

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

课件

笔记

代码

中英双语字幕

一键打包下载

官方笔记翻译

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习

Auto-encoder

生成式对抗网络

学习率

深度学习

卷积神经网络

GAN

自监督

自注意力机制

批次标准化

神经网络压缩

强化学习

元学习

Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**
称为 **AI 内容创作者?** 回复 [添砖加瓦]



Life Long Learning

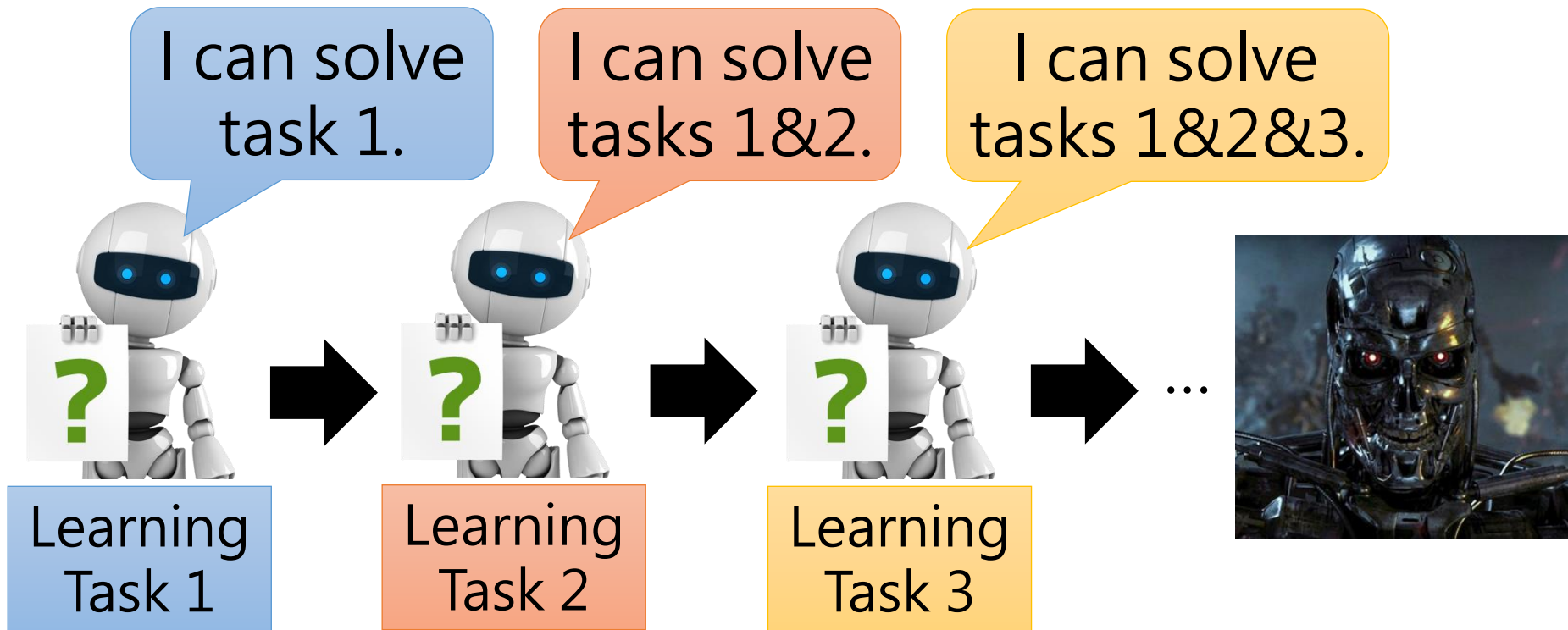
Hung-yi Lee
李宏毅

Life Long Learning (終身學習)



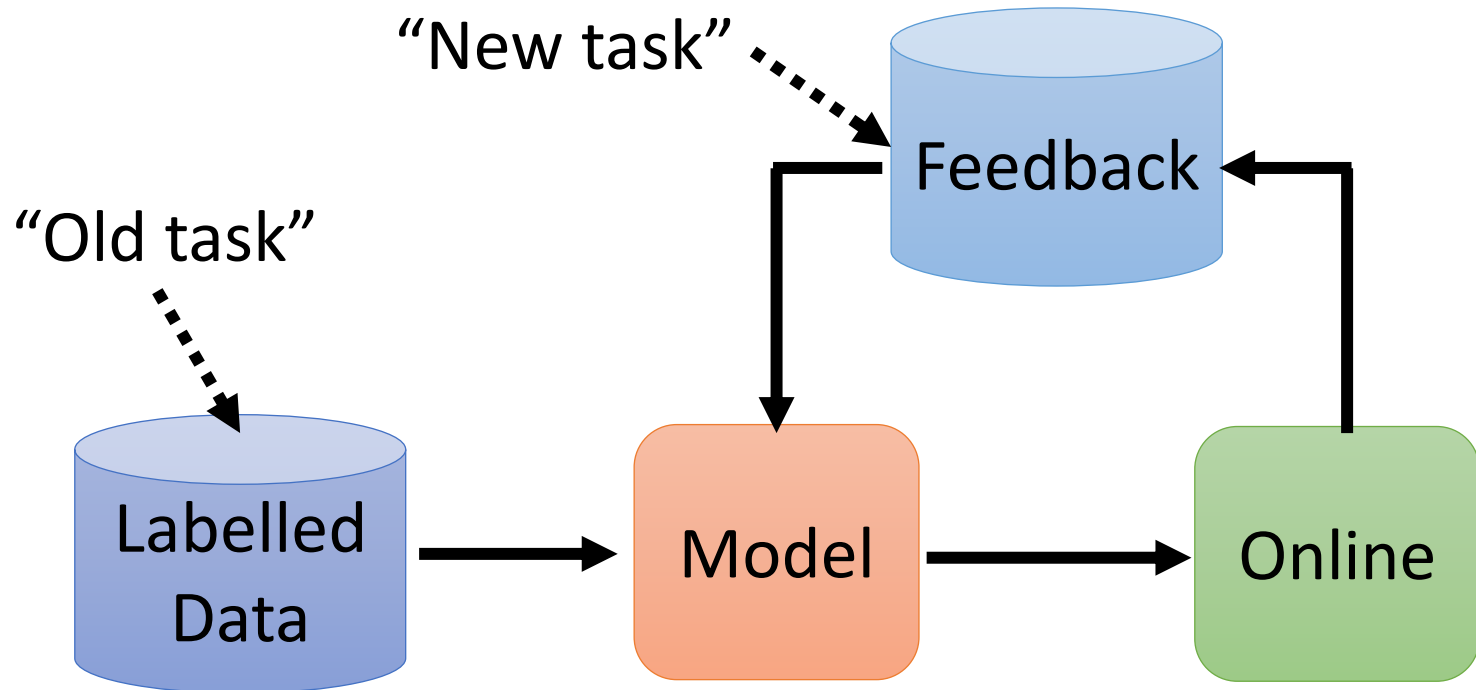
<https://world.edu/lifelong-learning-part-time-undergraduate-provision-crisis/>

What people think about AI ...



Life Long Learning (LLL), Continuous Learning,
Never Ending Learning, Incremental Learning

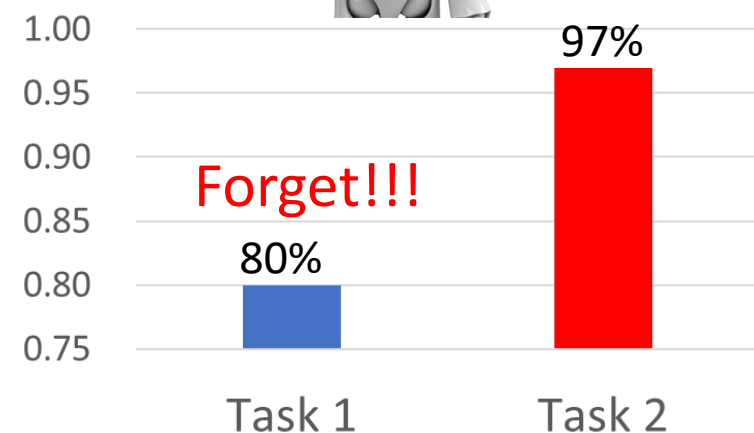
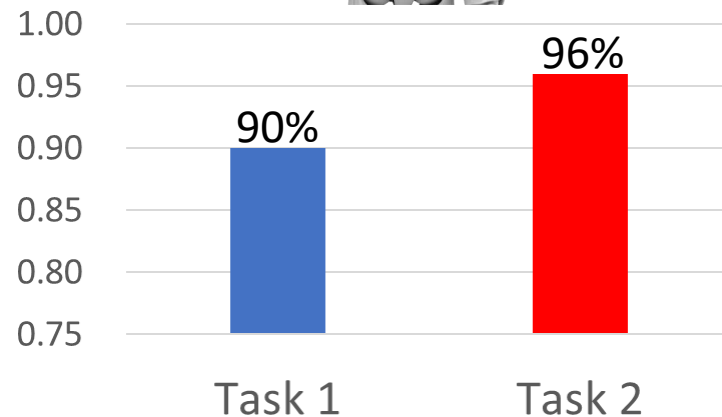
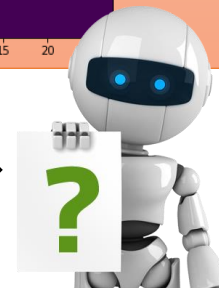
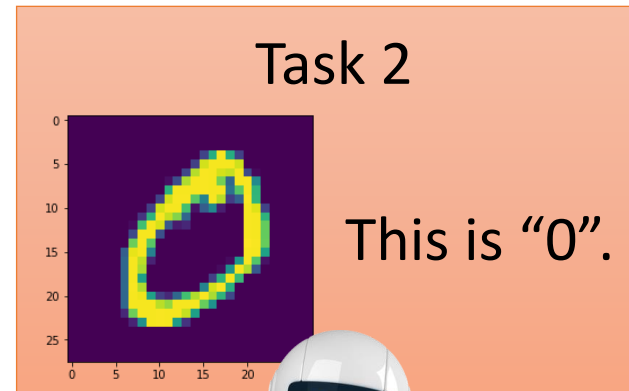
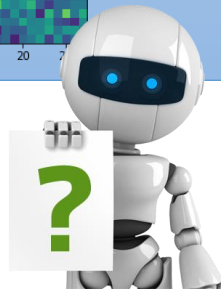
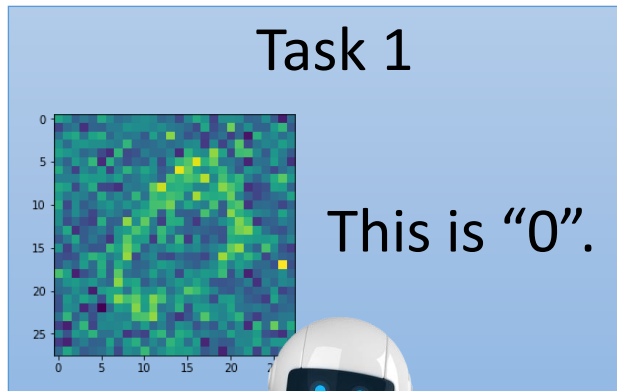
Life Long Learning in real-world applications

