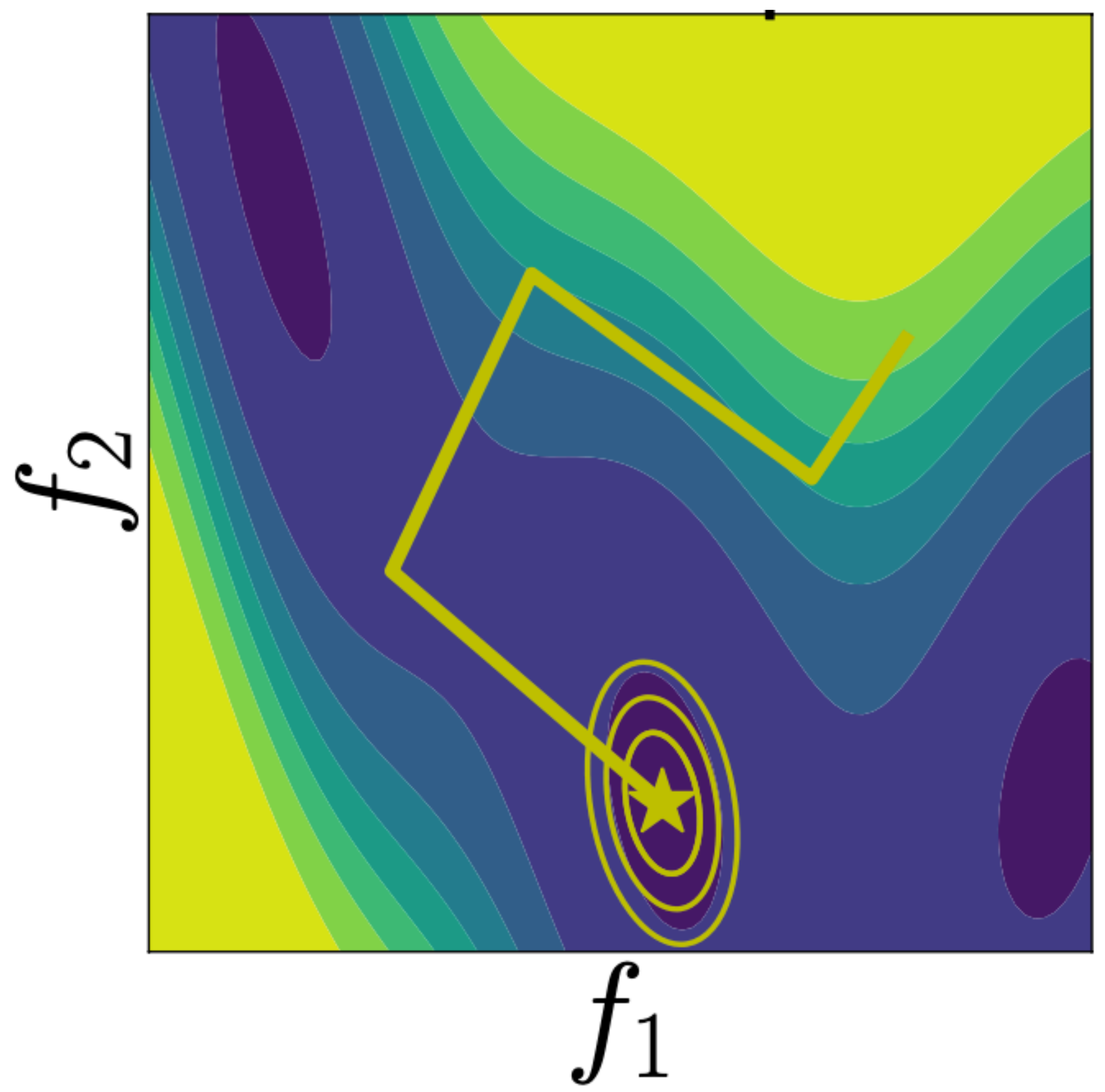
SFR: Sparse Function-space Representation

Weight-space NN training



Linearise to get kernel formulation

Function-space inference

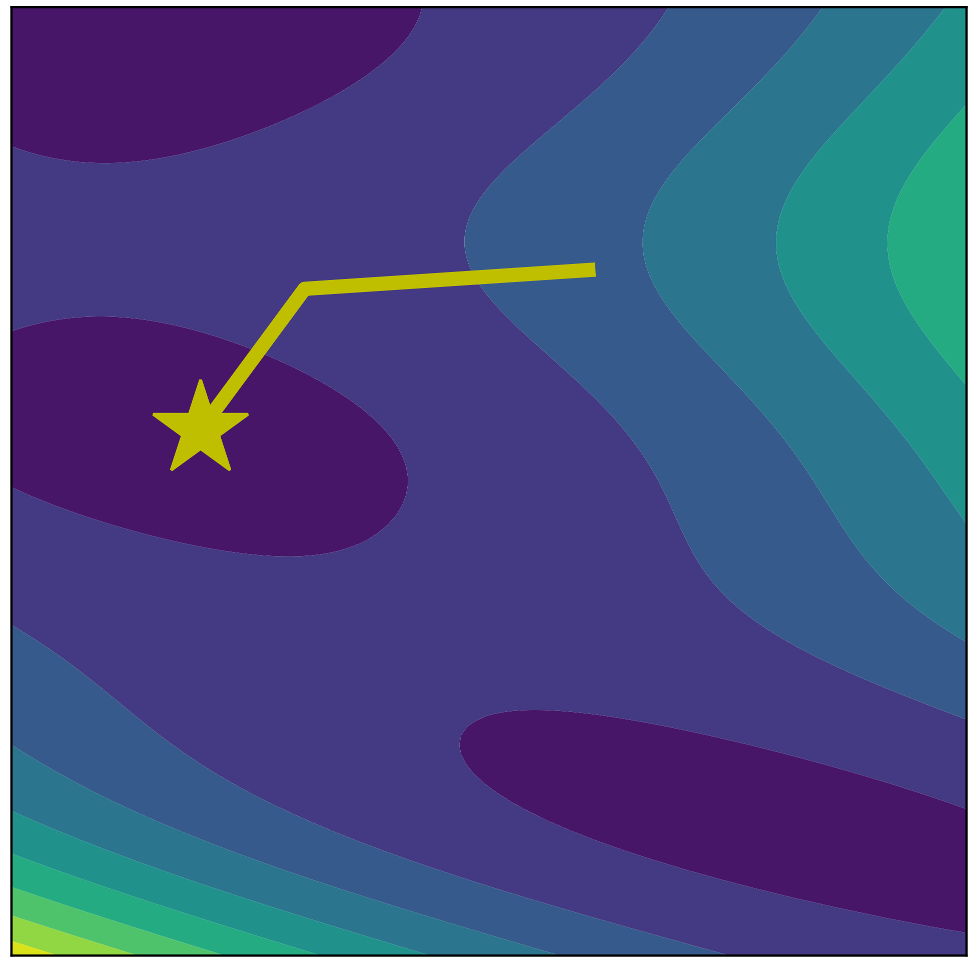


$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i|f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathcal{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \sum_{i=1}^{N} -\log p(y_i \mid f_{\mathbf{w}}(\mathbf{x}_i)) - \log p(\mathbf{f})$$

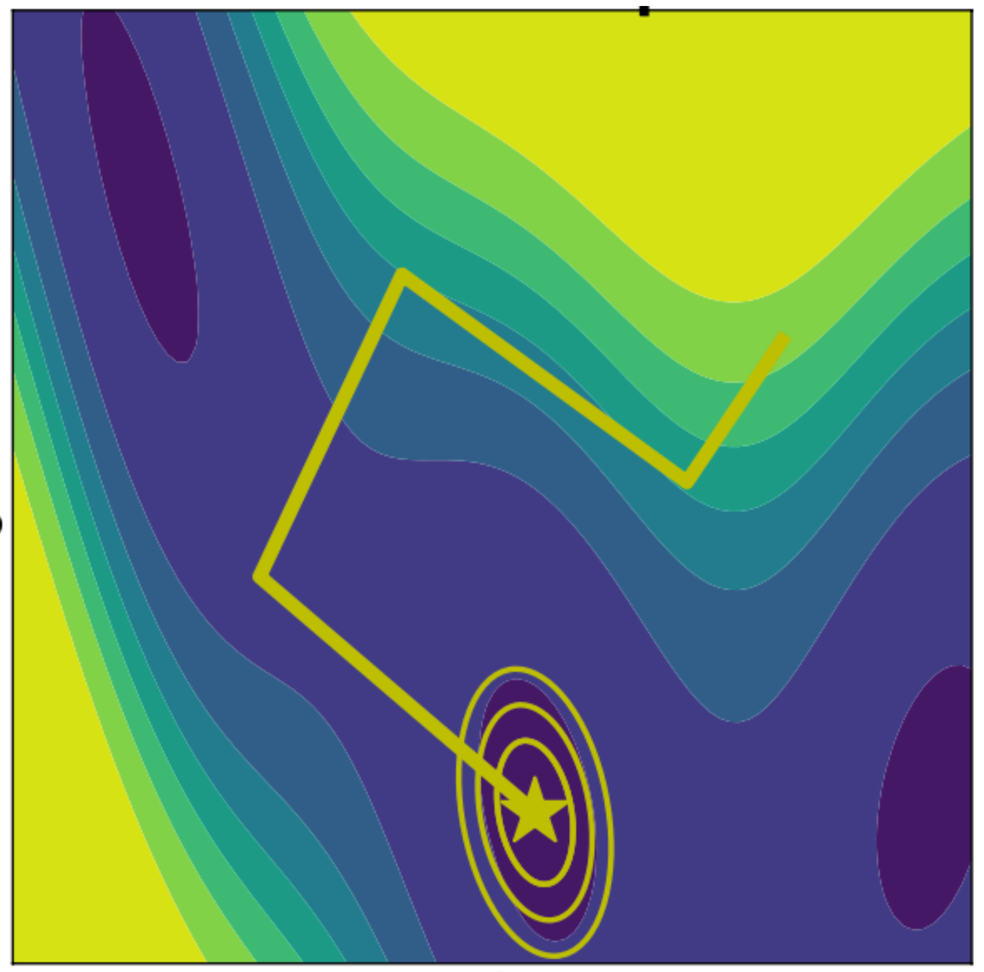
SFR: Sparse Function-space Representation

Weight-space NN training



Linearise to get kernel formulation

Function-space inference



$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{N} \underbrace{l(f_{\mathbf{w}}(\mathbf{x}_i), y_i)}_{-\log p(y_i|f_{\mathbf{w}}(\mathbf{x}_i))} + \underbrace{\mathcal{R}(\mathbf{w})}_{-\log p(\mathbf{w})}$$

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \sum_{i=1}^{N} -\log p(y_i \mid f_{\mathbf{w}}(\mathbf{x}_i)) - \log p(\mathbf{f})$$

$$\boxed{p(\mathbf{f} \mid \mathcal{D}) \approx q(\mathbf{f}) = \mathcal{N}(\mathbf{f} \mid \mathbf{f}^*, \mathbf{S}_{\mathbf{f}})}$$

Function-space Laplace approx.

FCAI

fcai.fi

SFR: Sparse Function-space Representation

$$\boxed{p(\mathbf{f} \mid \mathscr{D}) \approx q(\mathbf{f}) = \mathscr{N}(\mathbf{f} \mid \mathbf{f}^{*}, \mathbf{S}_{\mathbf{f}})}$$

Function-space Laplace approx.

