Welcome to AIP Open Seminar!

Here are some important notices about this seminar:

1. Reproduction is prohibited.

   複製を禁止します

2. Reproducing all or any part of the contents is prohibited without the author's permission.

   所有者の許可なくコンテンツまたはその一部を転載することを禁じます

3. This seminar will be recorded and the video will be available on the RIKEN AIP website later.

   本セミナーは録画いたします。後日、RIKEN AIPウエブサイトに掲載予定です

4. If you have any questions or comments, please scroll to the bottom of the page and click the **Q and A tab**.

   You can put questions or comments in the box.

   質問は画面下部のQ&A機能をご利用ください

Q and A session

If you have any questions or comments, Please scroll to the bottom of the page and click the Q and A tab.

(You can put comments or questions in the box.)

質問は、画面下部の「Q and A　機能」をご利用ください。

Welcome to AIP Open Seminar!

AIP Open Seminar Series website

https://aip.riken.jp/event-list/seminars/

Schedule

Attendance to the seminar is free of charge, but pre-registration is required from Doorkeeper to obtain the Zoom access link.

- March 17 at 15:00-17:00 JST
  Talks by Structured Learning Team (PI: Yoshinobu Kawahara)
- March 24 at 15:00-17:00 JST
  Talks by Mathematical Science Team (PI: Kenichi Bannai)
- March 31 at 15:00 – 17:00 JST
  Talks by Computational Learning Theory Team (PI: Kohei Hatano)
- April 7 at 15:00 – 17:00 JST
  Talks by Deep Learning Theory Team (PI: Taiji Suzuki)

Bayesian Principles for Learning-Machines

Mohammad Emtiyaz Khan
RIKEN Center for AI Project, Tokyo
http://emtiyaz.github.io

AI that learn like humans

Learn and adapt quickly throughout their lives

Human Learning at the age of 6 months.

Converged at the age of 12 months

Transfer skills

at the age of 14 months

Bayesian Principles

Our research

Human learning $\neq$ **Deep learning**

Life-long learning from small chunks of data in a non-stationary world

Bulk learning from a large amount of data in a stationary world

My current research focuses on reducing this gap!

1. Parisi, German I., et al. "Continual lifelong learning with neural networks: A review." *Neural Networks* (2019)

2. Geisler, W. S., and Randy L. D. "Bayesian natural selection and the evolution of perceptual systems." *Philosophical Transactions of the Royal Society of London. Biological Sciences* (2002)

Approximate Bayesian Inference Team

Emtiyaz Khan
Team Leader

Pierre Alquier
Research Scientist

Gian Maria Marconi
Postdoc

Thomas Möllenhoff
Postdoc

https://team-approx-bayes.github.io/

Wu Lin
PhD Student
University of British Columbia

Dharmesh Tailor
Research Assistant

Fariz Ikhwantri
Part-time Student
Tokyo Institute of Technology

Happy Buzaaba
Part-time Student
University of Tsukuba

Evgenii Egorov
Remote Collaborator
Skoltech

Siddharth Swaroop
Remote Collaborator
University of Cambridge

Dimitri Meunier
Remote Collaborator
ENSAE Paris

Peter Nickl
Remote Collaborator
TU Darmstadt

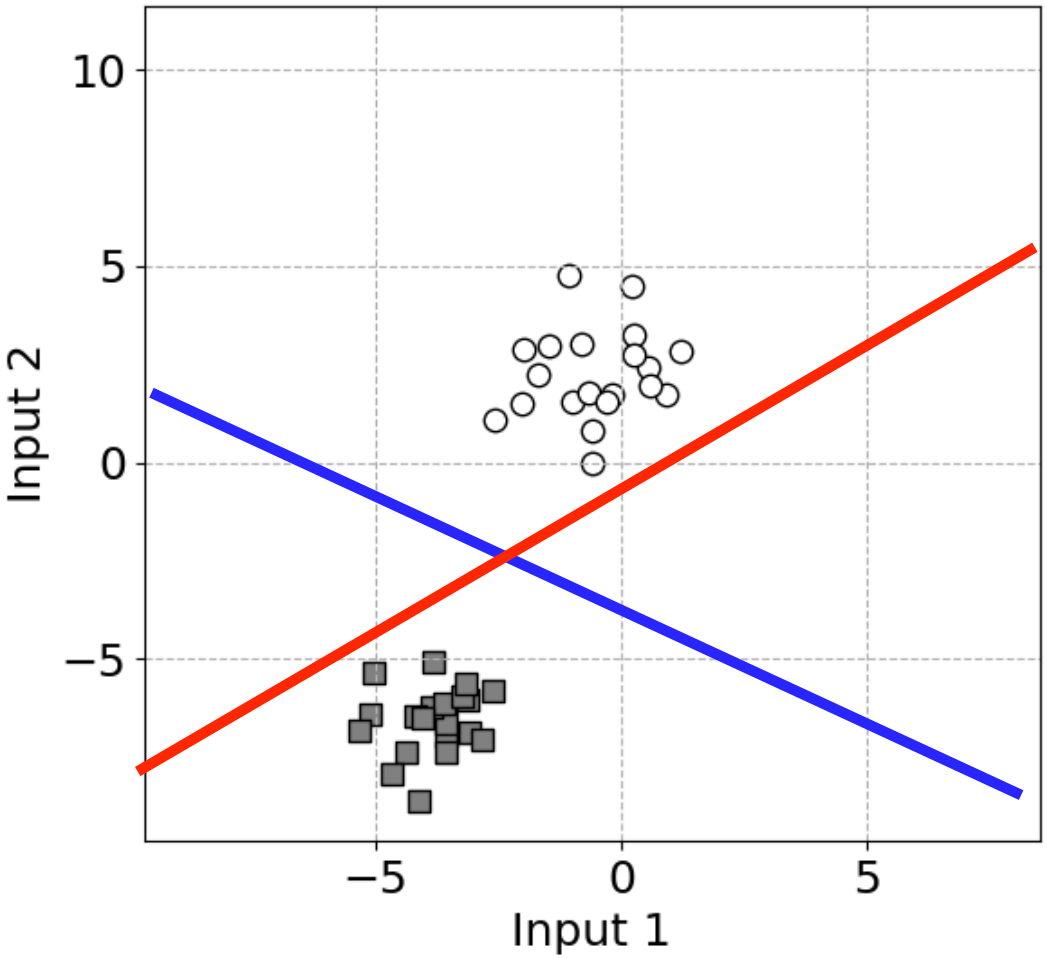
Erik Daxberger
Remote Collaborator
University of Cambridge

Alexandre Piché
Remote Collaborator
MILA

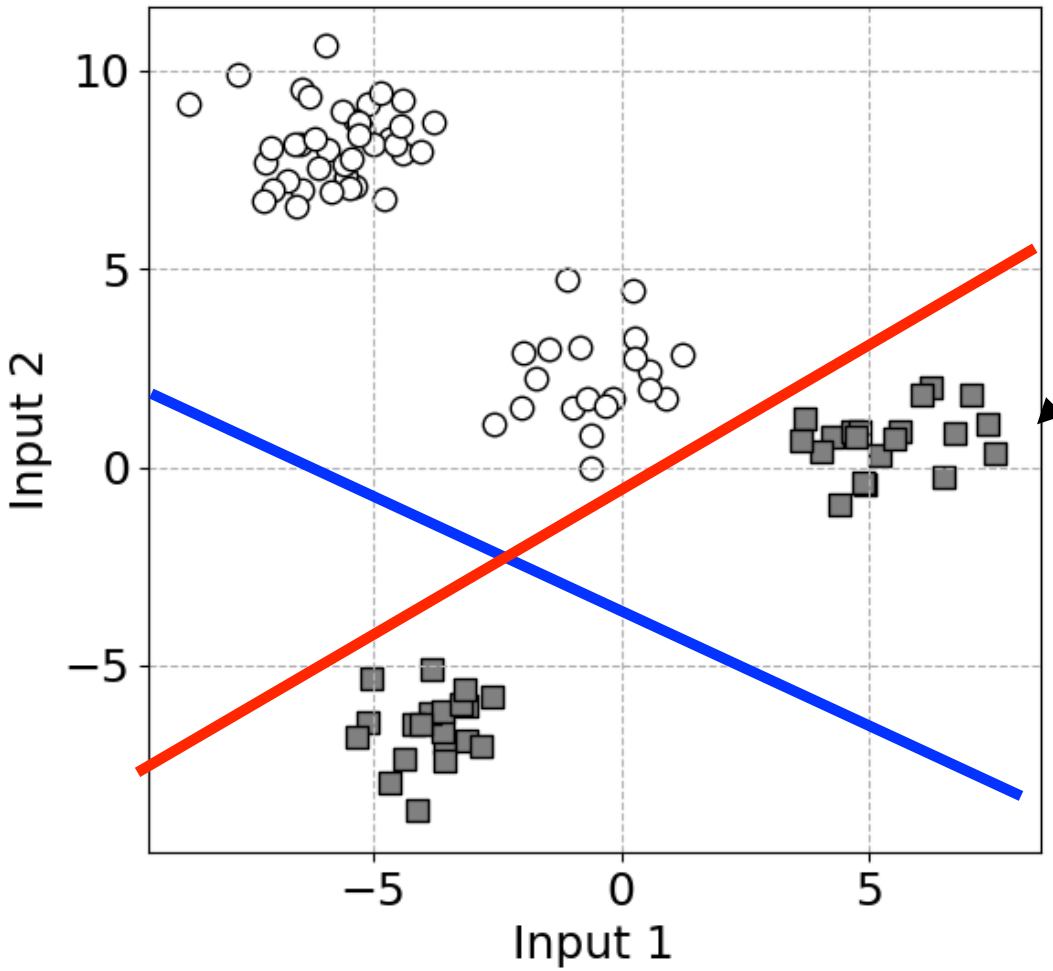
Bayesian (Principles for) Learning-Machines

- Uncertainty (Background)
  - What you don't know now, can hurt you later
- Learning (Past work)
  - Derive learning-algorithms from Bayes
- Knowledge (Current work)
  - Knowledge representation and its transfer
  - Memorable experiences (Dharmesh Tailor)
  - Continual learning (Emti)
  - Meta learning (Pierre Alquier)

Which is a good classifier?



Which is a good classifier?