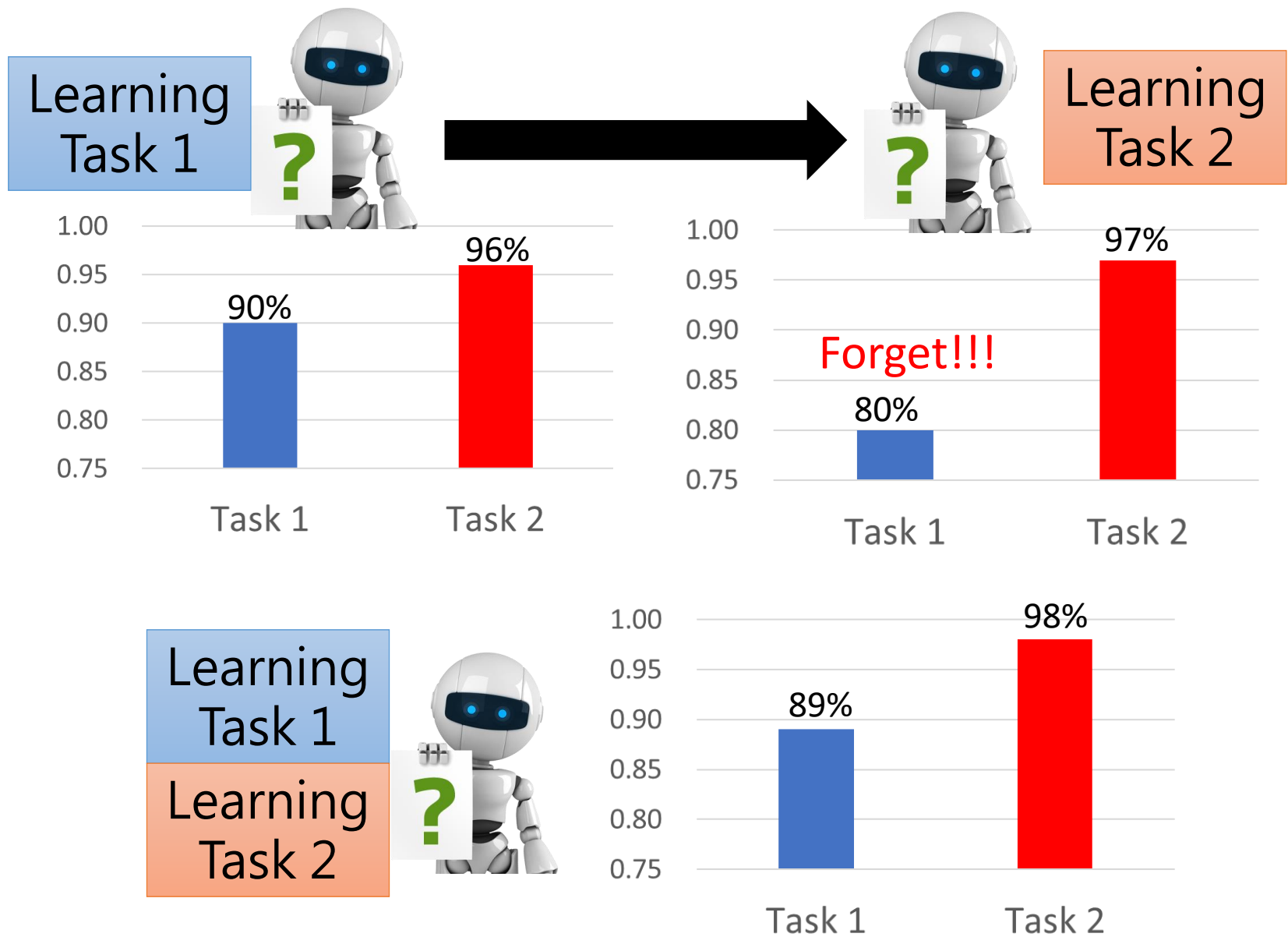


Example



= 3 layers, 50 neurons each





The network has enough capacity to learn both tasks.

Example

- QA: Given a document, answer the question based on the document.
- There are 20 QA tasks in bAbi corpus.

Task 5: Three Argument Relations

Mary gave the cake to Fred.

Fred gave the cake to Bill.

Jeff was given the milk by Bill.

Who gave the cake to Fred? A: Mary

Who did Fred give the cake to? A: Bill

Task 15: Basic Deduction

Sheep are afraid of wolves.

Cats are afraid of dogs.

Mice are afraid of cats.

Gertrude is a sheep.

What is Gertrude afraid of? A:wolves

- Train a QA model through the 20 tasks

Example

Task 5: Three Argument Relations

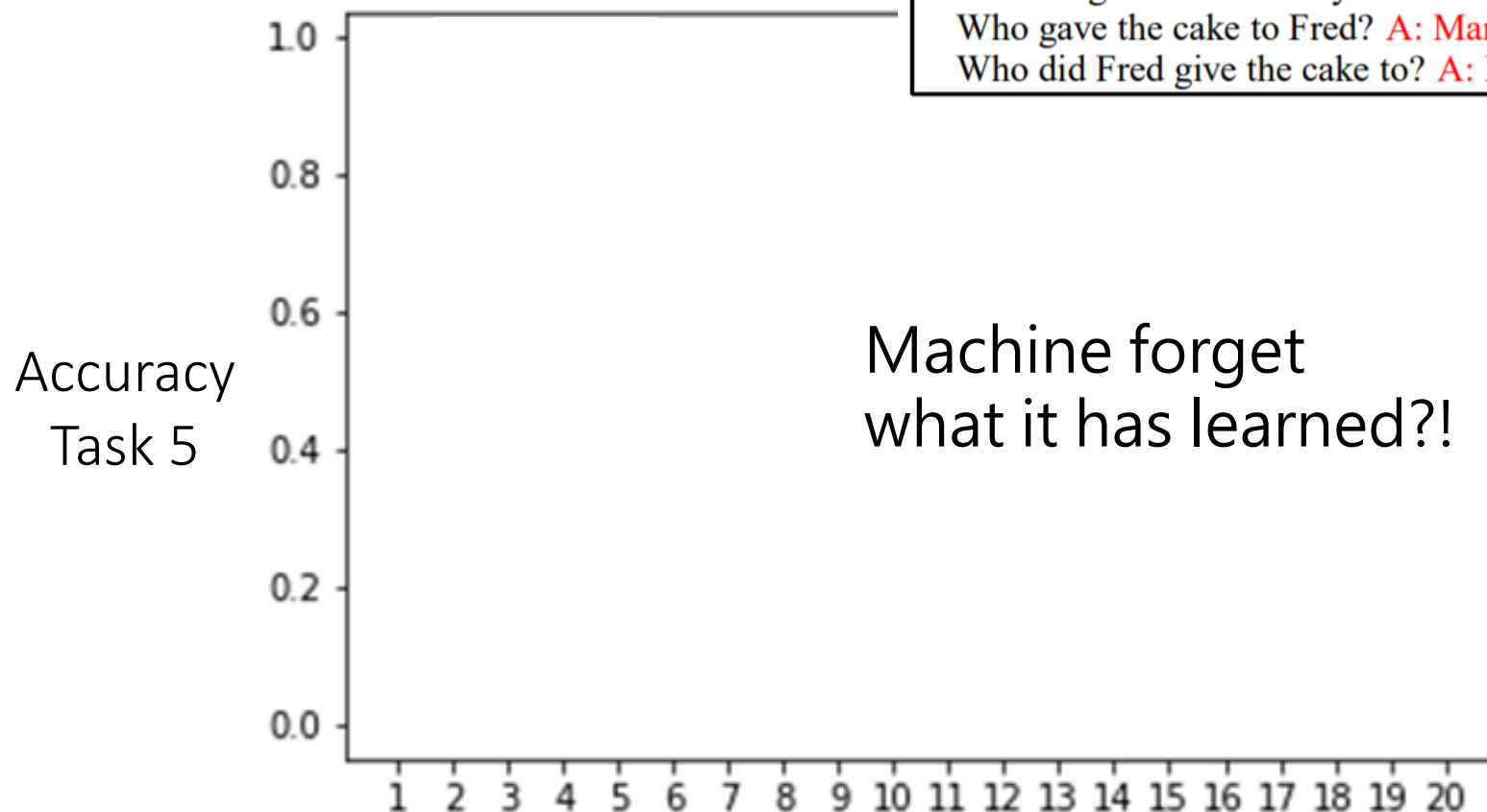
Mary gave the cake to Fred.

Fred gave the cake to Bill.

Jeff was given the milk by Bill.