FCAI

SFR: Sparse Function-space Representation

$$\boxed{p(\mathbf{f} \mid \mathcal{D}) \approx q(\mathbf{f}) = \mathcal{N}(\mathbf{f} \mid \mathbf{f}^*, \mathbf{S}_\mathbf{f})}$$

Function-space Laplace approx.

SFR: Sparse Function-space Representation

$$\boxed{p(\mathbf{f} \mid \mathcal{D}) \approx q(\mathbf{f}) = \mathcal{N}(\mathbf{f} \mid \mathbf{f}^{*}, \mathbf{S}_{\mathbf{f}})}$$

Function-space Laplace approx.

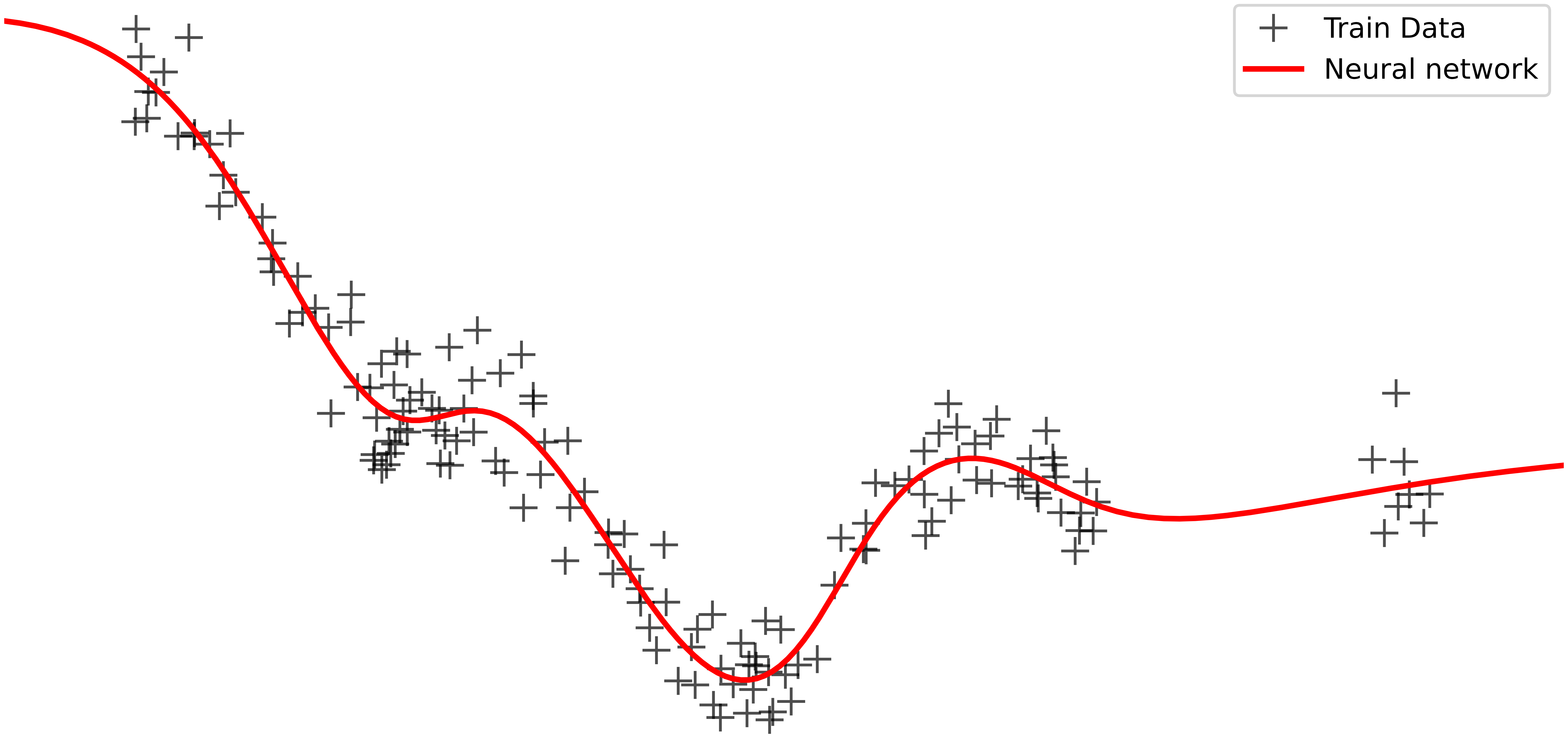
- ⚠️ **GP predictive posterior is computationally expensive**

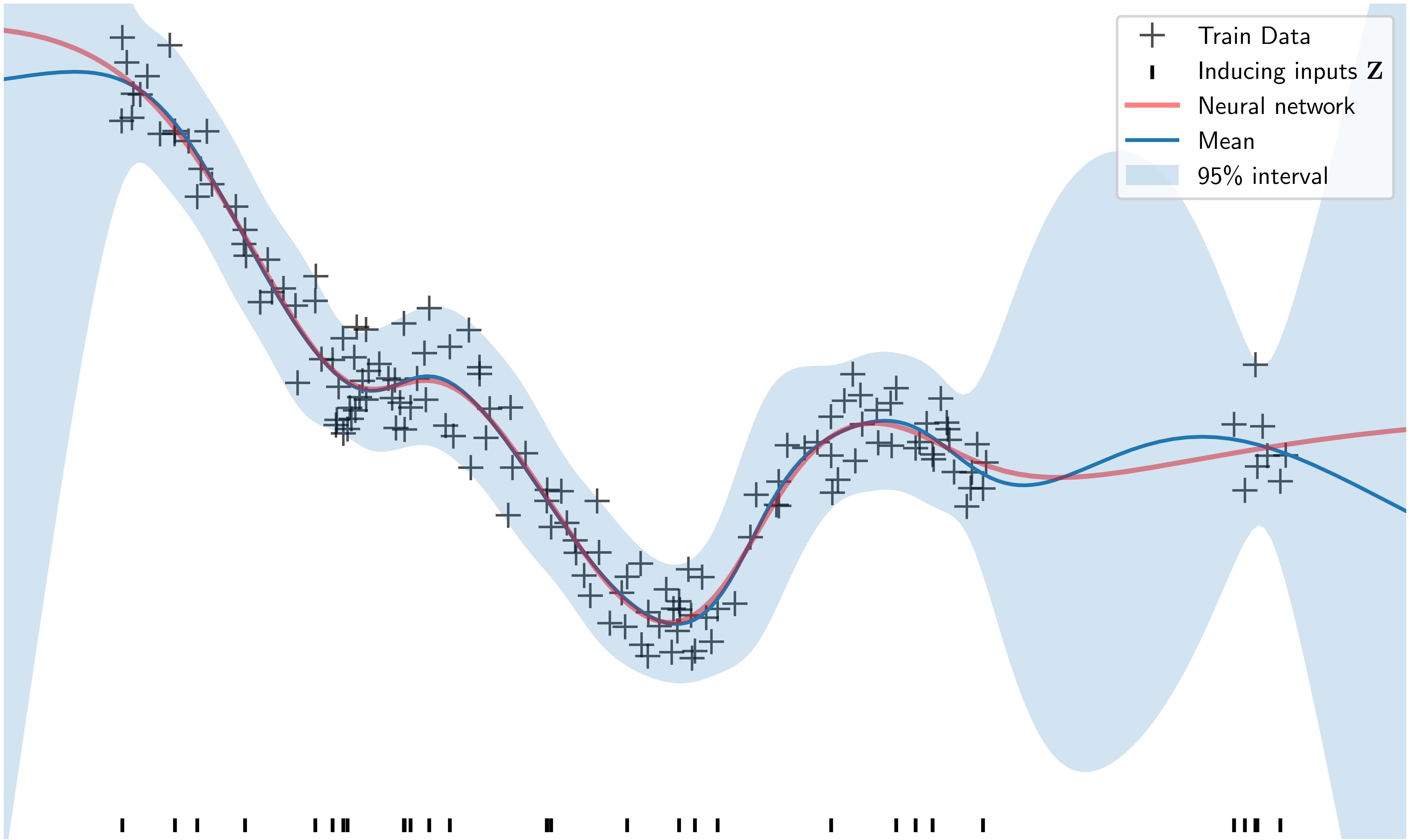
SFR: Sparse Function-space Representation

$$\boxed{p(\mathbf{f} \mid \mathcal{D}) \approx q(\mathbf{f}) = \mathcal{N}(\mathbf{f} \mid \mathbf{f}^{*}, \mathbf{S}_{\mathbf{f}})}$$

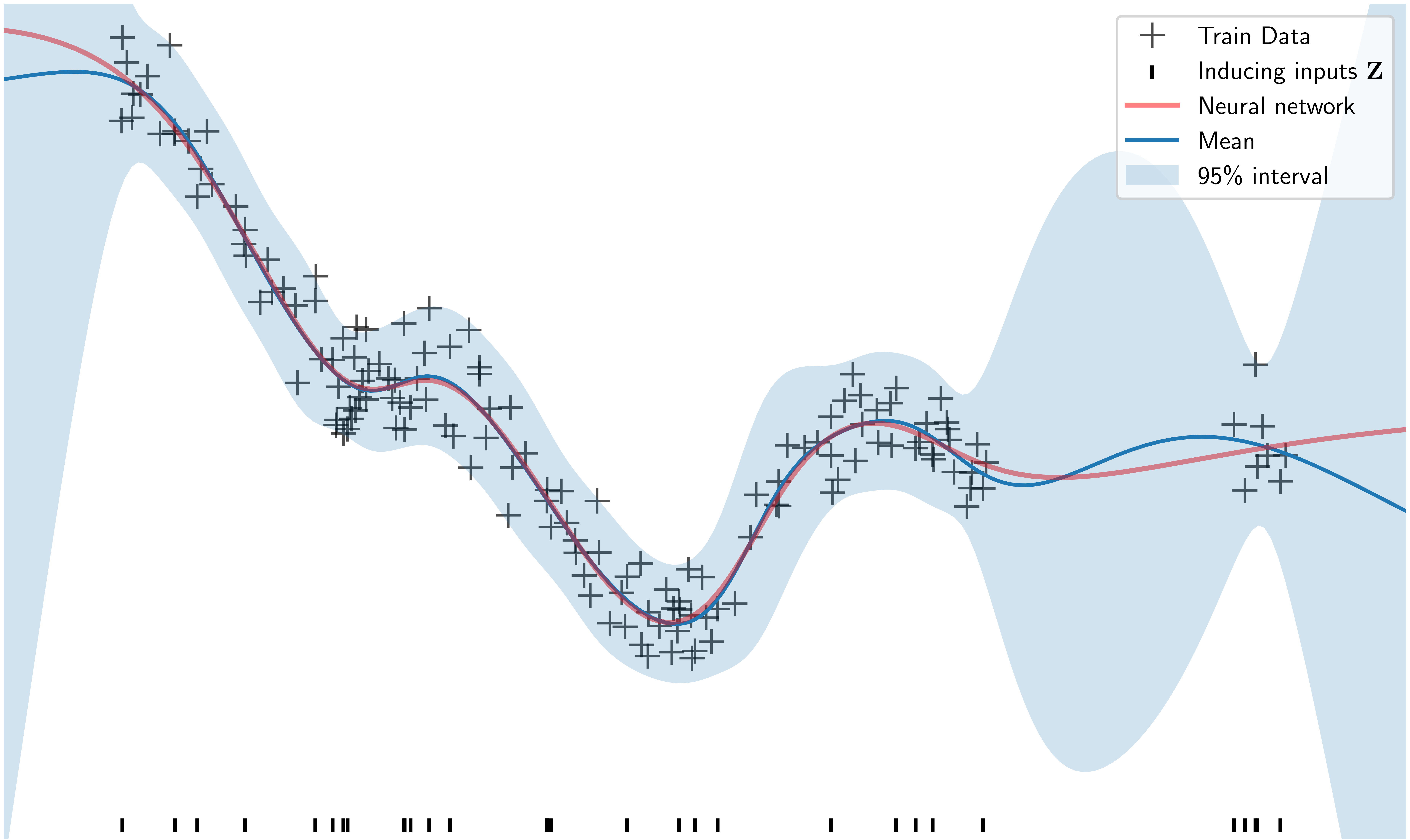
Function-space
Laplace approx.