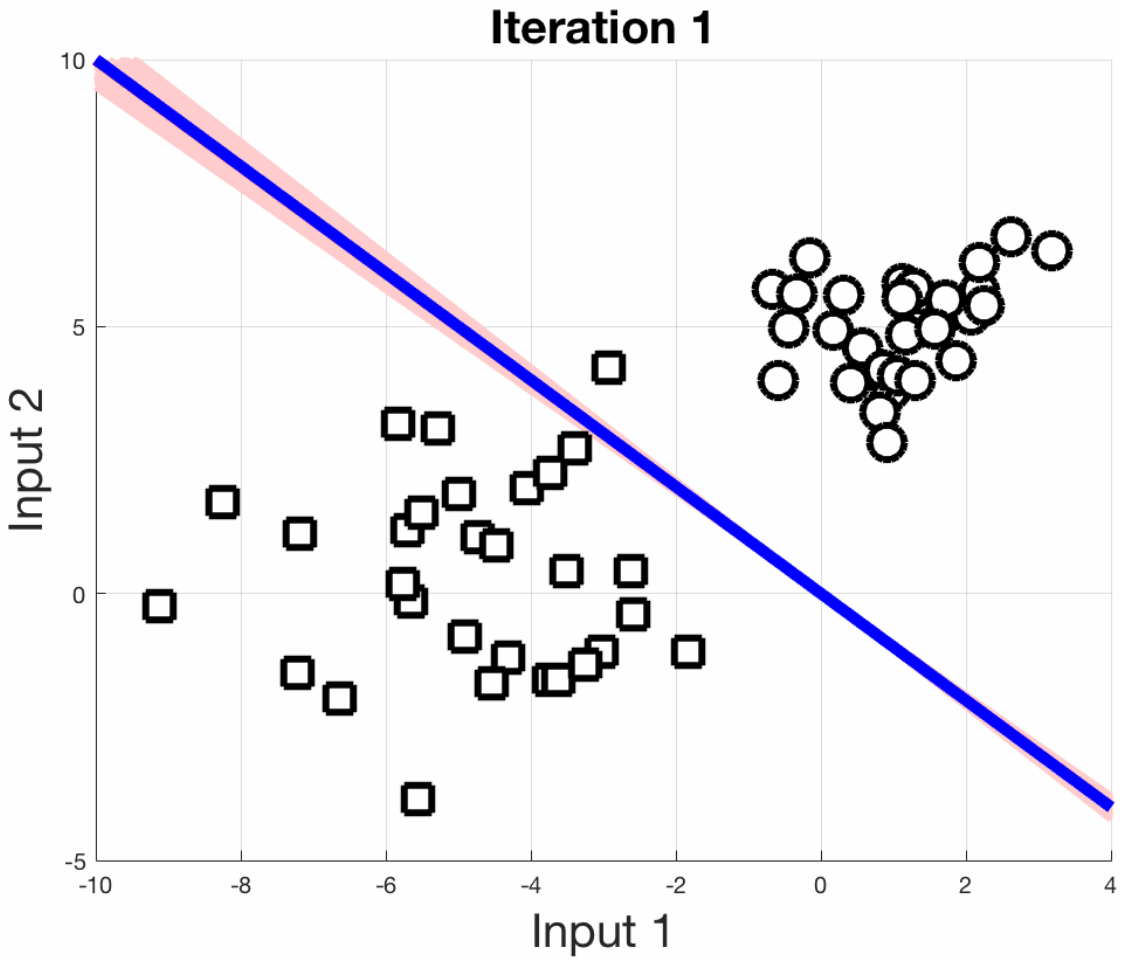


Misclassified by the red line, but not by the blue

What you don't know now, can hurt you later

"Uncertainty matters"

Bayesian Principles



Estimate a distribution over parameters (e.g., Gaussian)

Frequentist (Adam)

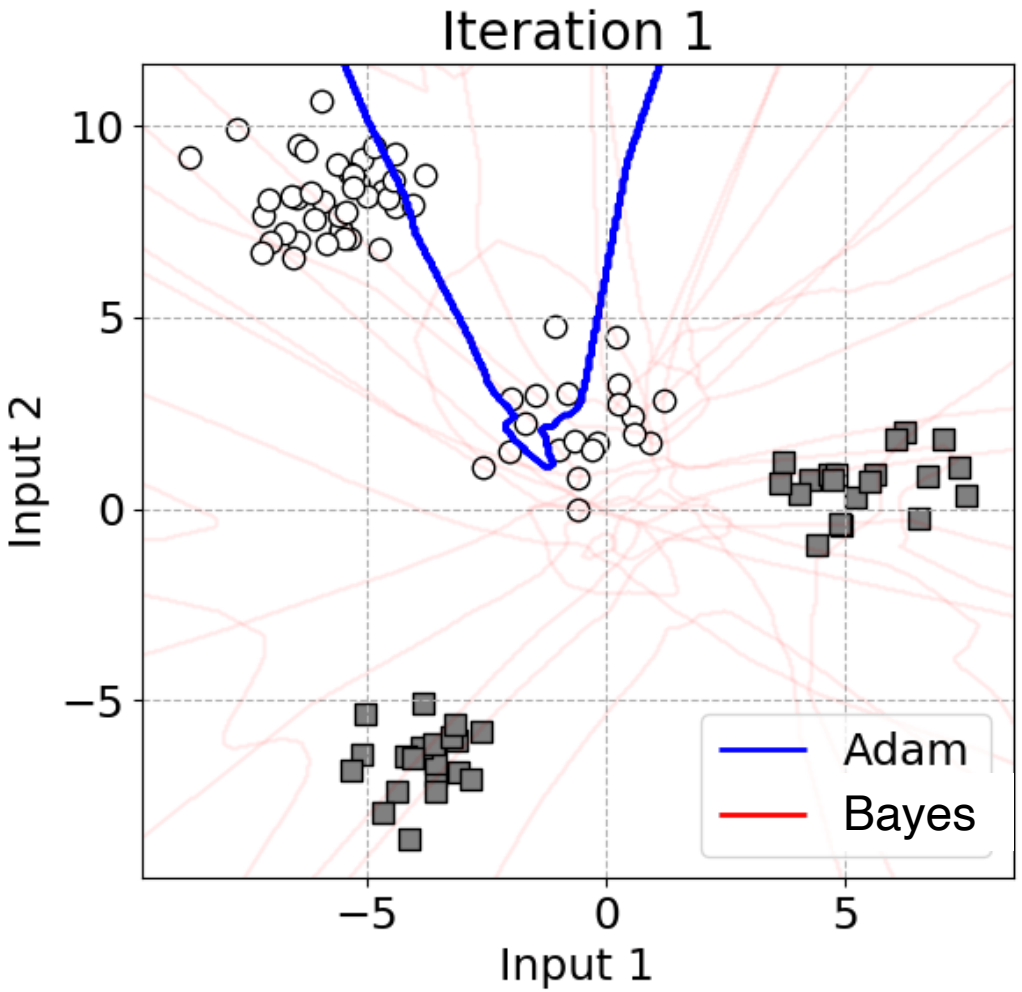
Bayes (VON, mean)

Bayes (samples)

Logistic Regression
Minibatch = 5,
Learning rates = (0.01, 0.01)

1. Khan, et al. "Fast and scalable Bayesian deep learning by weight-perturbation in Adam." *ICML* (2018).

Bayesian Principles



This can be used to quantify the uncertainty of deep networks

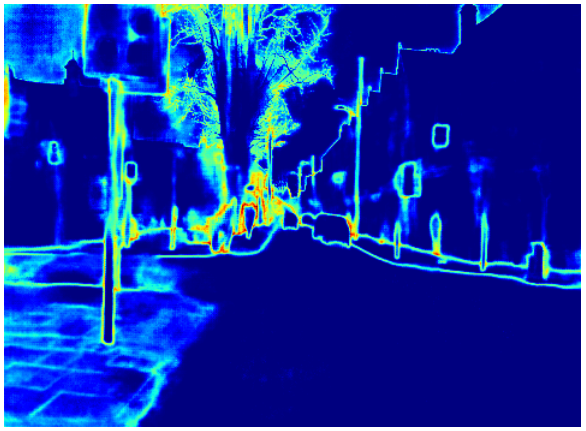
Get the code from https://github.com/team-approx-bayes/dl-with-bayes

1. Khan, et al. "Fast and scalable Bayesian deep learning by weight-perturbation in Adam." *ICML* (2018).
2. Osawa et al. “Practical Deep Learning with Bayesian Principles.” NeurIPS (2019).

