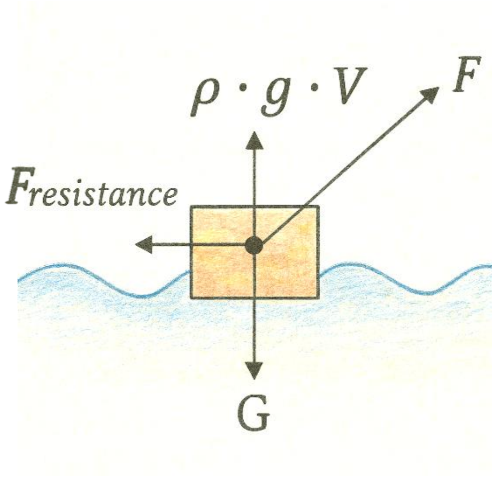
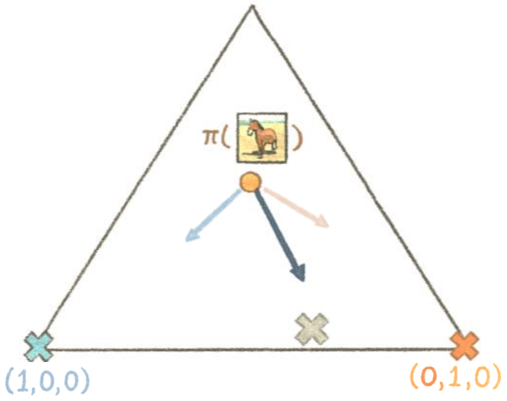
Learning Dynamics of Deep Learning

-- Force Analysis of Deep Neural Networks



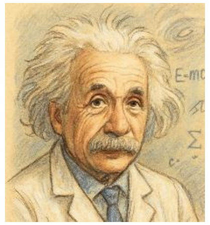


Yi (Joshua) Ren
Supervisor: Danica J. Sutherland

Motivation: understanding and controlling AI systems need theory

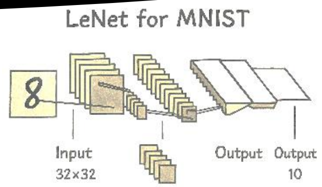
Model size
Reasoning depth
Problem facing
...

Complexity



It is the **theory** which decides what we can observe.
-- Albert Einstein

ML Theory



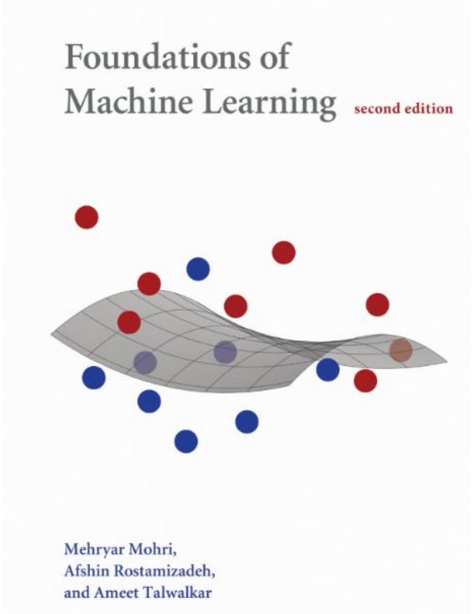
ChatGPT
Gemini
Claude

Year

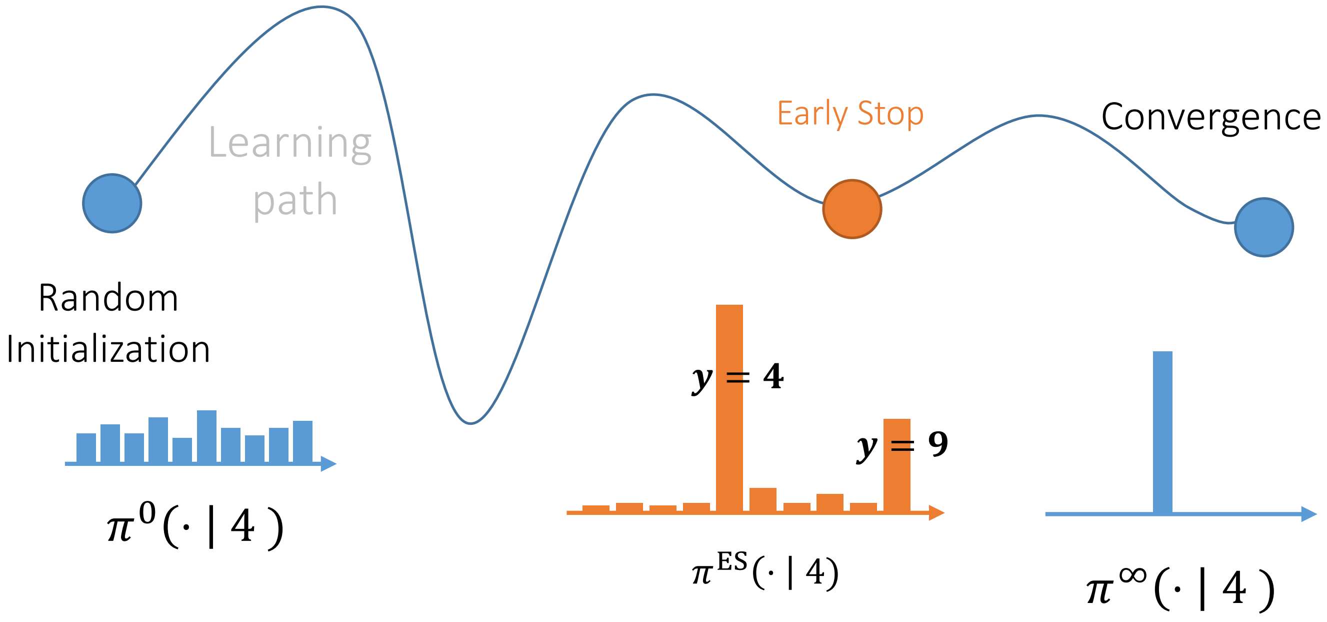
1998 2025 ???

