Presentation at the Conference on Lifelong Learning Agents 2024,
July 30, 2024

How to Build Machines That Adapt Quickly

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

Summary of recent research at https://emtiyaz.github.io/papers/symposium_2023.pdf
Slides available at https://emtiyaz.github.io/

Continual Lifelong Learning

Keep learning for a long time by observing, interacting, adapting, exploring the environment

Human Learning at the age of 6 months.

Converged at the age of 12 months

Transfer skills

at the age of 14 months

Current state of ML

https://www.youtube.com/watch?v=TxobtWAFh8o The video is from 2017

Continual Lifelong Adaptation

For sustainable, reliable, transparent AI

What are (some) Fundamental Principles of Continual Lifelong Learning?

Connecting, combining, and improving existing methods

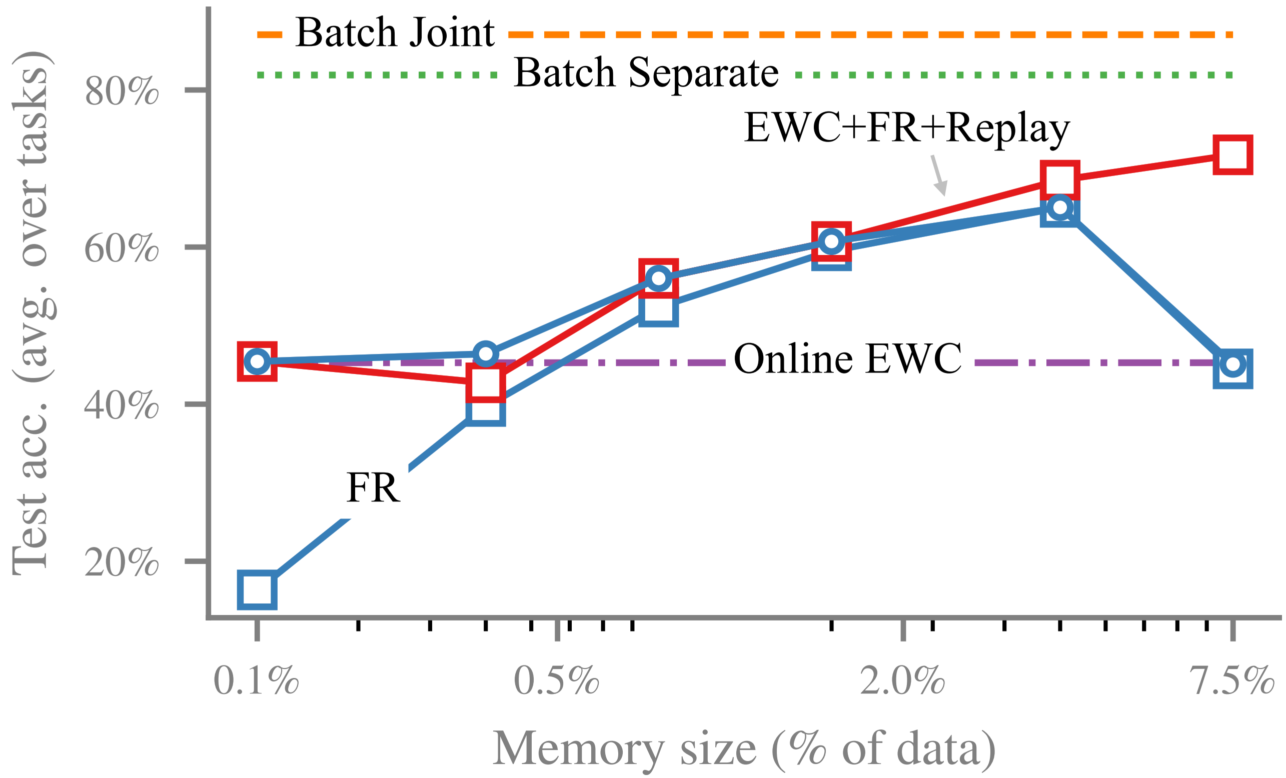
Outline of the Talk

- Distributed information over time and space [1] requires dealing with Interference between the past and future
  - "Gradient mismatch" [2] & "reconstruction" [3-5]
- Quick adaptation is possible when mismatches are caused by just a few examples
  - "Memorable Past" or Memory of models [4, 6]
- The difficulty of lifelong learning reduces to a faithful representation of the past

1. Khan and Rue, The Bayesian Learning Rule, JMLR (2023).
2. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).
3. Khan and Swaroop. Knowledge-Adaptation Priors, NeurIPS (2021).
4. Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020
5. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR (2023).
6. Nickl, Xu, Tailor, Moellenhoff, Khan, The memory-perturbation equation, NeurIPS (2023)

Results on ImageNet with ResNet-18

Obtain 78% accuracy with just 7.5% data by combining EWC, Functional Reg. & Replay.



See the poster #J6 today.

1. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR2023(CoLLAs 2024)

Distributed Information Processing over Time and Space

Continual Learning or Sequential Learning

$$
\begin{tikzcd}
\theta_1 \arrow[r] & \theta_{1+2} \arrow[r] & \theta_{1+2+3} \arrow[r] & {} \\
\mathscr{D}_1 \arrow[u] & \mathscr{D}_2 \arrow[u] & \mathscr{D}_3 \arrow[u] &
\end{tikzcd}
$$

Federated Learning or Model Merging

$$
\begin{tikzcd}
& \theta_{1+2} \arrow[dl, shift left] \arrow[dr, shift left] & \\
\theta_1 \arrow[ur, shift left] & & \theta_2 \arrow[ul, shift left] \\
\mathscr{D}_1 \arrow[u] & & \mathscr{D}_2 \arrow[u]
\end{tikzcd}
$$

For such problems, we must be able to distinguish the new information apart from the old information.

Note: These are not graphical models. Arrows represent the “flow” of information

The Intuition

If $\mathscr{D}_1$ and $\mathscr{D}_2$ are different from each other, then θ_{1+2} should also be different from θ_1 and θ_2.

The Bayesian way [1,2] is to define "new information" by measuring the gain/change in the posterior (or in θ_1 or its predictions $f_i(\theta_1)$)

$$KL(p_{1+2} \| p_1) \qquad \theta_{1+2} - \theta_1 \qquad f_i(\theta_{1+2}) - f_i(\theta_1)$$

I will present a simpler way to quantify $\theta_{1+2} - \theta_1$ in terms of "gradient mismatch", but remember that there is always an underlying Bayesian principle [3]

1. Jaynes, Information theory and statistical mechanics, 1957
2. Zellner, Optimal information processing and Bayes's theorem, The American Statistician, 1988.
3. Khan and Rue, The Bayesian Learning Rule, JMLR (2023).

Nico Daheim
(TUD)

Thomas Moellenhoff
(RIKEN)

Model Merging

Connecting inaccuracy of model merging to gradient mismatch

1. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).