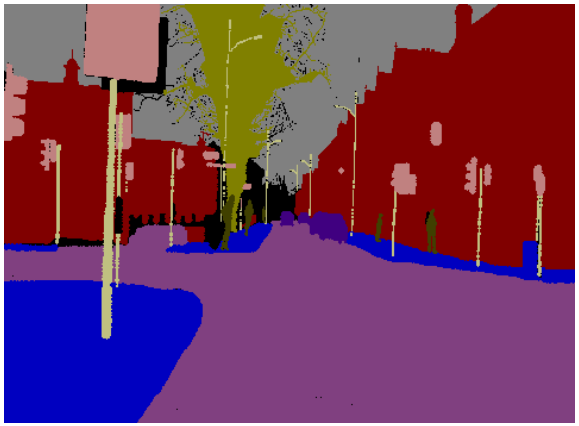
Image Segmentation

Image

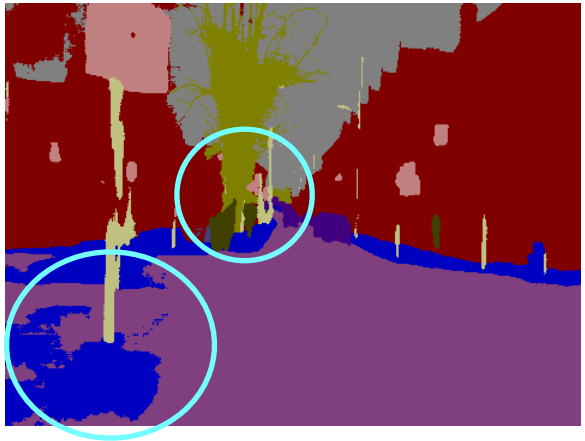
Uncertainty

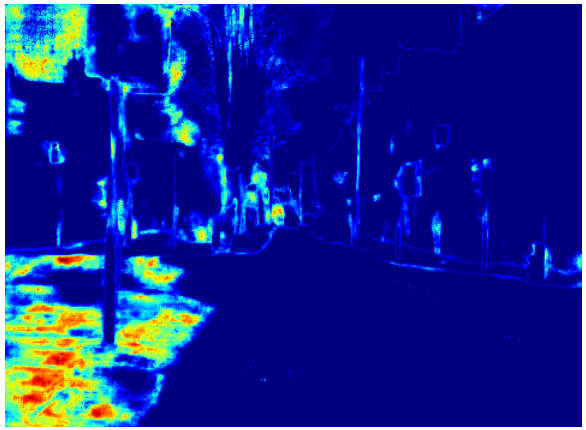


True Segments



Prediction

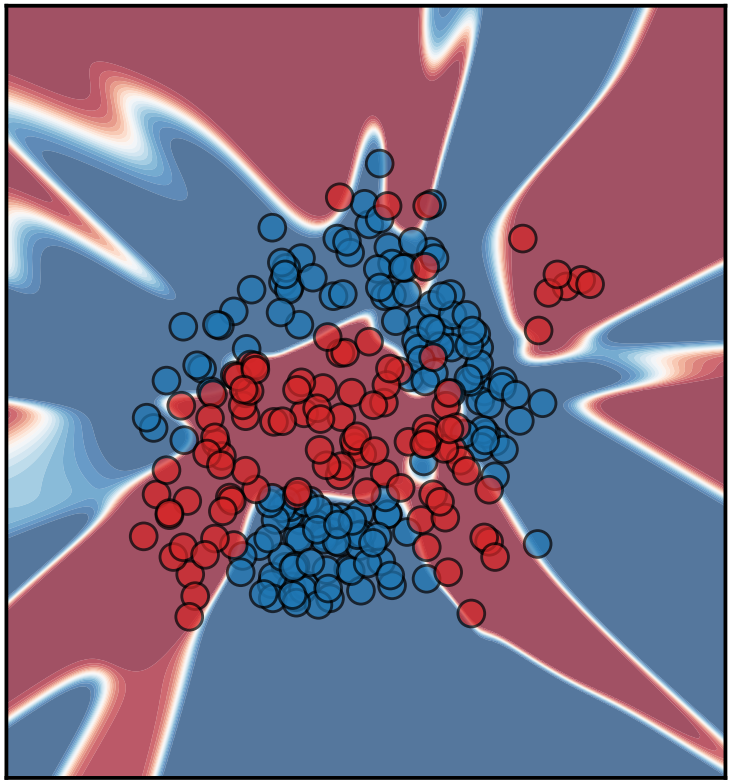




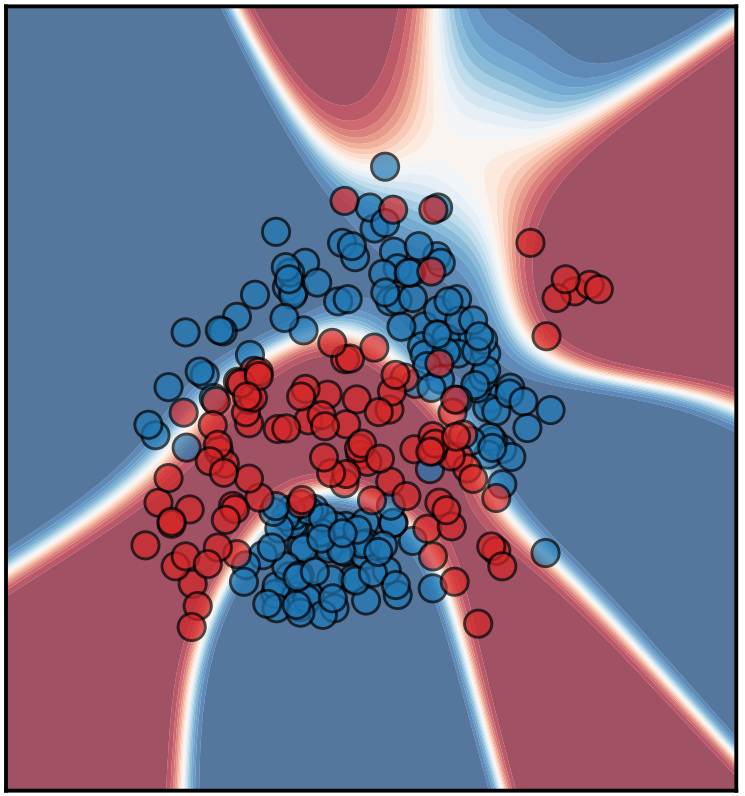
Kendall, Alex, Yarin Gal, and Roberto Cipolla. "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics." *CVPR*. 2018.

Reduce Overfitting

Standard DL



Bayesian DL



Left figure is cross-validation. Right figure is“Marginal Likelihoods”.

Bayesian (Principles for) Learning-Machines

- Uncertainty (Background)
  - What you don't know now, can hurt you later
- Learning (Past work)
  - Derive learning-algorithms from Bayes
- Knowledge (Current work)
  - Knowledge representation and its transfer
  - Memorable experiences (Dharmesh Tailor)
  - Continual learning (Emti)
  - Meta learning (Pierre Alquier)

Bayesian Principles

$$\min_{\theta} \ \ell(\theta) \quad \text{vs} \quad \min_{q \in \mathcal{Q}} \ \mathbb{E}_{q(\theta)}[\ell(\theta)] - \mathcal{H}(q)$$

Entropy

Posterior approximation

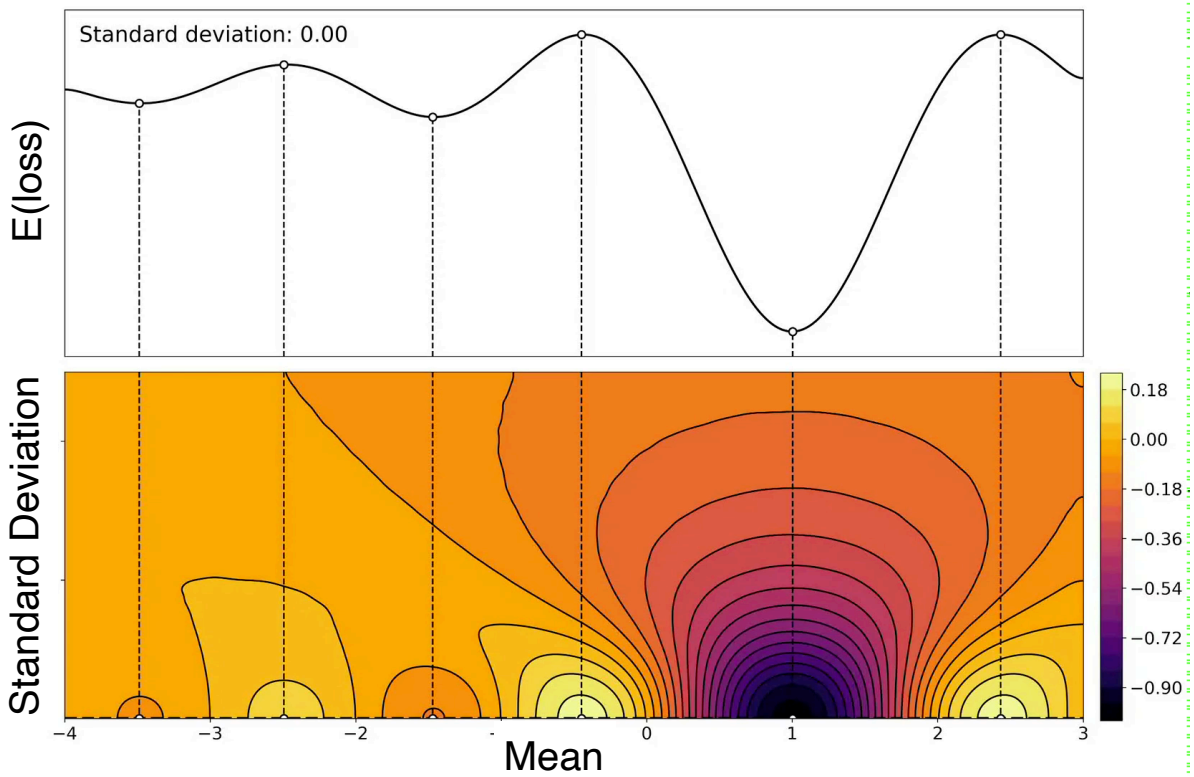
1. Zellner, A. "Optimal information processing and Bayes's theorem." *The American Statistician* (1988)

Bayesian Principles

$$\min_{\theta} \ \ell(\theta) \quad \text{vs} \quad \min_{q \in \mathcal{Q}} \ \mathbb{E}_{q(\theta)}[\ell(\theta)] - \mathcal{H}(q)$$

Entropy

Posterior approximation



Common in

- Search
- Inference
- Optimization
- Online learning
- Reinforcement learning

1. Zellner, A. "Optimal information processing and Bayes's theorem." *The American Statistician* (1988)
2. Huszar's blog, Evolution Strategies, Variational Optimisation and Natural ES (2017)
3. Khan et al. "Variational adaptive-Newton method for explorative learning." *arXiv* (2017).

Bayesian Learning Rule

$$\min_{\theta} \ \ell(\theta) \quad \text{vs} \quad \min_{q \in \mathcal{Q}} \ \mathbb{E}_{q(\theta)}[\ell(\theta)] - \mathcal{H}(q)$$

Exponential-family Approx.

Deep Learning algo: $\theta \leftarrow \theta - \rho H_{\theta}^{-1} \nabla_{\theta} \ell(\theta)$

Bayes learning rule: $\lambda \leftarrow \lambda - \rho \nabla_{\mu} \left(\mathbb{E}_q[\ell(\theta)] - \mathcal{H}(q) \right)$

Natural Gradient

Natural and Expectation parameters of an exponential family distribution q

By changing *Q*, we can recover DL algorithms (and more)

1. Khan and Lin. "Conjugate-computation variational inference: Converting variational inference in non-conjugate models to inferences in conjugate models." AIstats (2017).
2. Khan and Rue. "Learning-Algorithms from Bayesian Principles" (2020) (work in progress, an early draft available at https://emtiyaz.github.io/papers/learning_from_bayes.pdf)

Gradient Descent from Bayes

Gradient descent: $\theta \leftarrow \theta - \rho \nabla_\theta \ell(\theta)$

Bayes Learn Rule: $m \leftarrow m - \rho \nabla_m \ell(m)$

"Global" to "local" (delta approx)
$\mathbb{E}_q[\ell(\theta)] \approx \ell(m)$

$$m \leftarrow m - \rho \nabla_m \mathbb{E}_q[\ell(\theta)]$$

$$\lambda \leftarrow \lambda - \rho \nabla_\mu \left(\mathbb{E}_q[\ell(\theta)] - \mathcal{H}(q)\right)$$

Derived by choosing Gaussian with fixed covariance

Gaussian distribution	$q(\theta) := \mathcal{N}(m, 1)$
Natural parameters	$\lambda := m$
Expectation parameters	$\mu := \mathbb{E}_q[\theta] = m$
Entropy	$\mathcal{H}(q) := \log(2\pi)/2$

1. Khan and Rue. "Learning-Algorithms from Bayesian Principles" (2020) (work in progress, an early draft available at https://emtiyaz.github.io/papers/learning_from_bayes.pdf)

Bayesian learning rule: $\lambda \leftarrow \lambda - \rho \nabla_{\mu} \left(\mathbb{E}_q[\ell(\theta)] - \mathcal{H}(q) \right)$

