Uncertainty Quantification in Deep Learning

Martin Trapp
Aalto University
Finland

Vienna Deep Learning Meetup
September 11th, 2024

Outline

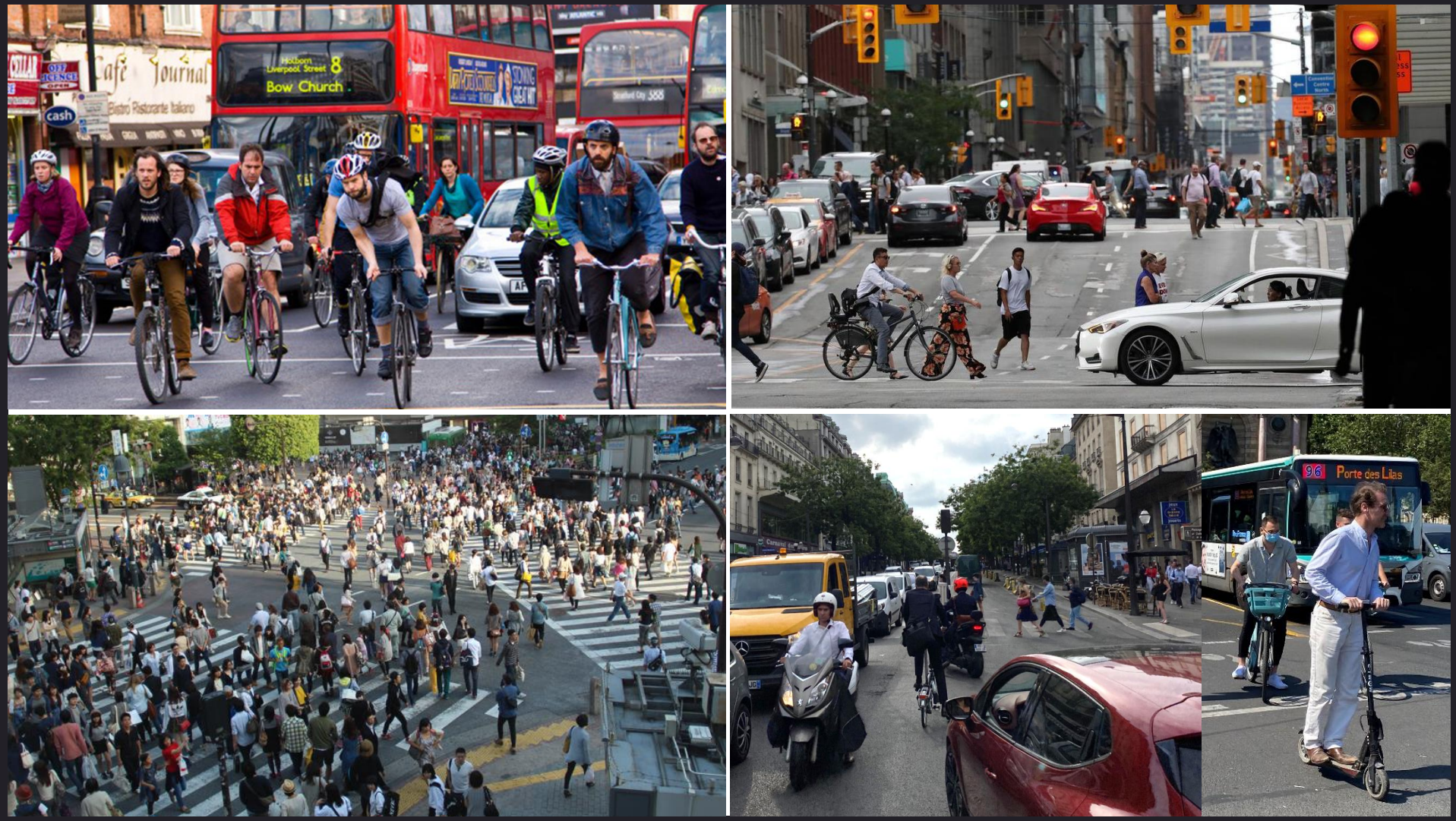
Why … do we need uncertainty quantification?

What … uncertainties are there?

How … can we quantify uncertainties?

Why do we need UQ?

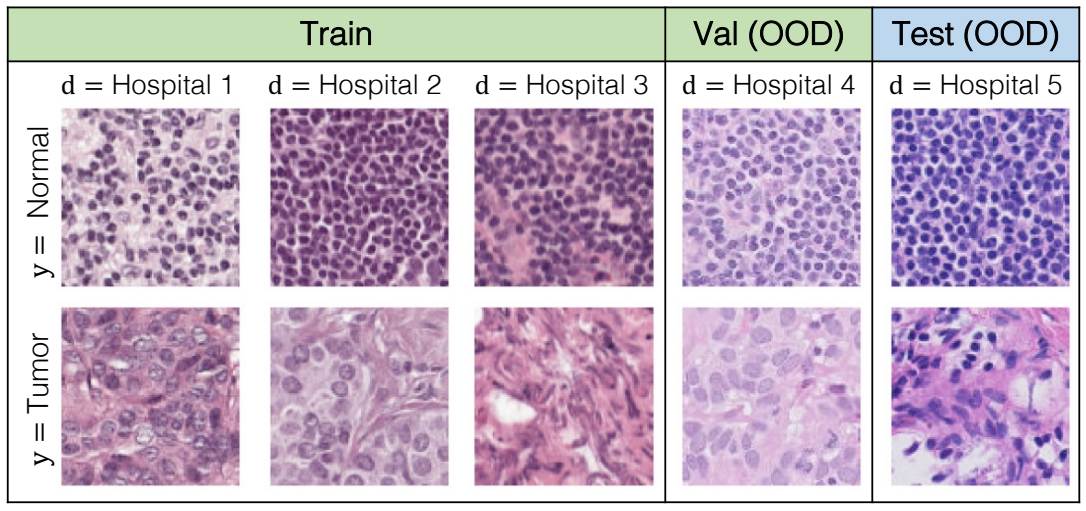
The Real World is Messy



Source: ICCV 2023 Tutorial on *The Many Faces of Reliability of Deep Learning for Real-World Deployment*

The Real World is Messy

Distribution Shifts



Source: P W Koh *et al.* WILDS: A Benchmark of in-the-Wild Distribution Shifts. In ICML, 2021.

Various Sources of Noise

Source: S Sabour *et al.* SpotlessSplats: Ignoring Distractors in 3D Gaussian Splatting. In ICML, 2021.

Aalto University
School of Science

Expectations for ML Systems

Our systems/models should be safe and trustworthy.

For this, we require they are:

- Accurate in their predictions
- Robust to input-pertubations (noise, adversarial attacks)
- "Know when they don't know" (recognize out-of-domain data)
- Act if they are uncertain (*e.g.*, active learning, reasoning)

However, ...

- Accurate Predictions
- Sensitive to pertubations
- Don't know when they don't know
- Overconfident in their predictions

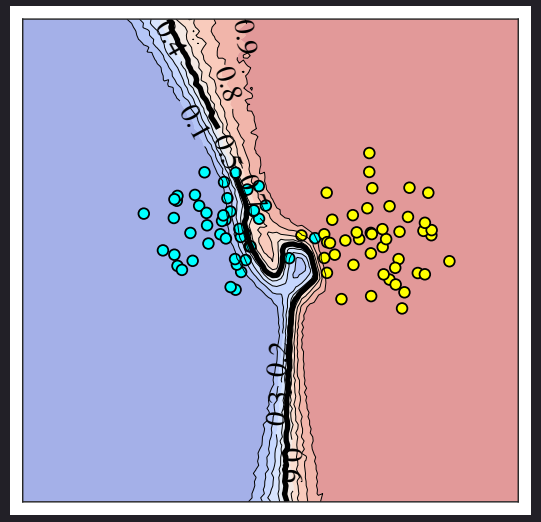
Often considered "unreliable"


WELL, WHAT DO
WE DO NOW?
imgflip.com

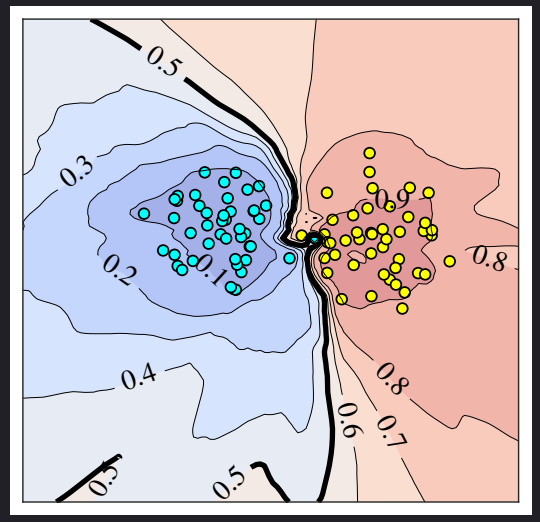
Uncertainty Quantification can Help...

Reduce Overconfidence

NN w/o UQ



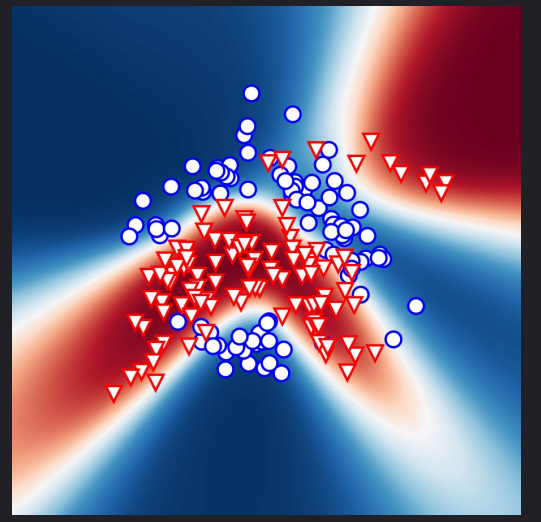
NN w UQ



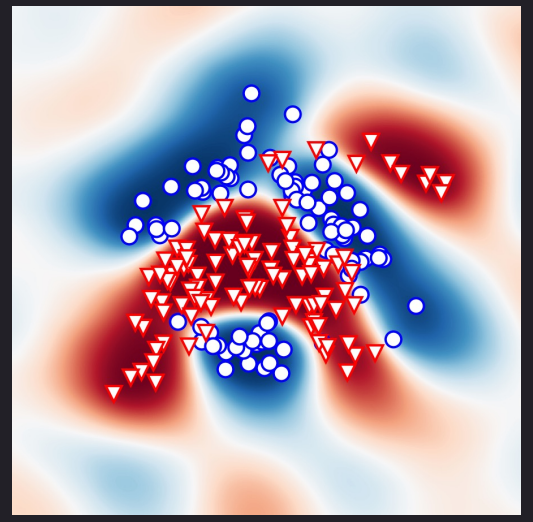
Source: A Kristiadi, M Hein, P Hennig. Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks. In ICML, 2020.