Motivation: ML theory needs diverse perspectives

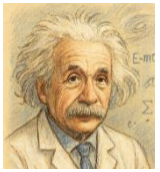
- PAC learning framework
  -- strict, elegant, global, and macroscopic



- But, I failed to use it understanding this emergent behavior:
  -- an interesting pairing effect emerges during training



Methodology: zoom in, in time and sample spaces



It is the theory which decides what we can **observe**.
-- Albert Einstein

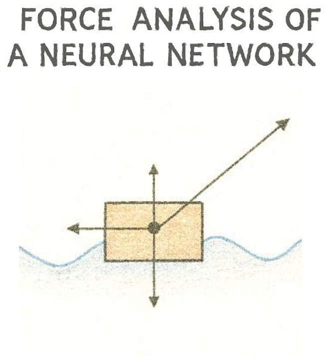
Force analysis of Neural Network (Learning Dynamics of Deep Learning)

A fine-grained, physics-inspired ML theoretical framework

Outline

Chapt. 1 & 2

Intro.
PHY & DNN



→

Chapt. 3 & 4

Warm up:
Fundamentals of
Learning Dynamics

Ren et al. ICLR 2022
Ren et al. ICLR2020
Guo, Ren et al. ICLR2024

→

Chapt. 5

Main Application:
LLM's fine-tuning

Ren et al. ICLR 2025
(Outstanding Paper Award)
Deng, Ren et al. ICML-W 2025

→

Chapt. 6 & 7

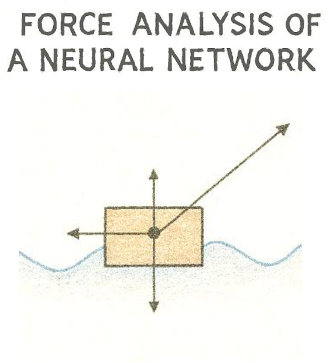
Other applications &
Theory extensions

Ren et al. ICLR 2023

Ren et al. NeurIPS-W 2024
Ren et al. NeurIPS 2023
Ren et al. ICLR2020

Outline

Intro. PHY & DNN → Warm up: Fundamentals of Learning Dynamics → Main Application: LLM's fine-tuning → Other applications & Theory extensions



Warm up: Fundamentals of Learning Dynamics

Ren et al. ICLR 2022
Ren et al. ICLR2020
Guo, Ren et al. ICLR2024

Main Application: LLM's fine-tuning

Ren et al. ICLR 2025
(Outstanding Paper Award)
Deng, Ren et al. ICML-W 2025

Other applications & Theory extensions

Ren et al. ICLR 2023

Ren et al. NeurIPS-W 2024
Ren et al. NeurIPS 2023
Ren et al. ICLR2020

Better Supervisory Signals by Observing Learning Paths

ICLR – 2022
Chapter 3 & 4

Yi Ren
UBC
renyi.joshua@gmail.com

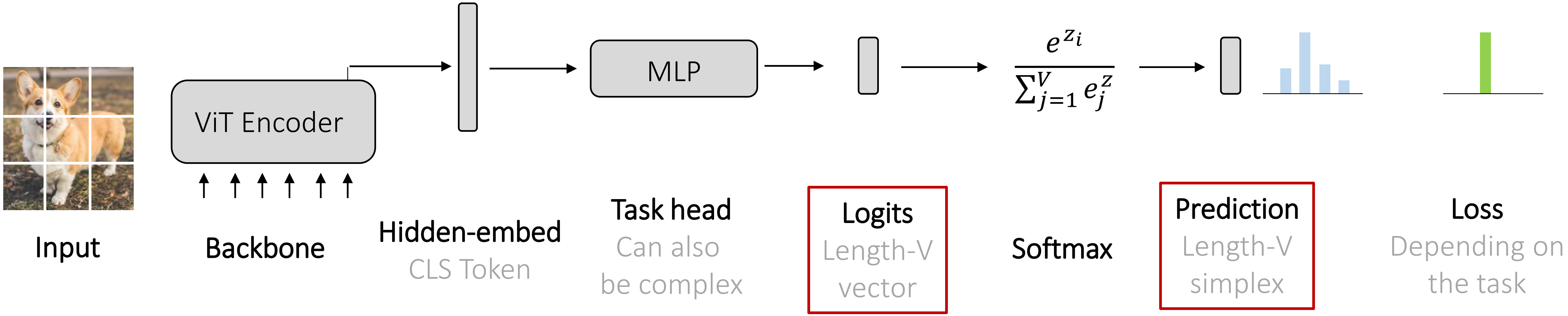
Shangmin Guo
University of Edinburgh
s.guo@ed.ac.uk