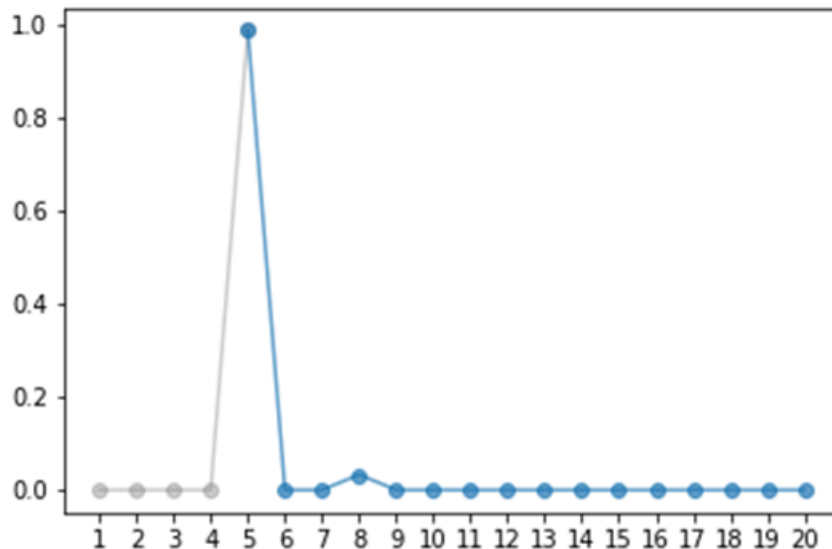
Who gave the cake to Fred? A: Mary

Who did Fred give the cake to? A: Bill



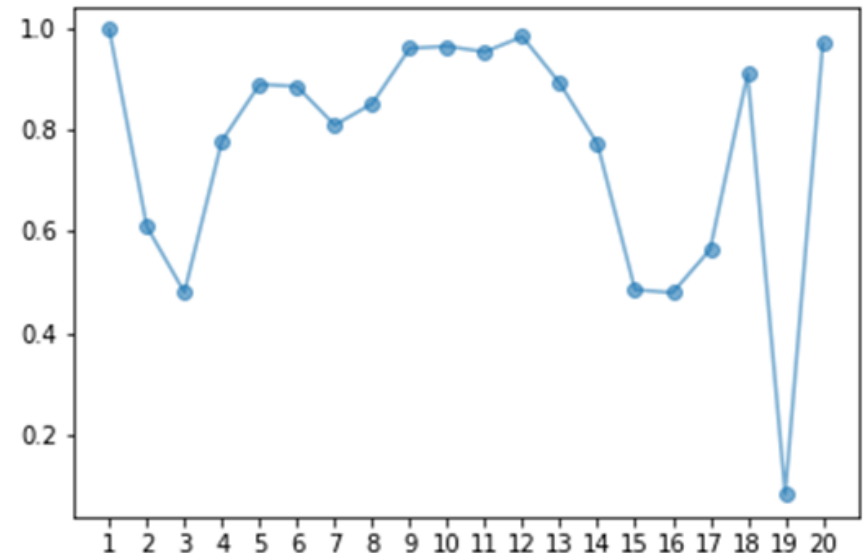
Example

Task 5 Accuracy



Learning 20 tasks
sequentially

Accuracy of all 20 tasks



Learning 20 tasks
simultaneously

Not because machine are not able to do it, but it just didn't do it.

是不為也 非不能也



Catastrophic
Forgetting

Learning
Task 1

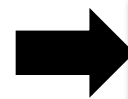
Learning
Task 2



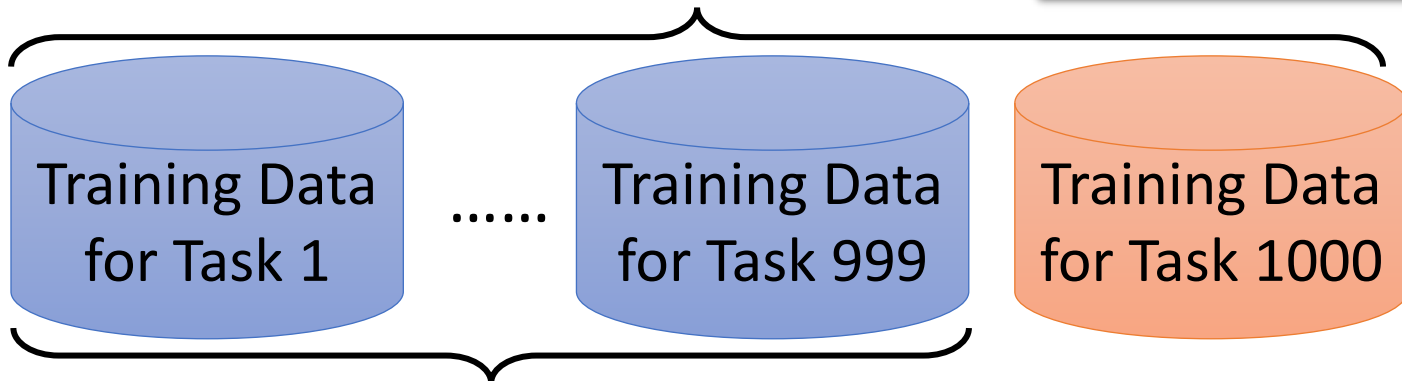
Wait a minute

- Multi-task training can solve the problem!

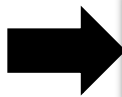
Using all the data for training



Computation issue



Always keep the data



Storage issue

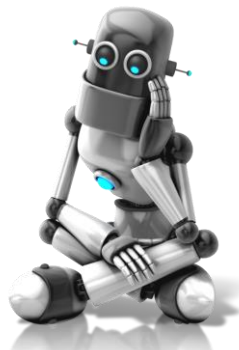
- Multi-task training can be considered as the upper bound of LLL.