Model Merging

$$
\begin{tikzcd}
 & \theta_{1+2} & \\
\theta_1 \arrow[ur] & & \theta_2 \arrow[ul] \\
\mathcal{D}_1 \arrow[u] & & \mathcal{D}_2 \arrow[u]
\end{tikzcd}
$$

Given θ_1 fine-tuned on $\mathcal{D}_1$ and θ_2 fine-tuned on $\mathcal{D}_2$, merge them (to estimate θ_{1+2}).

Simplest strategy is to use $\alpha_1\theta_1 + \alpha_2\theta_2$ for scalars α_1, α_2 [1]. The quality depends on the difference:

$$\theta_{1+2} - (\alpha_1\theta_1 + \alpha_2\theta_2)$$

For simplicity, I will assume $\alpha_1 = \alpha_2 = 1$. For the full version, see our paper [2].

1. Wortsman et al. Robust fine-tuning of zero-shot models, CVPR 2022
2. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).

A (dual) View: Parameters as Gradients

$$\theta_1 = \arg\min_\theta \ell_1(\theta) + \frac{1}{2}\|\theta\|^2 \implies 0 = \nabla \ell_1(\theta_1) + \theta_1$$

$$\implies \theta_1 = -\boxed{\nabla \ell_1(\theta_1)}$$

In other words, parameters are gradients.

$$\theta_2 = \arg\min_\theta \ell_2(\theta) + \frac{1}{2}\|\theta\|^2 \implies \theta_2 = -\boxed{\nabla \ell_2(\theta_2)}$$

$$\theta_{1+2} = \arg\min_\theta \ell_1(\theta) + \ell_2(\theta) + \frac{1}{2}\|\theta\|^2$$

$$\implies \theta_{1+2} = -\boxed{\nabla \ell_1(\theta_{1+2})} - \boxed{\nabla \ell_2(\theta_{1+2})}$$

1. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).

Parameter Change as Gradient Mismatch

$$\theta_{1+2} = -\boxed{\nabla \ell_1(\theta_{1+2})} - \boxed{\nabla \ell_2(\theta_{1+2})}$$

$$\theta_1 = -\boxed{\nabla \ell_1(\theta_1)}$$

$$\theta_2 = -\boxed{\nabla \ell_2(\theta_2)}$$

Subtract the last two equations from the first one.

$$\implies \theta_{1+2} - (\theta_1 + \theta_2)$$

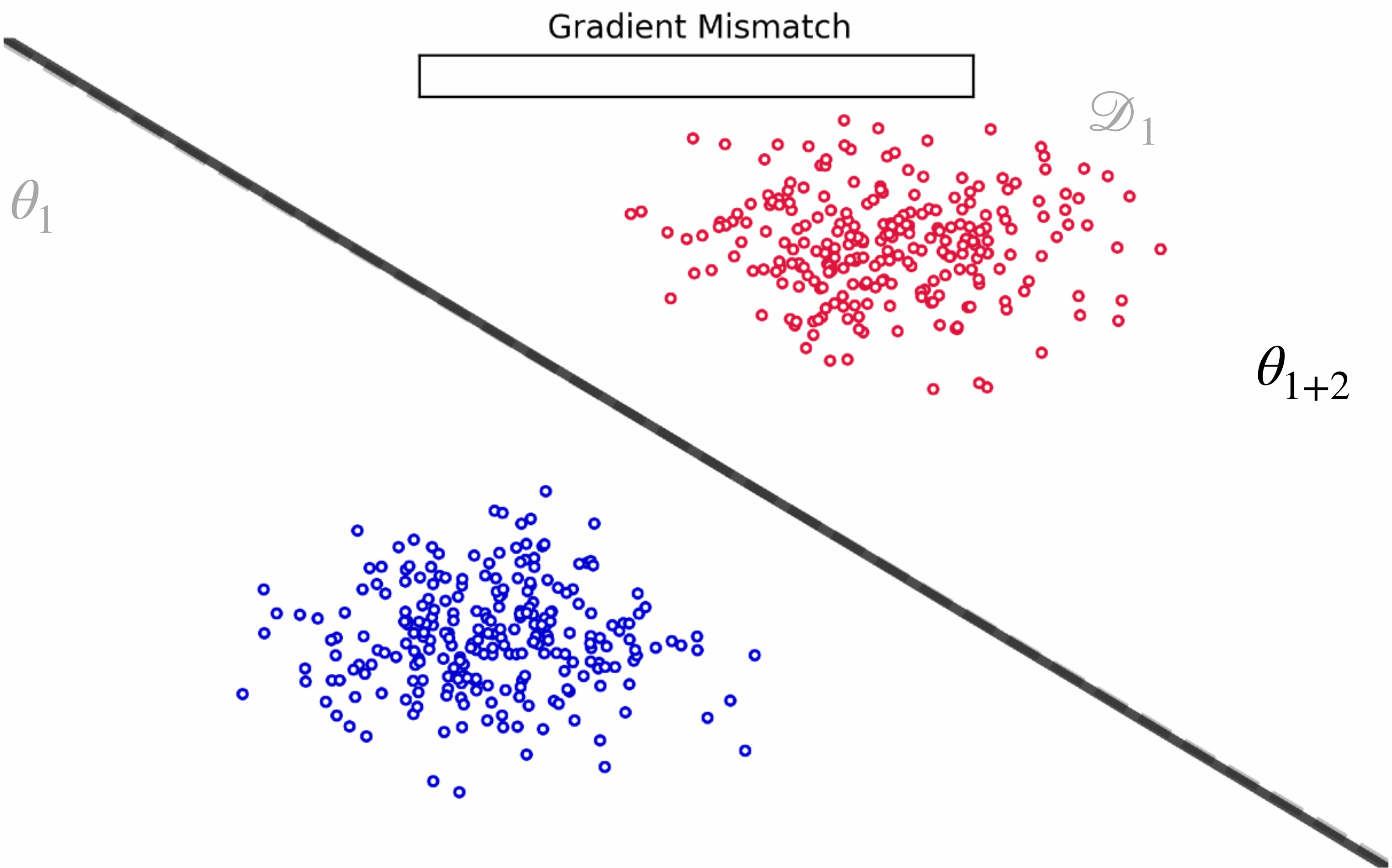
$$= -\Big[\underbrace{\overset{\text{New}}{\nabla \ell_1(\theta_{1+2})} - \overset{\text{Old}}{\nabla \ell_1(\theta_1)}}_{\text{Gradient Mismatch on } \mathscr{D}_1}\Big] - \Big[\underbrace{\overset{\text{New}}{\nabla \ell_2(\theta_{1+2})} - \overset{\text{Old}}{\nabla \ell_2(\theta_2)}}_{\text{Gradient Mismatch on } \mathscr{D}_2}\Big]$$

Gradient mismatch among new and old parameters!

1. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).

