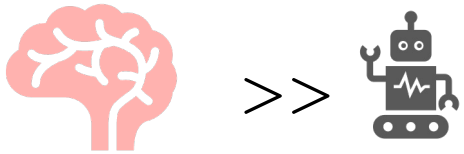
Maslow's Hammer for Catastrophic Forgetting: Node Re-Use vs. Node Activation

ICML 2022

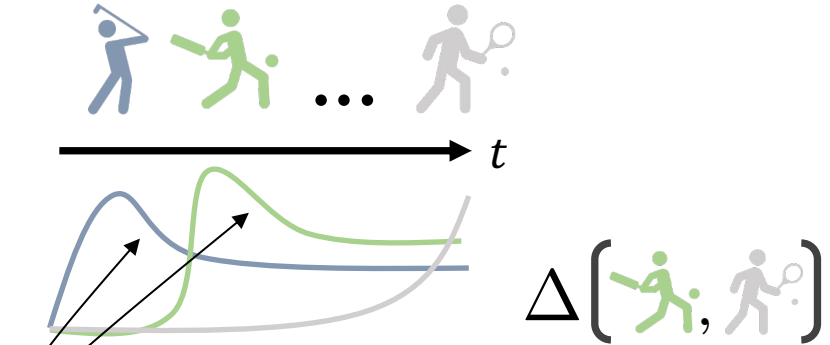
Sebastian Lee, Stefano Sarao Mannelli, Claudia Clopath, Sebastian Goldt, Andrew Saxe

Introduction

Naturalistic learning, "general" intelligence require multi-task & **continual learning**.