Learning Algorithm	Posterior Approx.	Algorithmic Approx.	Sec.
Optimization Algorithms			
Gradient Descent	Gaussian (fixed cov.)	Delta approx.	1.4
Newton's method	Gaussian	—"—	1.4
Multimodel optimization (New)	Mixture of Gaussians	—"—	3.2
Deep-Learning Algorithms			
Stochastic Gradient Descent	Gaussian (fixed cov.)	Delta approx., Stochastic approx.	4.1
RMSprop/Adam	Gaussian (diagonal cov.)	Delta approx., Stochastic approx., Hessian approx., Square-root scaling, Slow-moving scale vectors	4.2, 4.3
Dropout	Mixture of Gaussians	Delta approx., Stochastic approx., Responsibility approx.	4.4
STE	Bernoulli	Delta approx., Stochastic approx.	4.6
Online Gauss-Newton (OGN) (New)	Gaussian (diagonal cov.)	Gauss-Newton Hessian approx. in Adam & no square-root scaling	4.5
Variational OGN (New)	—"—	Remove Delta approx. from OGN	4.5
Bayesian Binary NN (New)	—"—	Remove Delta approx. from STE	4.6
Approximate Bayesian Inference Algorithms			
Conjugate Bayes	Exp-family	Set learning rate $\rho_t = 1$	5.1
Laplace's method	Gaussian	Delta approx.	5.2
Expectation-Maximization	Exp-Family + Gaussian	Delta approx. for the parameters	5.3
Stochastic VI (SVI)	Exp-family (mean-field)	Stochastic approx., local rate $\rho_t = 1$	5.4
VMP	—"—	Set learning rate $\rho_t = 1$	5.4
Non-Conjugate VMP	—"—	—"—	5.4
Non-Conjugate VI (New)	Mixture of Exp-family	None	5.5

Khan and Rue. "Learning-Algorithms from Bayesian Principles" (2020)

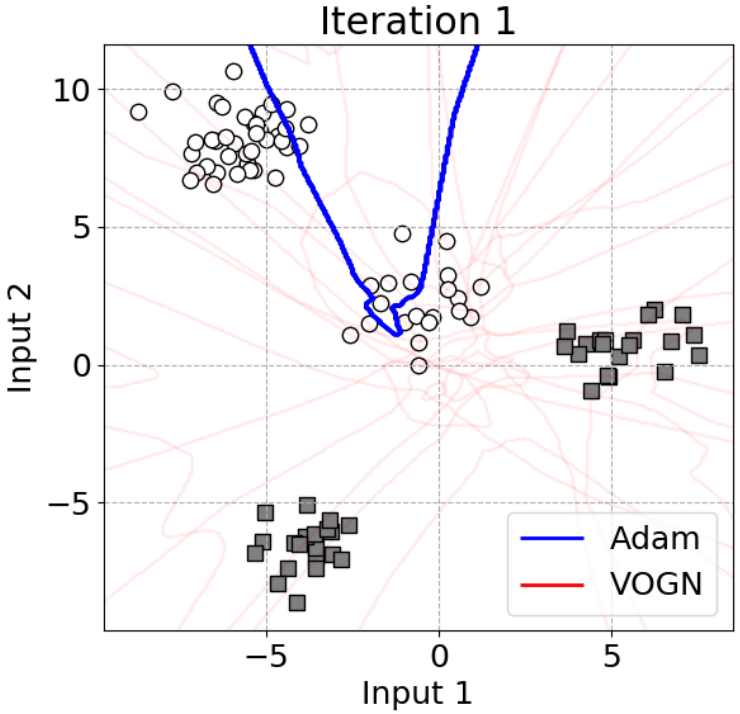
Work in progress (draft available at https://emtiyaz.github.io/papers/learning_from_bayes.pdf)

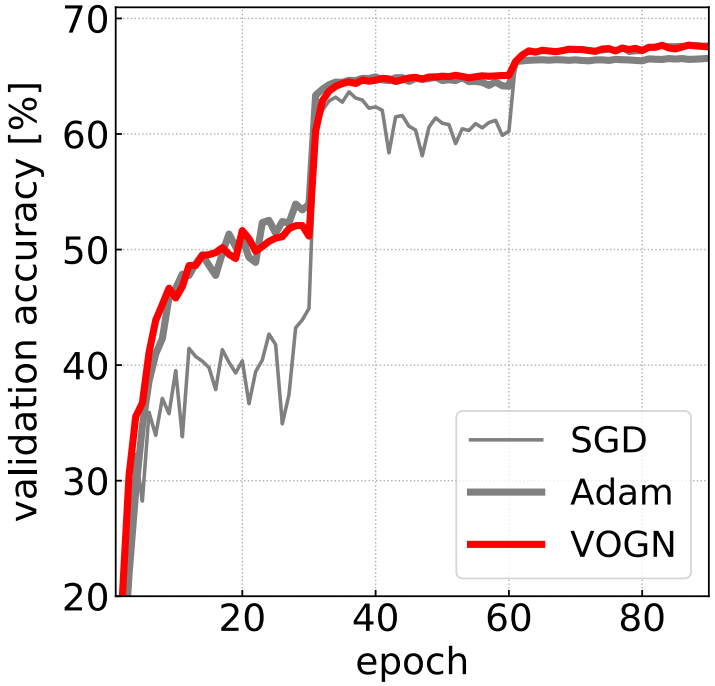
We can compute uncertainty using a variant of Adam.

1. Khan, et al. "Fast and scalable Bayesian deep learning by weight-perturbation in Adam." *ICML* (2018).
2. Osawa et al. "Practical Deep Learning with Bayesian Principles." NeurIPS (2019).

Uncertainty of Deep Nets

VOGN: A modification of Adam but match the performance on ImageNet





Code available at https://github.com/team-approx-bayes/dl-with-bayes

1. Khan, et al. "Fast and scalable Bayesian deep learning by weight-perturbation in Adam." *ICML* (2018).
2. Osawa et al. "Practical Deep Learning with Bayesian Principles." NeurIPS (2019).

Image Segmentation

Uncertainty (entropy of class probs)

(By Roman Bachmann)

Learning-Algorithms from Bayesian Principles

Mohammad Emtiyaz Khan
RIKEN center for Advanced Intelligence Project
Tokyo, Japan

Håvard Rue
CEMSE Division
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia

Version of November 3, 2020
DRAFT ONLY

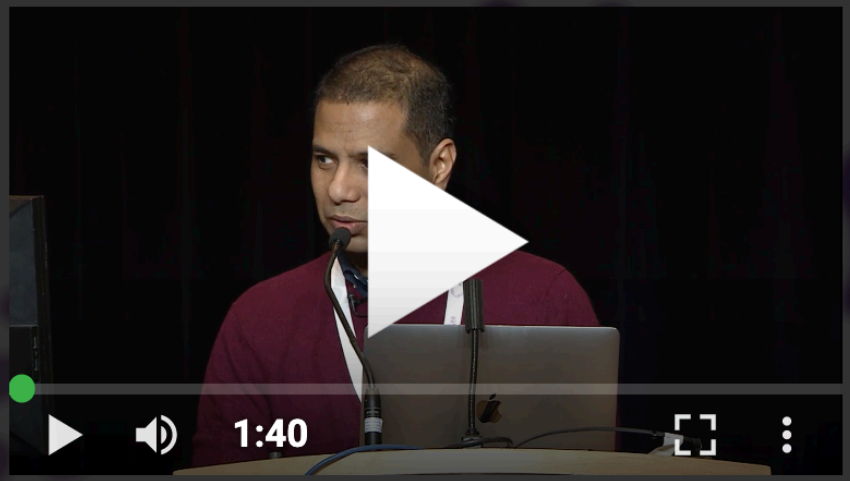
Abstract

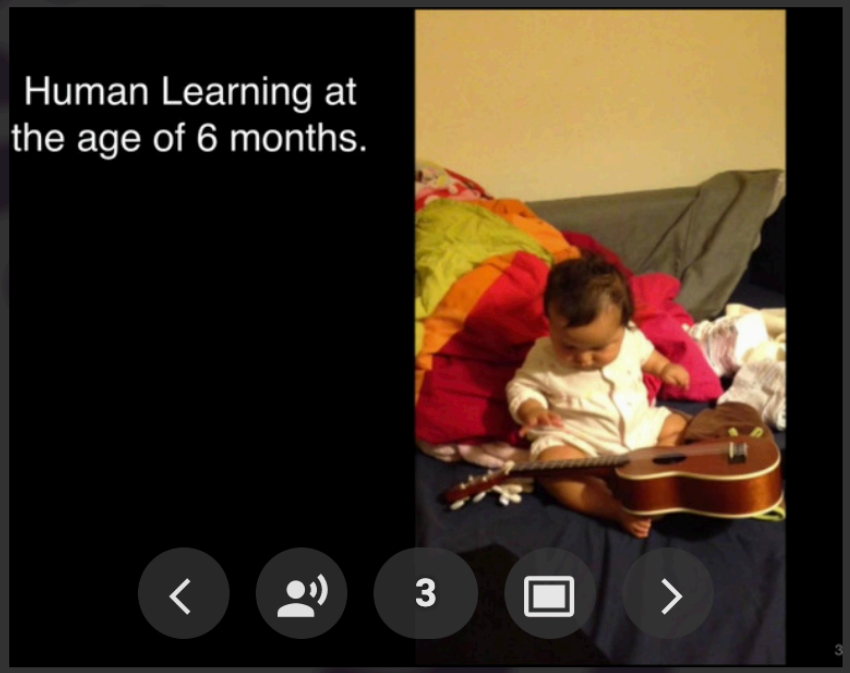
We show that many machine-learning algorithms are specific instances of a *single* algorithm called the Bayesian learning rule. The rule, derived from Bayesian principles, yields a wide-range of algorithms from fields such as optimization, deep learning, and graphical models. This includes classical algorithms such as ridge regression, Newton's method, and Kalman filter, as well as modern deep-learning algorithms such as stochastic-gradient descent, RMSprop, Adam, and Dropout. The key idea is to estimate posterior approximations using the Bayesian learning rule. Different approximations then result in different algorithms and further algorithmic approximations give rise to variants of those algorithms. Our work shows that Bayesian principles not only unify, generalize, and improve existing learning-algorithms, but also help us design new ones.

Available at https://emtiyaz.github.io/papers/learning_from_bayes.pdf

1. Olivier et al. "Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles", JMLR (2017).

NeurIPS 2019 Tutorial





Deep Learning with Bayesian Principles

by **Mohammad Emtiyaz Khan** · Dec 9, 2019 ·

#NeurIPS 2019 Follow

Views **151 807** Presentations **263** Followers **200**

Latest Popular ···

From System 1 Deep Learning to System 2 Deep Learning

by Yoshua Bengio

17,953 views · Dec 11, 2019

NeurIPS Workshop on Machine Learning for Creativity and Design...

by Aaron Hertzmann, Adam Roberts, ...