Gradient Mismatch

$$\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)$$



Reducing the Mismatch

$$\nabla \ell_1(\theta_{1+2}) \approx \nabla \ell_1(\theta_1) + H_1 \cdot (\theta_{1+2} - \theta_1)$$

$$\theta_{1+2} - (\theta_1 + \theta_2)$$

$$= - \left[\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1) \right] - \left[\nabla \ell_2(\theta_{1+2}) - \nabla \ell_2(\theta_2) \right]$$

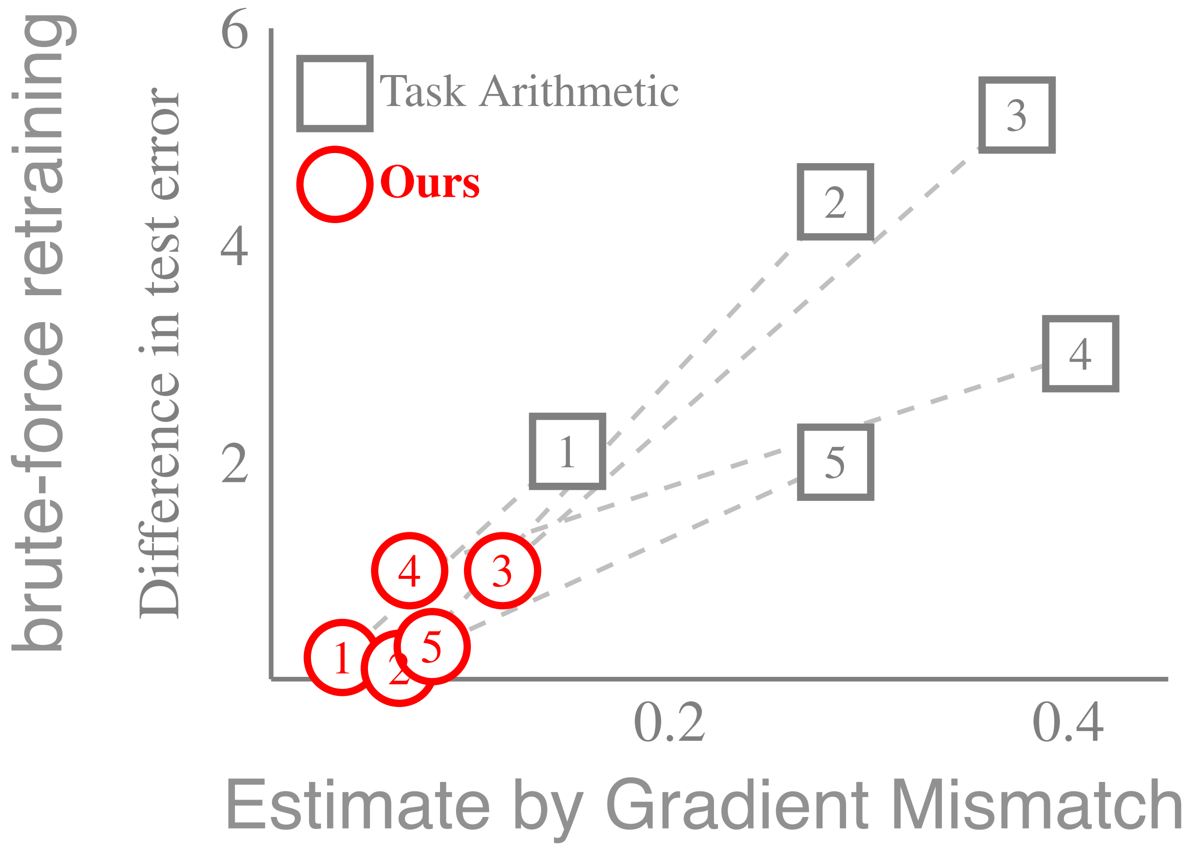
$$\approx - H_1 \cdot (\theta_{1+2} - \theta_1) \qquad - H_2 \cdot (\theta_{1+2} - \theta_2)$$

$$\implies \theta_{1+2} \approx \frac{H_1 + I}{H_1 + H_2 + I} \theta_1 + \frac{H_2 + I}{H_1 + H_2 + I} \theta_2$$

Hessian-based merging [2] reduces mismatch. More such results in [1], including task-arithmetic [3]

1. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).
2. Matena and Raffel. Merging models with Fisher-weighted averaging, NeurIPS 2022
3. Ilharco et al. Editing models with task arithmetic. ICLR 2023

Minimizing Gradient Mismatch Reduces Test Error



RoBERTa on IMDB

1. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).

$$
\begin{tikzcd}
\theta_1 \arrow[r] & \theta_{1+2} \arrow[r] & \theta_{1+2+3} \arrow[r] & {} \\
\mathscr{D}_1 \arrow[u] & \mathscr{D}_2 \arrow[u] & \mathscr{D}_3 \arrow[u] &
\end{tikzcd}
$$

Siddharth Swaroop
(U Cambridge,
Now in Harvard U)

Looking for a faculty
position in near
future.

Continual Learning

Gradient mismatch and its reconstruction

1. Khan and Swaroop. Knowledge-Adaptation Priors, NeurIPS (2021).
2. Daxberger, Swaroop, Osawa, Yokota, Turner, Hernandez-Lobato, Khan, Improving CL by Accurate Gradient Reconstruction of the Past, TMLR (2023) & CoLLAs (2024).

Gradient Mismatch in CL

$$\theta_{1+2} = -\nabla \ell_1(\theta_{1+2}) - \nabla \ell_2(\theta_{1+2})$$

$$\theta_1 = -\nabla \ell_1(\theta_1)$$

Subtract the 2nd eq. from the 1st eq.

$$\implies \theta_{1+2} - \theta_1$$

$$= -\Big[\underbrace{\overbrace{\nabla \ell_1(\theta_{1+2})}^{\text{New}} - \overbrace{\nabla \ell_1(\theta_1)}^{\text{Old}}}_{\text{Gradient Mismatch on } \mathscr{D}_1}\Big] - \underbrace{\nabla \ell_2(\theta_{1+2})}_{\text{New loss}}$$

Gradient Mismatch on the past data.

1. Khan and Swaroop. Knowledge-Adaptation Priors, NeurIPS (2021).

Knowledge-Adaptation Prior [1]

Find a regularizer that reconstructs the mismatch

$$\theta_{1+2} - \theta_1 + \left[\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)\right] + \nabla \ell_2(\theta_{1+2}) = 0$$

$$= \nabla D(\theta_{1+2} \| \theta_1)$$

Then, solve $\theta_{1+2} = \arg\min_{\theta} \; D(\theta \| \theta_1) + \ell_2(\theta)$

A wide-variety of adaptation methods can be seen as using different choices of D [2-9]

1. Khan and Swaroop. Knowledge-Adaptation Priors, NeurIPS (2021).
2. Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks. PNAS, 2017.
3. Benjamin et al. Measuring and regularizing networks in function space. ICLR 2019.
4. Hinton et al. Distilling the knowledge in a neural network, arXiv, 2015.
5. Buzzega et al. Dark experience for general continual learning: a strong, simple baseline. NeurIPS 2020.
6. Cauwenberghs and Poggio. Incremental and decremental SVM learning. NeurIPS, 2001.
7. Vapnik and Izmailov. Learning using privileged information: similarity control and …. JMLR, 2015.
8. Lopez-Paz and Ranzato. Gradient episodic memory for continual learning, NIPS'17
9. Csató and Opper. Sparse on-line Gaussian processes. Neural computation, 2002.