Wait a minute

- Train a model for each task



Learning
Task 1



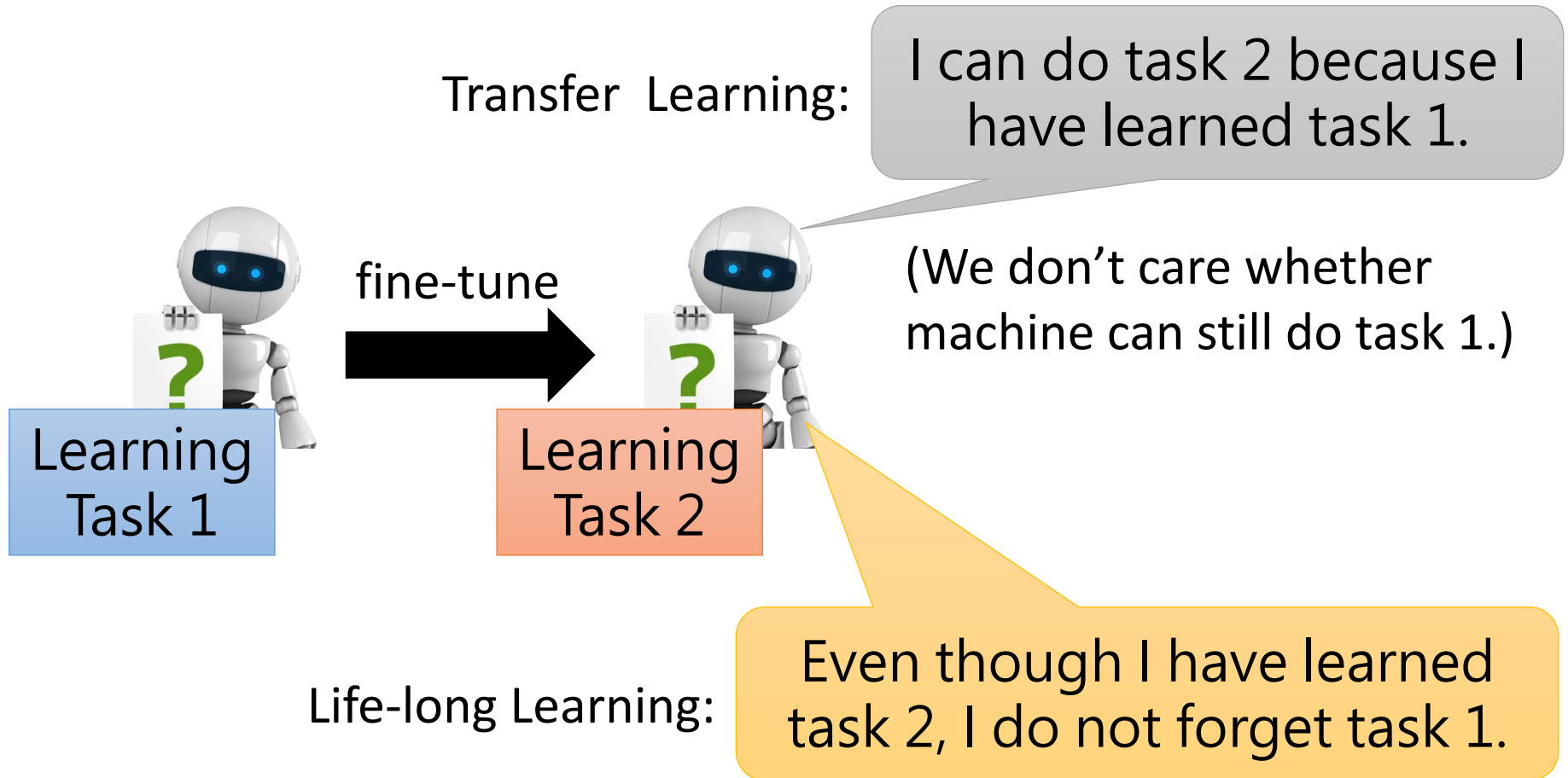
Learning
Task 2



Learning
Task 3

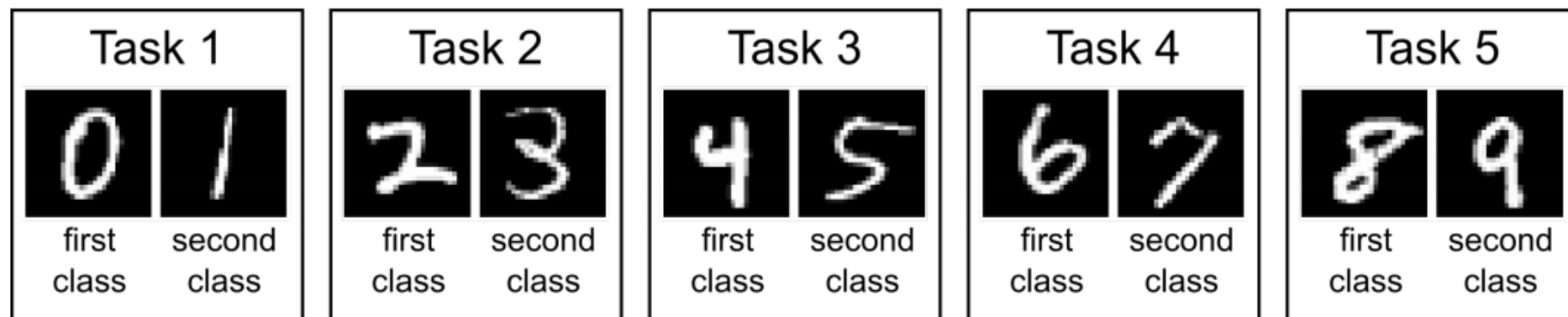
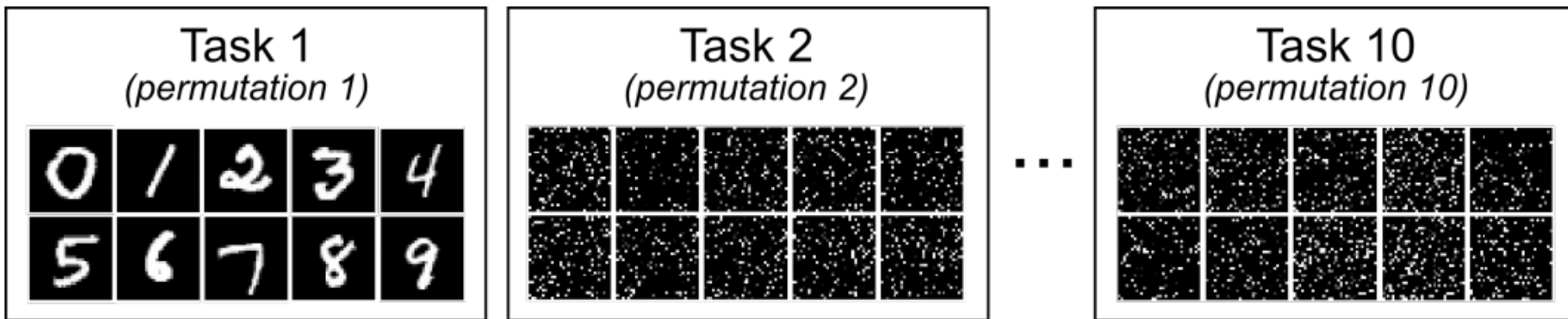
- Eventually we cannot store all the models ...
- Knowledge cannot transfer across different tasks

Life-Long v.s. Transfer



Evaluation

First of all, we need a sequence of tasks.



Evaluation

$R_{i,j}$: after training task i , performance on task j

If $i > j$,

After training task i , does task j be forgot

If $i < j$,

Can we transfer the skill of task i to task j

		Test on			
		Task 1	Task 2	Task T
Rand Init.		$R_{0,1}$	$R_{0,2}$		$R_{0,T}$
After Training	Task 1	$R_{1,1}$	$R_{1,2}$		$R_{1,T}$
	Task 2	$R_{2,1}$	$R_{2,2}$		$R_{2,T}$
	⋮				
	Task T-1	$R_{T-1,1}$	$R_{T-1,2}$		$R_{T-1,T}$
	Task T	$R_{T,1}$	$R_{T,2}$		$R_{T,T}$

$$\text{Accuracy} = \frac{1}{T} \sum_{i=1}^T R_{T,i}$$

$$\text{Backward Transfer} = \frac{1}{T-1} \sum_{i=1}^{T-1} R_{T,i} - R_{i,i}$$

(It is usually negative.)

Evaluation

$R_{i,j}$: after training task i , performance on task j


If $i > j$,

After training task i , does task j be forgot

If $i < j$,

Can we transfer the skill of task i to task j

		Test on			
		Task 1	Task 2	Task T
Rand Init.		$R_{0,1}$	$R_{0,2}$		$R_{0,T}$
After Training	Task 1	$R_{1,1}$	$R_{1,2}$		$R_{1,T}$
	Task 2	$R_{2,1}$	$R_{2,2}$		$R_{2,T}$
	:				
	Task T-1	$R_{T-1,1}$	$R_{T-1,2}$		$R_{T-1,T}$
	Task T	$R_{T,1}$	$R_{T,2}$		$R_{T,T}$



$$\text{Accuracy} = \frac{1}{T} \sum_{i=1}^T R_{T,i}$$

$$\text{Backward Transfer} = \frac{1}{T-1} \sum_{i=1}^{T-1} R_{T,i} - R_{i,i}$$

$$\text{Forward Transfer} = \frac{1}{T-1} \sum_{i=2}^T R_{i-1,i} - R_{0,i}$$

Research Directions

突觸的

可塑性

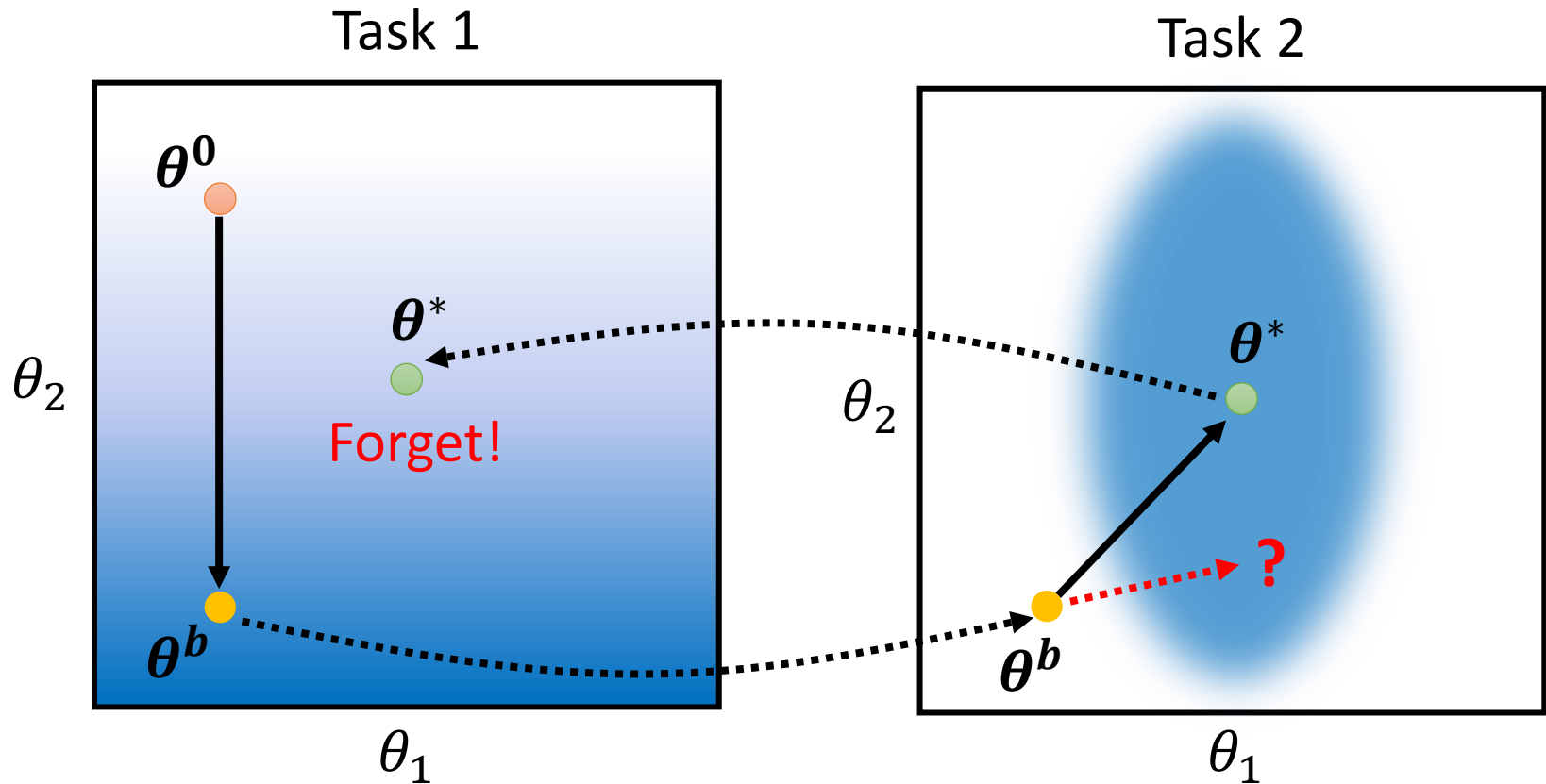
Regularization-
based Approach

Selective Synaptic Plasticity

Additional Neural Resource Allocation

Memory Reply

Why Catastrophic Forgetting?



The error surfaces of tasks 1 & 2.
(darker = smaller loss)

Selective Synaptic Plasticity

Basic Idea: Some parameters in the model are important to the previous tasks. Only change the unimportant parameters.

θ^b is the model learned from the previous tasks.

Each parameter θ_i^b has a “guard” b_i

Diagram illustrating the equation for selective synaptic plasticity, with annotations explaining the components:

$$\underline{L'(\boldsymbol{\theta})} = \underline{L(\boldsymbol{\theta})} + \lambda \sum_i \underline{b_i} (\underline{\theta_i} - \underline{\theta_i^b})^2$$

Annotations:

- $L'(\boldsymbol{\theta})$: Loss to be optimized
- $L(\boldsymbol{\theta})$: Loss for current task
- λ : Learning rate
- \sum_i : Sum over parameters
- b_i : How important this parameter is
- θ_i : Parameters to be learning
- θ_i^b : Parameters learned from previous task

Selective Synaptic Plasticity

Basic Idea: Some parameters in the model are important to the previous tasks. Only change the unimportant parameters.

θ^b is the model learned from the previous tasks.

Each parameter θ_i^b has a “guard” b_i

θ should be close to θ^b in certain directions.