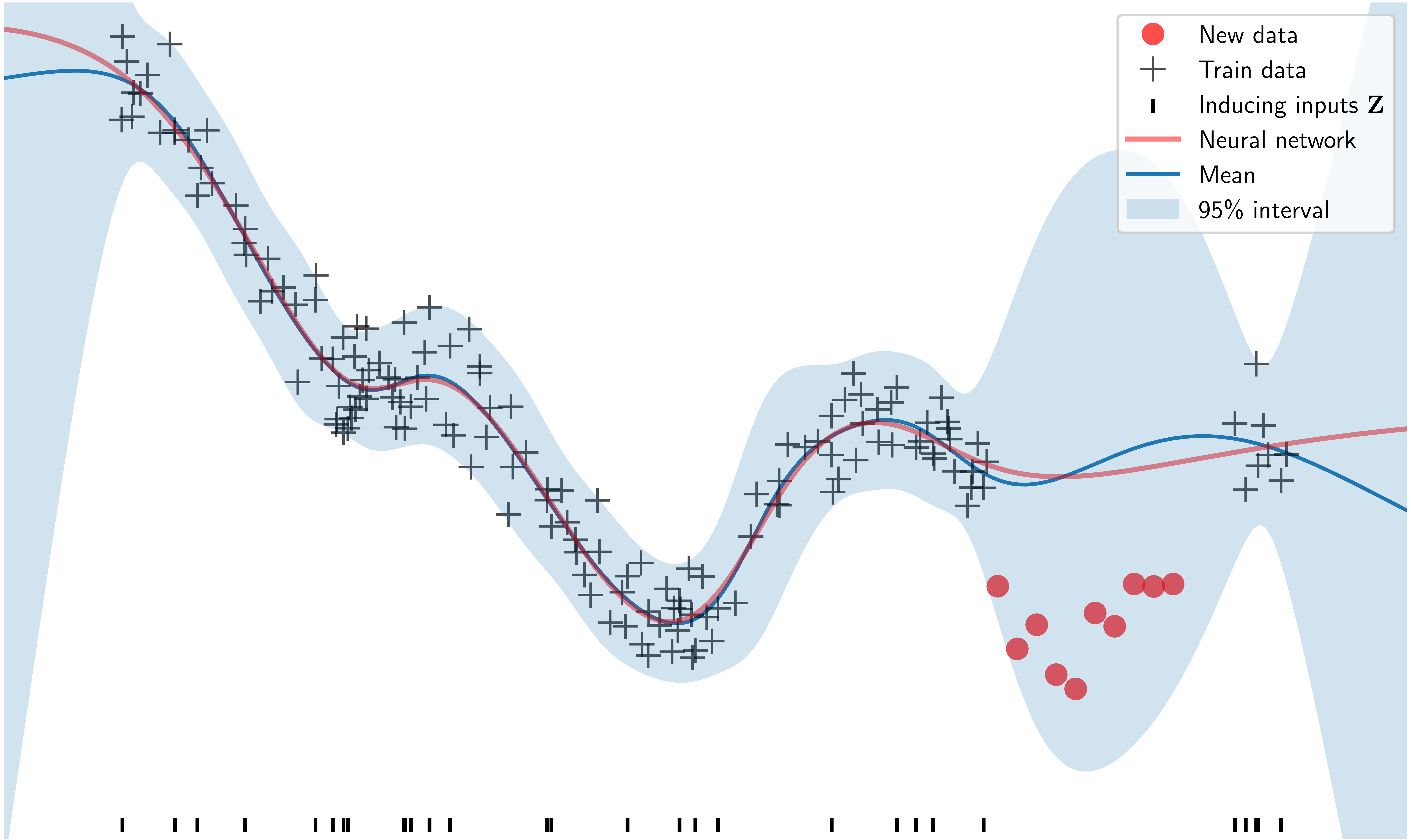
- ⚠️ GP predictive posterior is computationally expensive
- We sparsify the GP using a dual parameterization