Understand Inductive Biases

ReLU



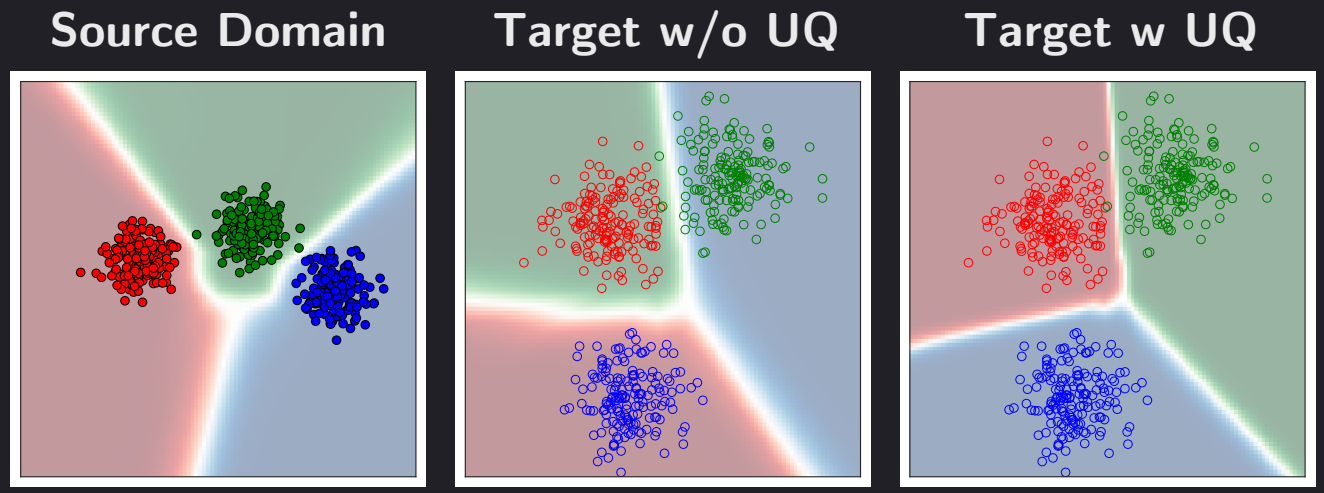
Periodic Function



Source: L Meronen, M Trapp, A Solin. Periodic activation functions induce stationarity. In NeurIPS, 2021.

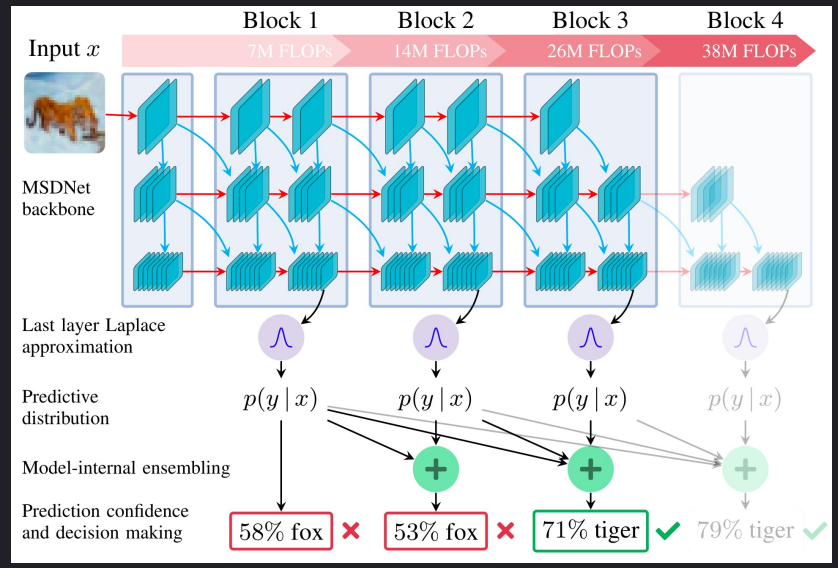
Uncertainty Quantification can Help...

Robustify our Methods



Source: S Roy *et al.* Uncertainty-guided Source-free Domain Adaptation. In ECCV, 2022.

Decide When to Act Differently

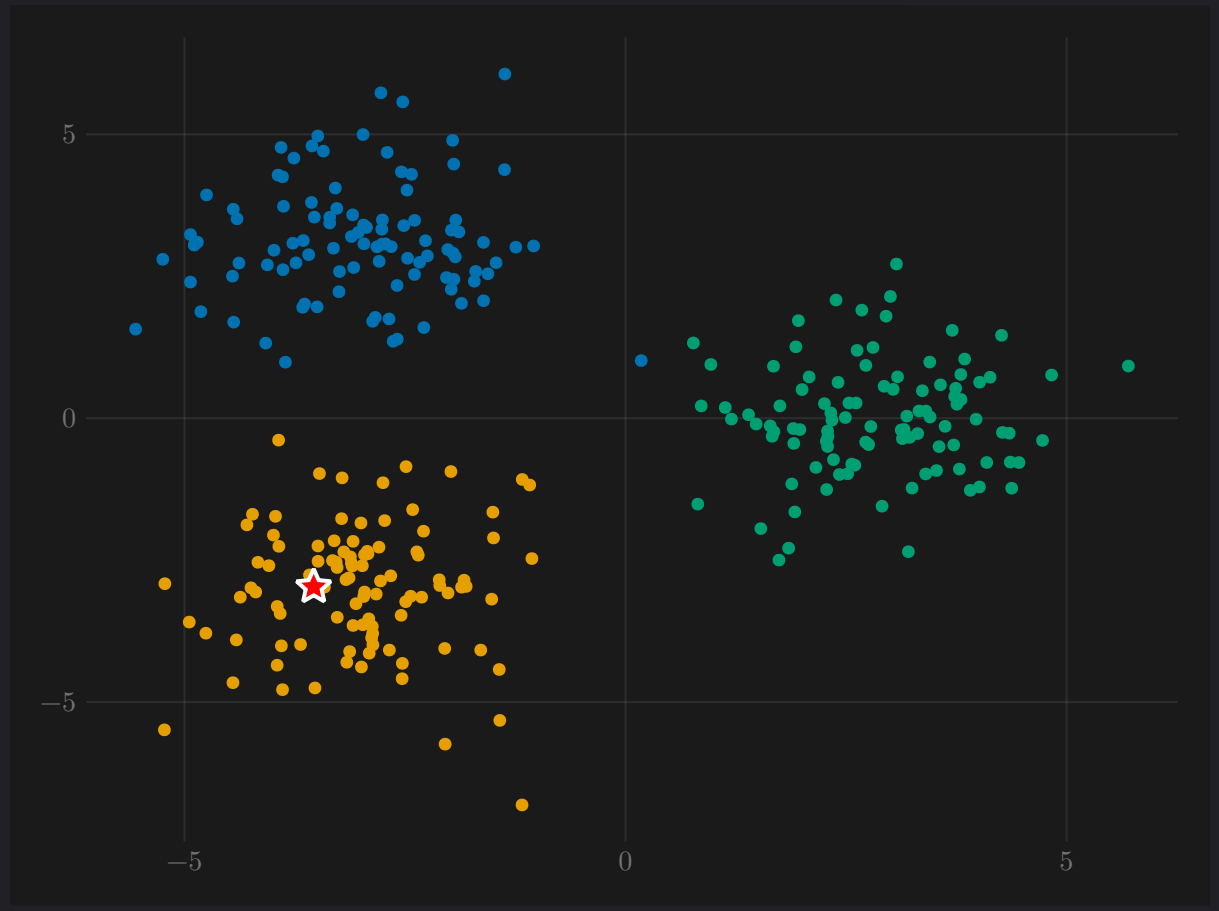


Source: L Meronen *et al.* Fixing Overconfidence in Dynamic Neural Networks. In WACV, 2024.