EWC as K-Priors

$$\left(\theta_{1+2} - \theta_1\right) + \left[\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)\right] + \nabla \ell_2(\theta_{1+2}) = 0$$

$$\approx H_1(\theta_{1+2} - \theta_1)$$

$$\implies (I + H_1)(\theta_{1+2} - \theta_1) + \nabla \ell_2(\theta_{1+2}) = 0$$

$$\implies \theta_{1+2} \approx \arg\min_{\theta} \; \frac{1}{2}\|\theta - \theta_1\|^2_{H_1+I} + \ell_2(\theta)$$

EWC reduces the mismatch by "reusing" θ_1 which is different from Experience Replay

$$\theta_{1+2} \approx \arg\min_{\theta} \; \hat{\ell}_1(\theta) + \ell_2(\theta)$$

1. Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks. PNAS, 2017.

Functional Regularizer (FR) as K-priors

For certain losses, gradient mismatch is equivalent to regularizing model "outputs/predictions".

$$\ell(\theta) = \sum_i \left[f_i(\theta) - y_i\right]^2/2 \qquad \nabla \ell(\theta) = \sum_i \phi_i \left[f_i(\theta) - y_i\right]$$

where $f_i(\theta) = \phi_i^\top \theta$ with ϕ_i being a feature vector.

$$\left(\theta_{1+2} - \theta_1\right) + \left[\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)\right] + \nabla \ell_2(\theta_{1+2}) = 0$$

$$\sum_{i \in \mathcal{D}_1} \phi_i \left[f_i(\theta_{1+2}) - f_i(\theta_1)\right]$$

$$\implies \theta_{1+2} = \arg\min_\theta \|\theta - \theta_1\|^2 + \sum_{i \in \mathcal{D}_1} \|f_i(\theta) - f_i(\theta_1)\|^2 + \ell_2(\theta)$$

1. Benjamin et al. Measuring and regularizing networks in function space. ICLR 2019.
2. Hinton et al. Distilling the knowledge in a neural network, arXiv, 2015.
3. Buzzega et al. Dark experience for general continual learning: a strong, simple baseline. NeurIPS 2020.

Knowledge Transfer in SVMs

It is also possible to rewrite entirely in function-space, but this is only exact for convex cases [1]

$$\arg\min_{\theta} \|\theta - \theta_1\|^2 + \|\Phi\theta - \Phi\theta_1\|^2 + \ell_2(\theta)$$

$$= \arg\max_{\alpha} \|\alpha - \alpha_1\|^2_{\Phi\Phi^\top + I} + \ell_2^*(\alpha)$$

where α is the dual variable; see [2-5].

Beware of the fully "function-space" methods; they assume linearity and ignore "label noise"!!!

1. Olivier Chapelle. Training a support vector machine in the primal. Neural Computation, 2007.
2. Cauwenberghs and Poggio. Incremental and decremental SVM learning. NeurIPS, 2001.
3. Vapnik and Izmailov. Learning using privileged information: similarity control and …. JMLR, 2015.
4. Lopez-Paz and Ranzato. Gradient episodic memory for continual learning, NIPS'17
5. Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020

How to Fix the FR methods

The problem: for neural-nets, features depend on θ

$$\nabla \ell(\theta) = \sum_i \nabla f_i(\theta) \underbrace{\boxed{\left[f_i(\theta) - y_i\right]}}_{:= e_i(\theta)}$$

But, we can fix this issue by using Replay [1]

$$\nabla \ell_1(\theta_1) - \nabla \ell_1(\theta_{1+2})$$

$$= \sum_i \nabla f_i(\theta_1) e_i(\theta_1) - \nabla f_i(\theta_{1+2})[f_i(\theta_{1+2}) - f_i(\theta_1) + f_i(\theta_1) - y_i]$$

$$= \sum_i \left[\nabla f_i(\theta_1) - \nabla f_i(\theta_{1+2})\right] e_i(\theta_1) - \nabla f_i(\theta_{1+2})[f_i(\theta_{1+2}) - f_i(\theta_1)]$$

$$= \nabla \ell_1(\theta_1) - \underbrace{\sum_i \nabla f_i(\theta_{1+2}) e_i(\theta_1)}_{\text{Replay}} - \underbrace{\sum_i \nabla f_i(\theta_{1+2})[f_i(\theta_{1+2}) - f_i(\theta_1)]}_{\text{Functional regularization}}$$

1. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR (2023).

Summary

- Gradient mismatch can be reduced
  - Weight regularizers (e.g., EWC)
  - Functional regularizers (& dual versions)
  - Replay.
- They are complementary and do different things.
  - Uncertainty in weights, predictions, & labels.
- Optimal combination depends on the task
- Are there general principles for their combination?
  - Look deeper into the sources of mismatch