9,654 views · Dec 14, 2019

Deep Learning with Bayesian Principles

by Mohammad Emtiyaz Khan

8,084 views · Dec 9, 2019

Efficient Processing of Deep Neural Network: from Algorithms to...

by Vivienne Sze

7,163 views · Dec 9, 2019

Past and New Work

- Natural Gradient Variational Inference
  1. Khan and Lin. "Conjugate-computation variational inference: Converting variational inference in non-conjugate models to inferences in conjugate models." AIstats (2017).
  2. Khan and Nielsen. "Fast yet simple natural-gradient descent for variational inference in complex models." (*2018) ISITA*.
- Mixture of Exponential family
  3. Lin et al. "Fast and Simple Natural-Gradient Variational Inference with Mixture of Exponential-family Approximations," ICML (2019).
- Generalization of natural gradients
  4. Lin et al. "Handling the Positive-Definite Constraint in the Bayesian Learning Rule", ICML (2020)
  5. Lin et al. "Tractable structured natural gradient descent using local parameterizations", under review, (2021)
- Gaussian approx <=> Newton-variants

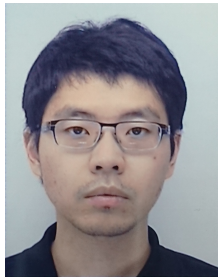
Wu Lin (UBC)

Mark Schmidt (UBC)

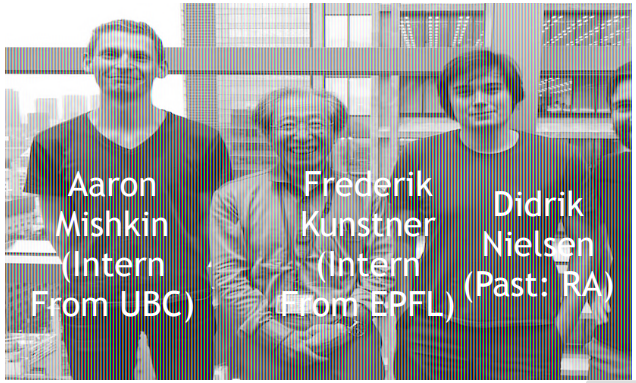
Frank Nielsen (Sony)

Gaussian Approximation and DL

1. Khan, et al. "Fast and scalable Bayesian deep learning by weight-perturbation in Adam." *ICML* (2018).
2. Mishkin et al. “SLANG: Fast Structured Covariance Approximations for Bayesian Deep Learning with Natural Gradient” NeurIPS (2018).
3. Osawa et al. “Practical Deep Learning with Bayesian Principles.” NeurIPS (2019).



Voot Tangkaratt
(Postdoc, RIKEN-AIP)



Yarin Gal
(UOxford)

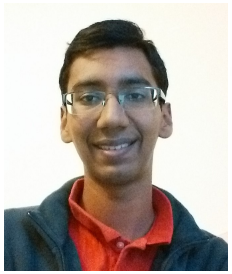
Akash Srivastava
(UEdinburgh)

Kazuki Osawa
(Tokyo Tech)

Rio Yokota
(Tokyo Tech)

Anirudh Jain
(Intern from
IIT-ISM, India)

Runa Eschenhagen
(Intern from
U Osnabruck)

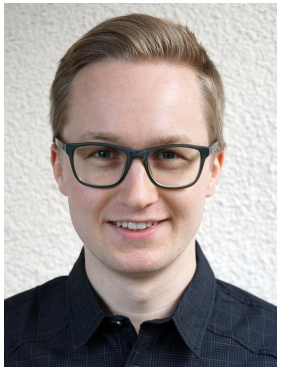


Siddharth
Swaroop
(UCambridge)

Rich Turner
(UCambridge)

Extensions

- Binary Neural Networks (Bernoulli approx)
  1. Meng, et al. "Training Binary Neural Networks using the Bayesian Learning Rule." *ICML* (2020).
- Gaussian Process
  2. Chang et al. “Fast Variational Learning in State-Space GP Models”, MLSP (2020)
  - For sparse GPs, BLR is a generalization of [1]

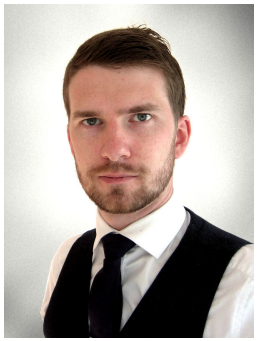


Roman Bachmann (Intern from EPFL)

Xiangming Meng (RIKEN-AIP)

Paul Chang (Aalto University)

W. J. Wilkinson (Aalto University)



Arno Solin (Aalto University)

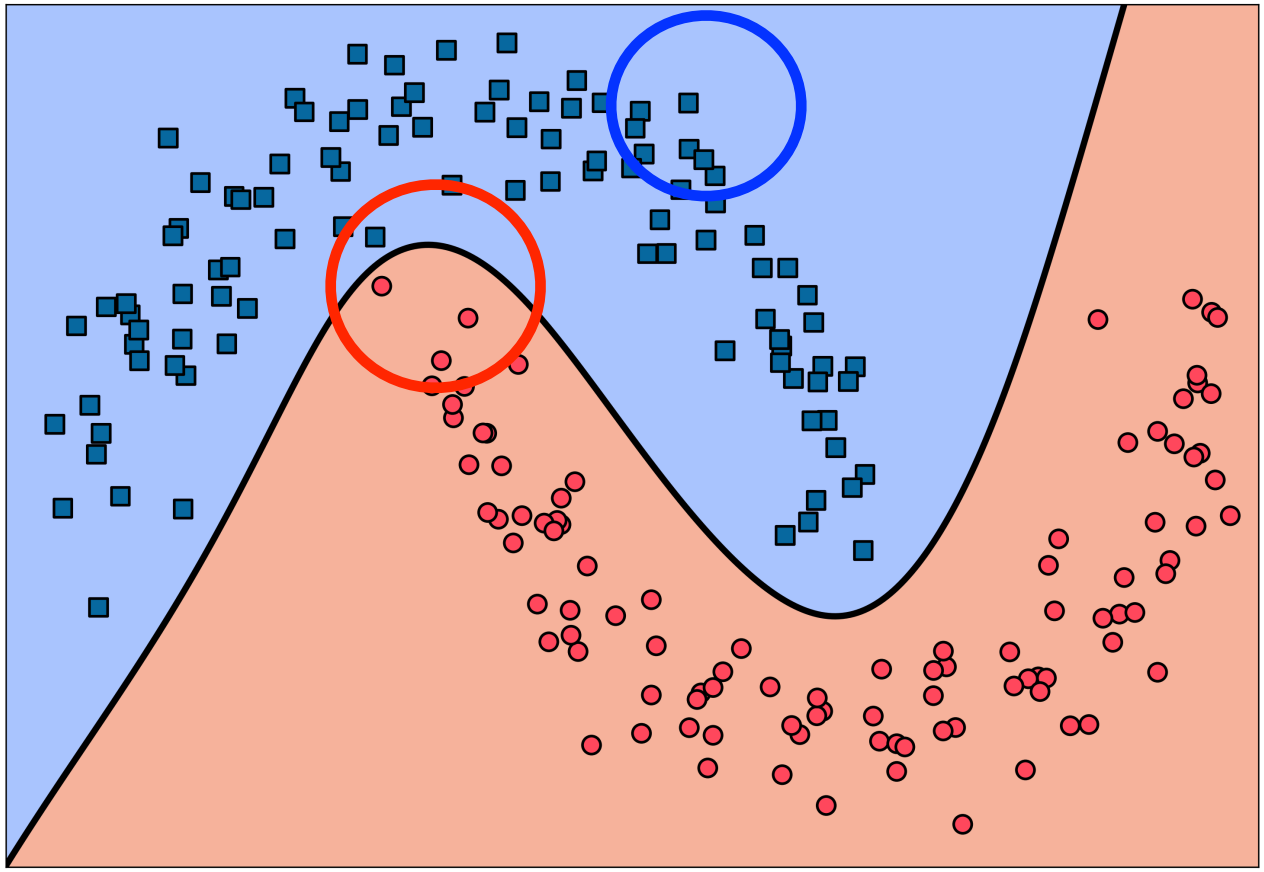
1. Hensman et al. “Gaussian Process for Big Data”, UAI (2013)

Bayesian (Principles for) Learning-Machines

- Uncertainty (Background, 10 mins)
  - What you don't know now, can hurt you later
- Learning (Past work, 20 mins)
  - Derive learning-algorithms from Bayes
- Knowledge (Current work)
  - Knowledge representation and its transfer
  - Memorable experiences (Dharmesh Tailor, 20 mins)
  - Continual learning (Emti, 15 mins)
  - Meta learning (Pierre Alquier, 30 mins)