Sources of Uncertainty

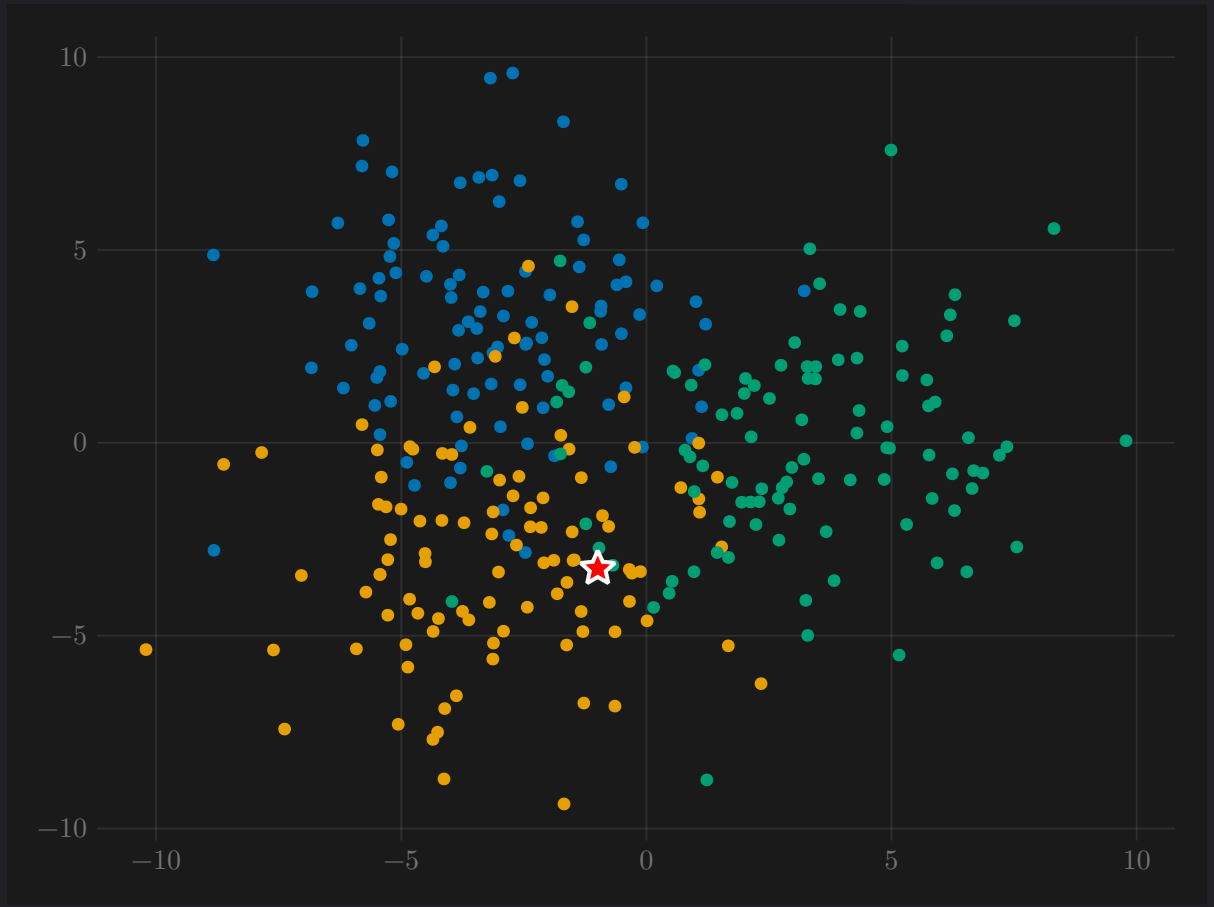
Aalto University
School of Science

Uncertainty Inherent to the Data

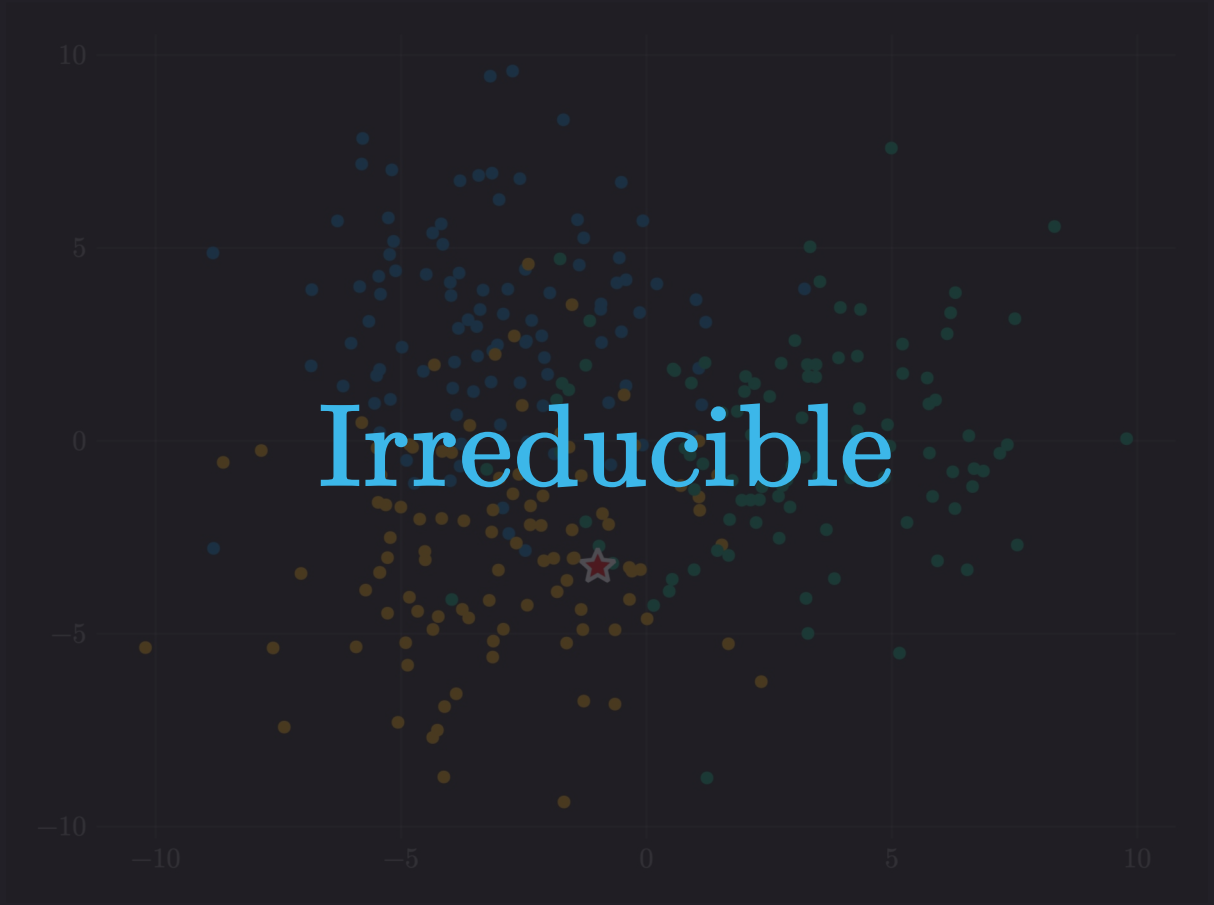


Aalto University
School of Science

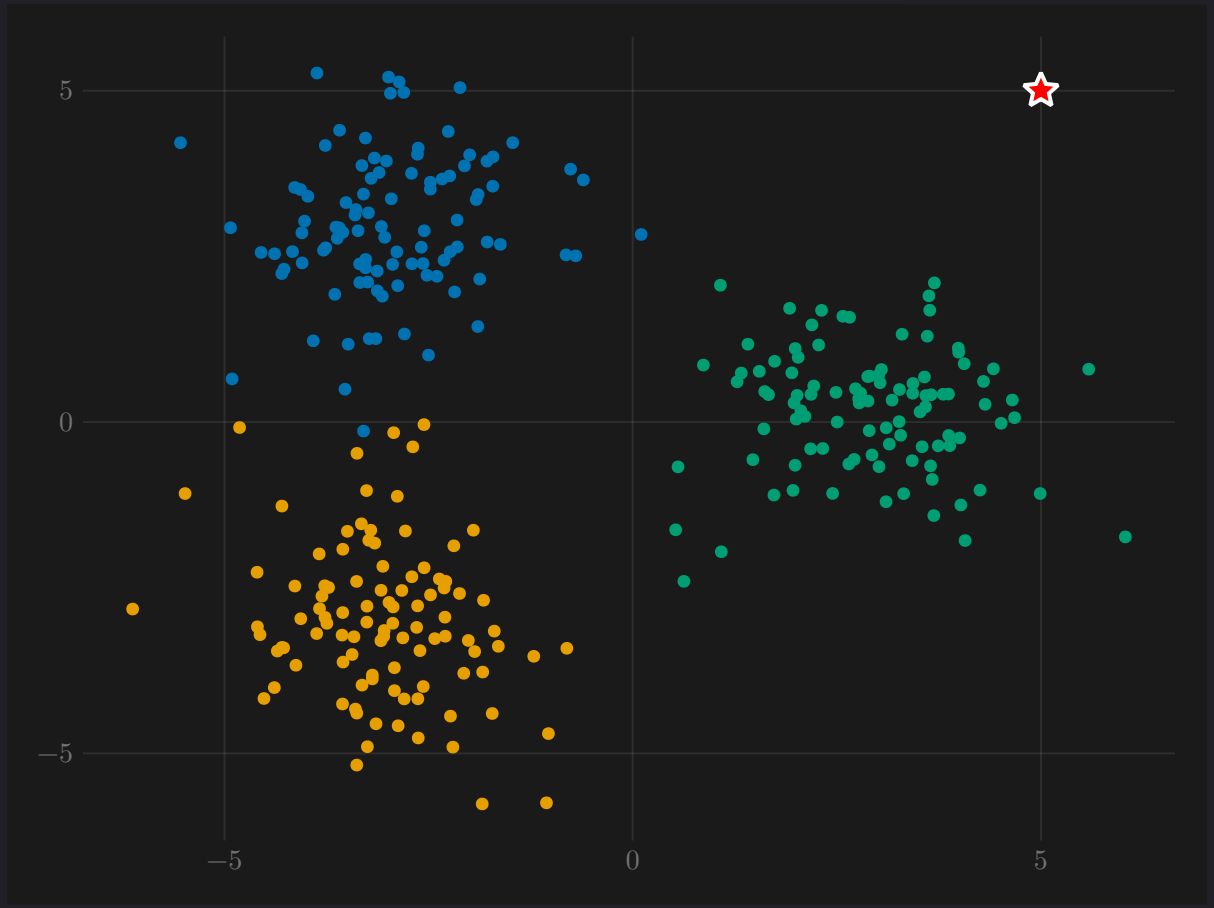
Uncertainty Inherent to the Data



Uncertainty Inherent to the Data



Uncertainty due to Lack of Knowledge

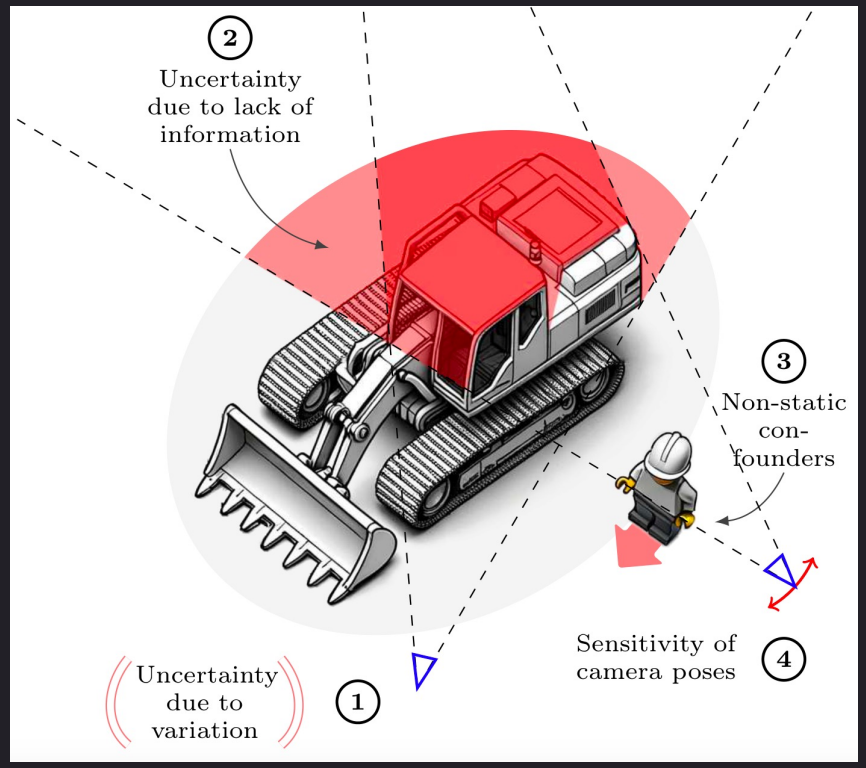


Uncertainty due to Lack of Knowledge

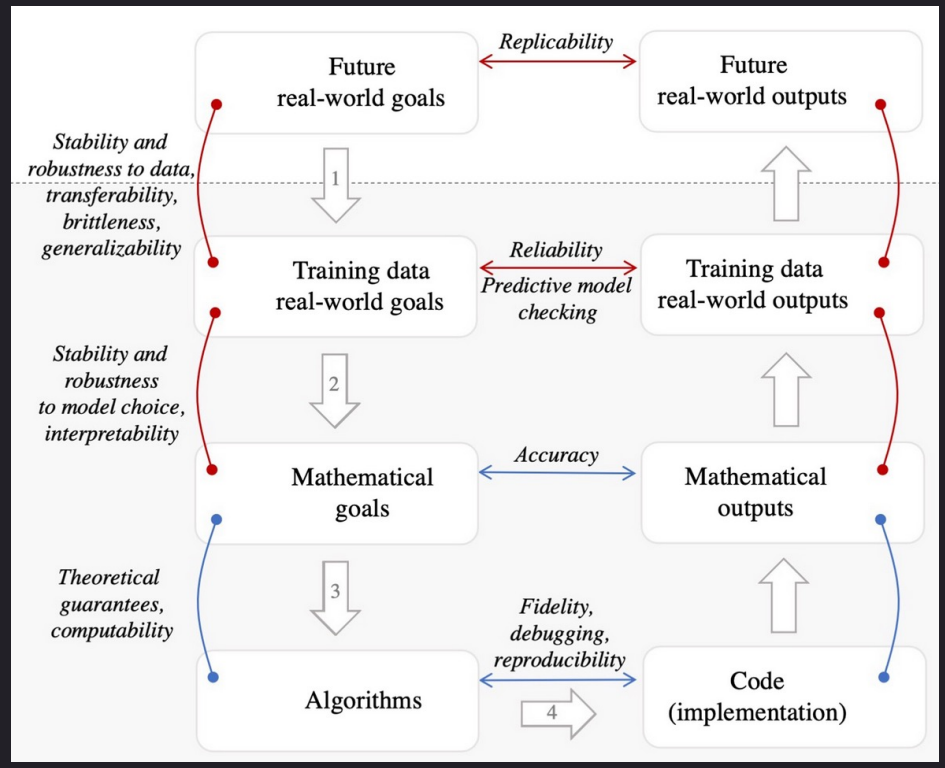
Reducible

Uncertainty due to Lack of Knowledge

Sources vary between Applications



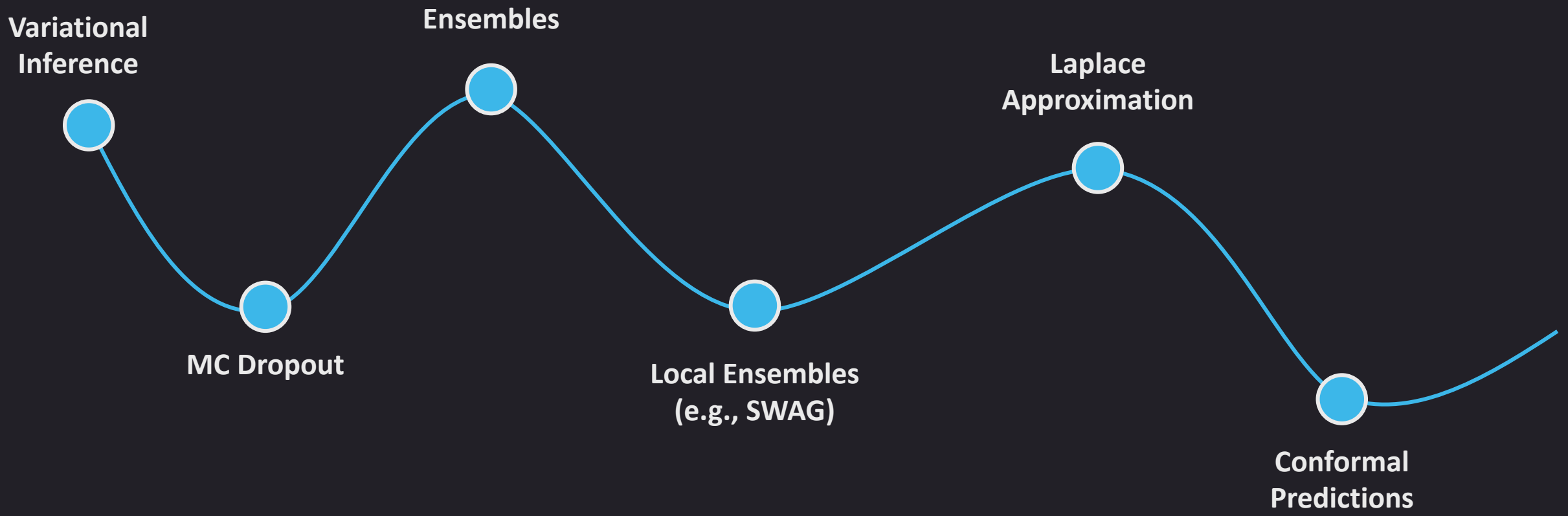
Source: M Klasson *et al.* Sources of Uncertainty in 3D Scene Reconstruction. In UNCV Workshop at ECCV, 2024.



Source: T Broederick *et al.* Toward a Taxonomy of Trust for Probabilistic Machine Learning. Science Advances 9, eabn399, 2023.