Peter Nickl[†]
peter.nickl@riken.jp

Lu Xu[*†]
lu.xu.sw@riken.jp

Dharmesh Tailor[*‡]
d.v.tailor@uva.nl

Thomas Möllenhoff[†]
thomas.moellenhoff@riken.jp

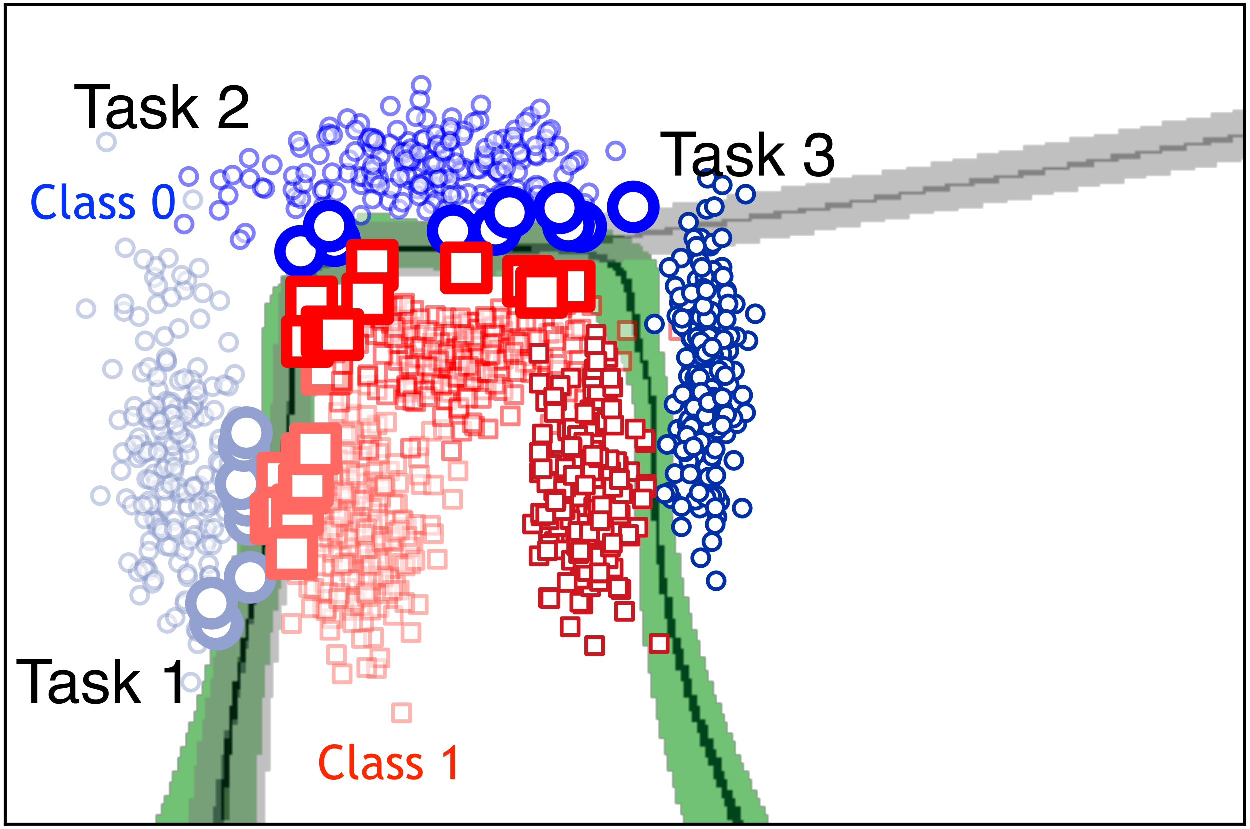
Memory

How to choose the examples to regularize appropriately? How to represent the past when the future is unknown?

1. Nickl, Xu, Tailor, Moellenhoff, Khan, The memory-perturbation equation, NeurIPS, 2023

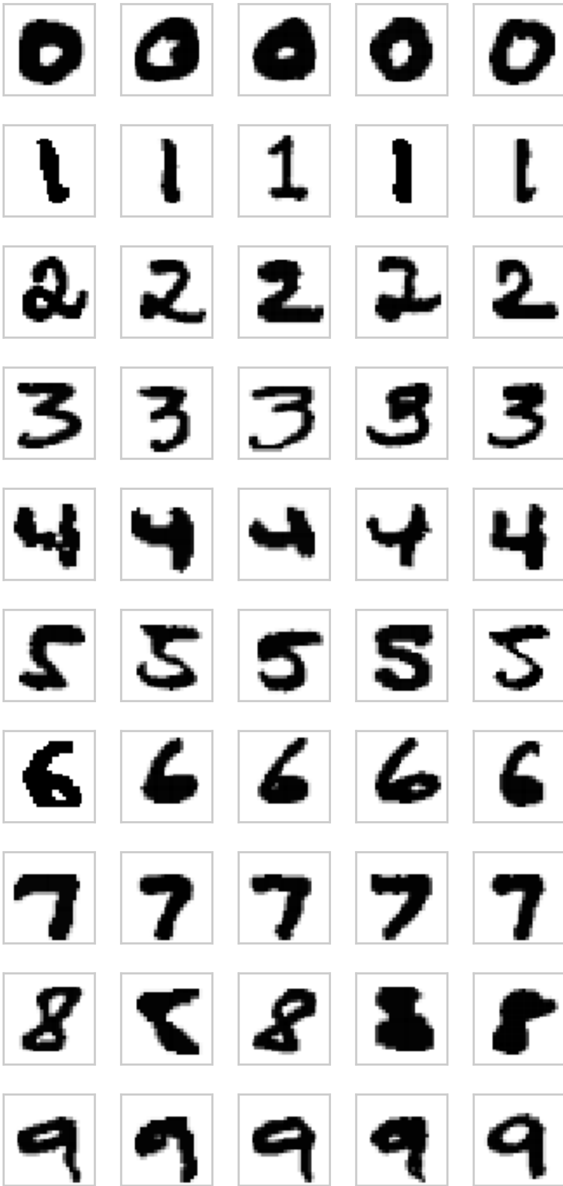
An Early Idea

Choose the memory at the boundary



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

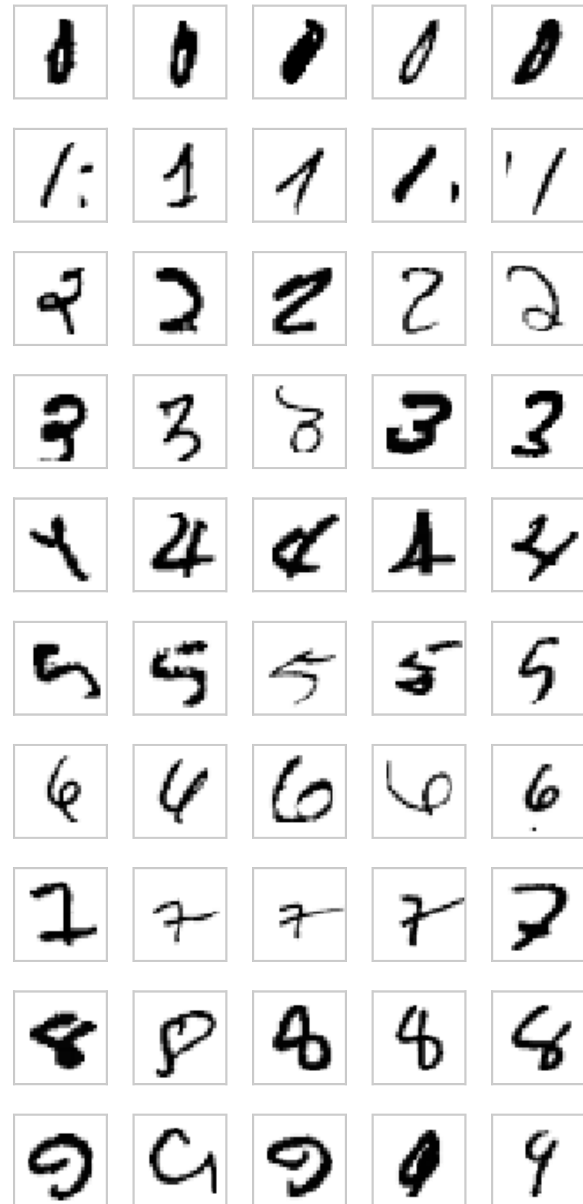
Less Memorable



$$\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1) \approx \sum_i \phi_i \beta_i \phi_i^\top (\theta_{1+2} - \theta_1)$$

Related to Leverage score and influence function.