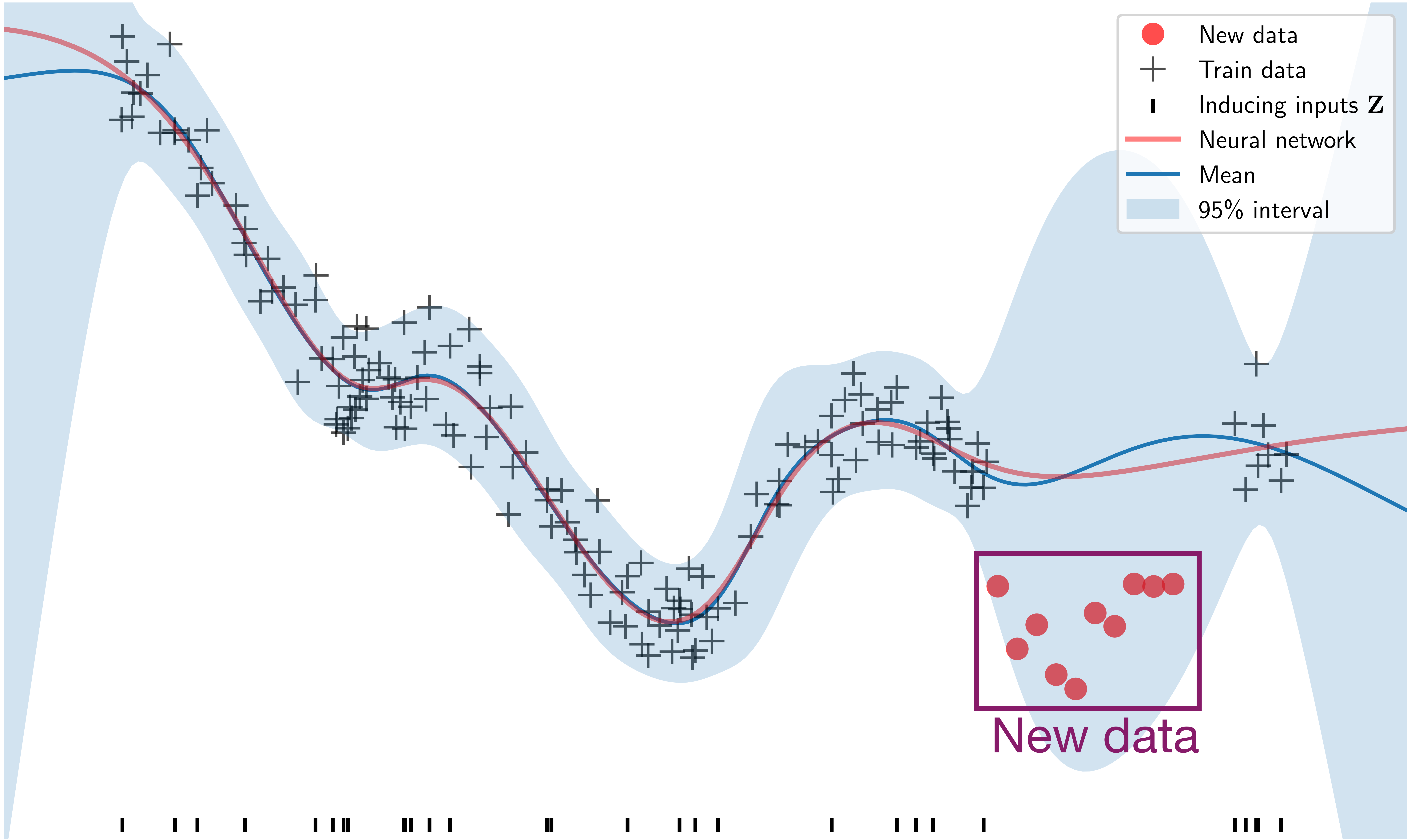

Train Data
Neural network

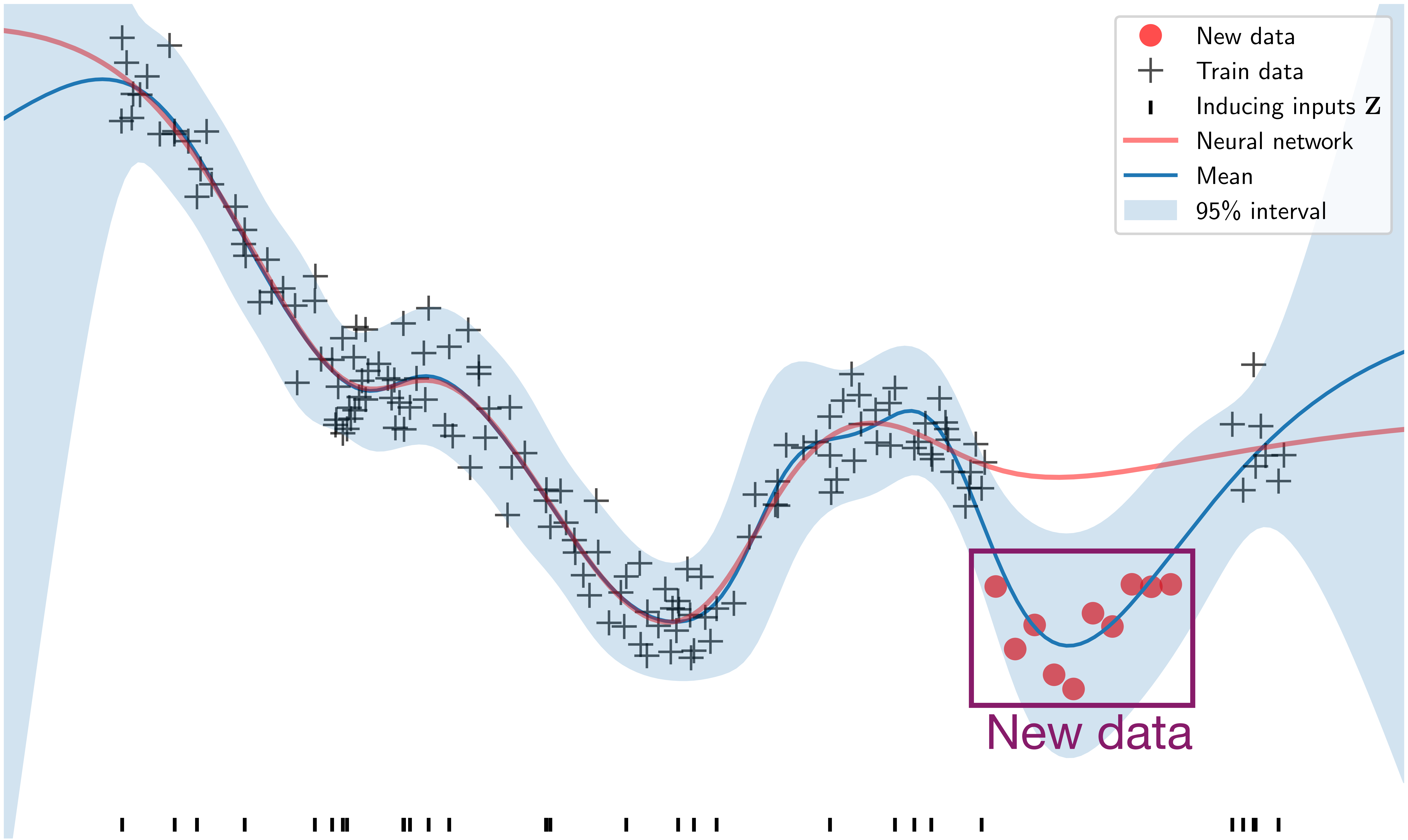

Train Data
Inducing inputs $\mathbf{Z}$
Neural network
Mean
95% interval

Uncertainty estimates ✅

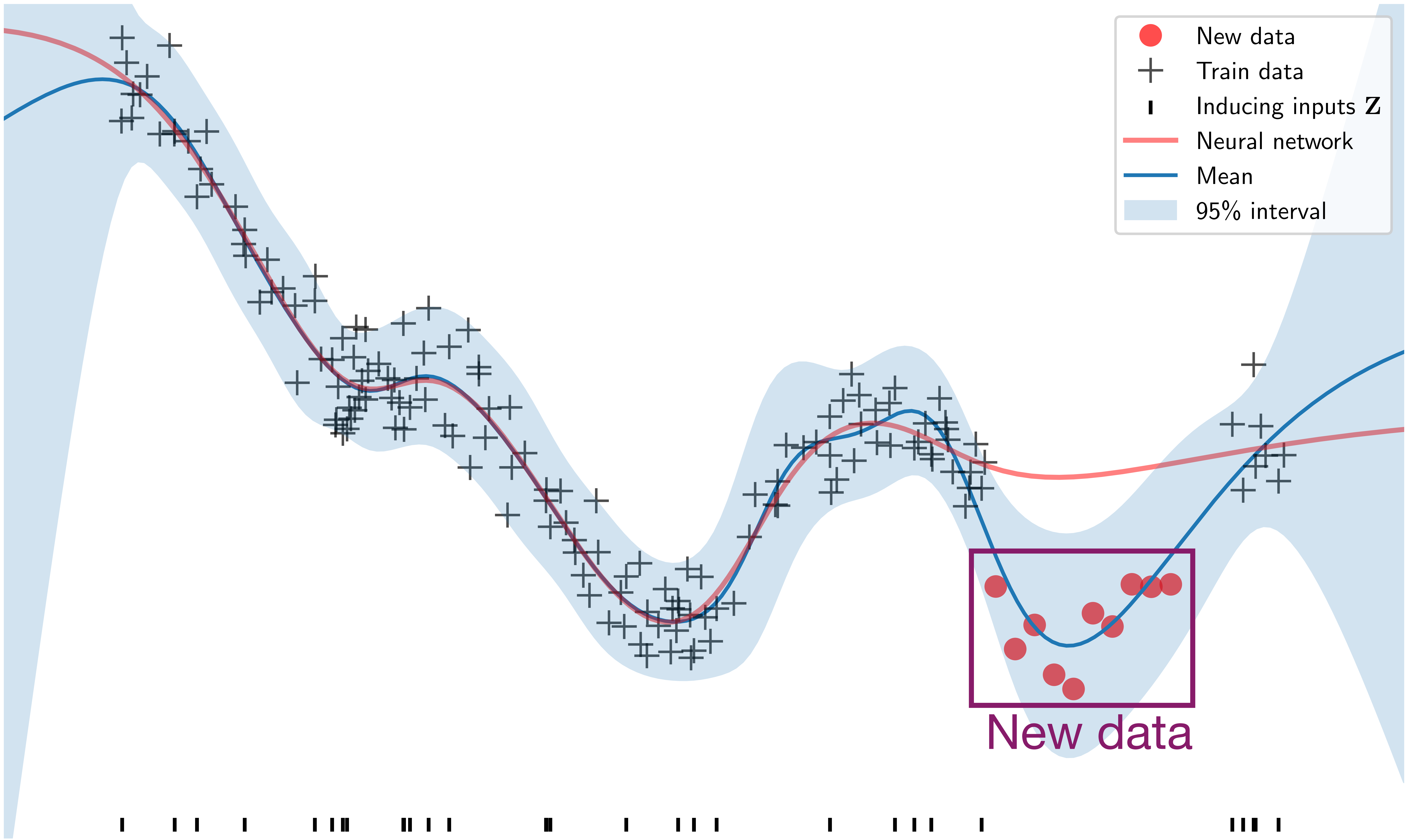



New data
Train data
Inducing inputs $\mathbf{Z}$
Neural network
Mean
95% interval


New data
Train data
Inducing inputs $\mathbf{Z}$
Neural network
Mean
95% interval
New data


New data
Train data
Inducing inputs $\mathbf{Z}$
Neural network
Mean
95% interval
New data