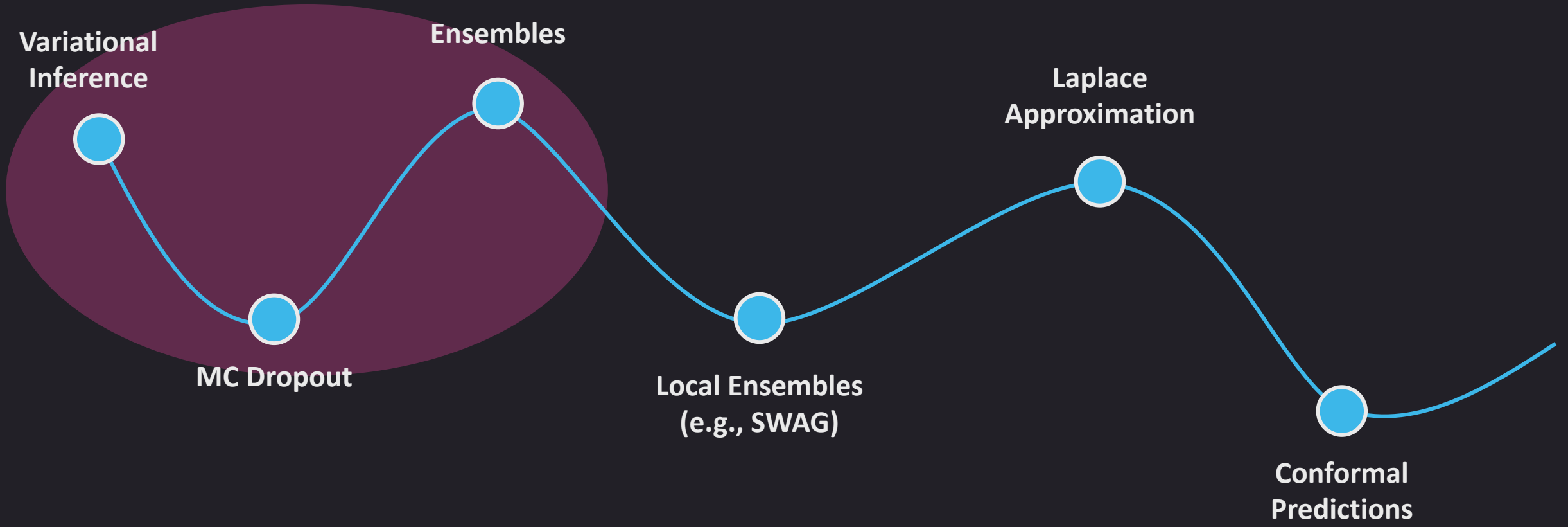
Methods for UQ in DL

Methods for Uncertainty Quantification



Aalto University
School of Science

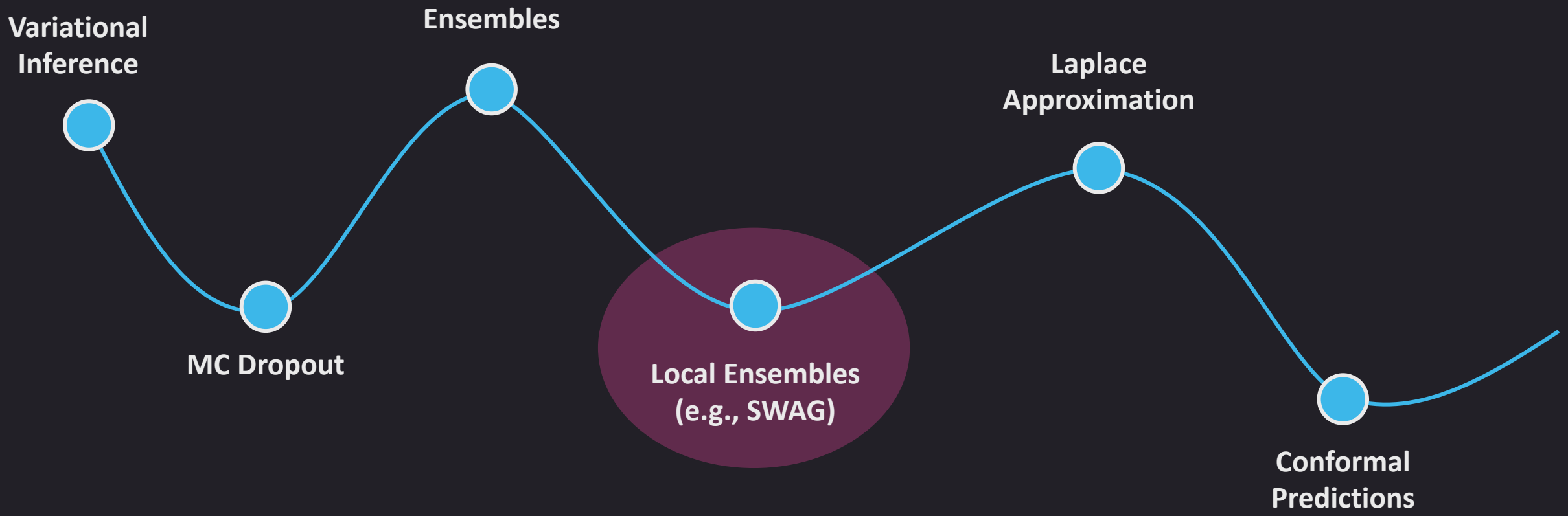
Methods for Uncertainty Quantification



Incorporate UQ into the Training

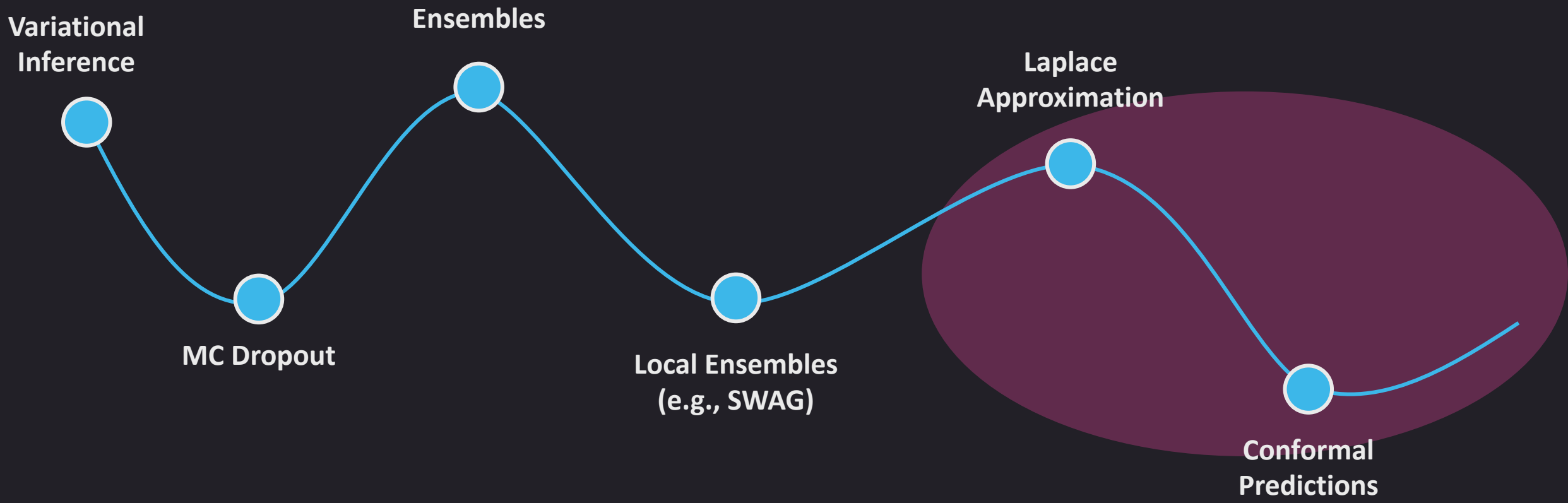
Aalto University
School of Science

Methods for Uncertainty Quantification



Moved toward post-hoc Methods

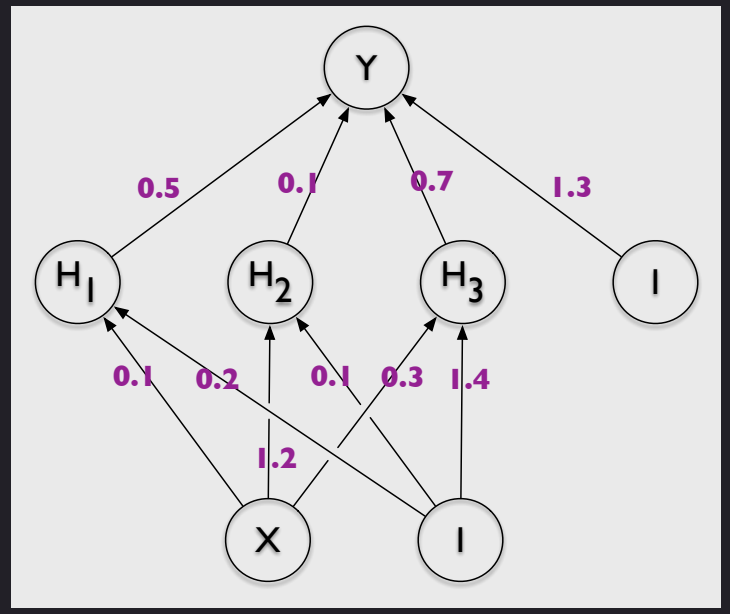
Methods for Uncertainty Quantification



Aera of post-hoc Methods

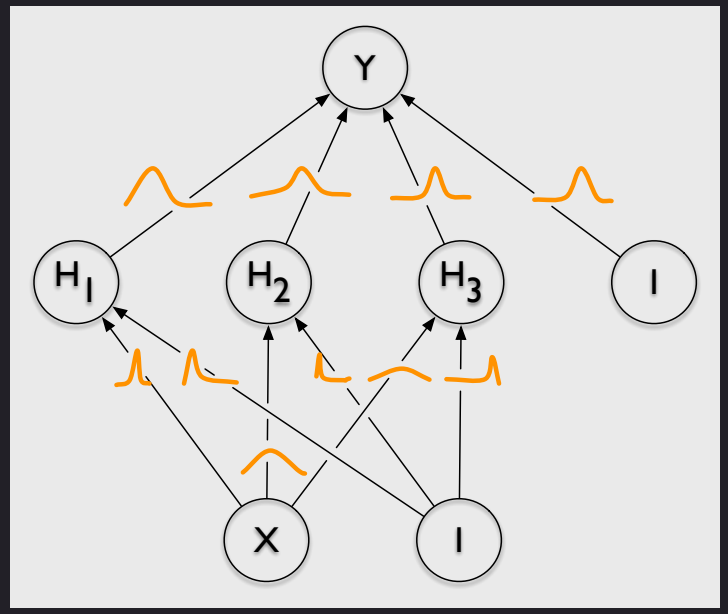
Bayesian Deep Learning

Standard NN



- Weights are learnable parameters
- Output is deterministically given for any input

Bayesian NN



- Weights are represented by probability distributions
- Parameters of probability distributions are learnable
- Output is a distribution and not a single value

Source: C Blundell *et al.* Weight Uncertainty in Neural Networks, In ICML 2015.