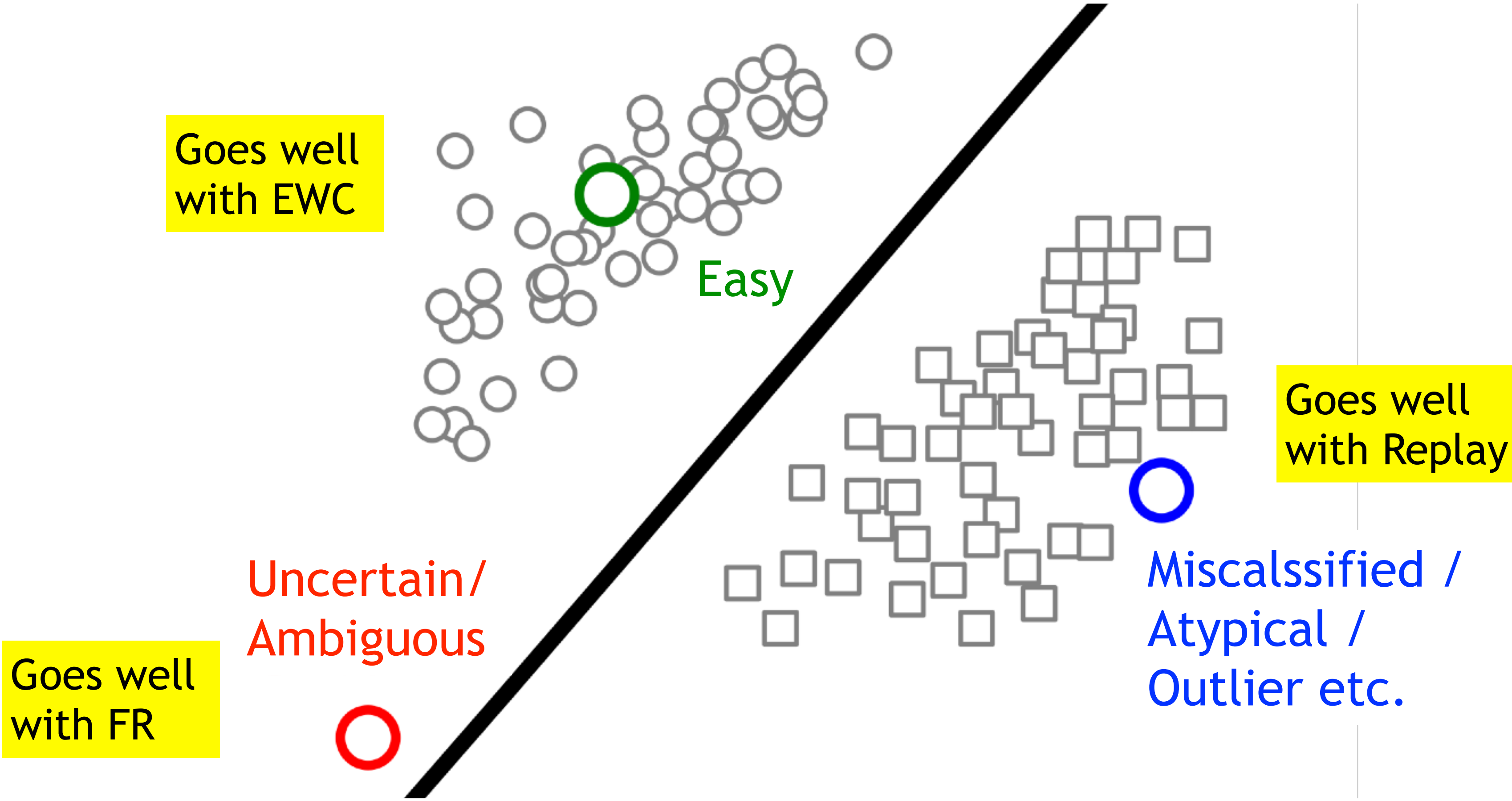
More memorable



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

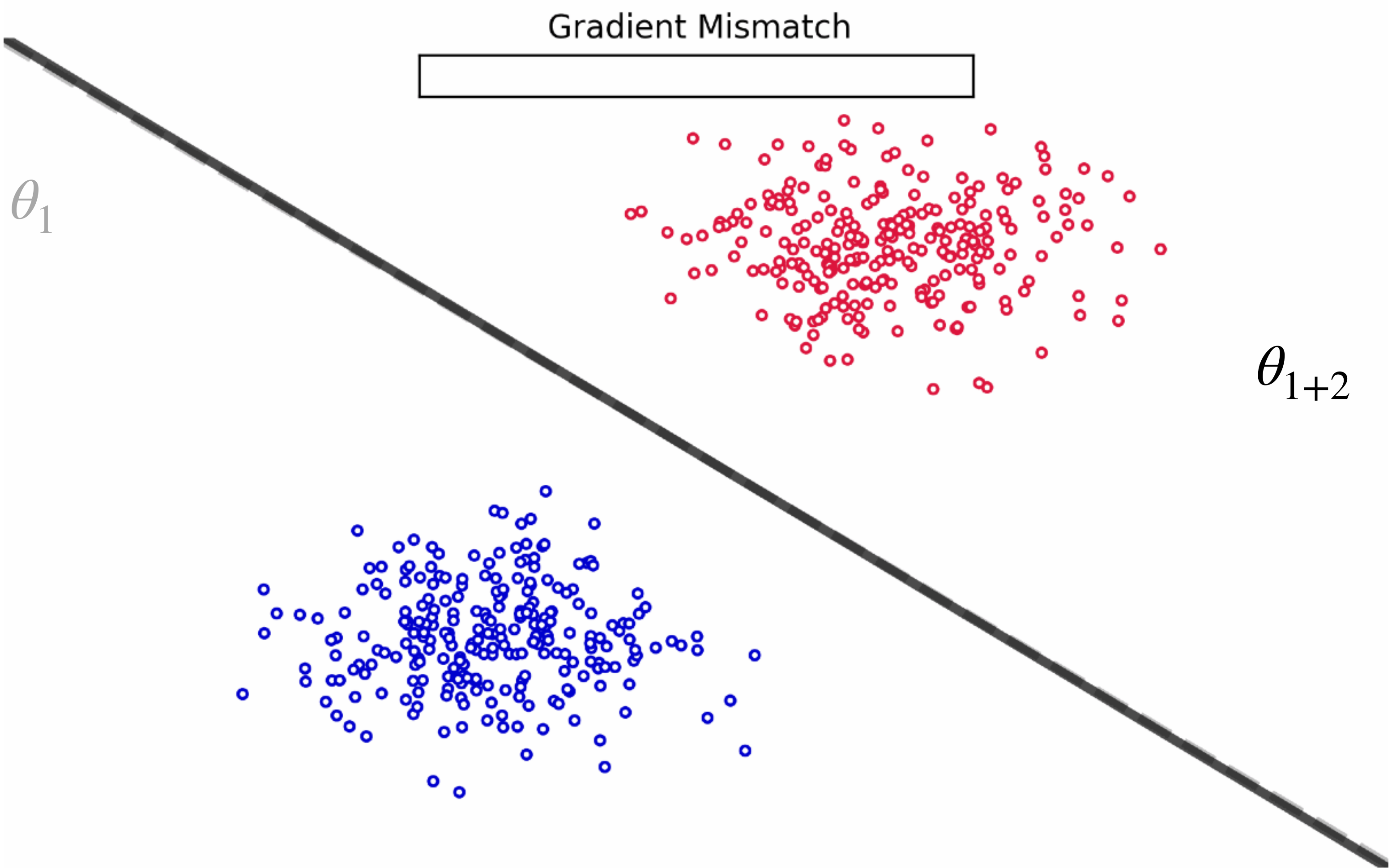
Three types of Examples

Very similar to Support Vectors!