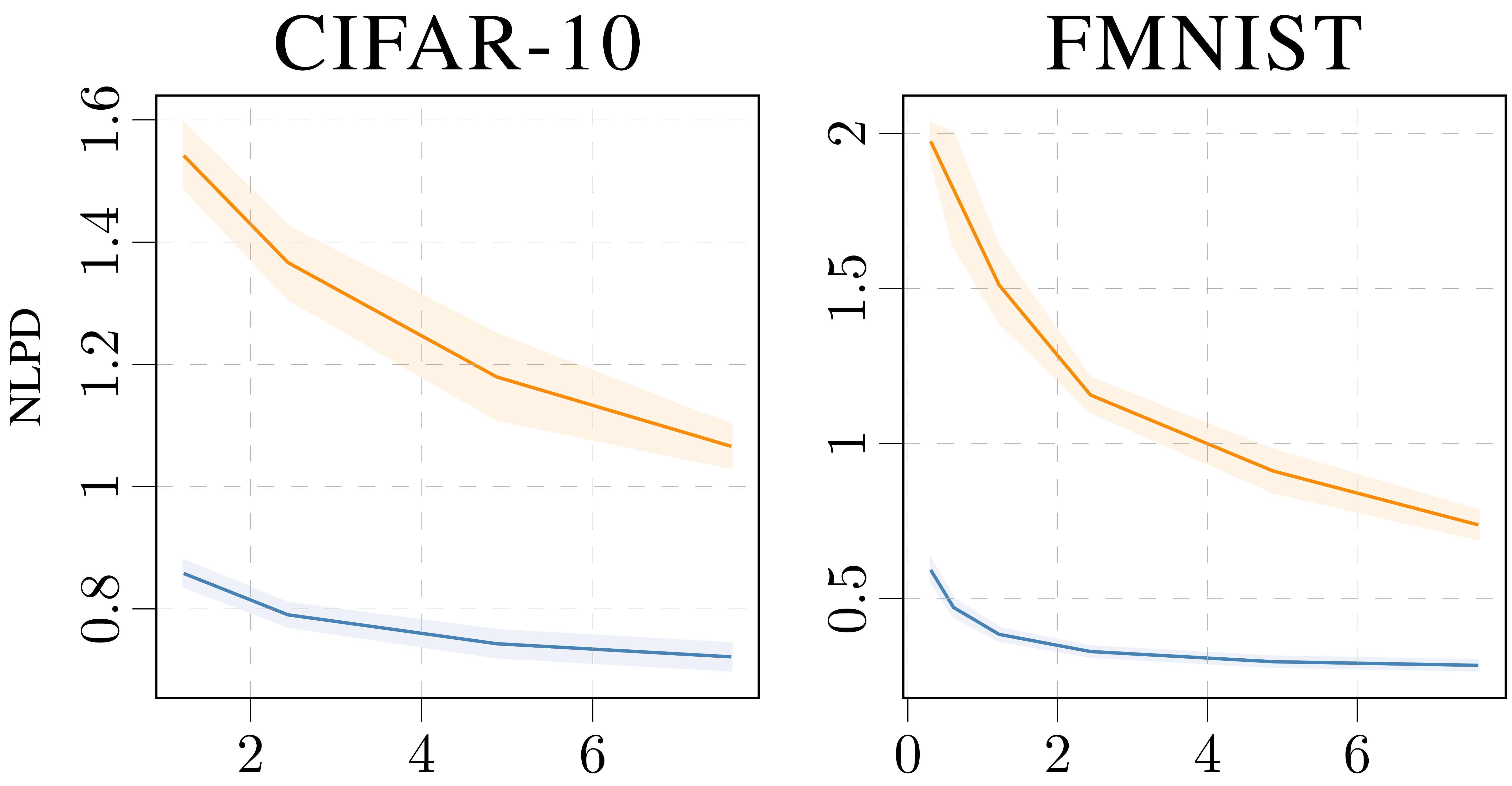
Incorporate new data fast ✅



FCAI

fcai.fi

Results: Effective Sparsification

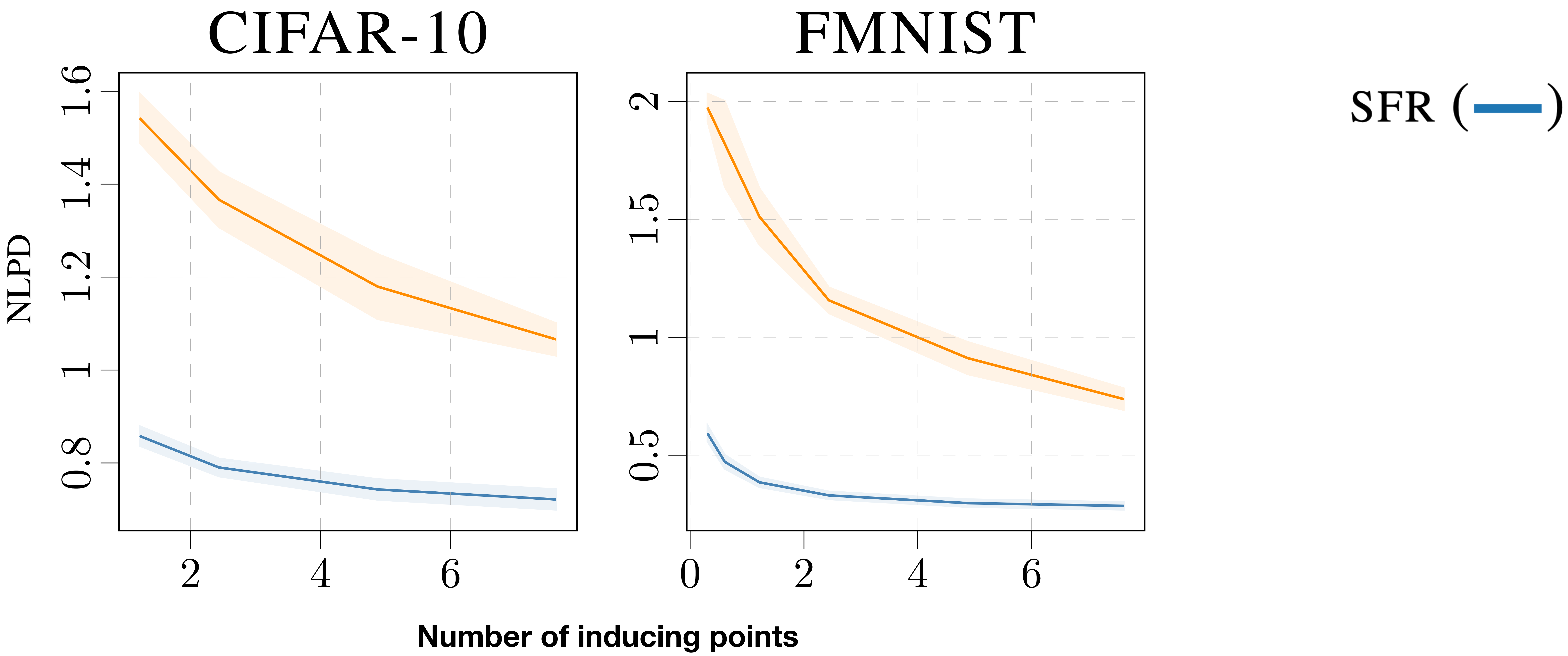


Number of inducing points

FCAI

fcai.fi

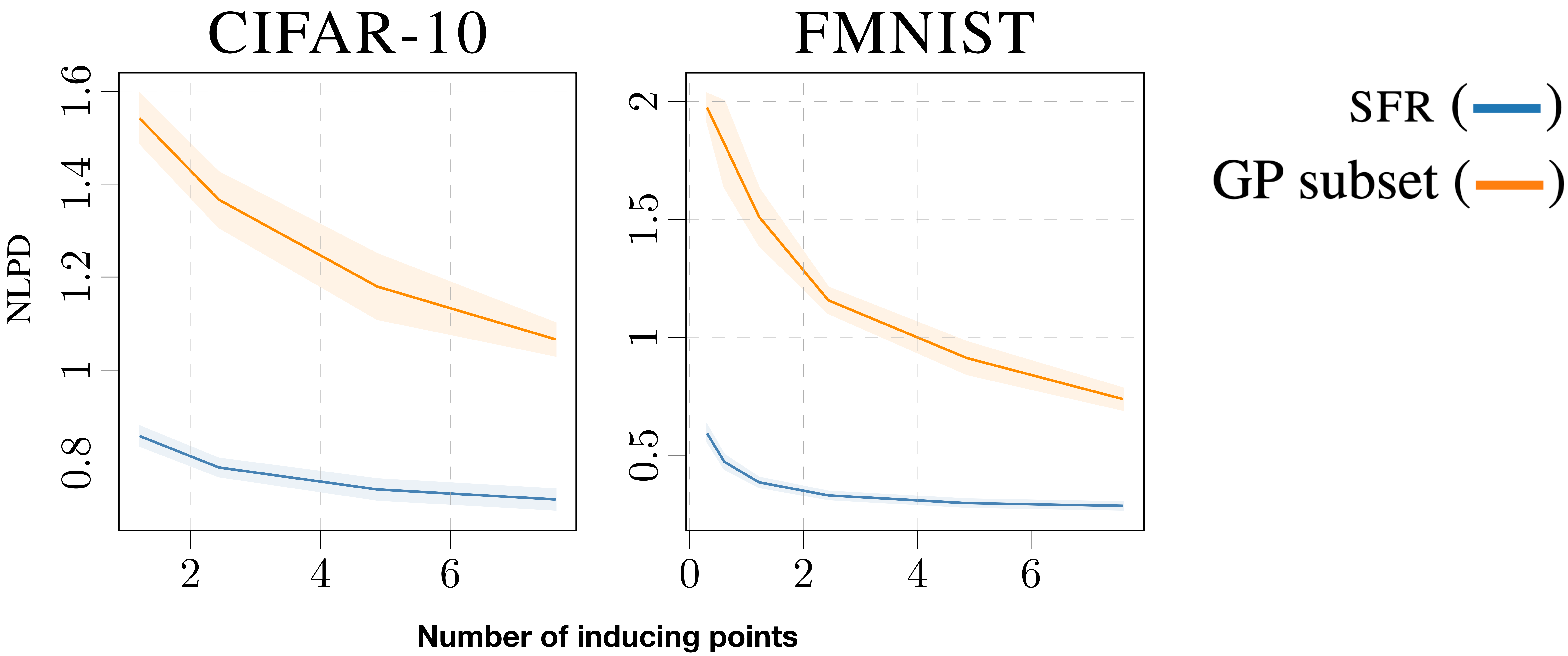
Results: Effective Sparsification



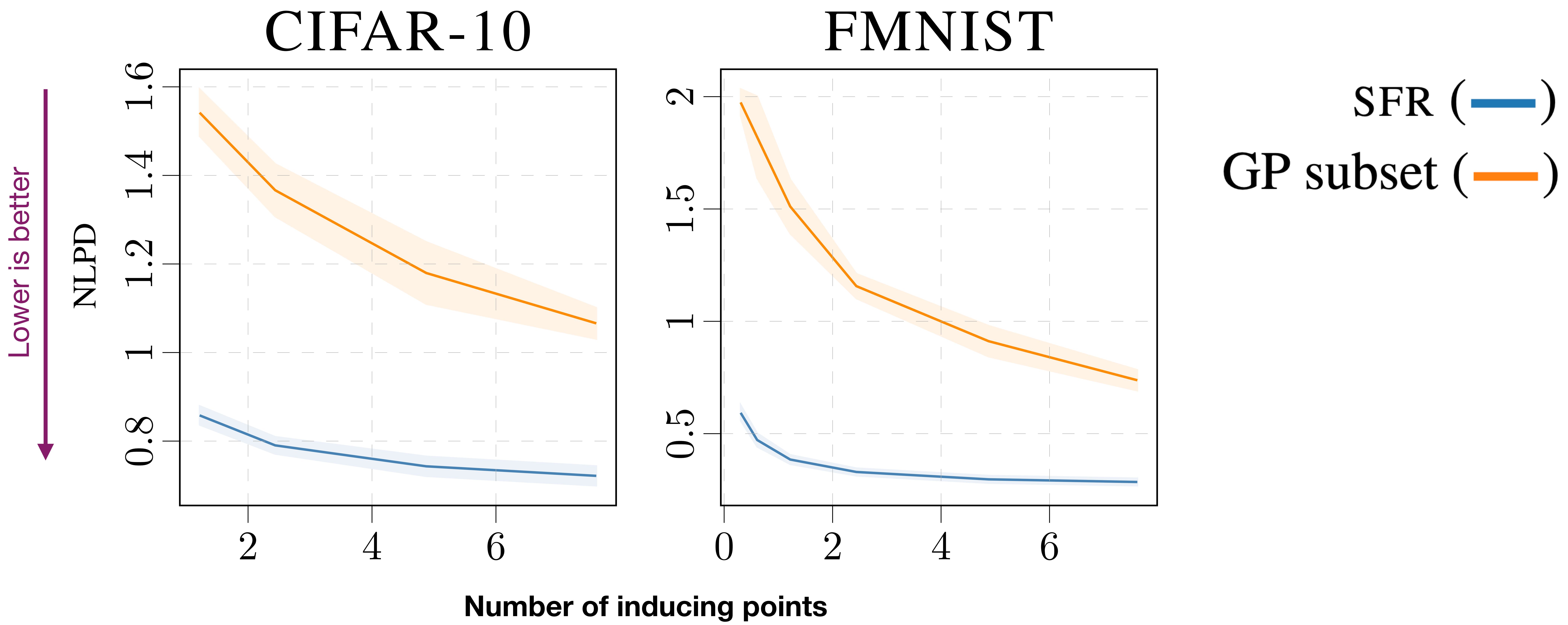
FCAI

fcai.fi

Results: Effective Sparsification



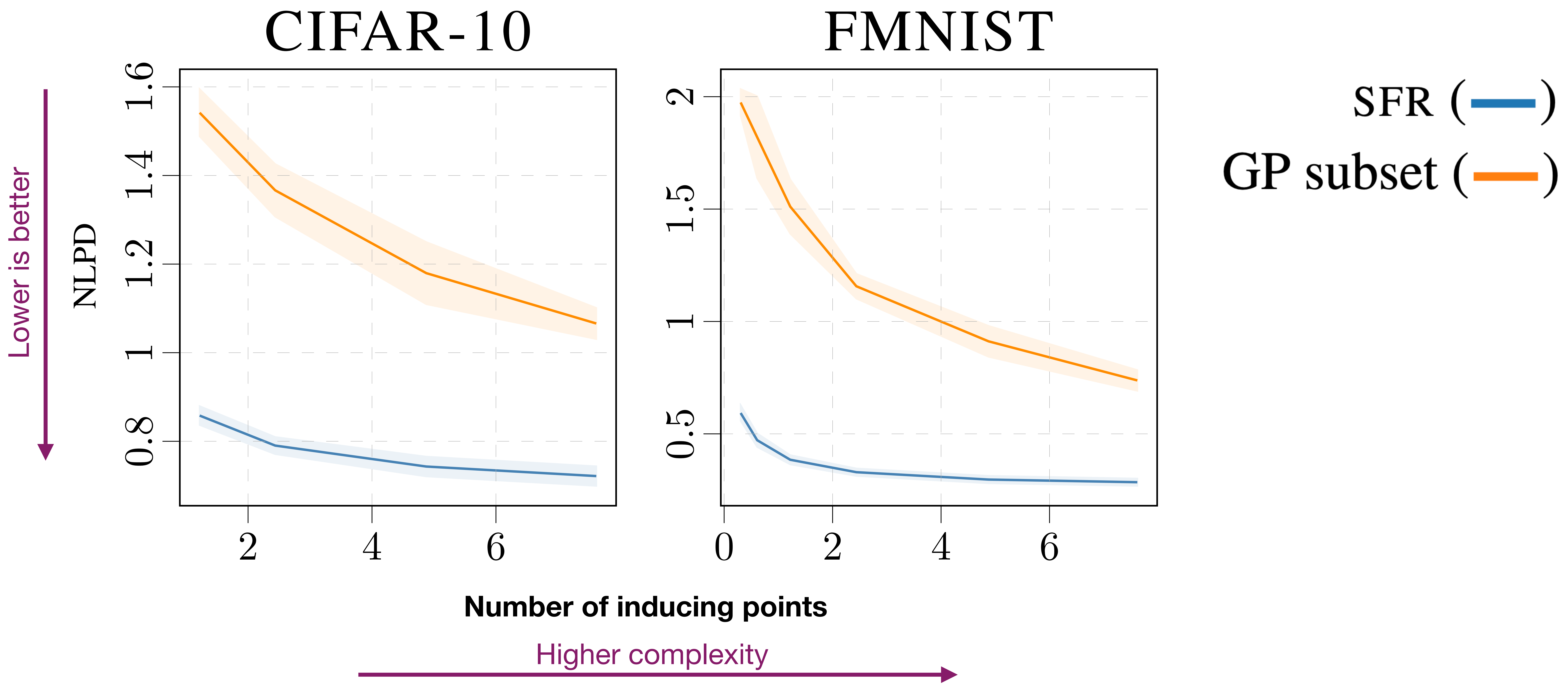