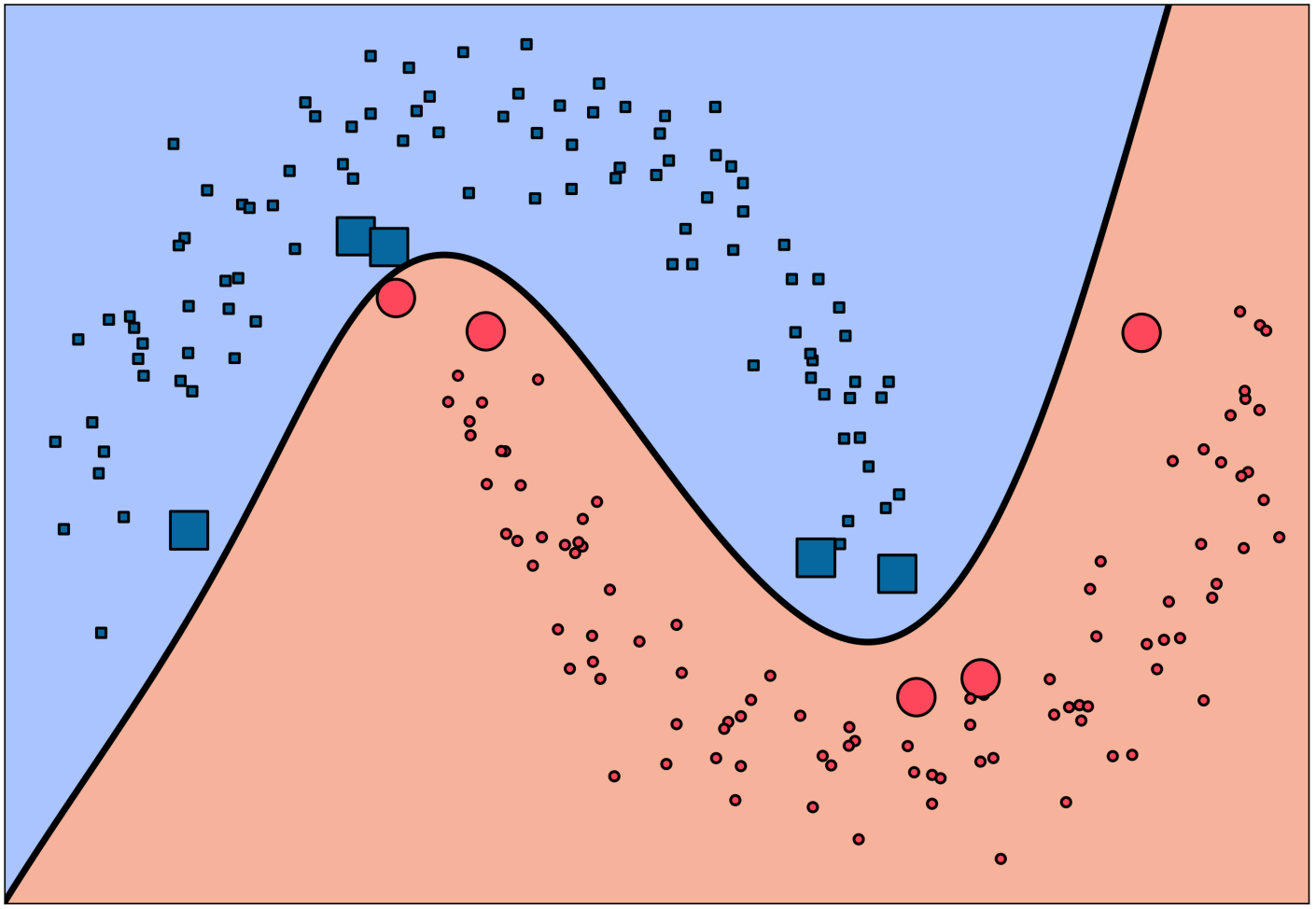
Relevance of Data Examples

Which examples are most relevant for the classifier? Red circle vs Blue circle.



Model view vs Data view

Bayes “automatically” defines data-relevance



Data view (Very much like SVMs)

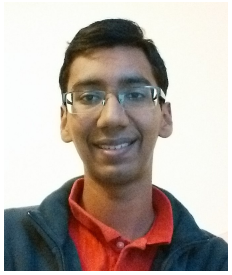
(By Roman Bachmann)

Bayesian (Principles for) Learning-Machines

- Uncertainty (Background)
  - What you don't know now, can hurt you later
- Learning (Past work)
  - Derive learning-algorithms from Bayes
- Knowledge (Current work)
  - Knowledge representation and its transfer
  - Memorable experiences (Dharmesh Tailor)
  - Continual learning (Emti)
  - Meta learning (Pierre Alquier)

Continual Learning with Bayes

PingBo Pan
(Intern from UT Sydney)



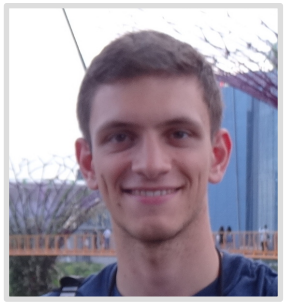
Siddharth Swaroop
(University of Cambridge)

Runa Eschenhagen
(Intern from University of Osnabruck)

Rich Turner
(University of Cambridge)

Alexander Immer
(Intern from EPFL)

Ehsan Abedi
(Intern from EPFL)



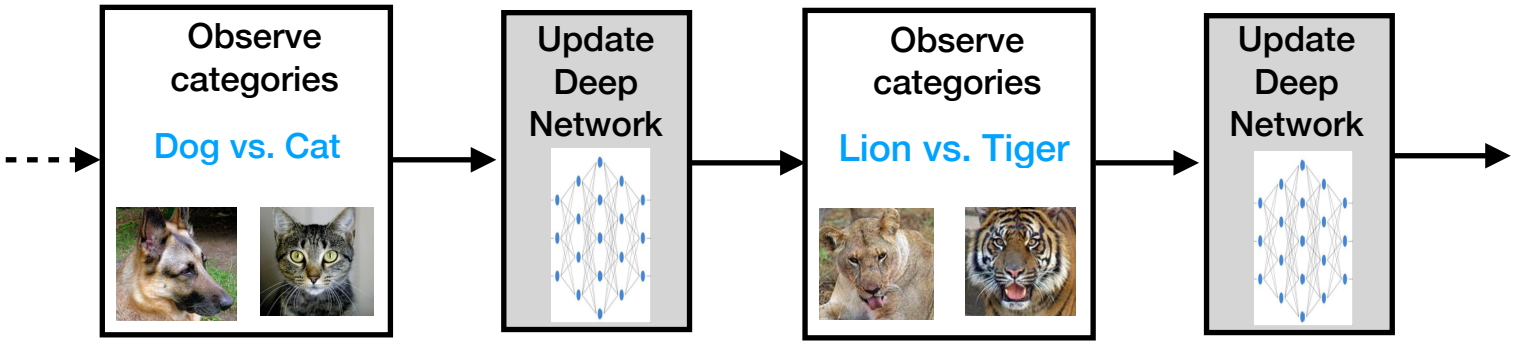
Maciej Korzepa
(Intern from DTU)

1. Khan et al. “Approximate Inference Turns Deep Networks into Gaussian Process”, NeurIPS, 2019
2. *Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020*

Continual Learning

Standard Deep Learning

Continual Learning: past classes never revisited



Standard training leads to catastrophic forgetting.

Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *Proceedings of the national academy of sciences* 114.13 (2017): 3521-3526.

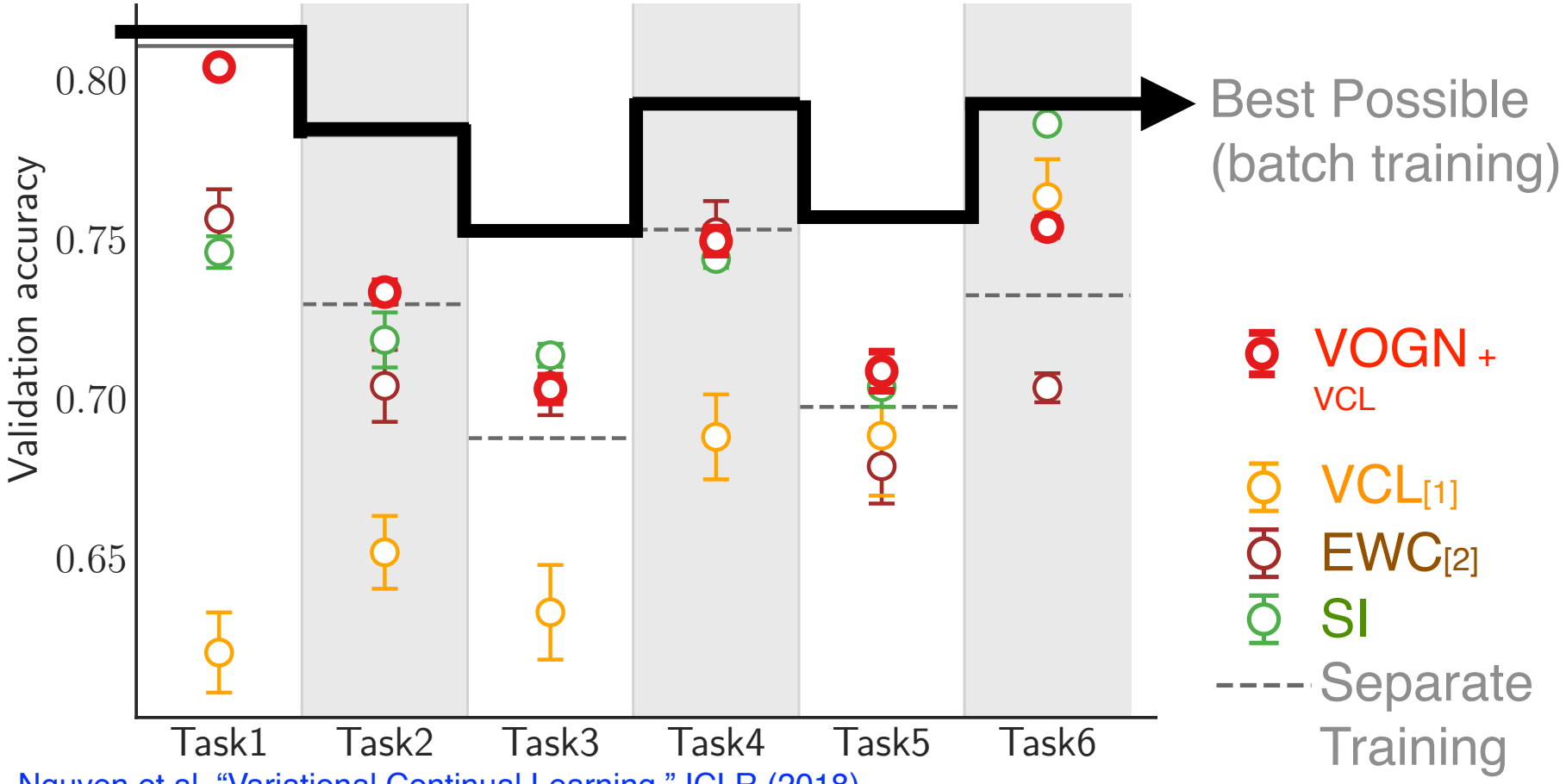
Existing Continual Learning Methods

- Weight regularization
  - Elastic-weight consolidation (EWC) [1]
  - Structured Laplace [2]
  - Synaptic Intelligence (SI) [3]
  - Variational Continual learning (VCL) [4]
- Memory-based
  - Learning without forgetting [5] (and many more..)
  - Gradient Episodic Memory [6] (and many more..)
- Functional Regularization [7]

1. Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *PNAS* (2017).
2. Ritter et al. "Online structured laplace … for overcoming catastrophic forgetting." *NeurIPs*. 2018.
3. Zenke et al. "Continual learning through synaptic intelligence." *ICML,* 2017.
4. Nguyen, Cuong V., et al. "Variational continual learning." *arXiv preprint arXiv:1710.10628* (2017).
5. Li and Hoem, “Learning without forgetting”, IEEE PAMI (2017)
6. Lopez-Paz, Ronzato, “Gradient episodic memory for continual learning”, NeurIPs (2017)
7. Titsias et al., “Functional Regularisation for Continual Learning with Gaussian Processes”, ICLR (2020)