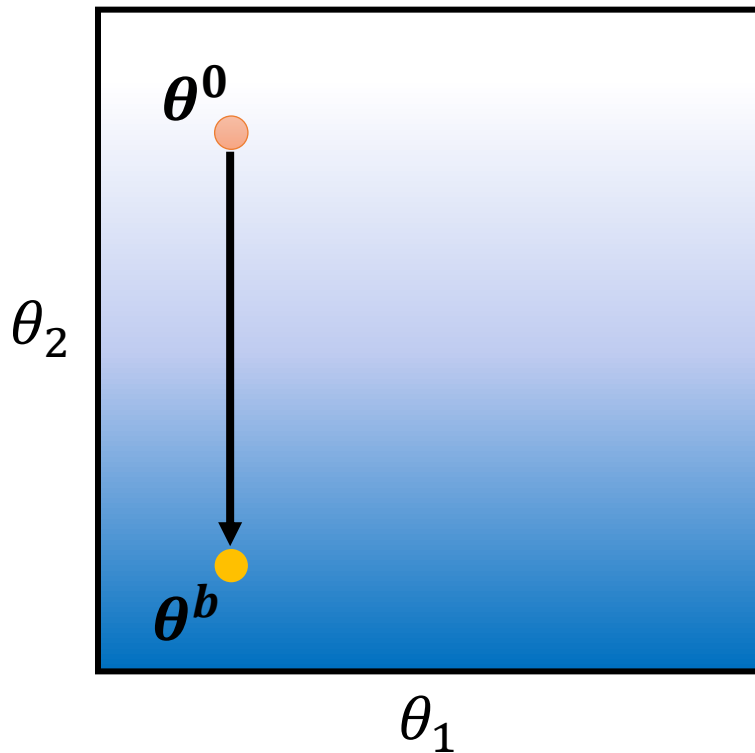
$$L'(\theta) = L(\theta) + \lambda \sum_i b_i (\theta_i - \theta_i^b)^2$$

If $b_i = 0$, there is no constraint on θ_i \longrightarrow Catastrophic Forgetting

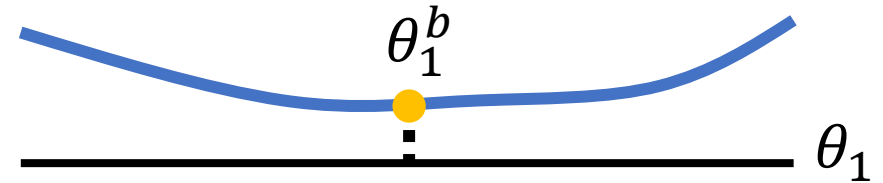
If $b_i = \infty$, θ_i would always be equal to θ_i^b \longrightarrow Intransigence

Selective Synaptic Plasticity

Task 1

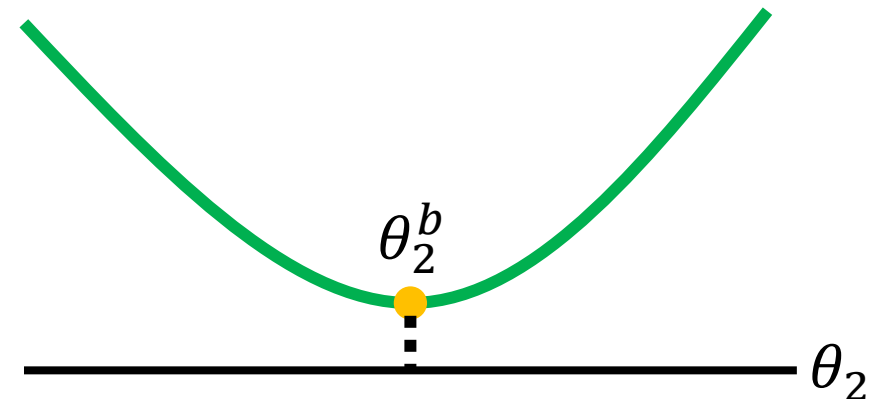


Each parameter has a
“guard” b_i



can be changed 😊

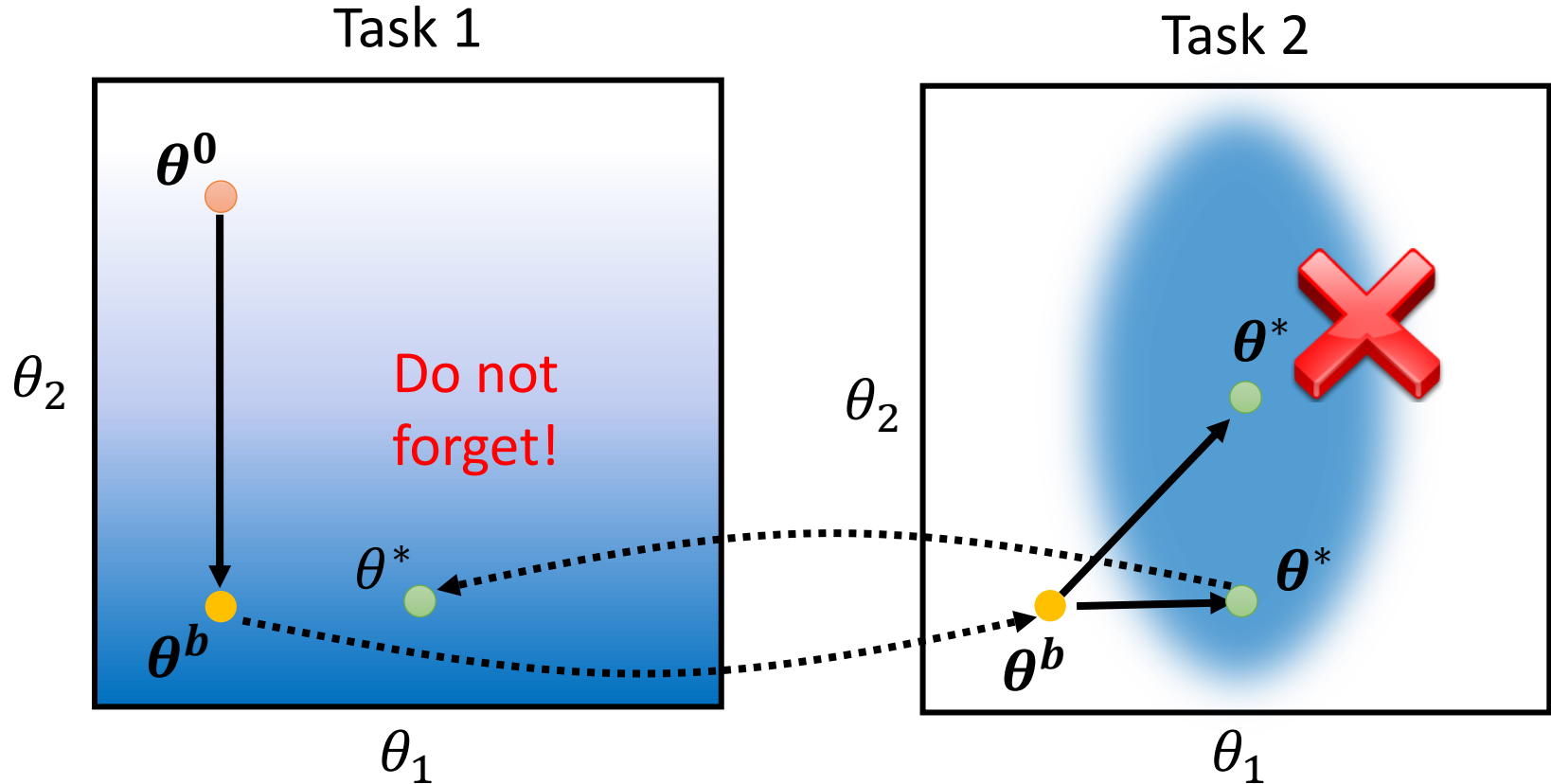
➡ b_1 is small



don't touch it!

➡ b_2 is large

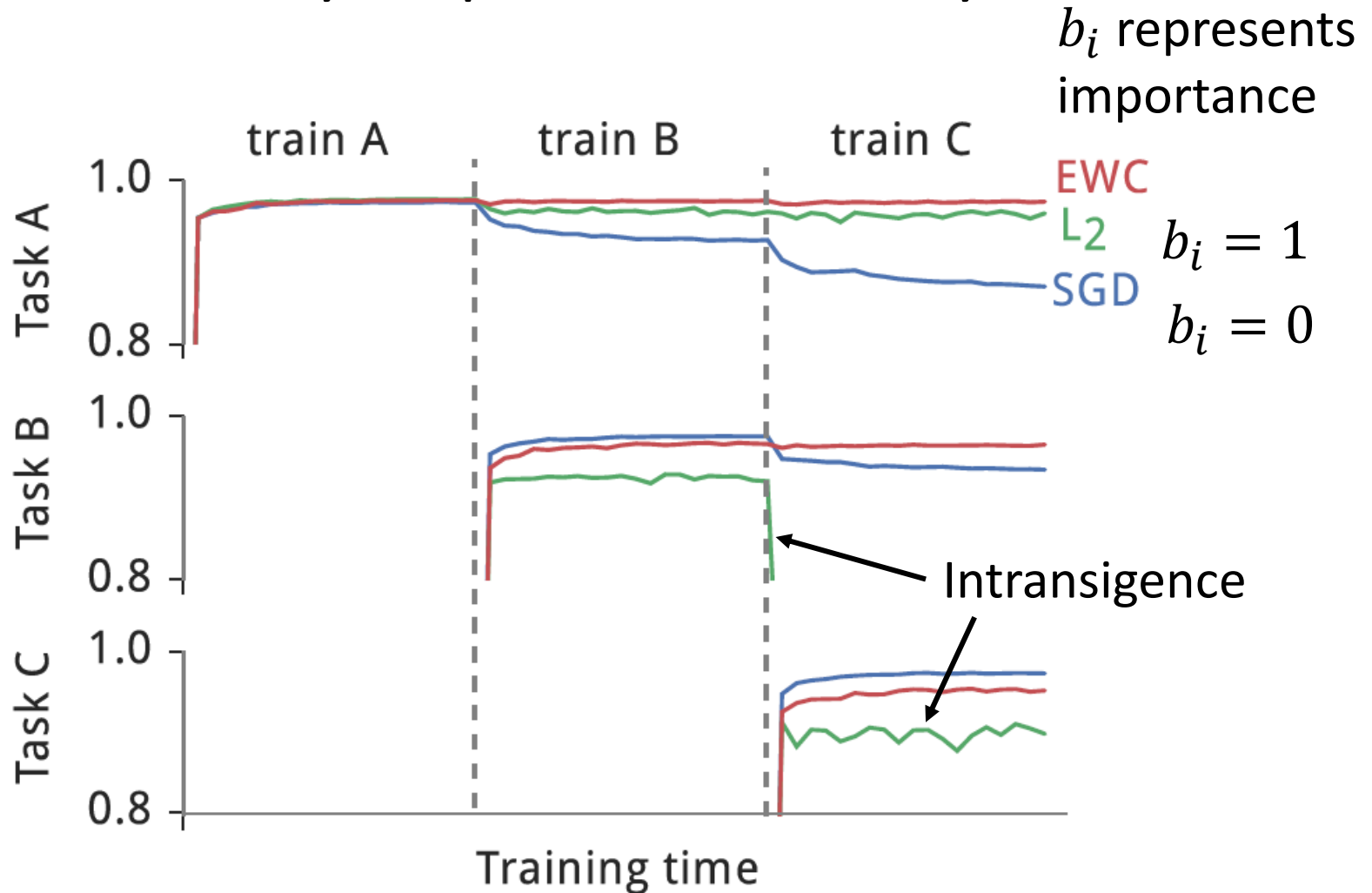
Selective Synaptic Plasticity



b_1 is small, while b_2 is large.