Mismatch Between the Past & Future

$$\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)$$



Combining CL Methods

Look deeper into the sources of mismatches

$$\left(\theta_{1+2} - \theta_1\right) + \left[\nabla \ell_1(\theta_{1+2}) - \nabla \ell_1(\theta_1)\right] + \nabla \ell_2(\theta_{1+2}) = 0$$

$$\sum_{i \in \mathscr{D}_1 \backslash (\mathscr{M}_1 \cup \mathscr{M}_2)} \ldots + \sum_{i \in \mathscr{M}_1} \ldots + \sum_{i \in \mathscr{M}_2} \ldots$$

Low mismatch points, approx by EWC

Some high mismatch points by FR

High mismatch with label-noise by Replay

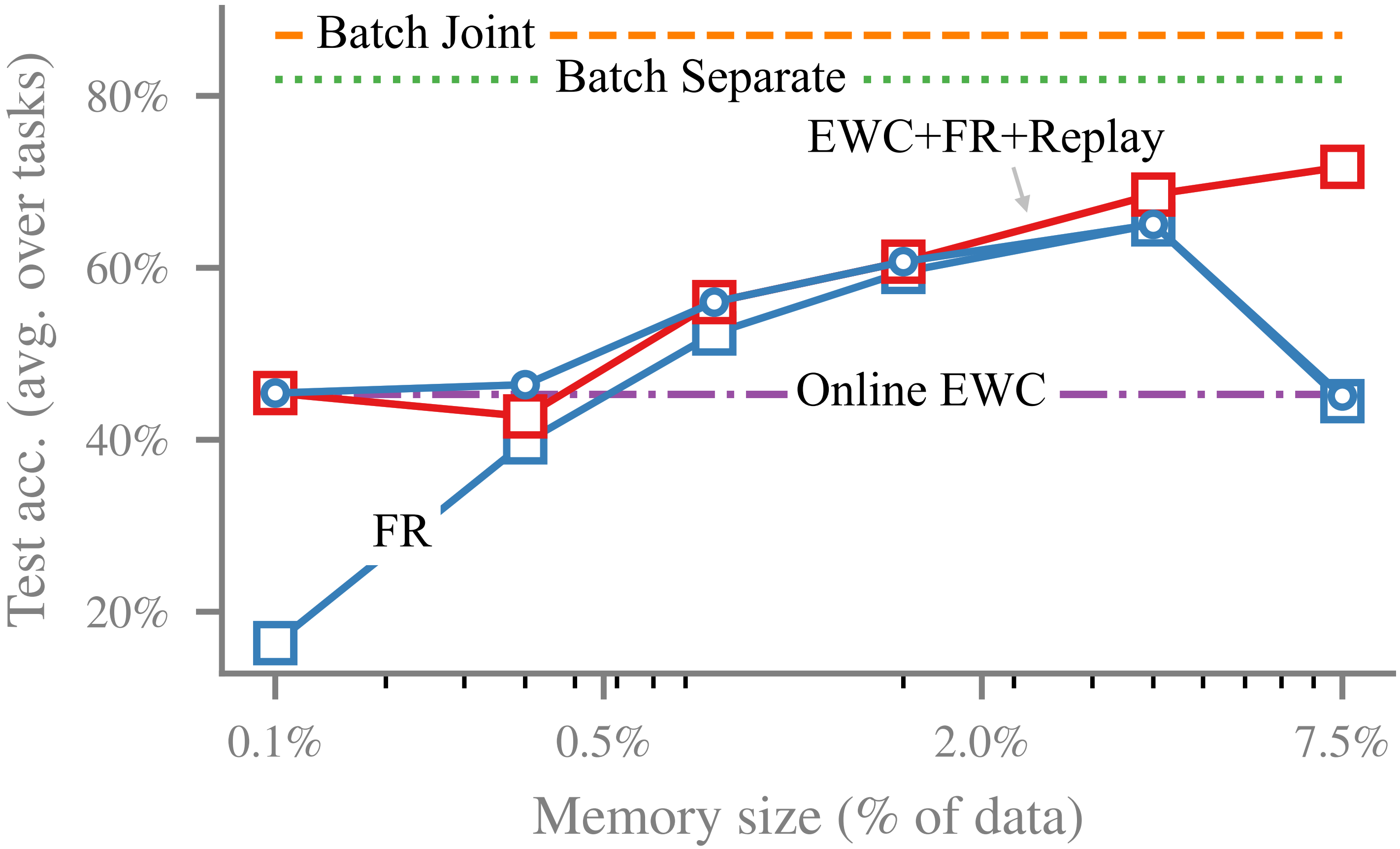
$$\|\theta - \theta_1\|^2_{H_1^{\backslash \mathscr{M}_1 \cup \mathscr{M}_2}} + \sum_{i \in \mathscr{M}_1} \|f_i(\theta) - f_i(\theta_1)\|^2 + \sum_{i \in \mathscr{M}_2} f_i(\theta) e_i(\theta_1)$$

But, θ_{1+2} is unknown so we can't choose well without assuming things about the future.

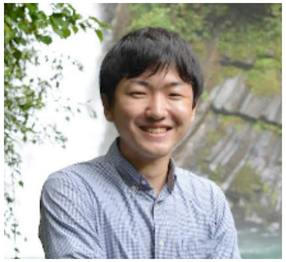
1. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR (2023).

Results with Random Memory on ImageNet with ResNet-18

Get 78% accuracy with 7.5% (random) memory



Erik Daxberger (U Cambridge, Now in Apple)



Kazuki Osawa (TokyoTech, now in DeepMind)

See the poster #J6 today.

1. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR 2023.

Memory = Sensitive Examples

The future is unknown, but we could "protect" θ_1 from "expected" changes, say by deleting data ($\mathscr{M}$)

$$\left(\theta_{-\mathscr{M}} - \theta_1\right) - [\nabla \ell_1(\theta_1) - \nabla \ell_1(\theta_{-\mathscr{M}})] - \nabla \ell_{\mathscr{M}}(\theta_{-\mathscr{M}}) = 0$$

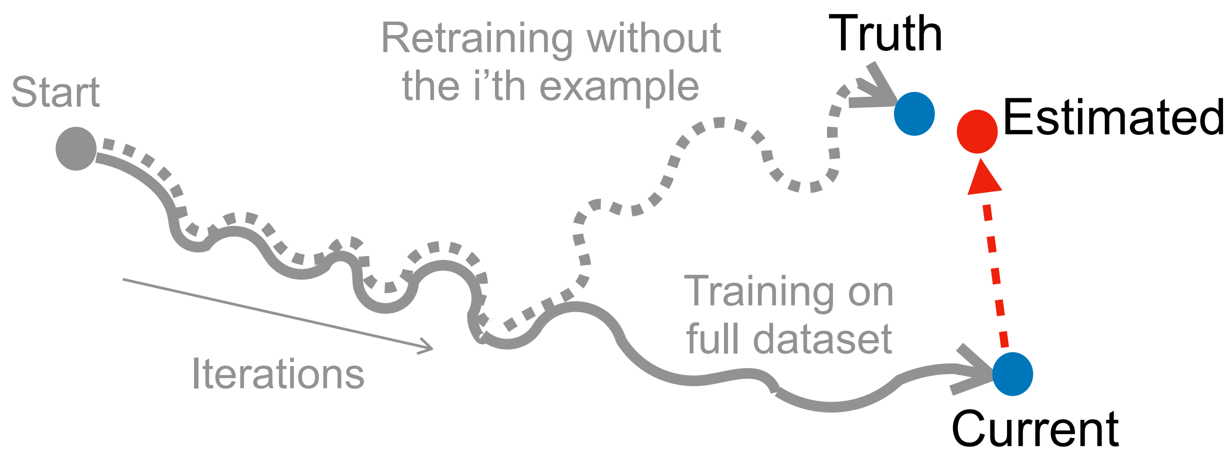
$$\approx H_1(\theta_1 - \theta_{-\mathscr{M}}) \qquad \approx \ell_{\mathscr{M}}(\theta_1)$$

$$\implies \theta_{-\mathscr{M}} - \theta_1 \approx (H_1 + I)^{-1} \nabla \ell_{\mathscr{M}}(\theta_1)$$

Coincides with Influence Measures!