Bayesian Deep Learning

Unfortunately, inference in Bayesian Neural Networks (BNNs) is notoriously hard!

In BNNs we need to perform Bayesian inference (conditioning):

Prior on the weights

Training Data

Likelihood function

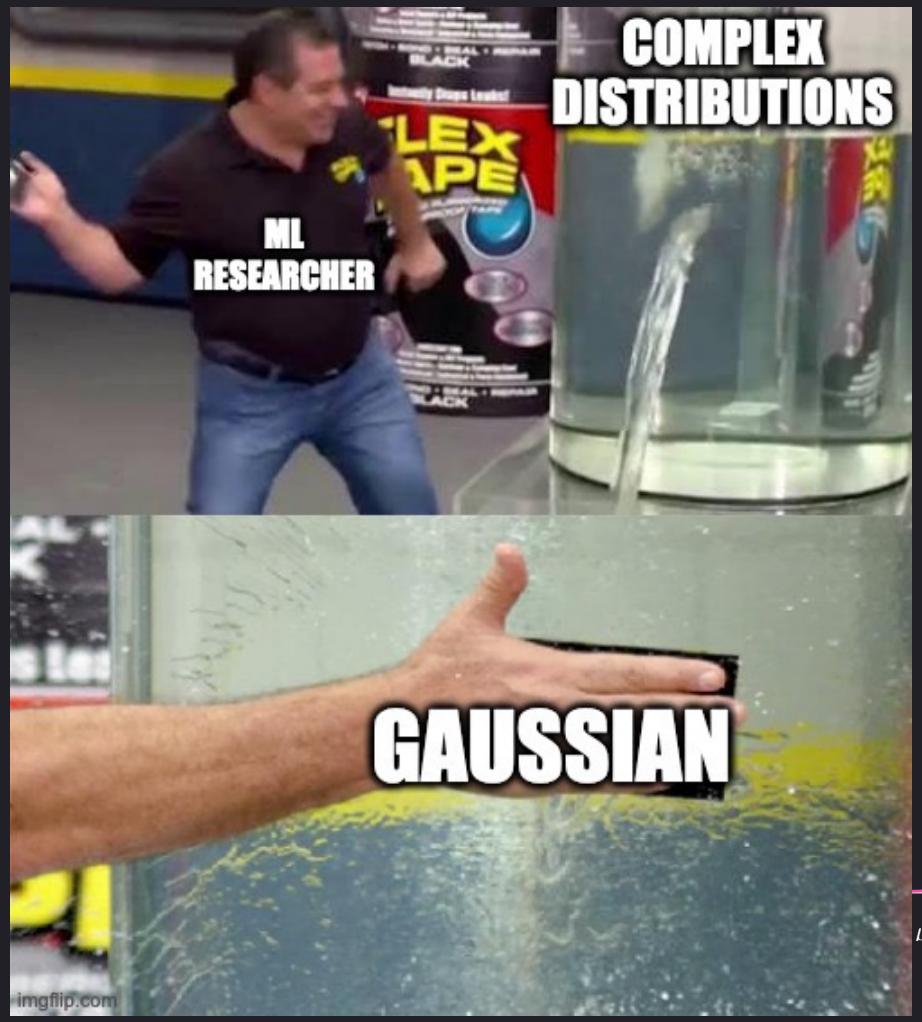
Posterior (conditional) Distribution over the Weights

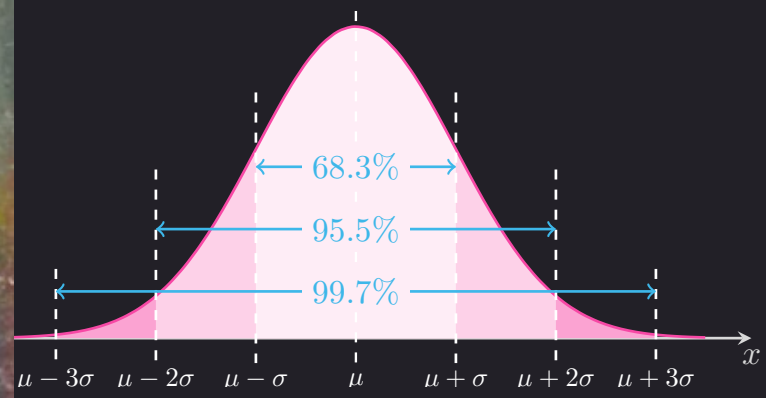
$$p(\boldsymbol{w} \mid \mathcal{D}) = \frac{p(\boldsymbol{w})p(\mathcal{D} \mid \boldsymbol{w})}{\int_{\boldsymbol{w}} p(\mathcal{D} \mid \boldsymbol{w}) \mathrm{d}\boldsymbol{w}}$$

We cannot compute this!

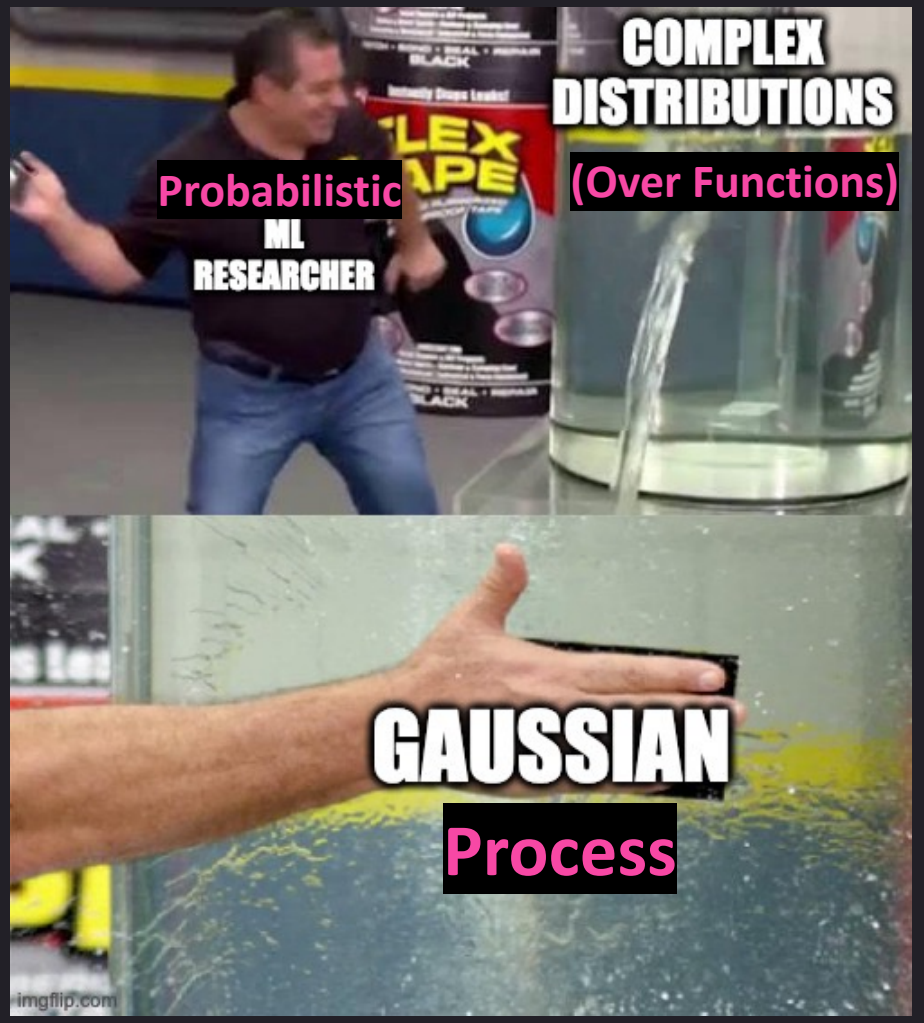

WELL, WHAT DO
WE DO NOW?
imgflip.com

Laplace Approximation





Laplace Approximation



Laplace Approximation

$$p(\boldsymbol{w} \mid \mathcal{D}) = \frac{p(\boldsymbol{w})p(\mathcal{D} \mid \boldsymbol{w})}{\int_{\boldsymbol{w}} p(\mathcal{D} \mid \boldsymbol{w})\mathrm{d}\boldsymbol{w}}$$