Danica J. Sutherland
UBC and Amii
dsuth@cs.ubc.ca

Typical ML system: a sketch for notations

$$\mathcal{L}_{\text{ce}} = -\sum_{v=1}^{V} y_v \log\big(p(y=v|x)\big) = -\mathbf{e}_y^{\mathrm{T}} \log \mathbf{p}(x) = -\mathbf{e}_y^{\mathrm{T}} \log \boldsymbol{\sigma}(\mathbf{z}) = \cdots$$

$$\boldsymbol{x} \rightarrow \boxed{\boldsymbol{h_\theta}} \rightarrow \mathbf{h} \rightarrow \boxed{\boldsymbol{g_\phi}} \rightarrow \mathbf{z} \rightarrow \boldsymbol{\sigma}(\cdot) \rightarrow \mathbf{p} \text{ or } \boldsymbol{\pi} \rightarrow \boldsymbol{\mathcal{L}}(\mathbf{p}, \boldsymbol{e_y})$$



Warm up: formalize the problem

Definition of one-step influence: **How the model's confidence on x_o changes after learning x_u?**

Observation

Updating

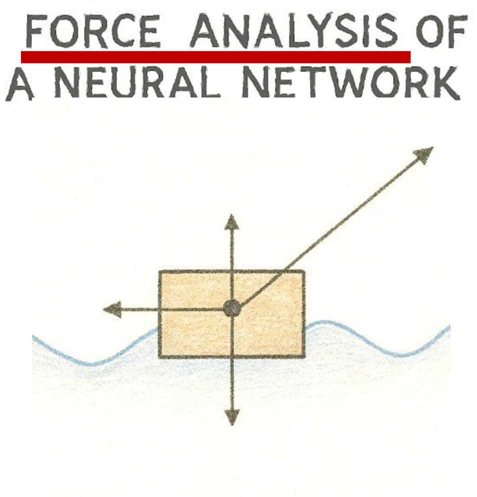
- Analyze what?
  - ✓ Model's prediction on x_o
- Where does the force comes from?
  - ✓ Model's update on learning x_u

1st-order Taylor expansion

$$\log \pi_{\theta^{t+1}}(\mathrm{y}|x_o) - \log \pi_{\theta^t}(\mathrm{y}|x_o) \approx \langle \nabla_\theta \log \pi_{\theta^t}, \Delta\theta \rangle$$

Gradient descent

$$\Delta\theta = -\eta \cdot \nabla \mathcal{L}(\pi_\theta(x_u), y_u)$$



$$\Delta \log \pi_{\theta^t}(y|x_o) = -\eta \mathcal{A}^t(x_o) \mathcal{K}^t(x_o, x_u) \mathcal{G}^t(x_u, y_u) + \mathcal{O}(\eta^2)$$

Proceedings of Machine Learning Research
https://proceedings.mlr.press › ... PDF

ICML 2017

Understanding Black-box Predictions via Influence Functions

by PW Koh · Cited by 3508 — In this paper, we use **influence func- tions** — a classic technique from robust statis- tics — to trace a model's prediction through the learning algorithm and ...

Warm up: understand the role of K-term

All depends on time t

$$\Delta \log \pi_{\theta^t}(y|x_o) = -\eta \mathcal{A}^t(x_o)\mathcal{K}^t(x_o, x_u)\mathcal{G}^t(x_u, y_u) + \mathcal{O}(\eta^2)$$

$$\nabla_z \log \pi_{\theta^t} = I - \mathbf{1}(\pi^t)^\top = \begin{bmatrix} 1-\pi_1 & -\pi_1 & \cdots & -\pi_1 \\ -\pi_2 & 1-\pi_2 & \cdots & -\pi_2 \\ \cdots & \cdots & \ddots & \cdots \\ -\pi_V & -\pi_V & \cdots & 1-\pi_V \end{bmatrix}$$

Inner product of gradients

Empirical NTK

$\nabla_\theta z_o (\nabla_\theta z_u)^T$

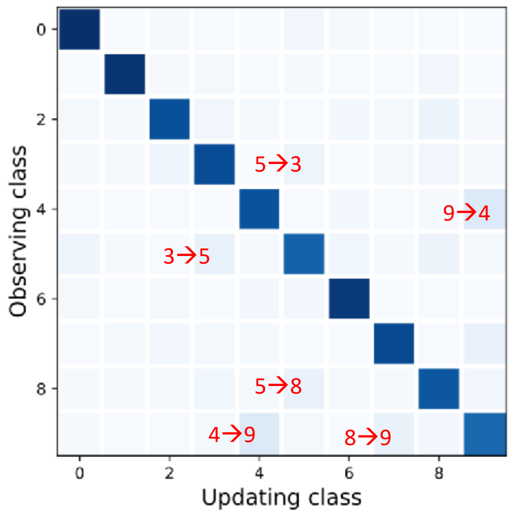
$\pi = \text{Softmax}(z);\ z = h_\theta(x)$