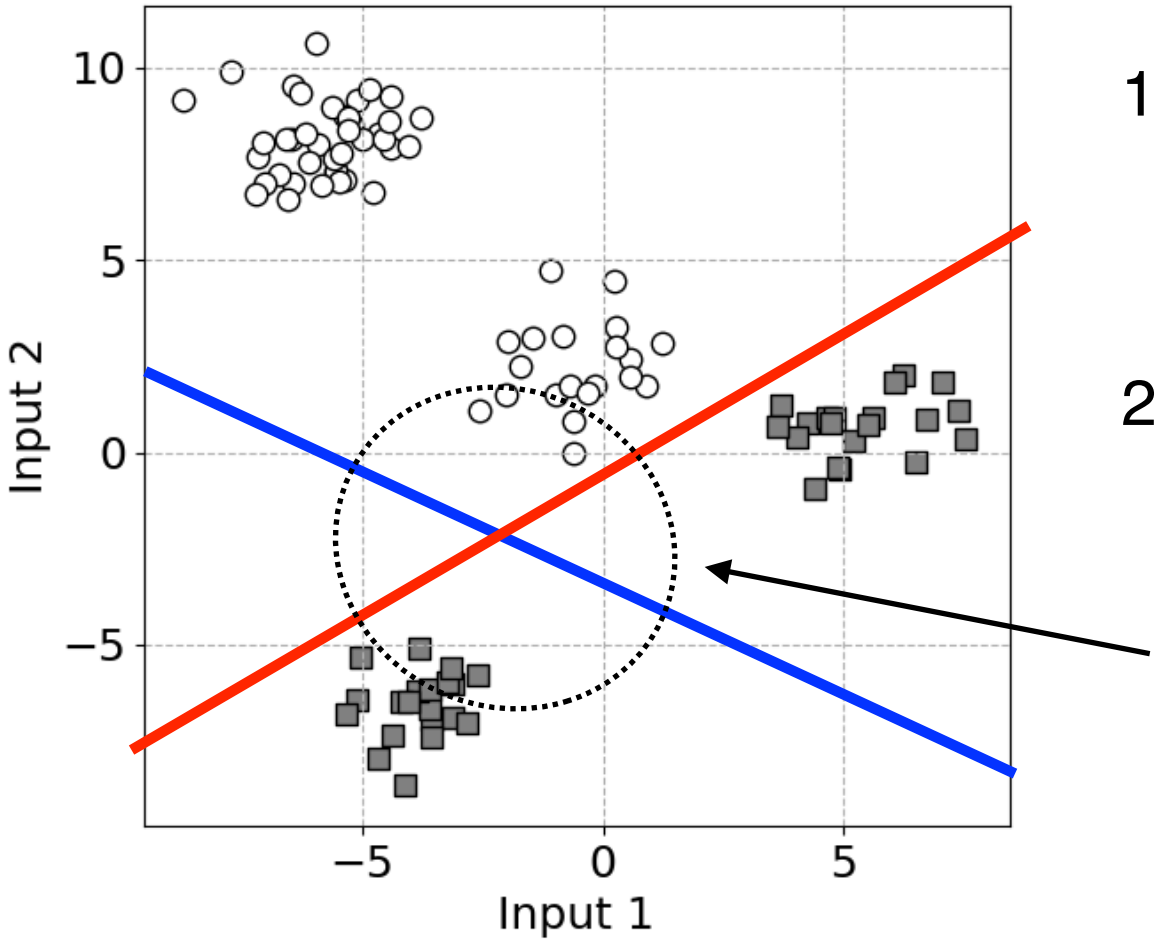
Continual Learning CIFAR-100

Bayesian Learning Rule improves the state-of-the-art,
but there is room for improvement.



1. Nguyen et al. “Variational Continual Learning.” ICLR (2018).
2. Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *PNAS 2017*

Continual Learning with Bayes

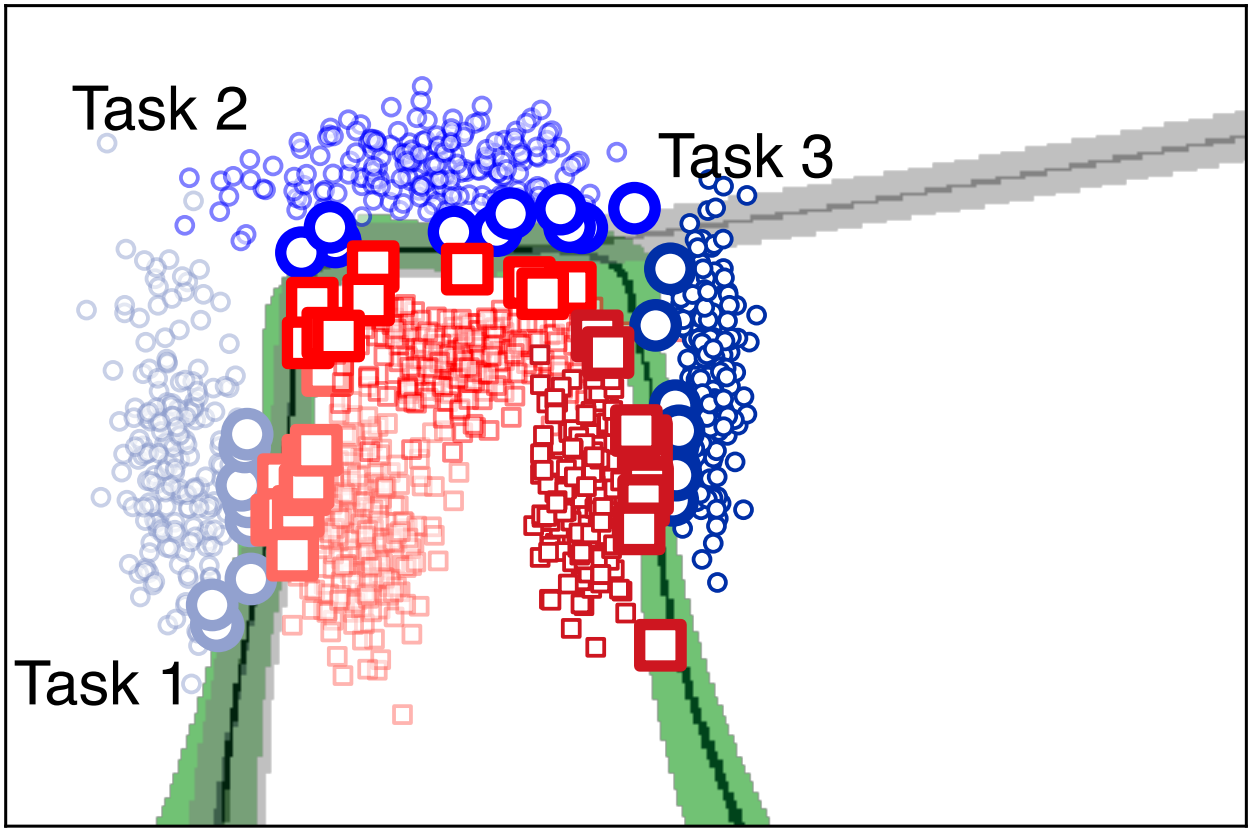


1. Identify Memorable past examples
   - Use Bayesian Learning Rule
2. Regularize network outputs
   - Make sure the predictions don't change too much

1. Khan et al. "Approximate Inference Turns Deep Networks into Gaussian Process", NeurIPS, 2019
2. *Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020*

Functional Regularization of Memorable Past (FROMP)

Regularize the function outputs.
Simply adds an additional term in Adam.



1. Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020

From Neural Net to Gaussian Process

Bayes learning rule: $\lambda \leftarrow \lambda - \rho \nabla_{\mu} \left(\mathbb{E}_q[\ell(\theta)] - \mathcal{H}(q) \right)$

Optimal natural parameters = natural gradients [1]

$$\lambda_* = \nabla_{\mu} \mathbb{E}_{q_*}[\ell(\theta)]$$

$$m_* = \sum_i \mathbb{E}_{q_*}[\nabla_\theta \ell_i(\theta)] = J_*^\top \alpha_*$$

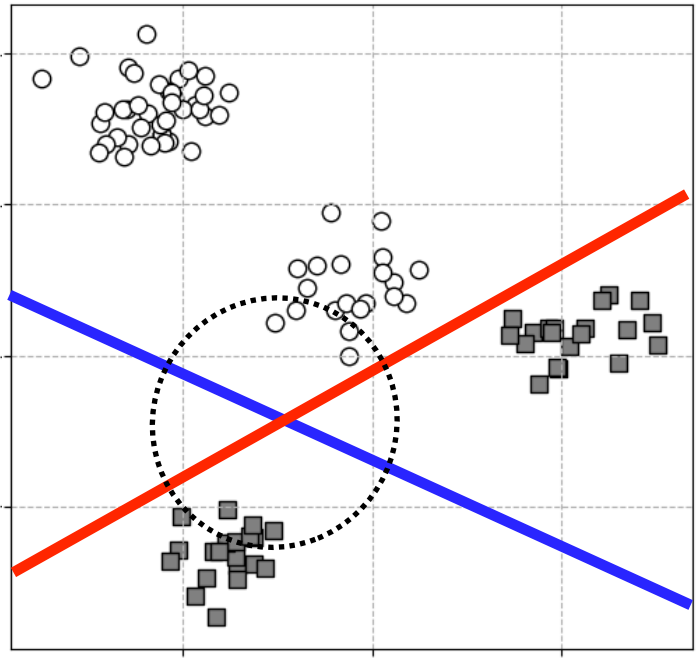
$$\Sigma_*^{-1} = \sum_i \mathbb{E}_{q_*}[\nabla^2 \ell_i(\theta)] + I \approx J_* \Lambda_* J_*^\top + I$$

Extract dual variables (α_*, Λ_*) from natural gradients!