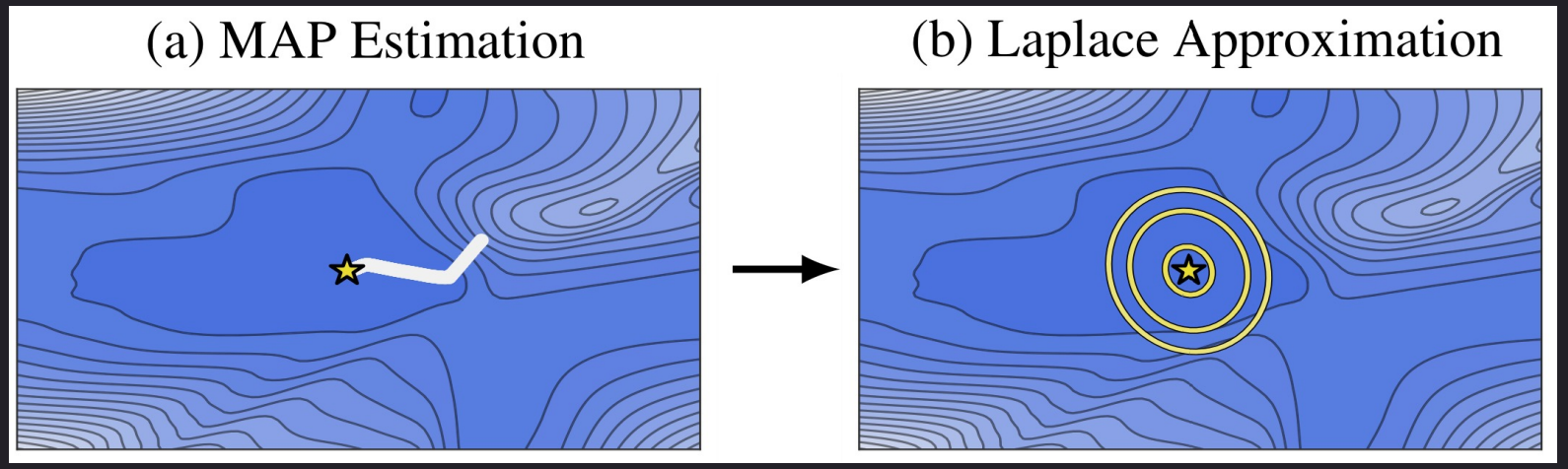
$$\log p(\boldsymbol{w} \mid \mathcal{D}) = \overbrace{\underbrace{\log p(\boldsymbol{w})}_{\text{L2 regularization}} + \underbrace{\log p(\mathcal{D} \mid \boldsymbol{w})}_{\text{Cross-Entropy}}}^{\text{Negated Loss}} - \underbrace{C}_{\text{Unknown}}$$

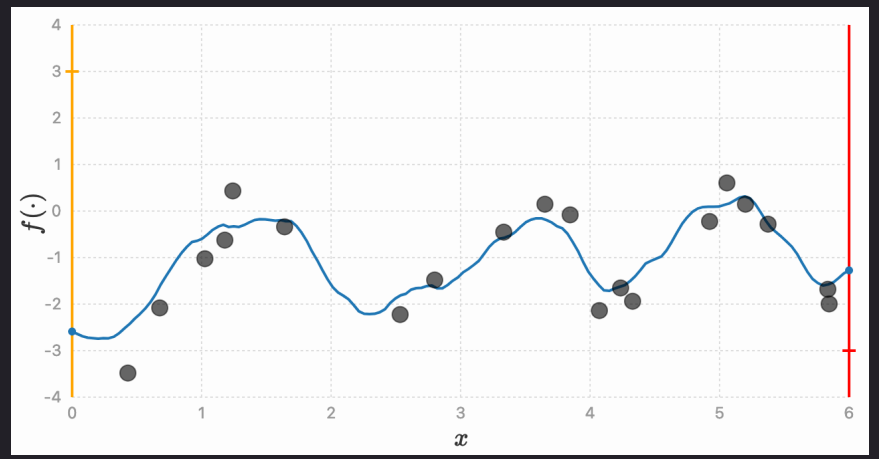
$$\underbrace{\log p(\boldsymbol{w}) + \log p(\mathcal{D} \mid \boldsymbol{w})}_{\ell(\boldsymbol{w})} \approx \ell(\boldsymbol{w}^*) + \boldsymbol{J}_{\ell|_{\boldsymbol{w}=\boldsymbol{w}^*}}(\boldsymbol{w} - \boldsymbol{w}^*) - \frac{1}{2}\boldsymbol{H}_{\ell|_{\boldsymbol{w}=\boldsymbol{w}^*}}(\boldsymbol{w} - \boldsymbol{w}^*)^2$$

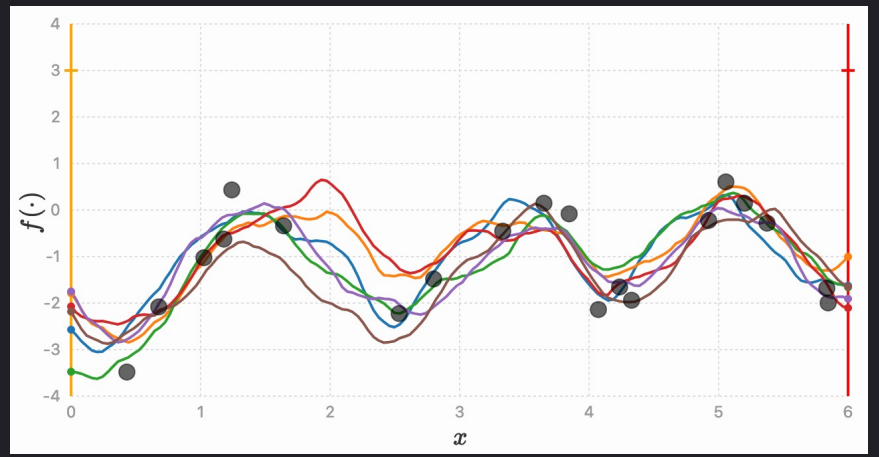
$$= \ell(\boldsymbol{w}^*) - \frac{1}{2}\boldsymbol{H}_{\ell|_{\boldsymbol{w}=\boldsymbol{w}^*}}(\boldsymbol{w} - \boldsymbol{w}^*)^2$$

$$\propto \log \mathcal{N}(\boldsymbol{w}^*, \boldsymbol{H}^{-1}_{\ell|_{\boldsymbol{w}=\boldsymbol{w}^*}})$$

Laplace Approximation

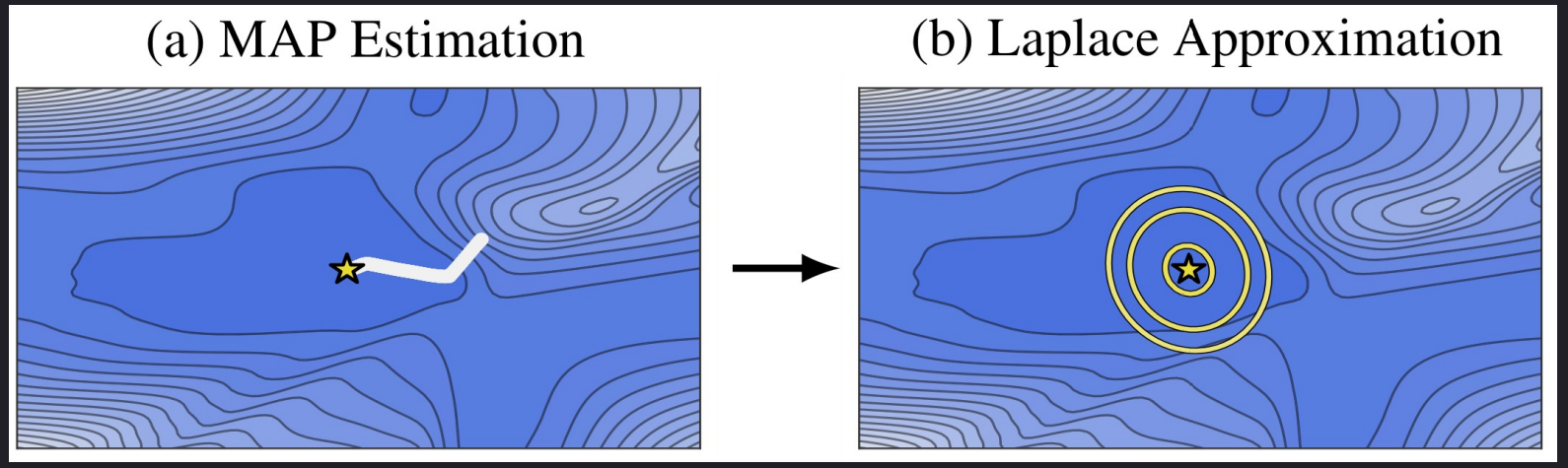


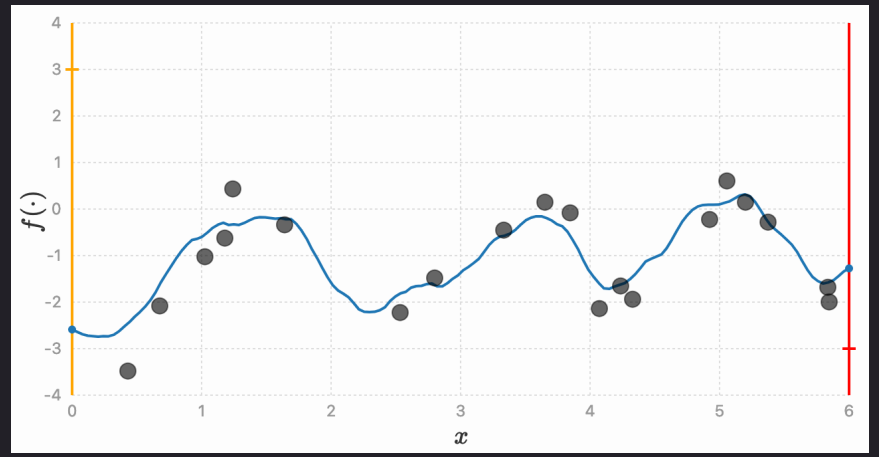


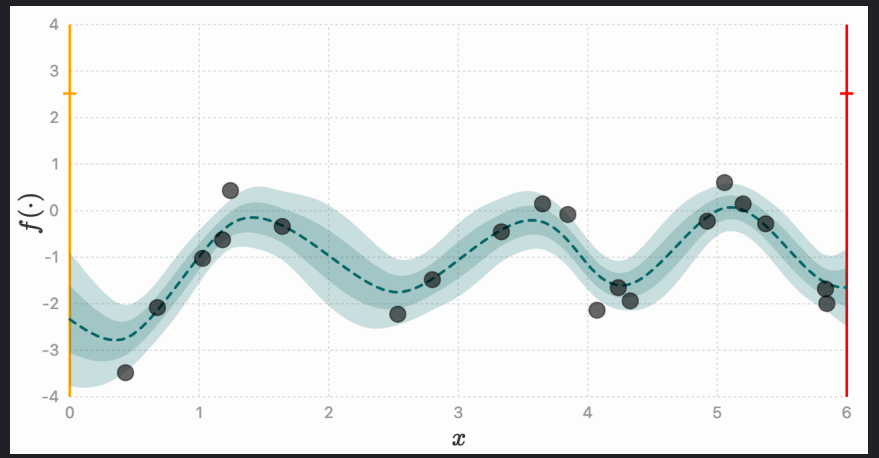


Source: E Daxberger *et al.* Laplace Redux - Effortless Bayesian Deep Learning. In NeurIPS 2021.
Source: Interactive Gaussian Process Visualization by Ti John (http://www.infinitecuriosity.org/vizgp/)