Results: Effective Sparsification



FCAI

fcai.fi

Results: Effective Sparsification

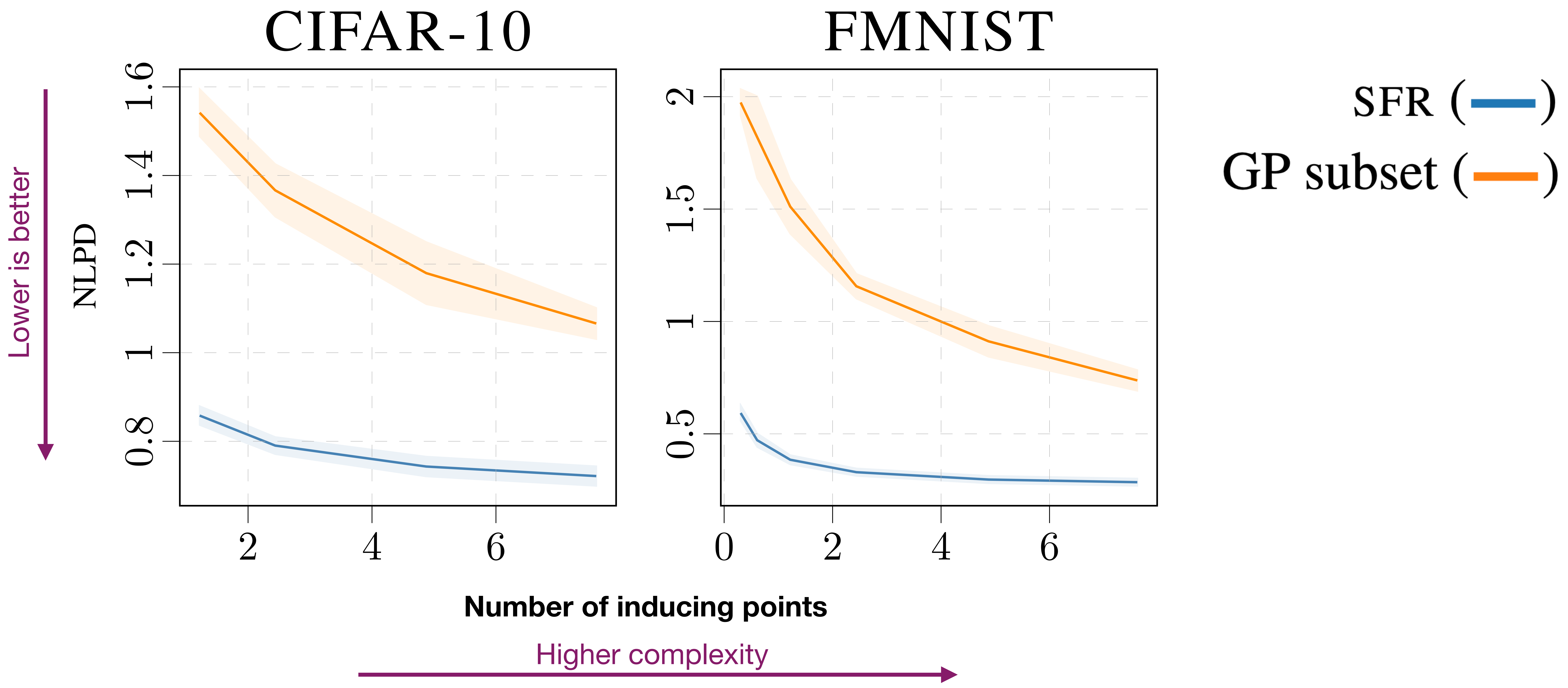


FCAI

fcai.fi

Results: Effective Sparsification

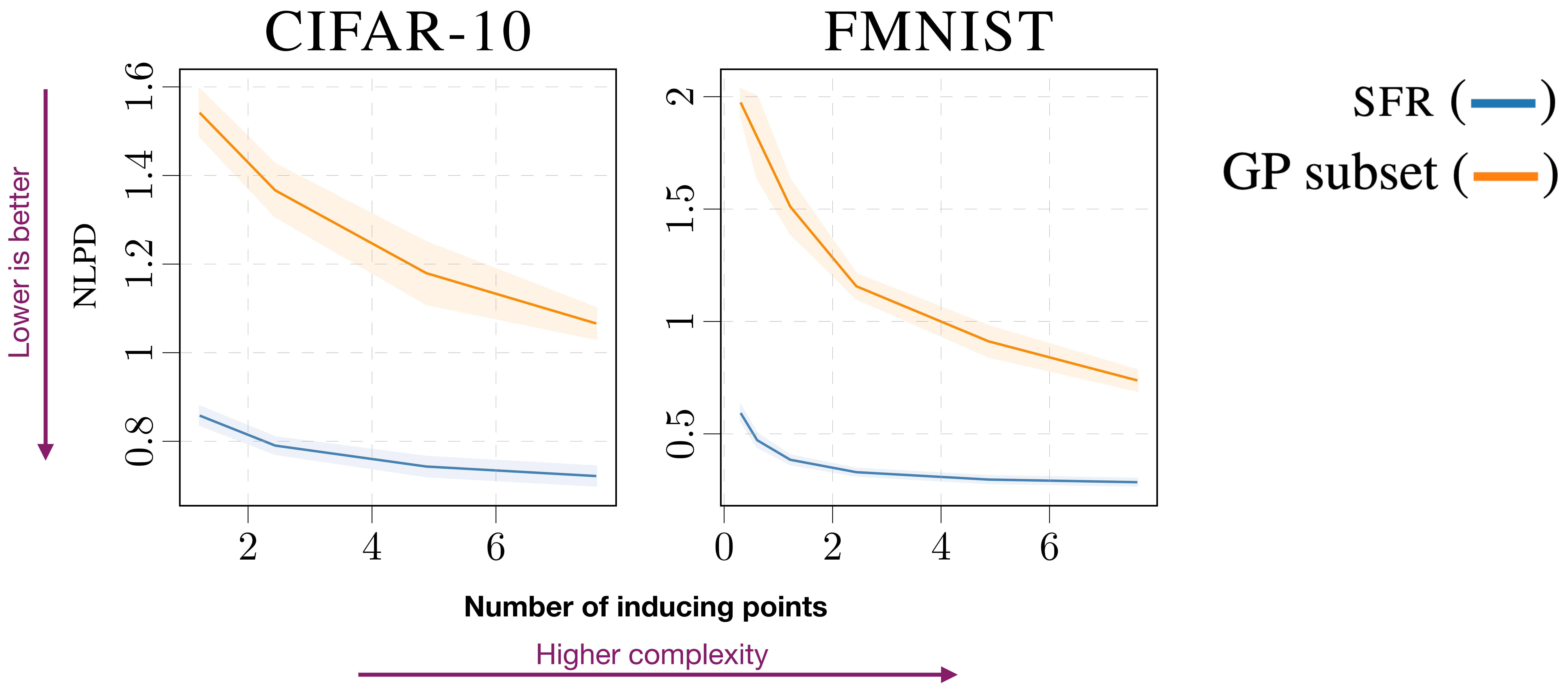
SFR makes better predictions with lower complexity



Scales to image inputs ✅

Results: Effective Sparsification

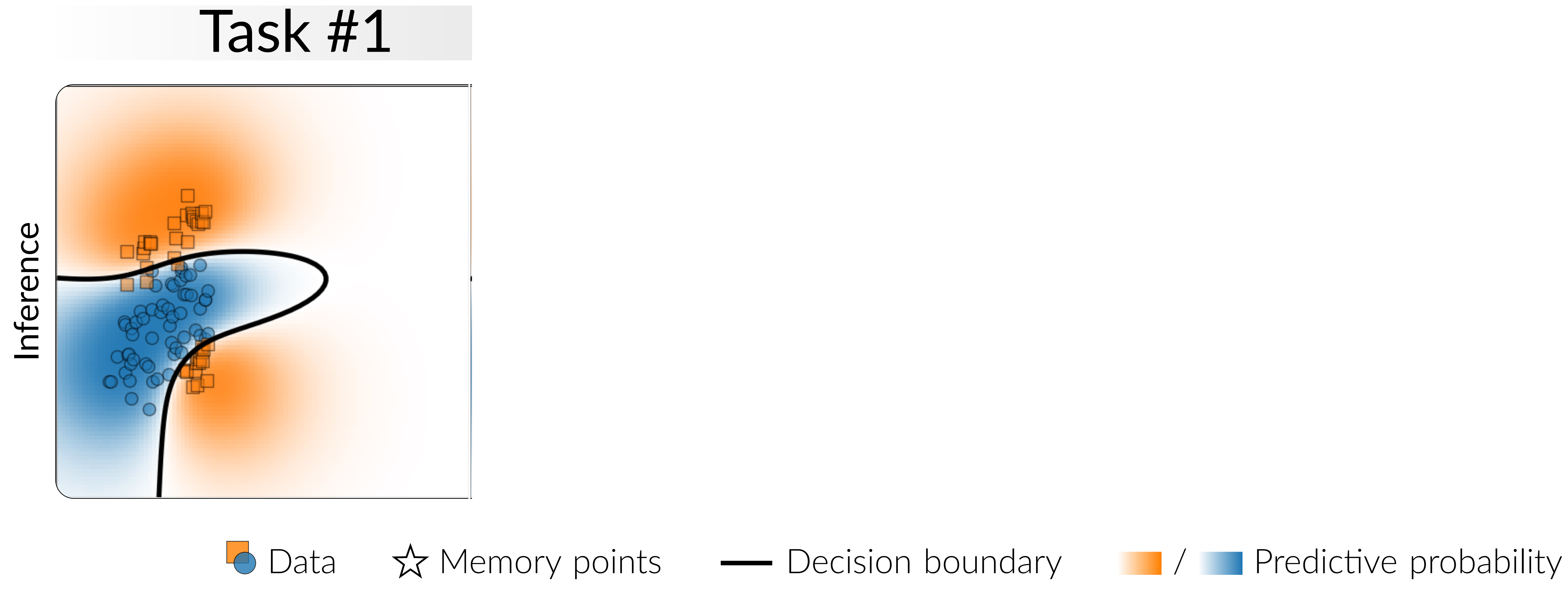
SFR makes better predictions with lower complexity