DNN2GP [2] uses this to express both the "iterates" and "solutions" as Gaussian process (also see [3])

1. Khan and Nielsen. "Fast yet simple natural-gradient descent for variational inference in complex models." ISITA, 2018
2. Khan et al. "Approximate Inference Turns Deep Networks into Gaussian Process", NeurIPS, 2019
3. Khan et al. "Fast Dual variational inference fo non-conjugate LGMs", ICML, 2013

Continual Learning with GPs



Weights Regularization [1]

$$(\theta - \theta_{old})^{\top}\Sigma_{old}^{-1}(\theta - \theta_{old})$$

Functional Regularization [2]

$$KL(p(\theta)||q(\theta)) \approx KL(p(f)||q(f))$$

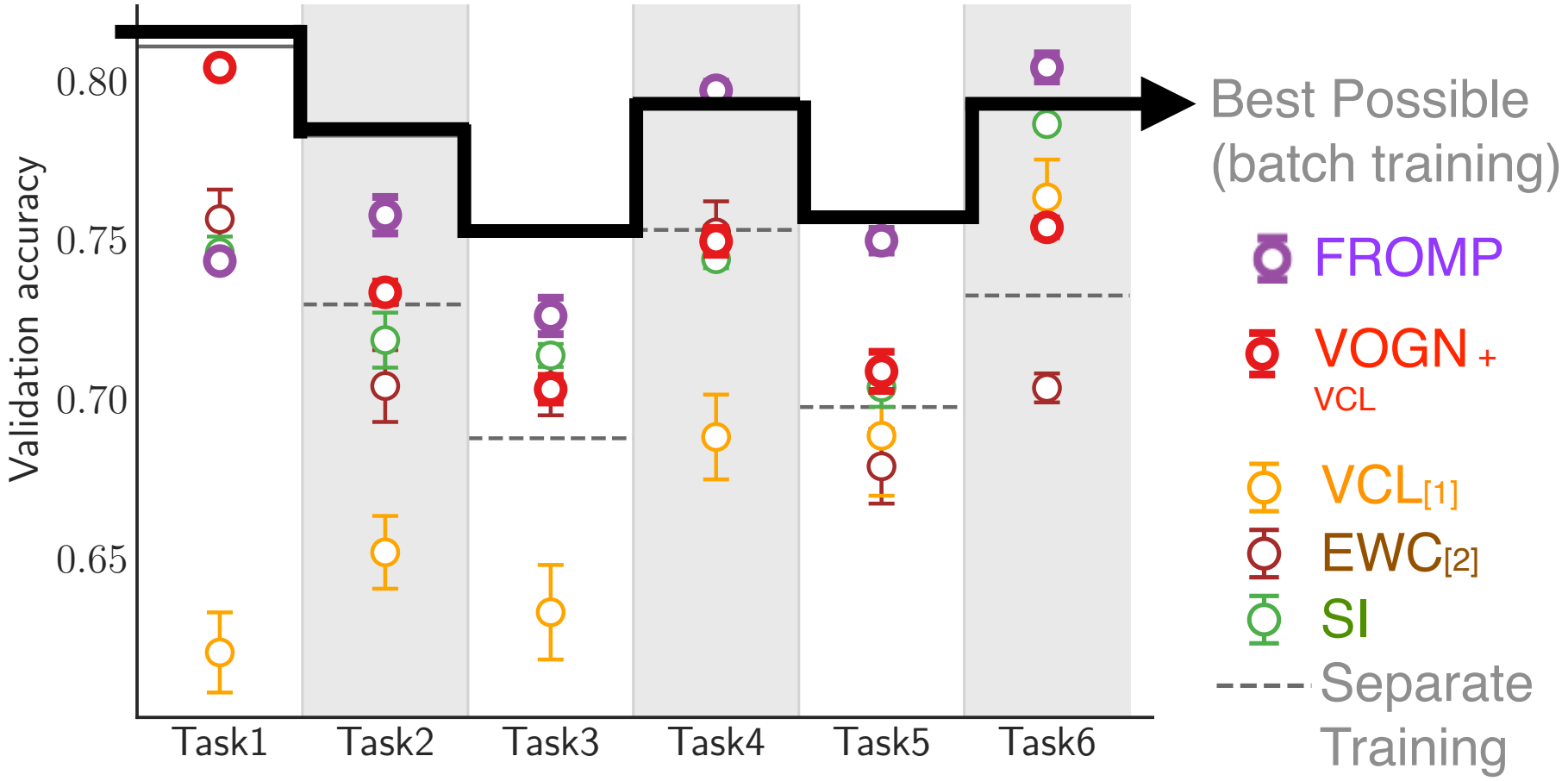
$$\left[f(X_m) - f_{old}(X_m)\right]^{\top} K_{old}(X_m, X_m)^{-1}\left[f(X_m) - f_{old}(X_m)\right]$$

Upcoming result (coming soon): Such regularizers enable optimal knowledge transfer!!!

1. Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks." *PNAS 2017*
2. *Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020*

Continual Learning: Improving Bayes

FROMP uses a GP prior in “function-space” over the “memorable pasts” and improves the performance.



1. Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020

FROMP results (MNIST)

Method	Permuted	Split
DLP [32]	82%	61.2%
EWC [18]	84%	63.1%
SI [37]	86%	98.9%
Improved VCL [33]	$93 \pm 1\%$	$98.4 \pm 0.4\%$
+ random Coreset	$\mathbf{94.6} \pm 0.3\%$	$98.2 \pm 0.4\%$
[1] FRCL-RND [34]	$94.2 \pm 0.1\%$	$97.1 \pm 0.7\%$
FRCL-TR [34]	$94.3 \pm 0.2\%$	$97.8 \pm 0.7\%$
FRORP-L_2	$87.9 \pm 0.7\%$	$98.5 \pm 0.2\%$
FROMP-L_2	$94.6 \pm 0.1\%$	$98.7 \pm 0.1\%$
FRORP	$94.6 \pm 0.1\%$	$\mathbf{99.0} \pm 0.1\%$
FROMP	$\mathbf{94.9} \pm 0.1\%$	$\mathbf{99.0} \pm 0.1\%$

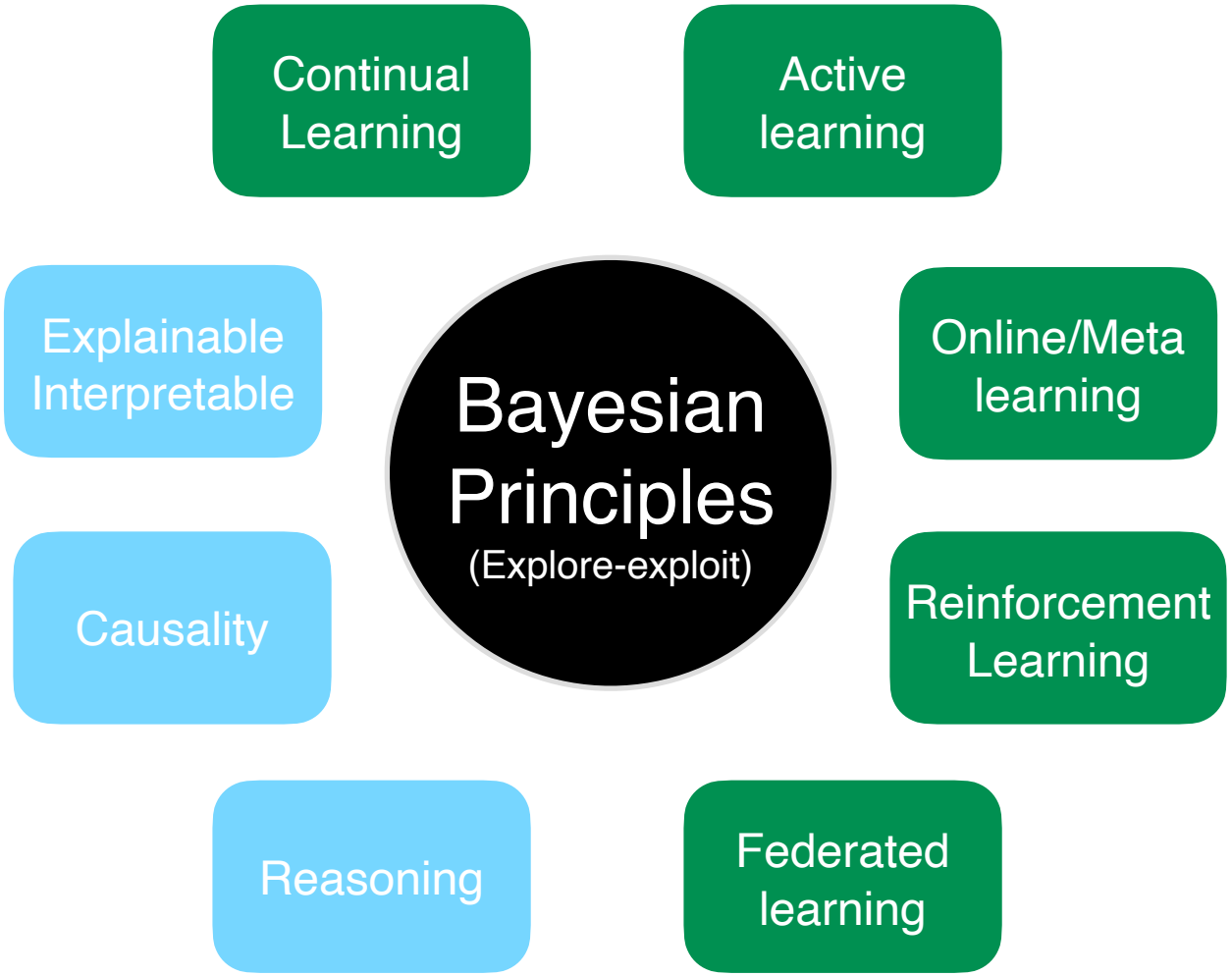
(a) MNIST comparisons: for Permuted, we use 200 examples as memorable/coreset/inducing points. For Split, we use 40.

1. Titsias et al. Functional Regularisation for Continual Learning with Gaussian Processes, *ICLR* (2020)

Bayesian (Principles for) Learning-Machines

- Uncertainty (Background)
  - What you don't know now, can hurt you later
- Learning (Past work)
  - Derive learning-algorithms from Bayes
- Knowledge (Current work)
  - Knowledge representation and its transfer
  - Memorable experiences (Dharmesh Tailor)
  - Continual learning (Emti)
  - Meta learning (Pierre Alquier)

Current Work



How to design AI that learn like us?

- Uncertainty -> Learning -> Knowledge
- Three questions
  - Q1: What do we know? (model)
  - Q2: What do we not know? (uncertainty)
  - Q3: What do we need to know? (action & exploration)
- Posterior approximation is the key
  - (Q1) Models == representation of the world
  - (Q2) Posterior approximations == representation of the model
  - (Q3) The Bayes-dual will enable Knowledge representation, transfer, and collection.

Approximate Bayesian Inference Team

Emtiyaz Khan
Team Leader

Pierre Alquier
Research Scientist

Gian Maria Marconi
Postdoc

Thomas Möllenhoff
Postdoc

https://team-approx-bayes.github.io/

Wu Lin
PhD Student
University of British Columbia

Dharmesh Tailor
Research Assistant

Fariz Ikhwantri
Part-time Student
Tokyo Institute of Technology

Happy Buzaaba
Part-time Student
University of Tsukuba

Evgenii Egorov
Remote Collaborator
Skoltech

Siddharth Swaroop
Remote Collaborator
University of Cambridge

Dimitri Meunier
Remote Collaborator
ENSAE Paris

Peter Nickl
Remote Collaborator
TU Darmstadt

Erik Daxberger
Remote Collaborator
University of Cambridge

Alexandre Piché
Remote Collaborator
MILA