Laplace Approximation

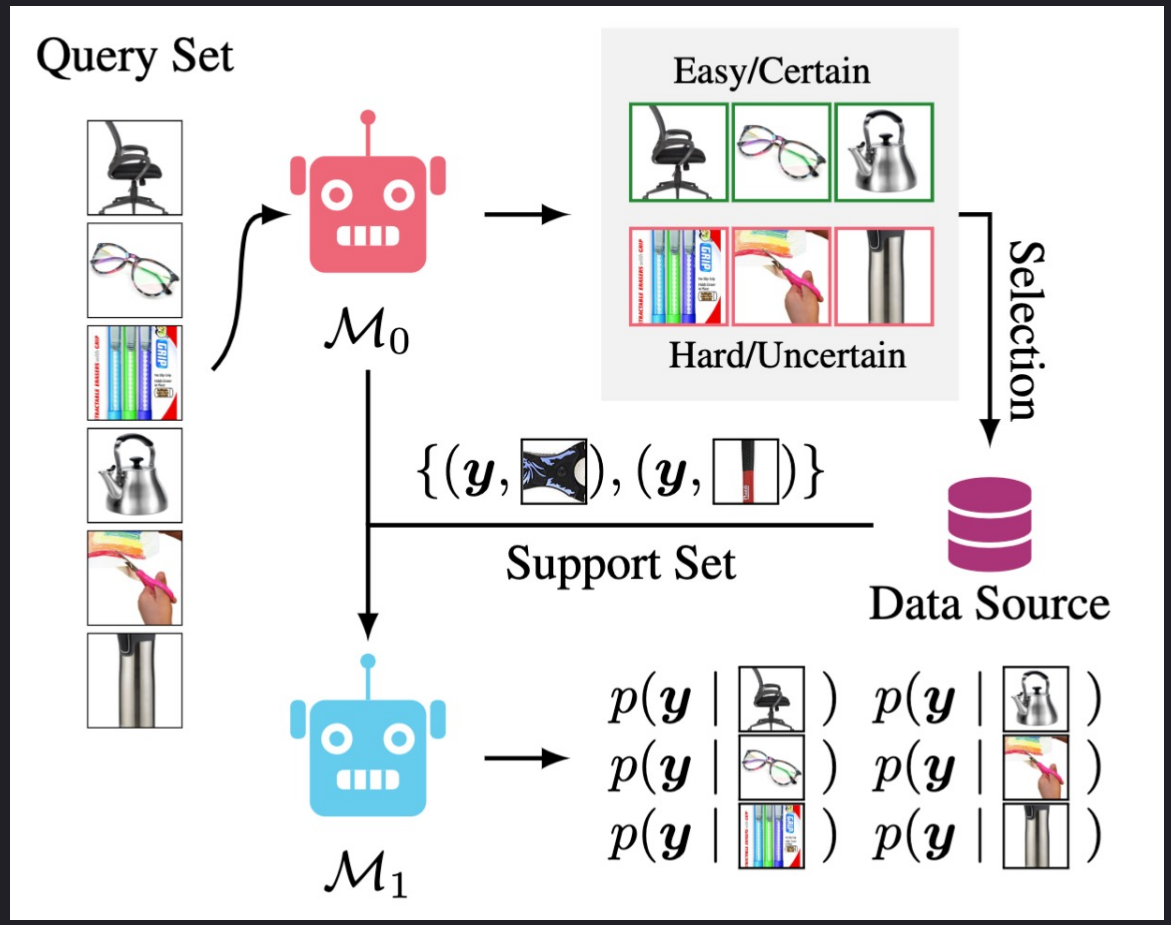






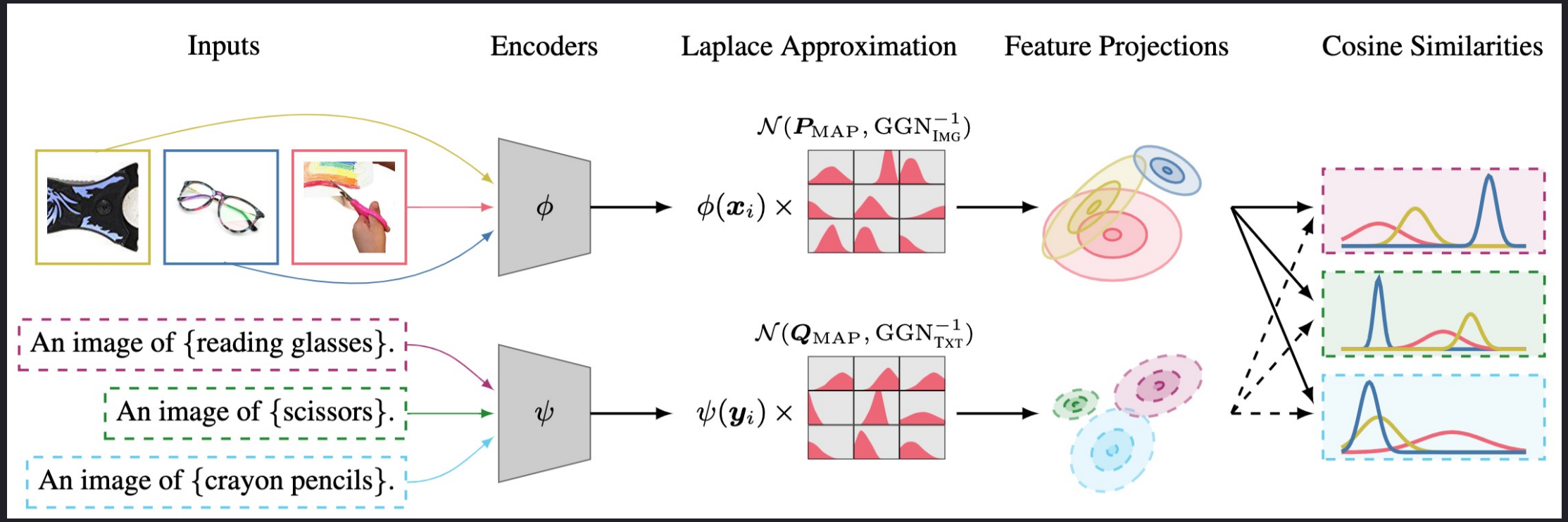
Source: E Daxberger *et al.* Laplace Redux - Effortless Bayesian Deep Learning. In NeurIPS 2021.
Source: Interactive Gaussian Process Visualization by Ti John (http://www.infinitecuriosity.org/vizgp/)

Probabilistic Active Learning in VLMs



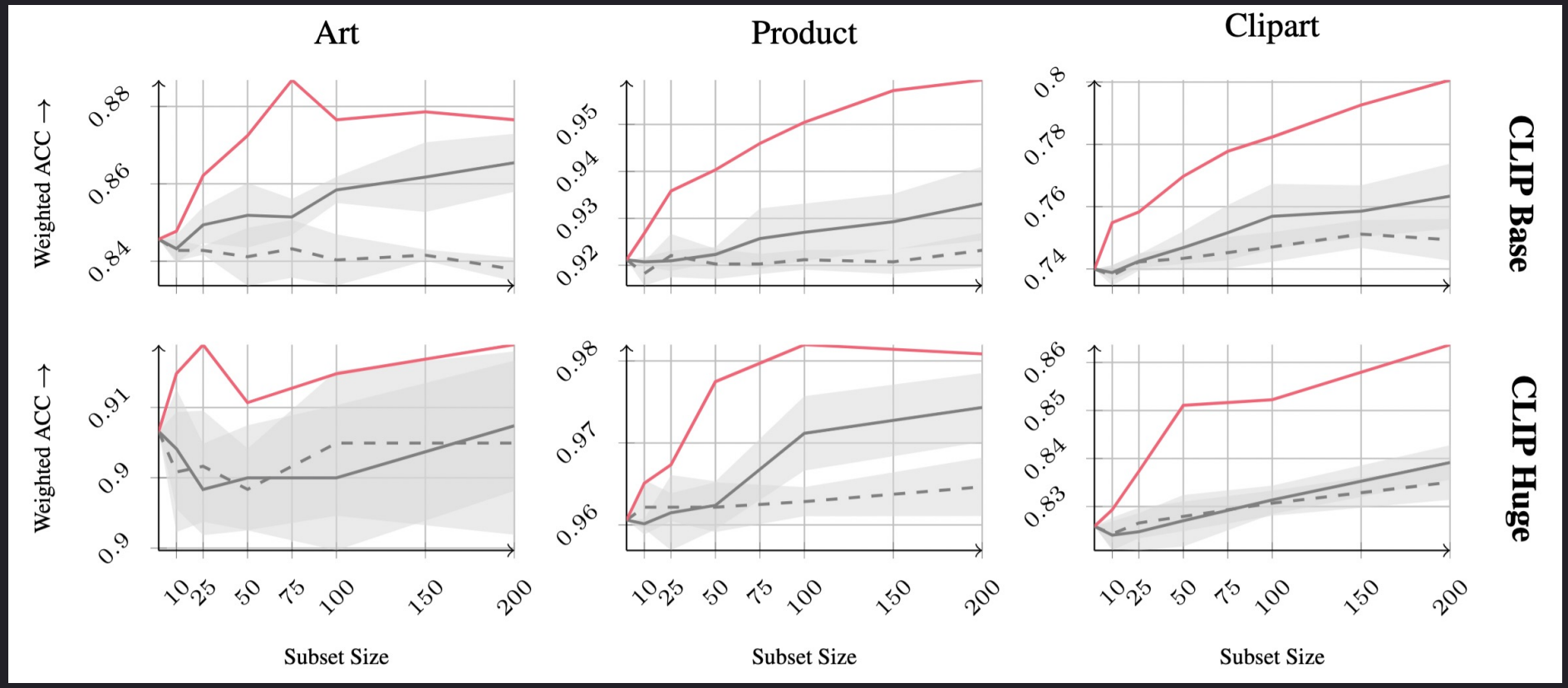
Source: A Baumann *et al.* Probabilistic Active Few-Shot Learning in Vision-Language Models. Under review.