SFR for Sequential Learning

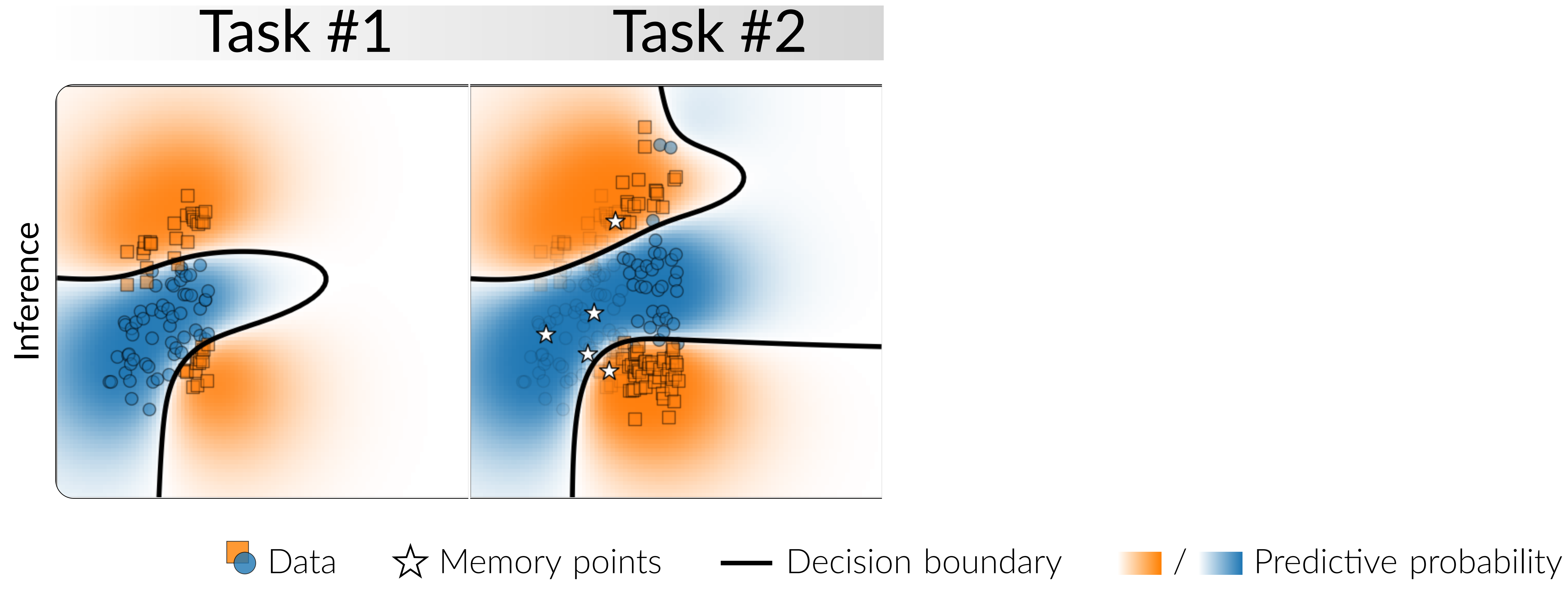
Continual Learning

SFR maintains a **functional representation** for continual learning



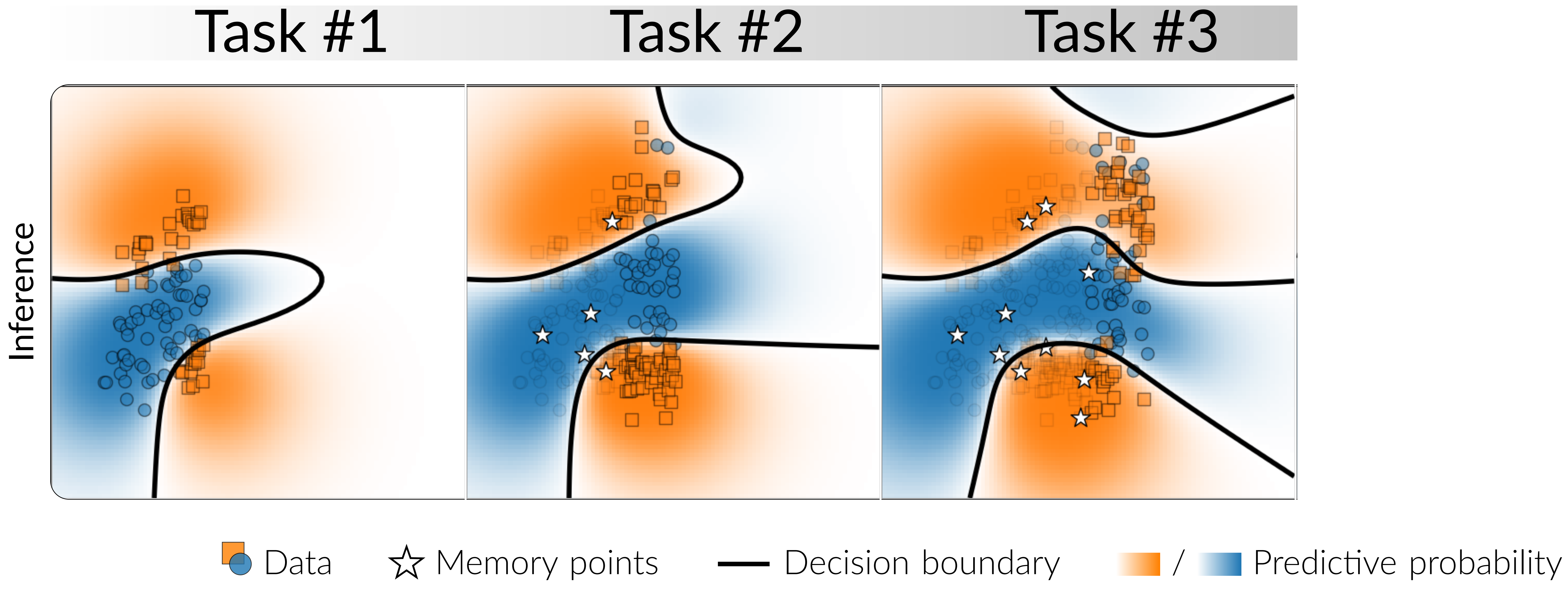
Continual Learning

SFR maintains a **functional representation** for continual learning



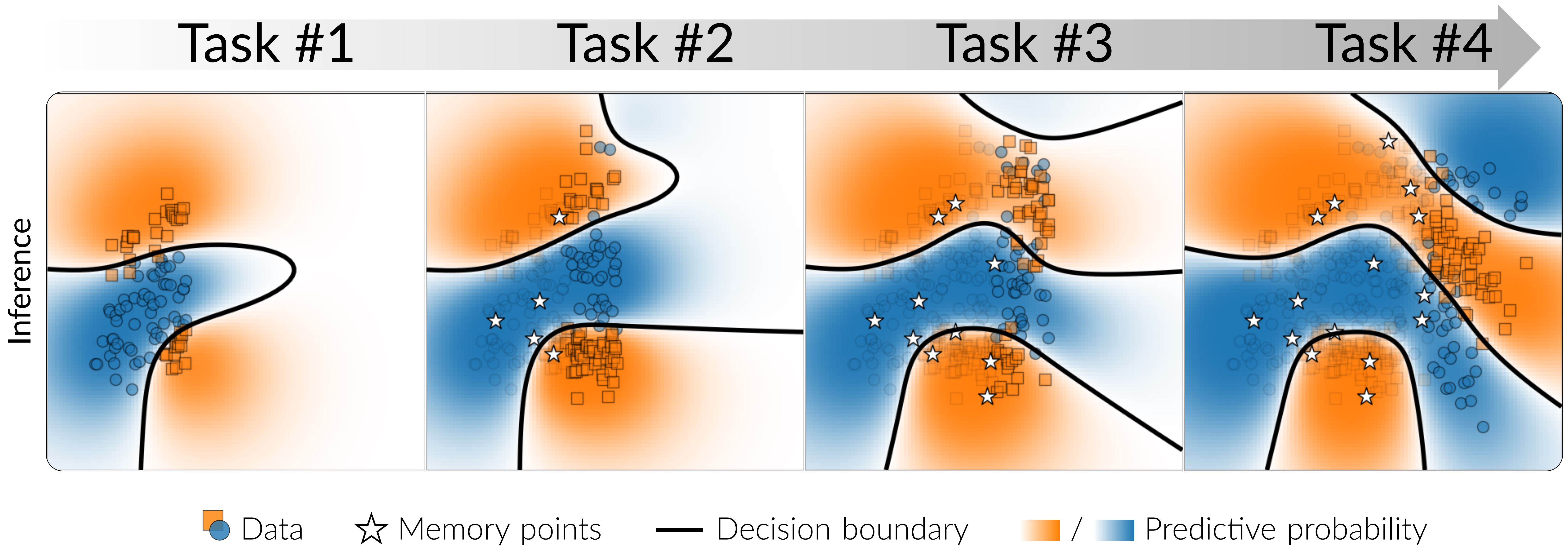
Continual Learning

SFR maintains a **functional representation** for continual learning



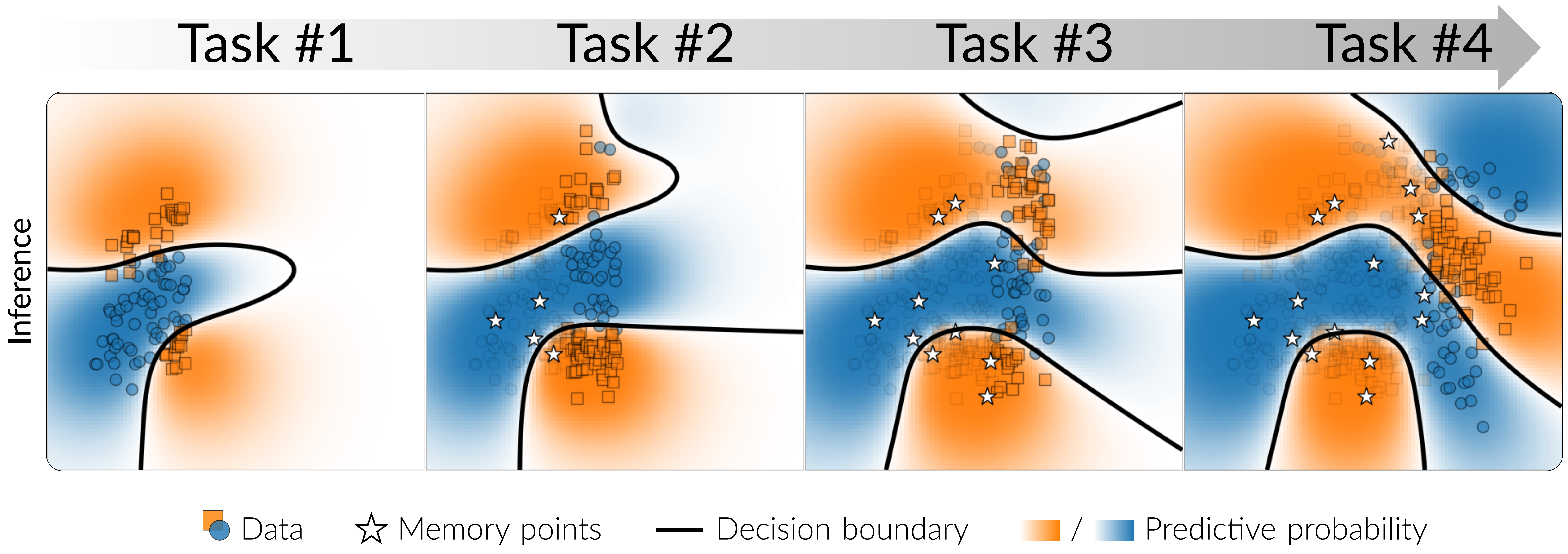
Continual Learning

SFR maintains a **functional representation** for continual learning



Continual Learning

SFR maintains a **functional representation** for continual learning



Method	S-MNIST (SH) 40 pts./task	S-MNIST (SH) 200 pts./task	S-FMNIST (SH) 200 pts./task	P-MNIST (SH) 200 pts./task
DER	85.26$\pm$0.54	92.13$\pm$0.45	**82.03**$\pm$0.57	93.08$\pm$0.11
FROMP	75.21$\pm$2.05	89.54$\pm$0.72	78.83$\pm$0.46	94.90$\pm$0.04
S-FSVI	84.51$\pm$1.30	92.87$\pm$0.14	77.54$\pm$0.40	**95.76**$\pm$0.02
SFR (Ours)	**89.22**$\pm$0.76	**94.19**$\pm$0.26	**81.96**$\pm$0.24	95.58$\pm$0.08