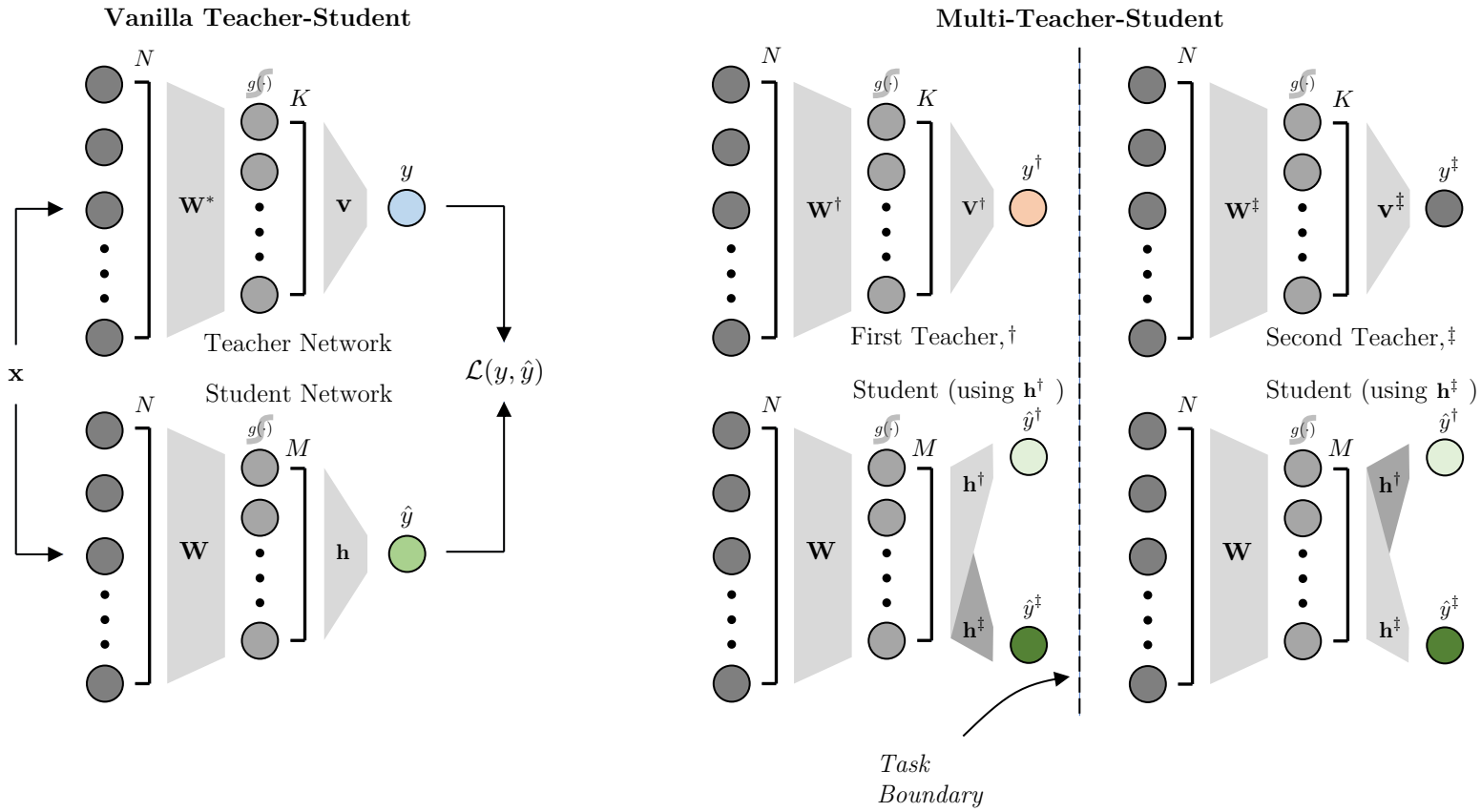


Animals are very good continual learners. Artificial agents are not.

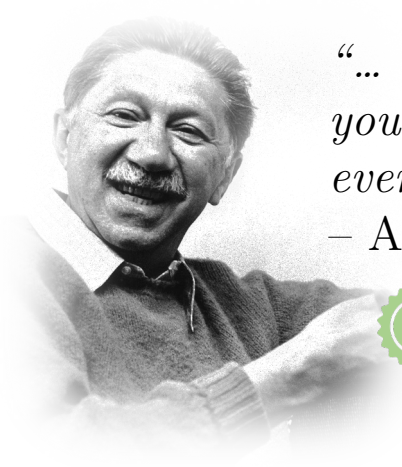


Aim: Better understand **catastrophic forgetting** in deep networks, especially in terms of **task similarity**.