Memory Perturbation Equation

Past that has the most influence on the present



Choose memory based on the following criteria:
Prediction Error x Prediction Variance

1. Nickl, Xu, Tailor, Moellenhoff, Khan, The memory-perturbation equation, NeurIPS, 2023

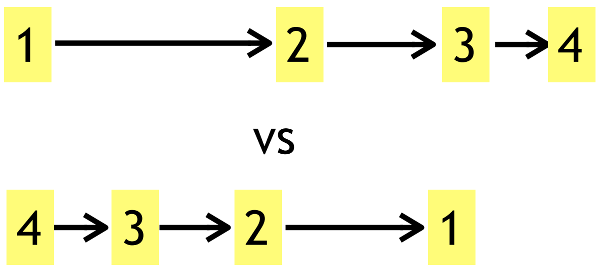
Outline of the Talk

- Distributed information over time and space [1] requires dealing with Interference between the past and future
  – "Gradient mismatch" [2] & "reconstruction" [3-5]
- Quick adaptation is possible when mismatches are caused by just a few examples
  – "Memorable Past" or Memory of models [4, 6]
- The difficulty of lifelong learning reduces to a faithful representation of the past

1. Khan and Rue, The Bayesian Learning Rule, JMLR (2023).
2. Daheim et al. Model merging by uncertainty-based gradient matching, ICLR (2024).
3. Khan and Swaroop. Knowledge-Adaptation Priors, NeurIPS (2021).
4. Pan et al. Continual Deep Learning by Functional Regularisation of Memorable Past, NeurIPS, 2020
5. Daxberger et al. Improving CL by Accurate Gradient Reconstruction of the Past, TMLR (2023).
6. Nickl, Xu, Tailor, Moellenhoff, Khan, The memory-perturbation equation, NeurIPS (2023)

Future of Continual Lifelong Learning

- Lifelong learning is possible only when each subtasks allows quick adaptation
  - Order matters!!!



- Revisit and fix mistakes
- Reduce revisiting frequency
  - e.g., linear to log-linear, worst case = batch
- Memorable past matter
  - Harder problems requires larger memory
  - But larger memory make the problem easier

The webpage is available at https://bayesduality.github.io/, and Twitter account @BayesDuality

The Bayes-Duality Project

Toward AI that learns adaptively, robustly, and continuously, like humans

Emtiyaz Khan

Research director
(Japan side)

Approx-Bayes team at
RIKEN-AIP and OIST

Julyan Arbel

Research director
(France side)

Statify-team, Inria
Grenoble Rhône-Alpes

Kenichi Bannai

Co-PI (Japan side)

Math-Science Team at
RIKEN-AIP and Keio
University

Rio Yokota

Co-PI
(Japan side)

Tokyo Institute of
Technology

Received total funding of JPY 220M + EUR 500K through the CREST-ANR grant! Thanks to JST for their generous funding!

Bayes-Duality Workshop

https://bayesduality.github.io/workshop_2024.html


Adam White
University of Alberta, Canada


Alexander Immer
ETH, Switzerland


Arindam Banerjee
University of Illinois Urbana-Champaign, US


Daiki Chijiwa
NTT Corporation, Japan

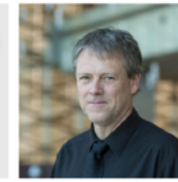

Ehsan Amid
Google DeepMind, US


Eugene Ndiaye
Apple, France


Frank Nielsen
Sony Computer Science Laboratories, Japan


Jonghyun Choi
Seoul National University, South Korea


Juho Lee
KAIST, South Korea


Haavard Rue
KAUST, Saudi Arabia


Hossein Mobahi
Google Research, US

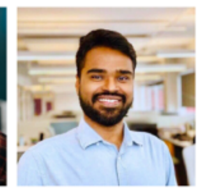

Martin Mundt
TU Darmstadt, Germany

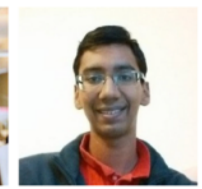

Matt Jones
University of Colorado, US


Nico Daheim
TU Darmstadt, Germany


Razvan Pascanu
Google DeepMind, US

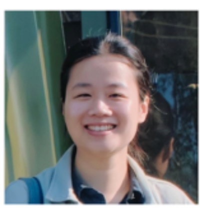

Rupam Mahmood
University of Alberta, Canada


Sarath Chandar
École Polytechnique de Montréal, Canada


Siddharth Swaroop
Harvard University, US


Tom Rainforth
University of Oxford, UK


Vincent Fortuin
Helmholtz AI, Germany


Yingzhen Li
Imperial College London, UK


Zelda Mariet
Bioptimus, US

Every year in June in Tokyo

Attendees are from a diverse research interests: Bayes, Duality, Continual/ Federated/Active learning, RL, Experiment Design etc.

Team Approx-Bayes

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

Emtiyaz Khan
Team Leader

Thomas Möllenhoff
Research Scientist

Keigo Nishida
Special Postdoctoral Resesarcher
RIKEN BDR

Hugo Monzón Maldonado
Postdoctoral Researcher

Pierre Alquier
Visiting Scientist
ESSEC Business School

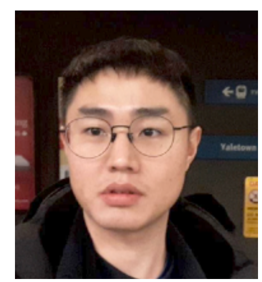
Dharmesh Tailor
Remote Collaborator
University of Amsterdam

Zhedong Liu
Postdoctoral Researcher

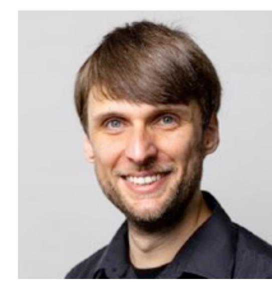
Anita Yang
Research Part-time Worker
The University of Tokyo

Clément Bazan
Intern
Tokyo Institute of Technology

Joseph Austerweil
Visiting Scientist
University of Winsconsin-Madison



Yohan Jung
(Started in July)



Christopher Anders
(Started in July)