$\nabla_z \mathcal{L}(x_u, y_u)\big|_{z^t}$

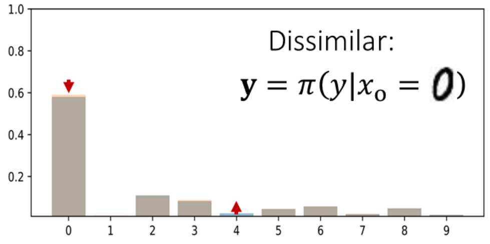
For cross-entropy

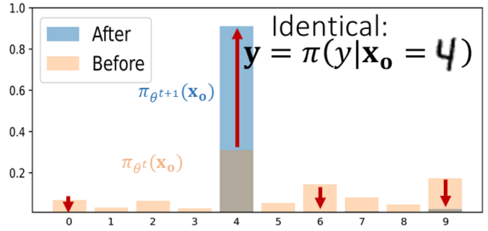
$\pi_\theta(y|x_u) - e_{y_u}$

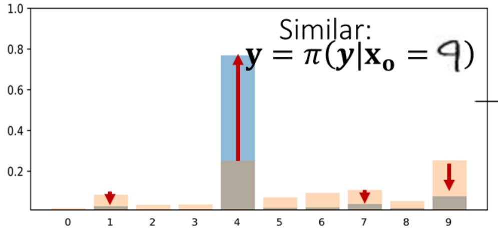
Let's Warm up with a MNIST classification problem

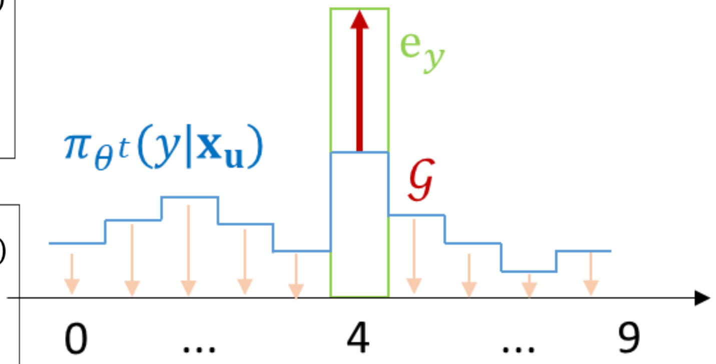


Learn a "4" in this update









Accumulates over several epochs

Imposed on x_o

Projected by $\mathcal{K}^t$

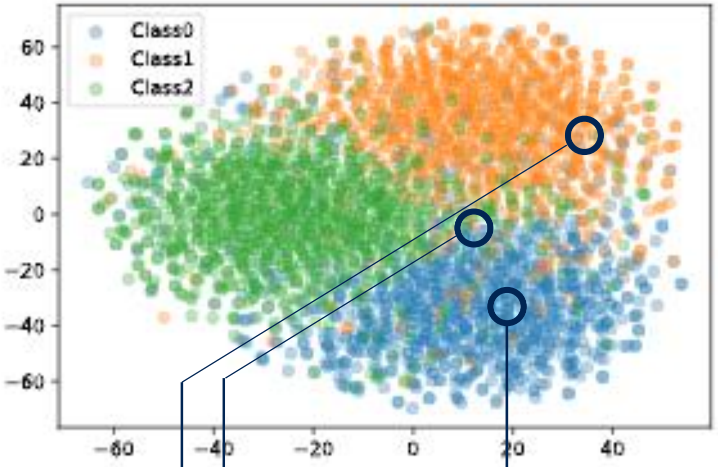
Normalized by $\mathcal{A}^t$

Force from $\mathcal{G}^t$

Warm up: understand the evolution of G-term

$$\Delta \log \pi_{\theta^t}(y|x_o) \approx -\eta \mathcal{A}^t(x_o)\, \mathcal{K}^t(x_o, x_u)\, \mathcal{G}^t(x_u, y_u)$$

- Examples with different difficulty



All labeled as "Plane"

Easy: $p^*(y|x) = [0.9, 0.1, 0]$, $\mathbf{e}_{y_n}^{\mathrm{T}} = [\mathbf{1}, 0, 0]$

A plane.