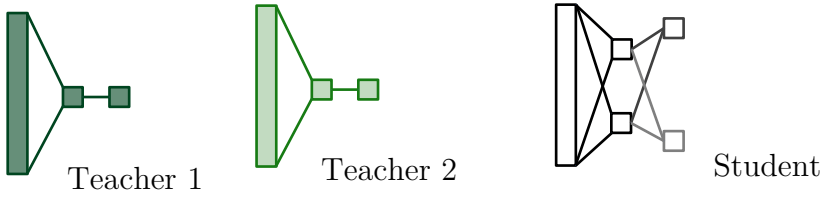
Teacher-Student Framework

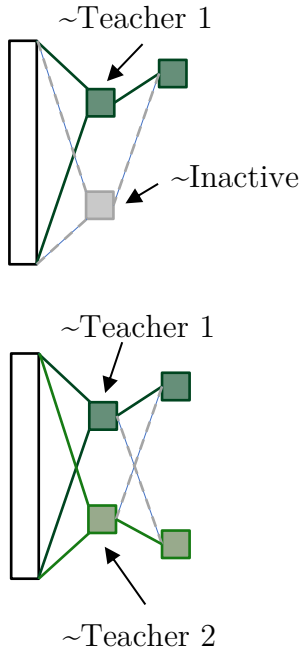


Maslow's Hammer Hypothesis

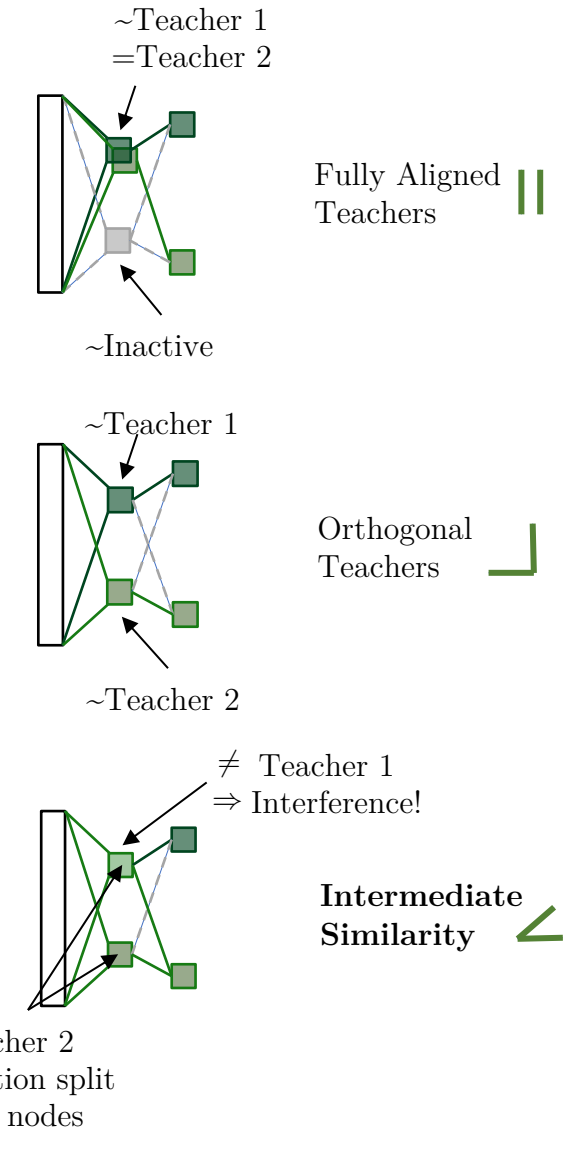


"... it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail."
– Abraham Maslow, 1966





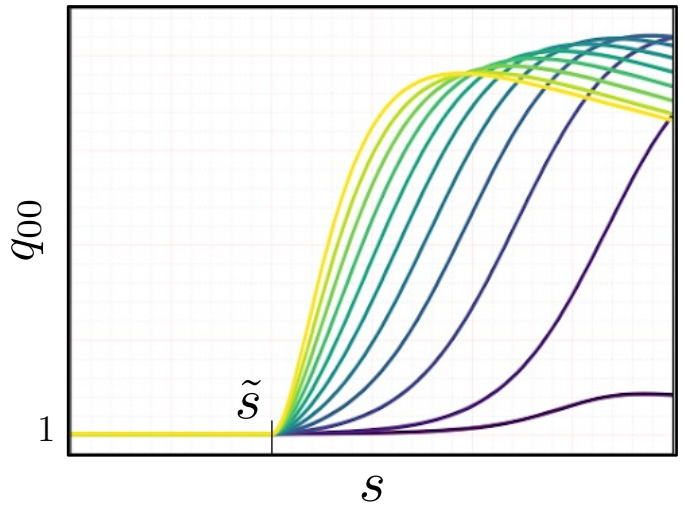
(Hypothetical)
minimal interference
solution



Teacher 2
solution split
over nodes

Evidence (Synthetic Data)

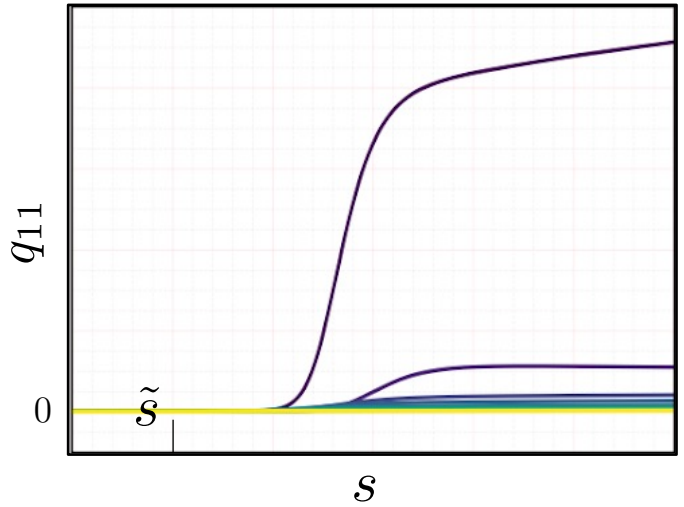
Specialised Node



Rate of movement from fixed point *increases* monotonically with teacher-teacher similarity.

$\Longrightarrow$ tendency for node re-use when tasks highly similar.

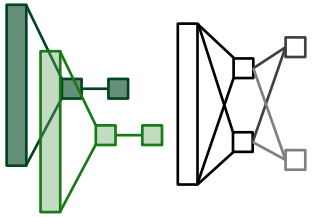
Inactive Node



Rate of movement from fixed point *decreases* monotonically with teacher-teacher similarity.

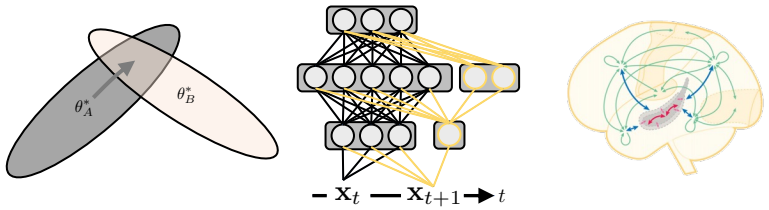
$\Longrightarrow$ tendency for node activation when tasks highly dissimilar.

Further Details



Teacher-Student

Fashion MNIST



Combatting Methods