(We can modify θ_1 , but do not change θ_2 .)

Selective Synaptic Plasticity



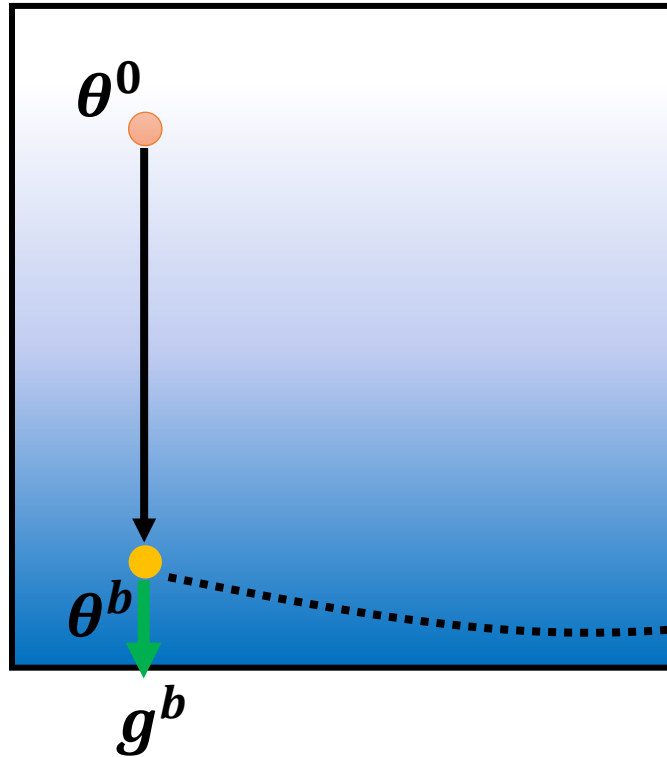
Selective Synaptic Plasticity

- Elastic Weight Consolidation (EWC)
 - <https://arxiv.org/abs/1612.00796>
- Synaptic Intelligence (SI)
 - <https://arxiv.org/abs/1703.04200>
- Memory Aware Synapses (MAS)
 - <https://arxiv.org/abs/1711.09601>
- RWalk
 - <https://arxiv.org/abs/1801.10112>
- Sliced Cramer Preservation (SCP)
 - <https://openreview.net/forum?id=BJge3TNKwH>

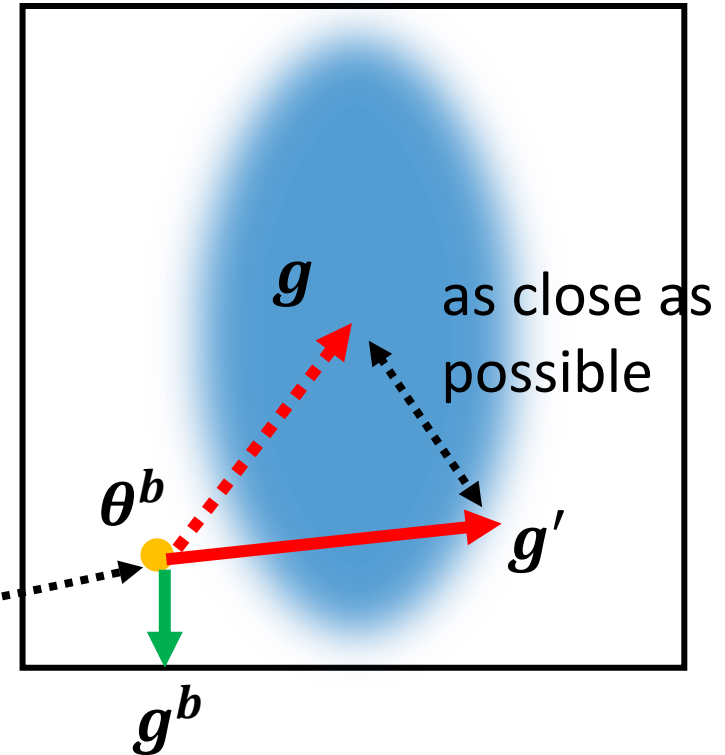
Gradient Episodic Memory (GEM)

<https://arxiv.org/abs/1706.08840>

Task 1



Task 2



$$g' \cdot g^b \geq 0$$

- ⋯→ : negative gradient of current task
- : negative gradient of previous task
- : update direction