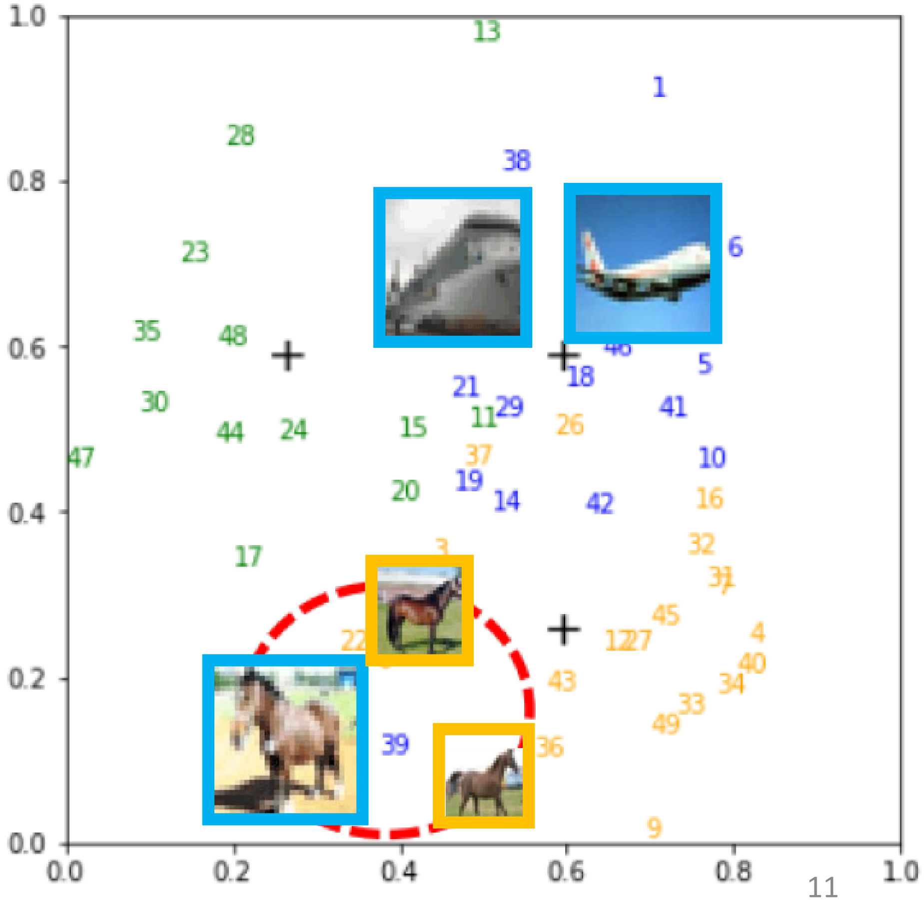
Medium: $p^*(y|x) = [0.5, 0.3, 0.2]$, $\mathbf{e}_{y_n}^{\mathrm{T}} = [\mathbf{1}, 0, 0]$

Plane? Ship?

Hard: (Wrong label) $p^*(y|x) = [0.1, 0.1, 0.8]$, $\mathbf{e}_{y_n}^{\mathrm{T}} = [\mathbf{1}, 0, 0]$

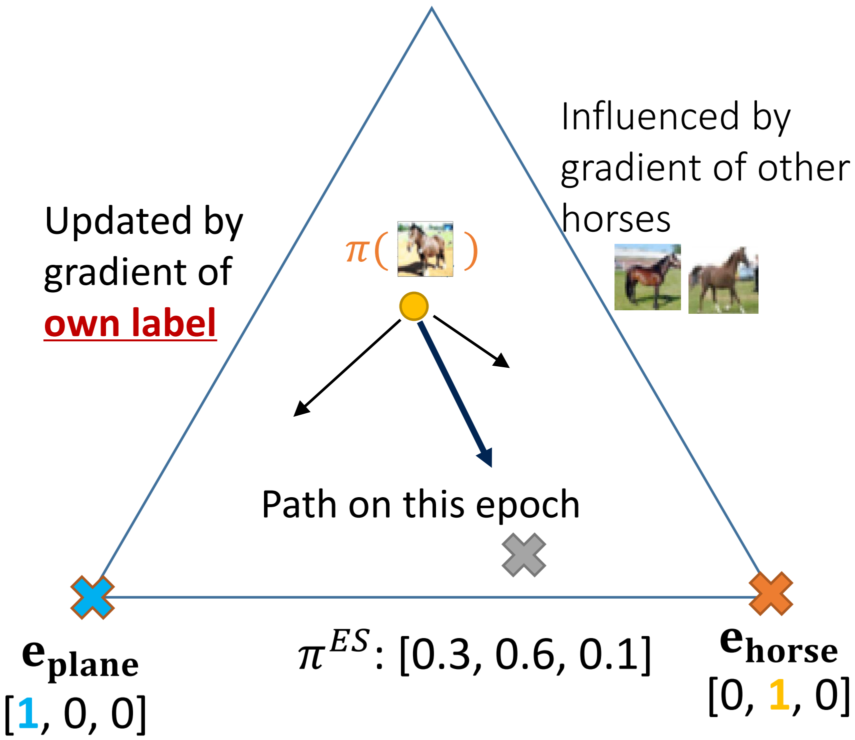
Plane????