Probabilistic Active Learning in VLMs



Source: A Baumann *et al.* Probabilistic Active Few-Shot Learning in Vision-Language Models. Under review.

Probabilistic Active Learning in VLMs



selection informed by uncertainties

uninformed random/informed random selection

Source: A Baumann *et al.* Probabilistic Active Few-Shot Learning in Vision-Language Models. Under review.

Streamlining Predictions in BDL

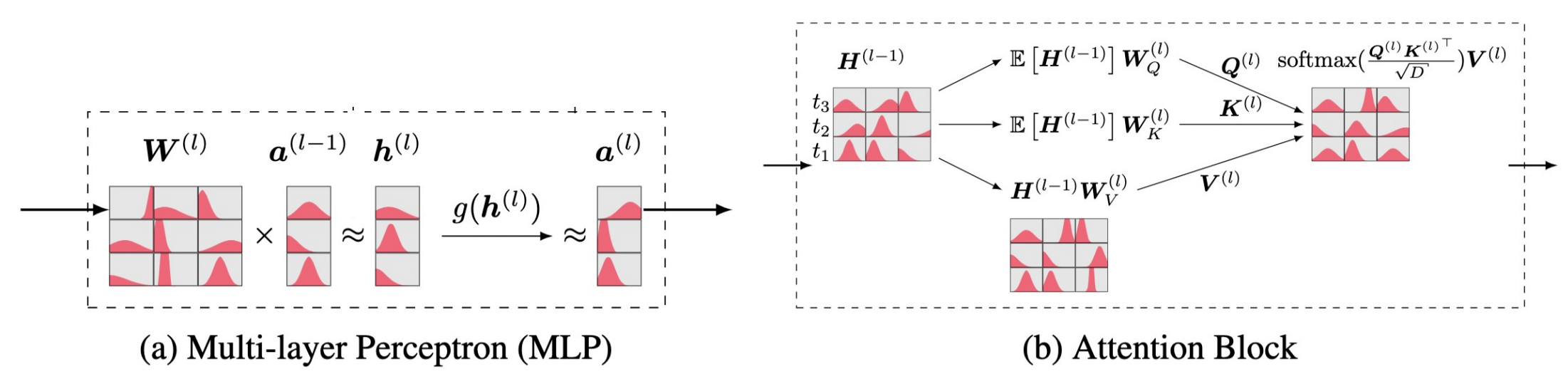
$$p(\boldsymbol{w} \mid \mathcal{D}) = \frac{p(\boldsymbol{w})}{\int}$$

Laplace
Variational Approximation

$$w)\mathrm{d}\boldsymbol{w}$$

$$p(y = c \mid \boldsymbol{x}^*, \mathcal{D}) = \int_{\boldsymbol{w}} p(\boldsymbol{w} \mid \mathcal{D}) p(\boldsymbol{x}^* \mid \boldsymbol{w}) \mathrm{d}\boldsymbol{w}$$

Streamlining Predictions in BDL



(a) Multi-layer Perceptron (MLP)

(b) Attention Block

Source: R Li *et al.* Posterior Inferred, Now What? Streamlining Prediction in Bayesian Deep Learning. Under review.

Challenges

- Laplace approximations are local approximations
- Estimating the Hessian matrix is computational demanding (requires approximations)
- Predictions under the linearized Laplace can be expensive to perform
- Somewhat ad-hoc, requires tuning of the prior variance.
- Unclear, what is a good prior on the weights.

Despite those challenges, current tools can already provide useful estimates in many scenarios.

Ready-Made Libraries

Python (PyTorch)

- Laplace Redux: https://github.com/aleximmer/Laplace

Julia

- LaplaceRedux.jl: https://github.com/JuliaTrustworthyAI/LaplaceRedux.jl

Conclusion

Aalto University
School of Science

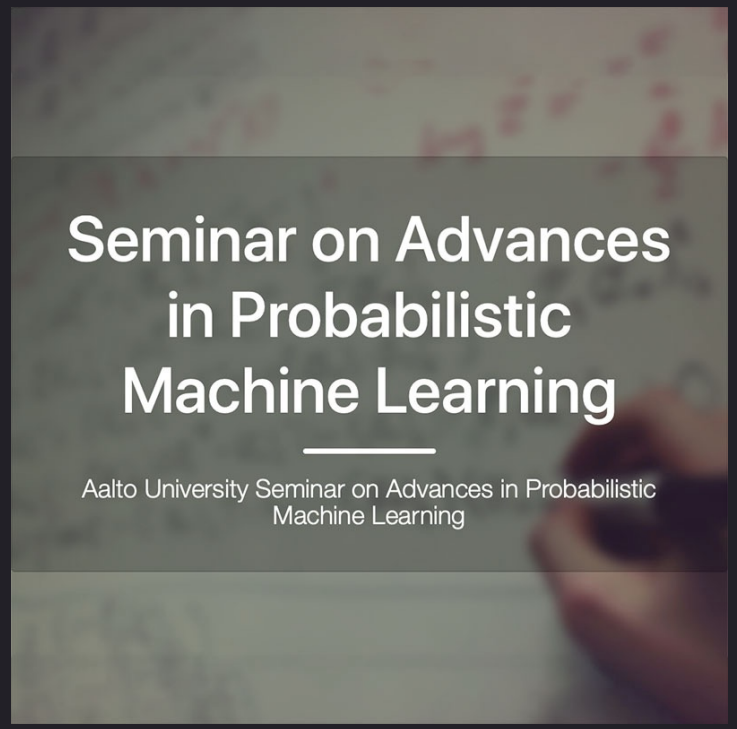
Conclusion

- To get toward reliable ML, we need to account for uncertainties!
- We need to move away from looking only at accuracy!
- Modern (post-hoc) tools can provide useful estimates.

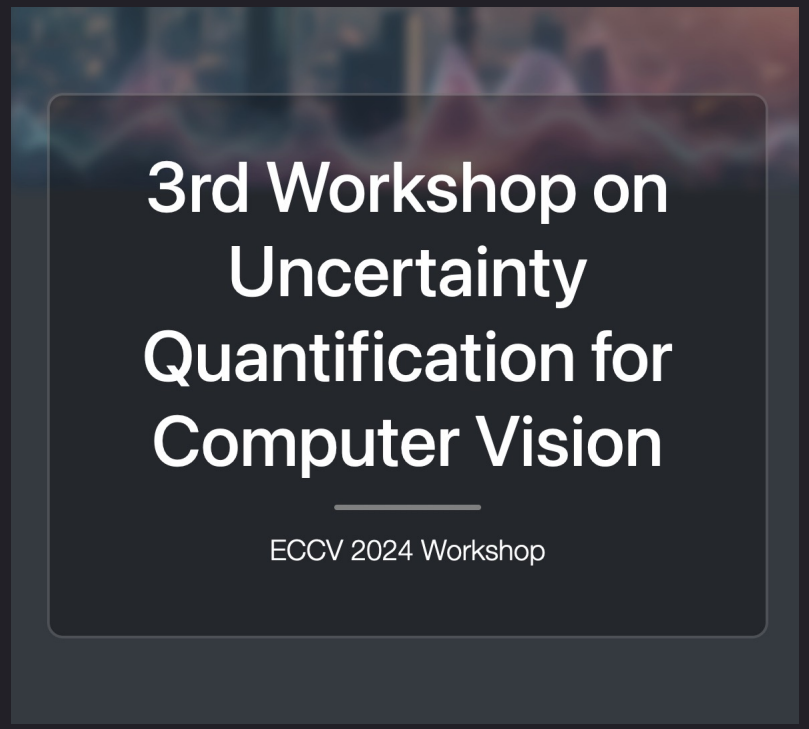
Various open challenges, but the future looks bright!

Aalto University
School of Science

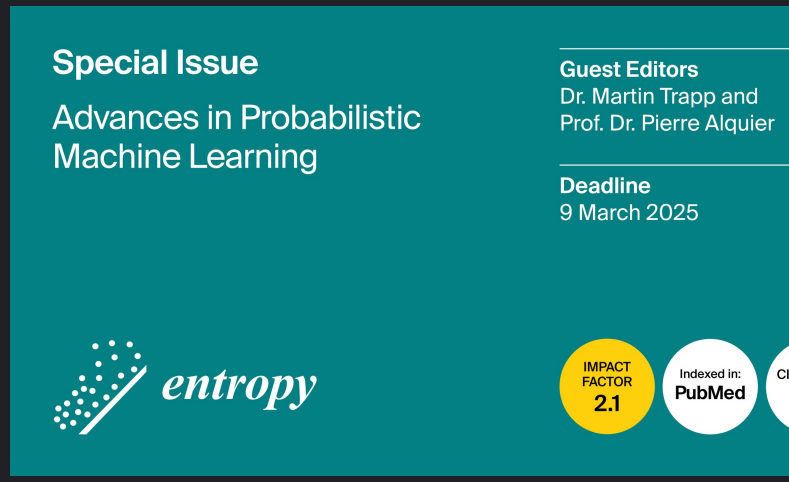
Some Advertisement



https://aaltoml.github.io/apml/



https://uncertainty-cv.github.io



References (order of appearance) 1/2

- P W Koh *et al.* **WILDS: A Benchmark of in-the-Wild Distribution Shifts**. In ICML, 2021. *(https://arxiv.org/abs/2012.07421)*
- S Sabour e*t al.* **SpotlessSplats: Ignoring Distractors in 3D Gaussian Splatting**. In ICML, 2021. *(https://arxiv.org/abs/2406.20055)*
- A Kristiadi, M Hein, P Hennig. **Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks.** In ICML, 2020. *(https://arxiv.org/abs/2002.10118)*
- L Meronen, M Trapp, A Solin. **Periodic activation functions induce stationarity.** In NeurIPS, 2021. *(https://arxiv.org/abs/2110.13572)*
- S Roy *et al.* **Uncertainty-guided Source-free Domain Adaptation.** In ECCV, 2022. *(https://arxiv.org/abs/2208.07591)*
- L Meronen *et al.* **Fixing Overconfidence in Dynamic Neural Networks.** In WACV, 2024. *(https://arxiv.org/abs/2302.06359)*

References (order of appearance) 2/2

- M Klasson *et al.* **Sources of Uncertainty in 3D Scene Reconstruction**. In UNCV Workshop at ECCV, 2024. *(https://www.arxiv.org/abs/2409.06407)*
- T Broederick *et al.* **Toward a Taxonomy of Trust for Probabilistic Machine Learning**. Science Advances 9, eabn399, 2023. *(https://arxiv.org/abs/2112.03270)*
- C Blundell *et al.* **Weight Uncertainty in Neural Networks**, In ICML 2015. *(https://arxiv.org/abs/1505.05424)*
- E Daxberger *et al.* **Laplace Redux - Effortless Bayesian Deep Learning**. In NeurIPS 2021. *(https://arxiv.org/abs/2106.14806)*
- A Baumann *et al.* **Probabilistic Active Few-Shot Learning in Vision-Language Models.** Under review.
- R Li *et al.* **Posterior Inferred, Now What? Streamlining Prediction in Bayesian Deep Learning.** Under review.