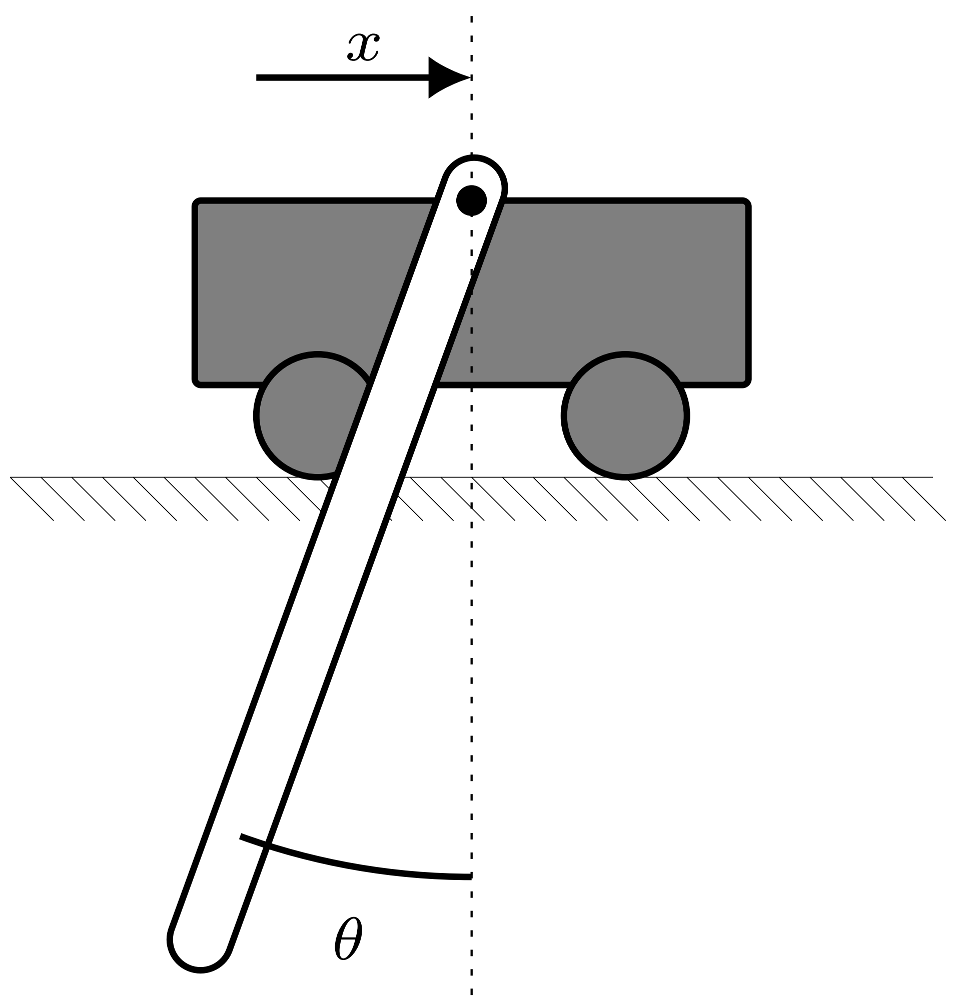
Reinforcement Learning

SFR's uncertainty can guide exploration in model-based RL

Reinforcement Learning

SFR's uncertainty can guide exploration in model-based RL

Cartpole swingup setup

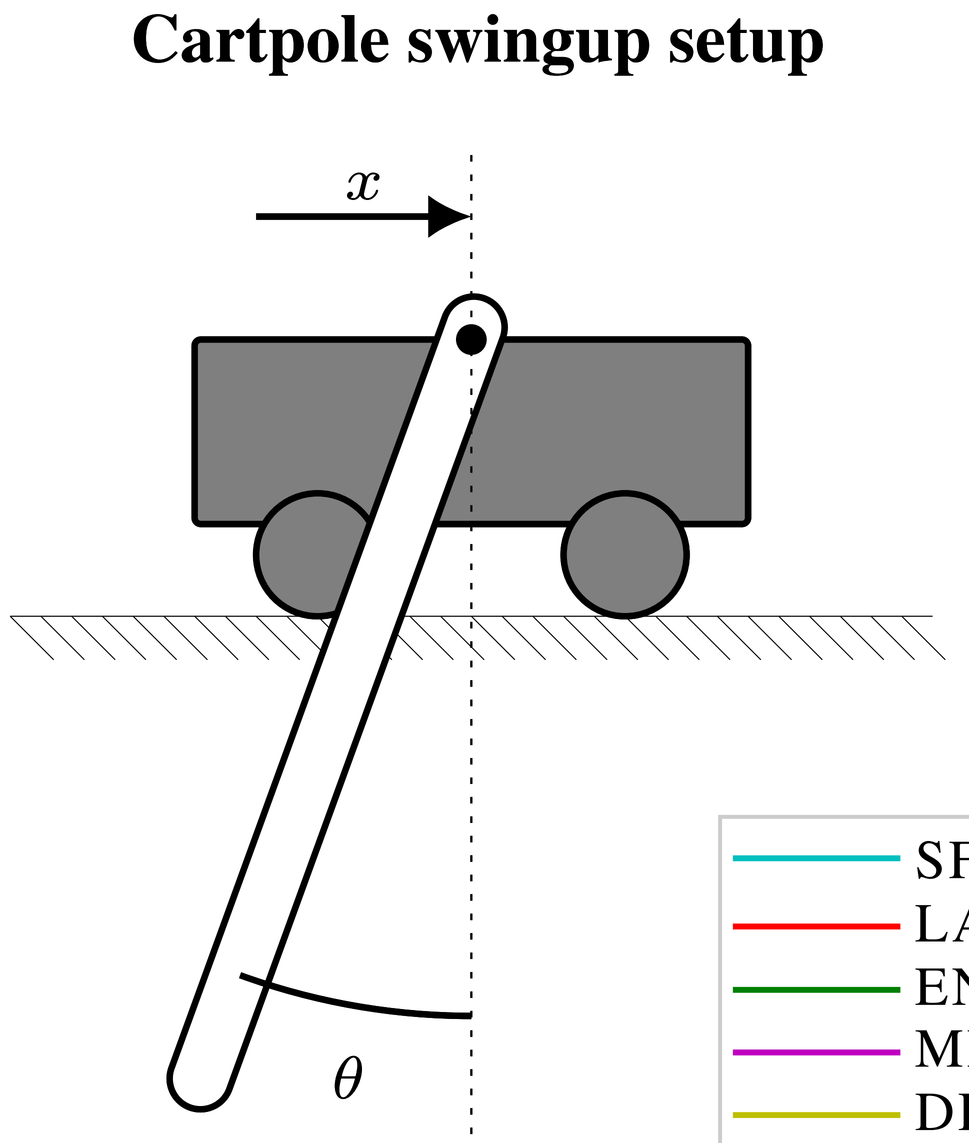


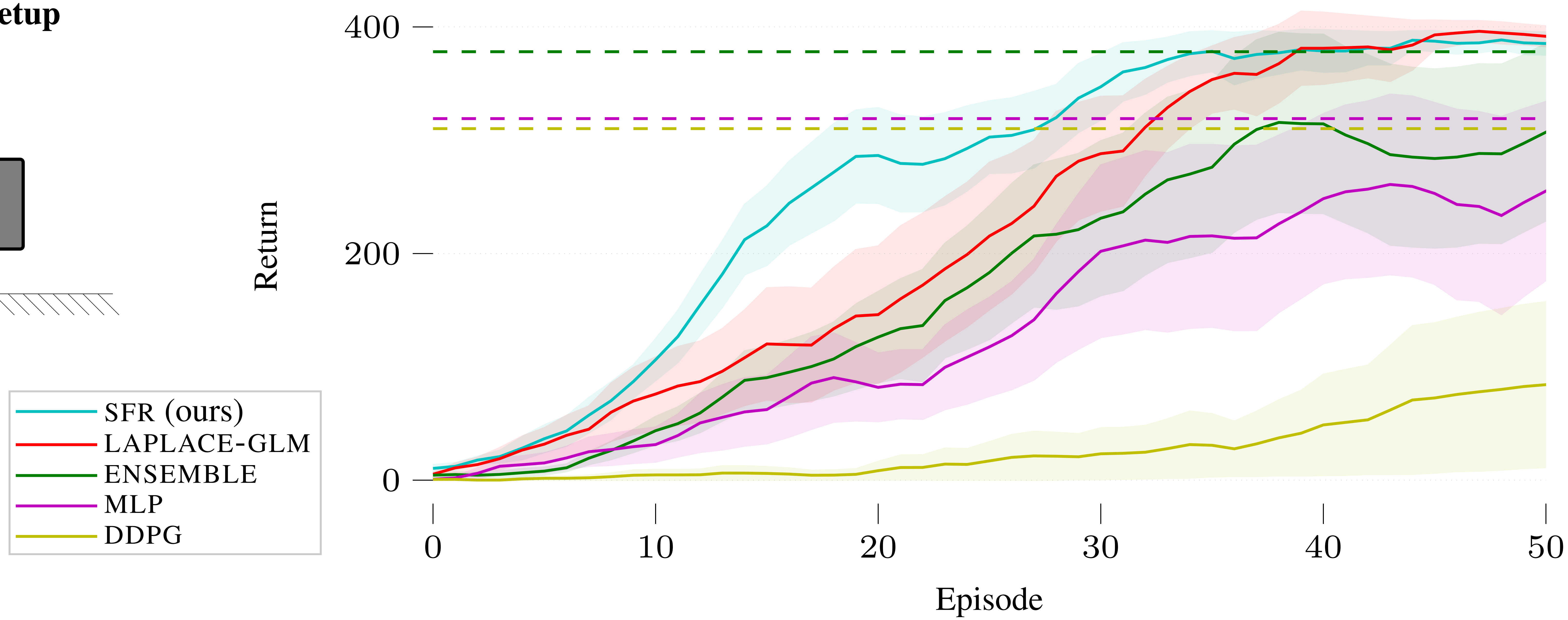
FCAI

fcai.fi

Reinforcement Learning

SFR's uncertainty can guide exploration in model-based RL





Overview

Overview

	SFR (Ours)	GP	NN

Overview

	SFR (Ours)	GP	NN
Uncertainty estimates	✅	✅	❌

Overview

	SFR (Ours)	GP	NN
Uncertainty estimates	✅	✅	❌
Image inputs	✅	❌	✅

Overview

	SFR (Ours)	GP	NN
Uncertainty estimates	✅	✅	❌
Image inputs	✅	❌	✅
Large data	✅	❌	✅

Overview

	SFR (Ours)	GP	NN
Uncertainty estimates	✅	✅	❌
Image inputs	✅	❌	✅
Large data	✅	❌	✅
Incorporate new data fast	✅	✅	❌

Thanks!

Project Website

https://aaltoml.github.io/sfr/

Code

Aidan Scannell*, Riccardo Mereu*, Paul Edmund Chang, Ella Tamir, Joni Pajarinen, Arno Solin

**Equal contribution*

FCAI ICLR fcai.fi