Need the data from the previous tasks

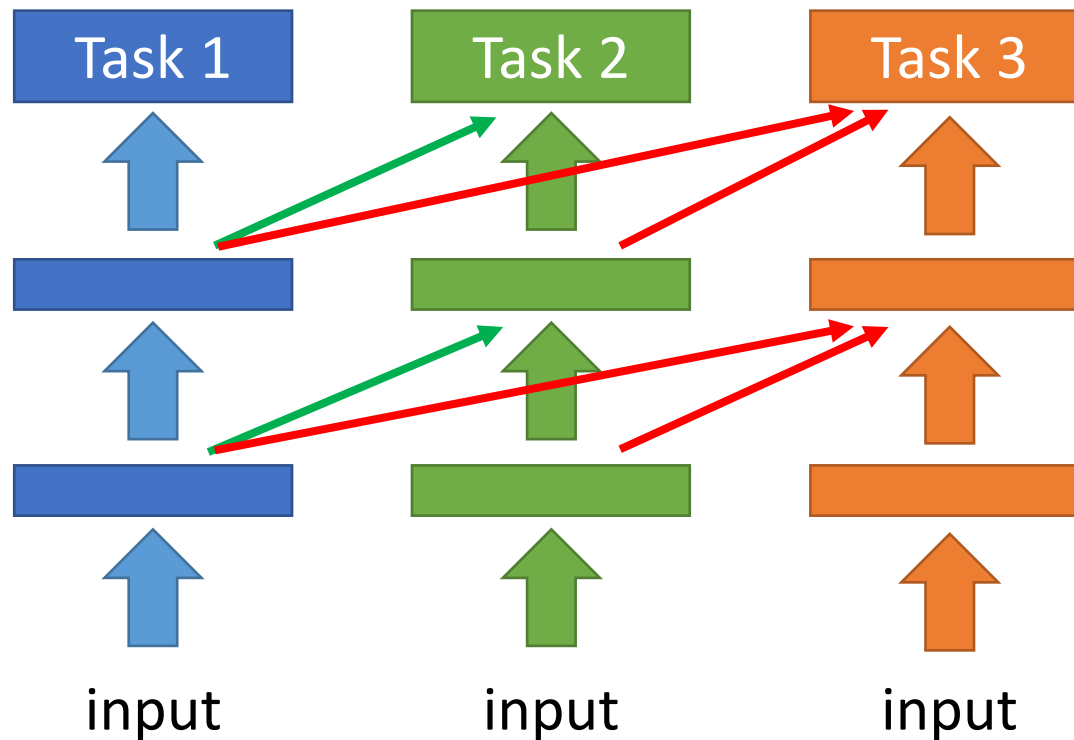
Research Directions

Selective Synaptic Plasticity

Additional Neural Resource Allocation

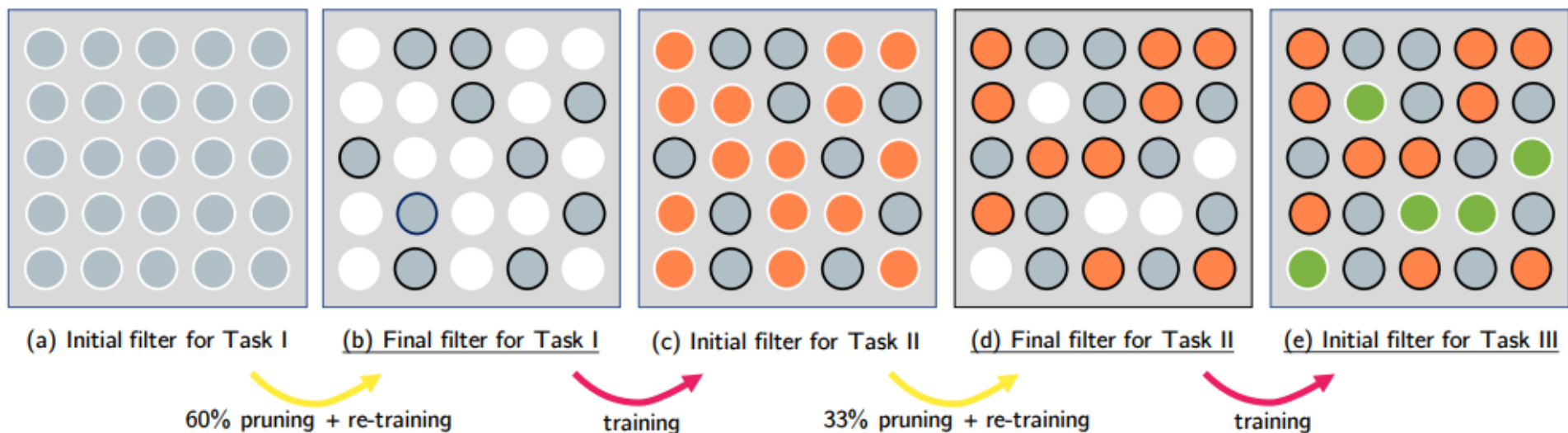
Memory Reply

Progressive Neural Networks

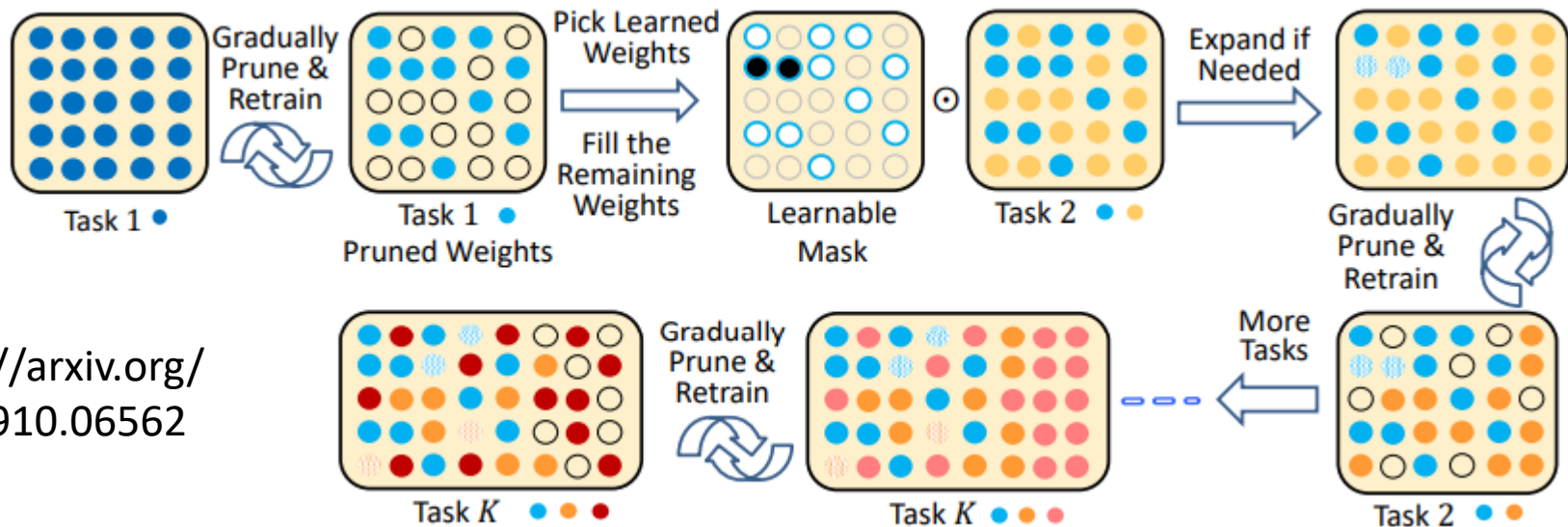


PackNet

<https://arxiv.org/abs/1711.05769>



Compacting, Picking, and Growing (CPG)



<https://arxiv.org/abs/1910.06562>

Research Directions

Selective Synaptic Plasticity

Additional Neural Resource Allocation

Memory Reply

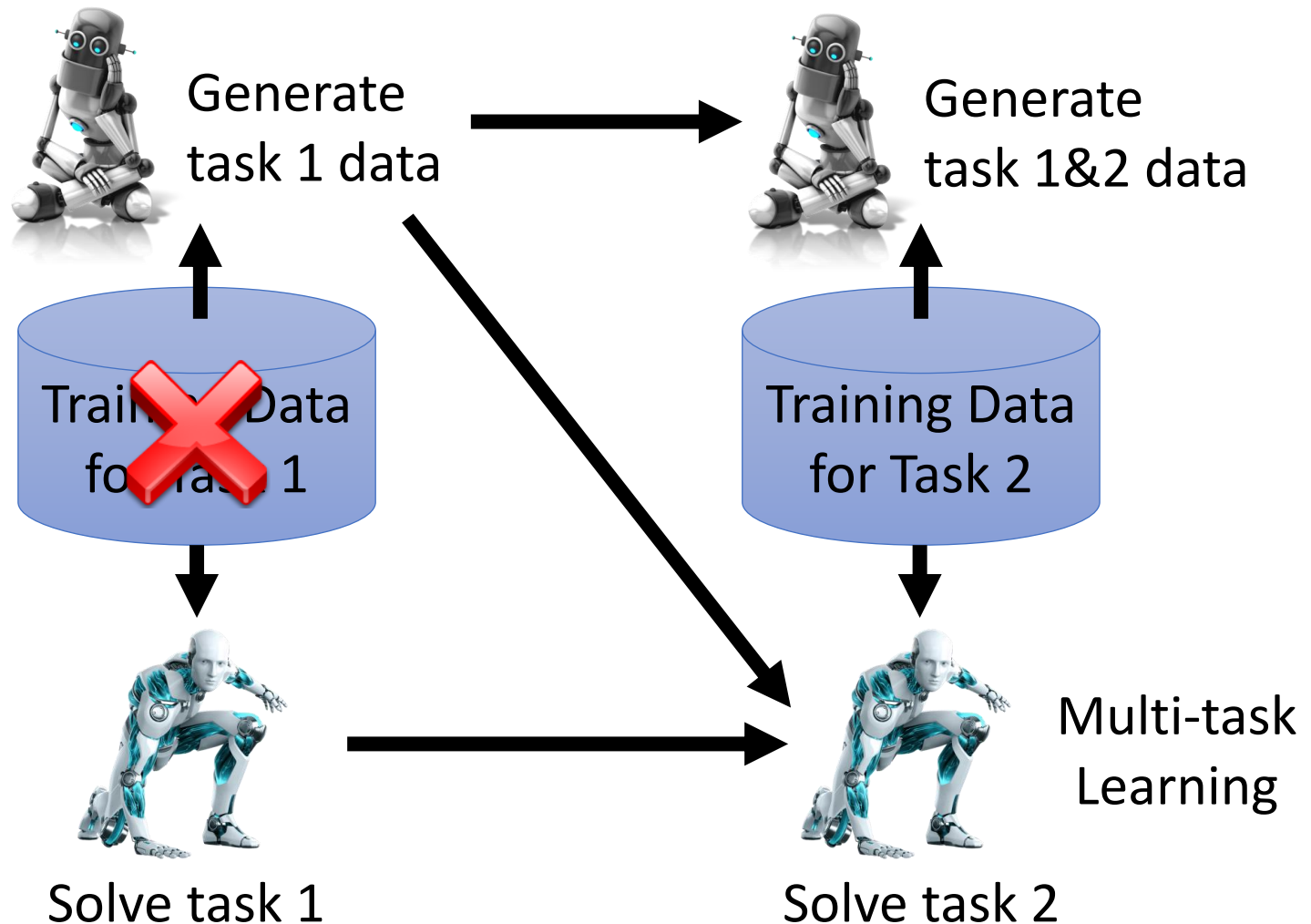
Generating Data

<https://arxiv.org/abs/1705.08690>

<https://arxiv.org/abs/1711.10563>

<https://arxiv.org/abs/1909.03329>

- Generating pseudo-data using generative model for previous tasks



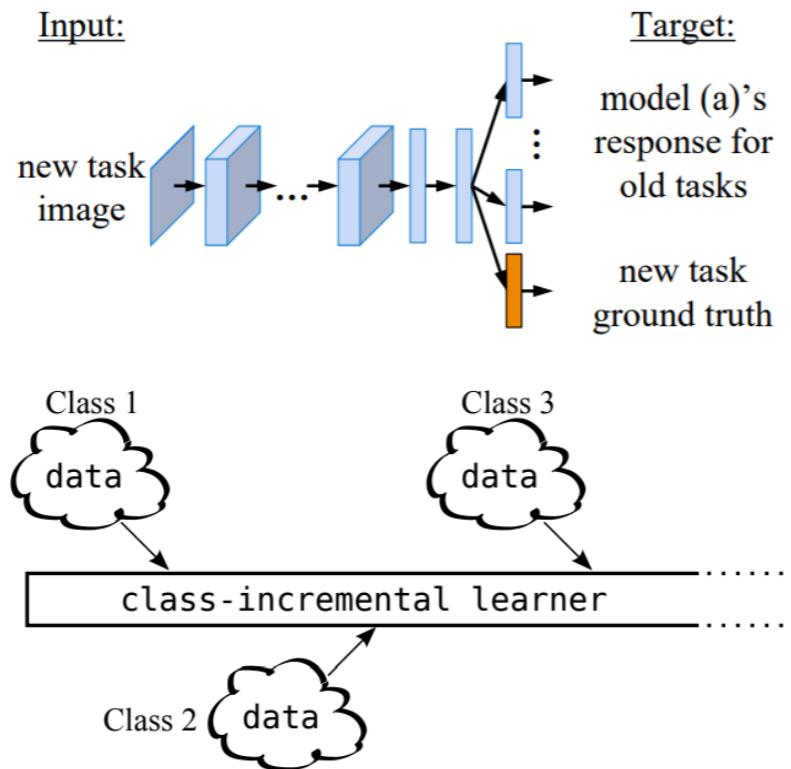
Adding new classes

Learning without forgetting (LwF)