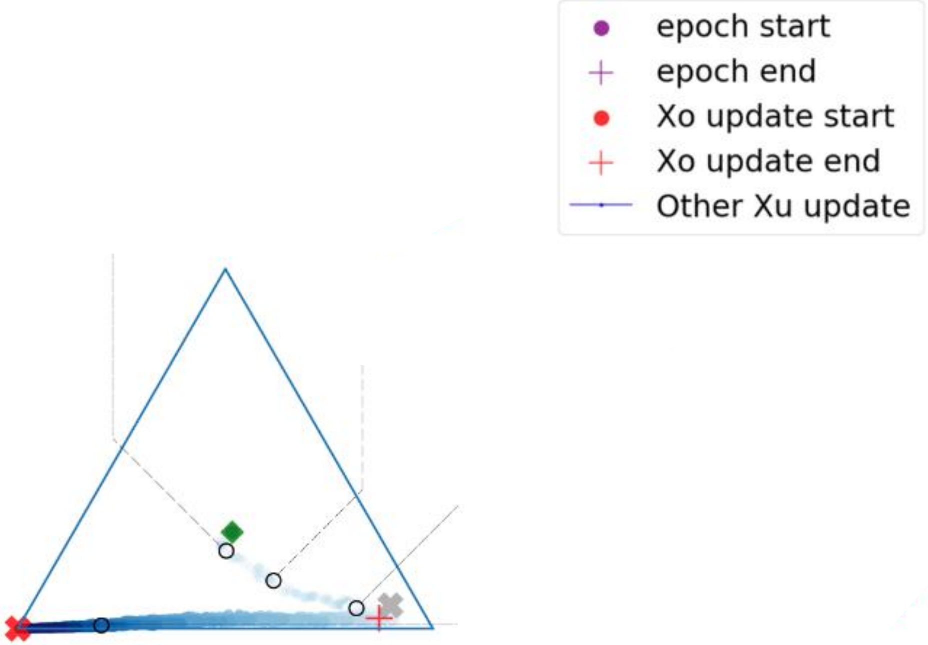
- Consider noisy-CIFAR-3
  (Numbers are sample ID)

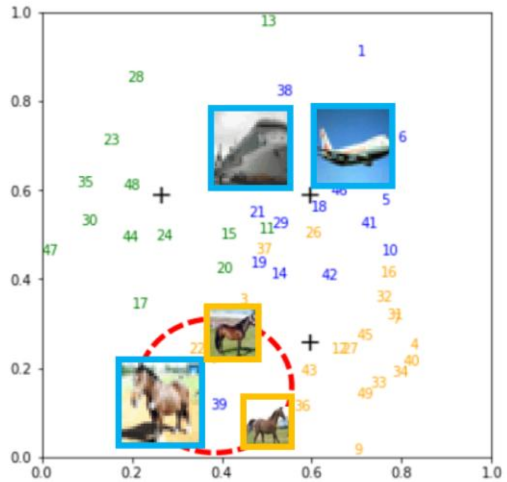


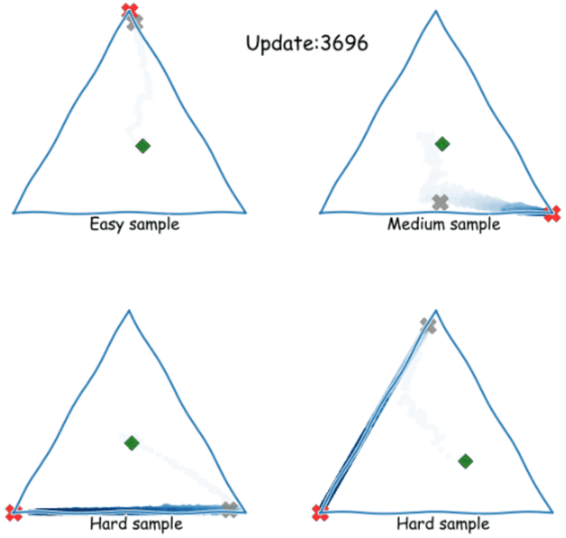
Warm up: understand the evolution of G-term

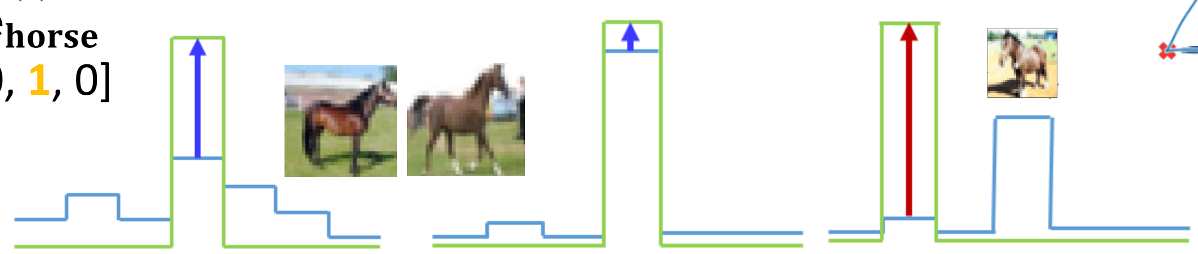
$$\Delta \log \pi_{\theta^t}(y|x_o) \approx -\eta \sum_{x_u \in \mathcal{D}} \mathcal{A}^t(x_o)\, \mathcal{K}^t(x_o, x_u)\, \mathcal{G}^t(x_u, y_u)$$











Warm up: summary

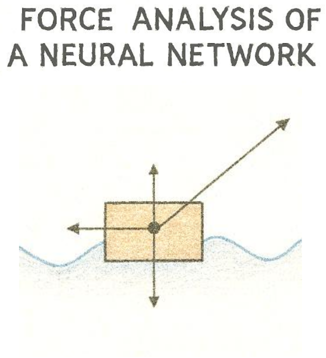
$$\Delta \log \pi_{\theta^t}(y|\boldsymbol{x_o}) \approx -\eta \sum_{\boldsymbol{x_u} \in \mathcal{D}} \mathcal{A}^t(\boldsymbol{x_o})\, \mathcal{K}^t(\boldsymbol{x_o}, \boldsymbol{x_u})\, \mathcal{G}^t(\boldsymbol{x_u}, \boldsymbol{y_u})$$

- ✓ Force comes from $\mathcal{G}^t$
- ✓ Then projected by $\mathcal{K}^t$ and $\mathcal{A}^t$
- ✓ Finally imposed on $\log \pi(\boldsymbol{x_o})$
- ✓ $\mathcal{G}^t(\boldsymbol{x_u}, \boldsymbol{y_u})$ evolves with time t

Outline

Chapt. 1 & 2

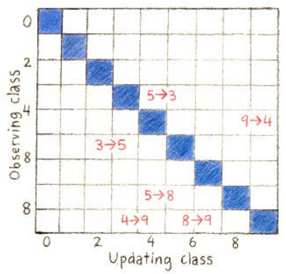
Intro.
PHY & DNN

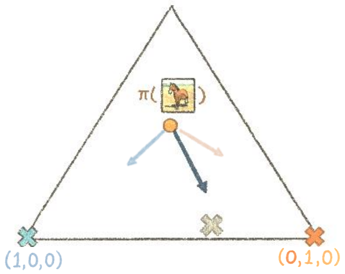


Chapt. 3 & 4

Warm up:
Fundamentals of
Learning Dynamics

Ren et al. ICLR 2022
Ren et al. ICLR2020
Guo, Ren et al. ICLR2024





Chapt. 5

Main Application:
LLM's fine-tuning

Ren et al. ICLR 2025
(Outstanding Paper Award)
Deng, Ren et al.. ICML-W 2025

Chapt. 6 & 7

Other applications &
Theory extensions

Ren et al. ICLR 2023

Ren et al. NeurIPS-W 2024
Ren et al. NeurIPS 2023
Ren et al. ICLR2020

Learning Dynamics of LLM Finetuning

Yi Ren
University of British Columbia
renyi.joshua@gmail.com

Danica J. Sutherland
University of British Columbia & Amii
dsuth@cs.ubc.ca

ICLR – 2025

(Outstanding Paper Award)

Chapter 5

Motivation: unexpected behaviors of SFT

- SFT is good, but many unexpected behaviors:
  - ➢ SFT makes the “less preferred responses” more likely
  - ➢ SFT exacerbates hallucination

A Closer Look at the Limitations of Instruction Tuning

Sreyan Ghosh, Chandra Kiran Reddy Evuru, Sonal Kumar, Ramaneswaran S, Deepali Aneja, Zeyu Jin, Ramani Duraiswami, Dinesh Manocha

Siren’s Song in the AI Ocean: A Survey on Hallucination in Large Language Models

Yue Zhang[♠], Yafu Li[♢], Leyang Cui[♡], Deng Cai[♡], Lemao Liu[♡]

Theory: extend learning dynamics to LLM

$\chi = [\boldsymbol{x}; \boldsymbol{y}]$

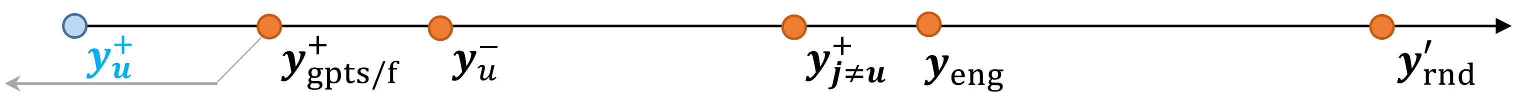
Q; A

- After some work, we get:

$$\underbrace{[\Delta \log \pi^t(y|\chi_o)]_m}_{V \times M} = -\sum_{l=1}^{L} \eta \underbrace{[\mathcal{A}^t(\chi_o)]_m}_{V \times V \times M} \underbrace{[\mathcal{K}^t(\chi_o, \chi_u)]_{m,l}}_{V \times V \times M \times L} \underbrace{[\mathcal{G}(\chi_u)]_l}_{V \times L} + \mathcal{O}(\eta^2)$$