<https://arxiv.org/abs/1606.09282>

iCaRL: Incremental Classifier and Representation Learning

<https://arxiv.org/abs/1611.07725>



Three scenarios for continual learning

<https://arxiv.org/abs/1904.07734>

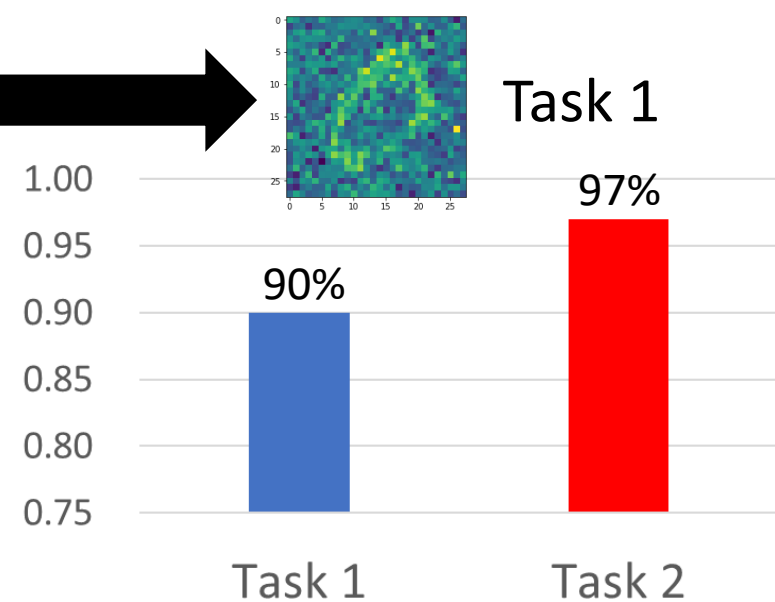
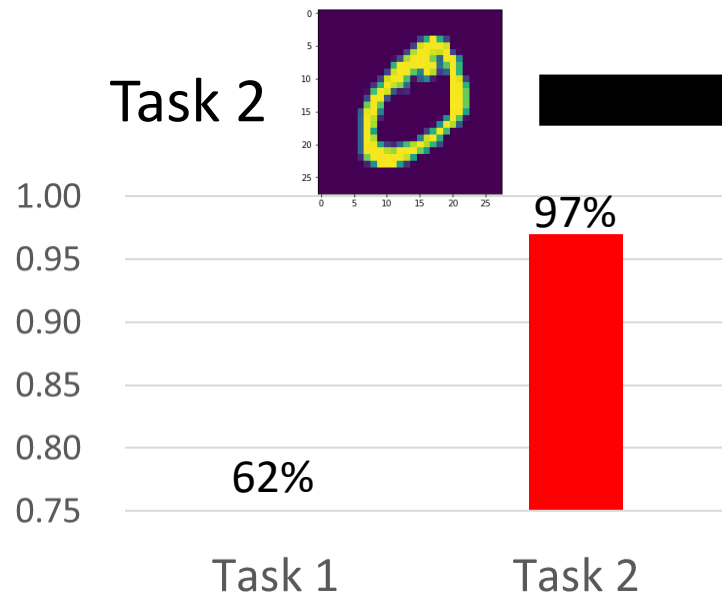
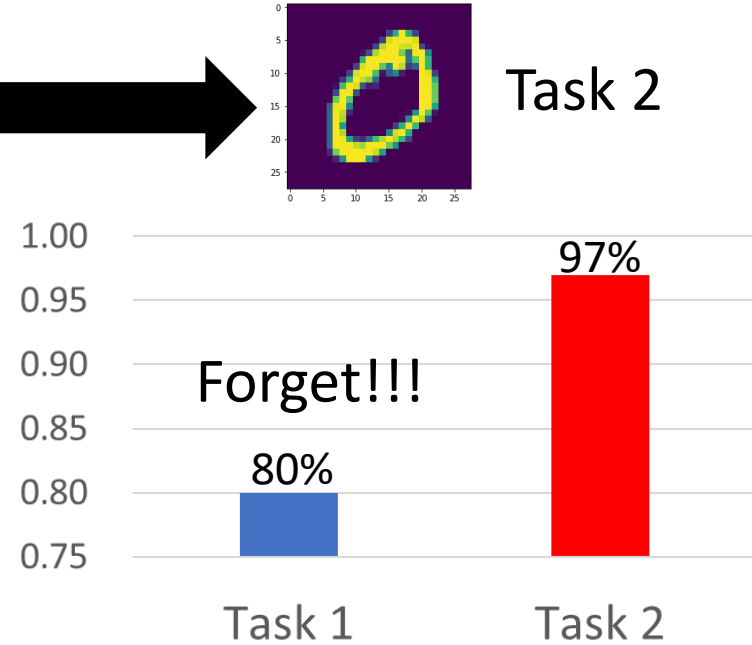
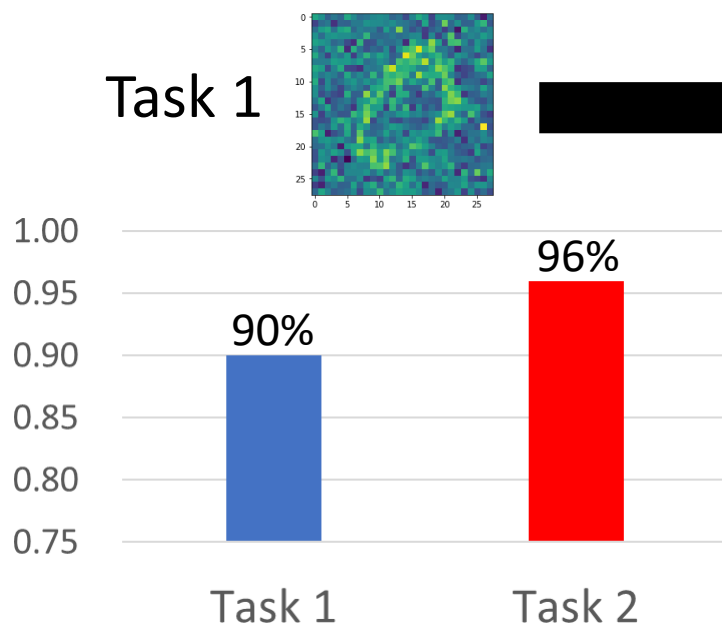
Concluding Remarks

Memory Reply

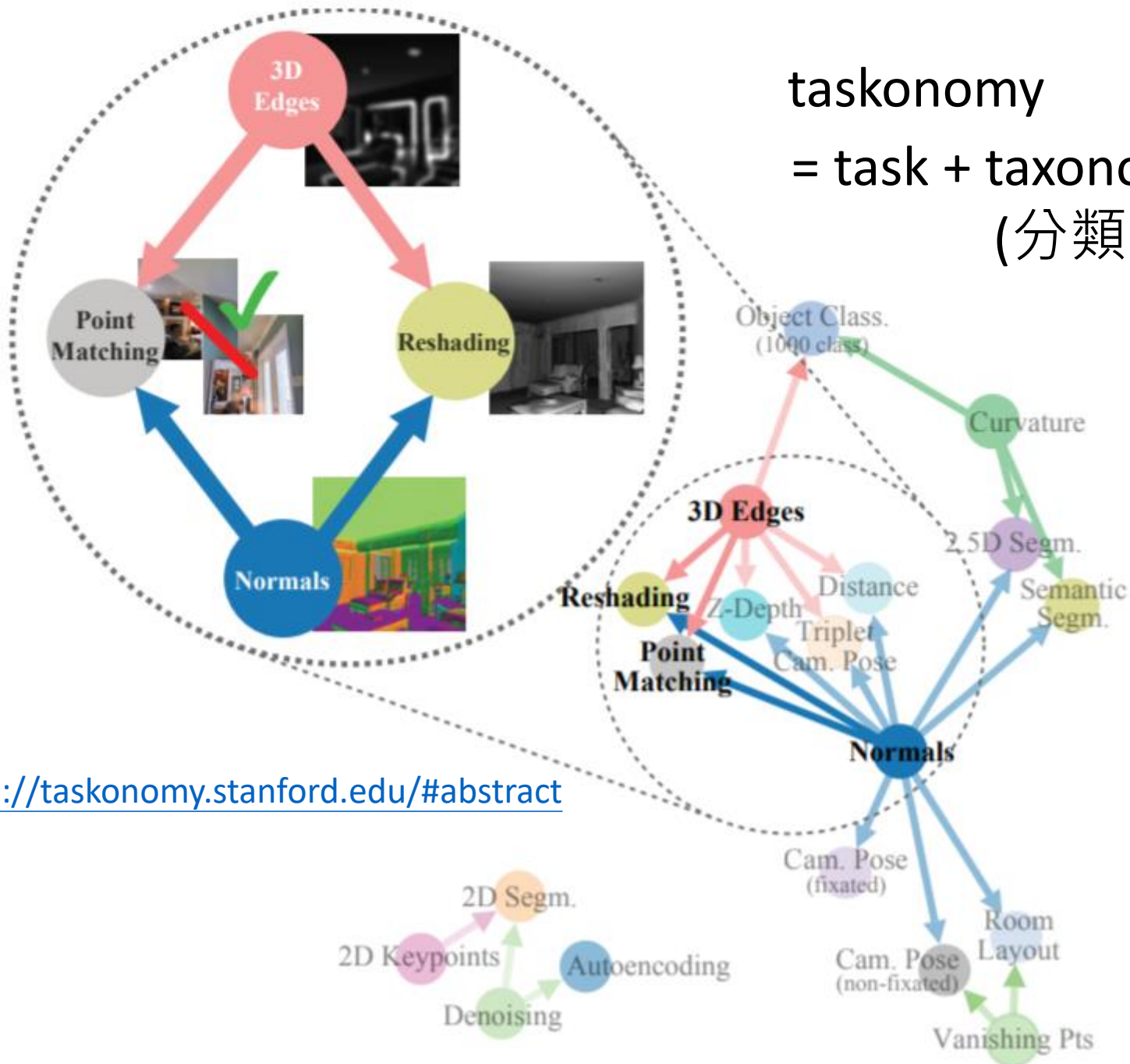
Additional Neural Resource Allocation

Selective Synaptic Plasticity

Curriculum Learning : what is the proper learning order?



taskonomy
= task + taxonomy
(分類學)



<http://taskonomy.stanford.edu/#abstract>

李宏毅 (Hung-yi Lee) · HYLEE | Machine Learning (2021)

HYLEE(2021) · 课程资料包 @ShowMeAI



视频

课件

笔记

代码

中英双语字幕

一键打包下载

官方笔记翻译

作业项目解析



视频 · B 站 [扫码或点击链接]

<https://www.bilibili.com/video/BV1fM4y137M4>



课件 & 代码 · 博客 [扫码或点击链接]

<http://blog.showmeai.tech/ntu-hylee-ml>

机器学习

Auto-encoder

生成式对抗网络

学习率

深度学习

卷积神经网络

GAN

自监督

自注意力机制

批次标准化

神经网络压缩

强化学习

元学习

Transformer

Awesome AI Courses Notes Cheatsheets 是 [ShowMeAI](#) 资料库的分支系列，覆盖最具知名度的 **TOP50+** 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

点击课程名称，跳转至课程**资料包**页面，**一键下载**课程全部资料！

机器学习	深度学习	自然语言处理	计算机视觉
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n
# Awesome AI Courses Notes Cheatsheets · 持续更新中			
知识图谱	图机器学习	深度强化学习	自动驾驶
Stanford · CS520	Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094



微信公众号

资料下载方式 2: 扫码点击**底部菜单栏**

称为 **AI 内容创作者**? 回复 [添砖加瓦]