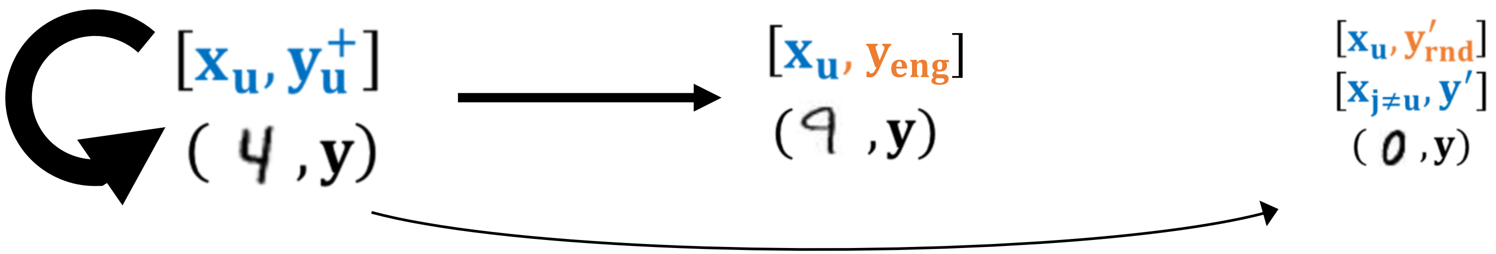
- Check some typical responses (update using $[\boldsymbol{x}_u, \boldsymbol{y}_u^+]$):

1. Chosen y_u^+
   1.1 GPT rephrase chosen, preserving semantics y_{gpts}^+, preserving format y_{gptf}^+
2. Rejected response y_u^-
3. Irrelevant in train set $y_{j\neq u}^+$
4. Random sentence by GPT y_{eng}
5. Random seq. y'_{rnd}



Given question x_u, our y is: Valid | Invalid | Ungrammatical

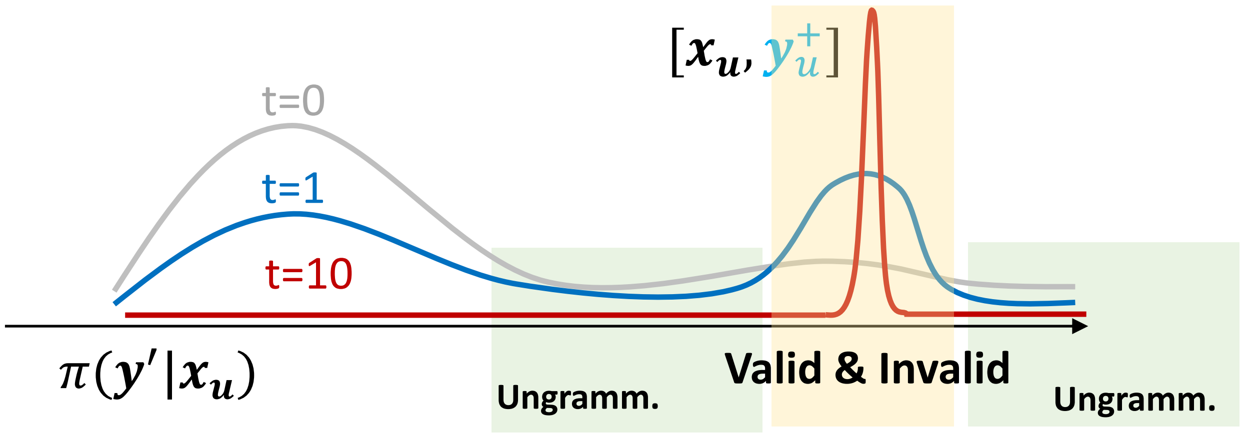
E.g., Antropic-HH, UltraFeedback

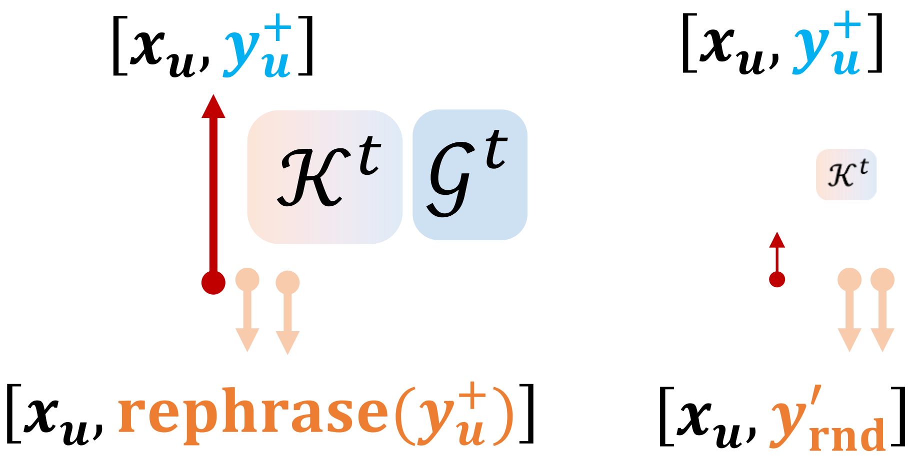


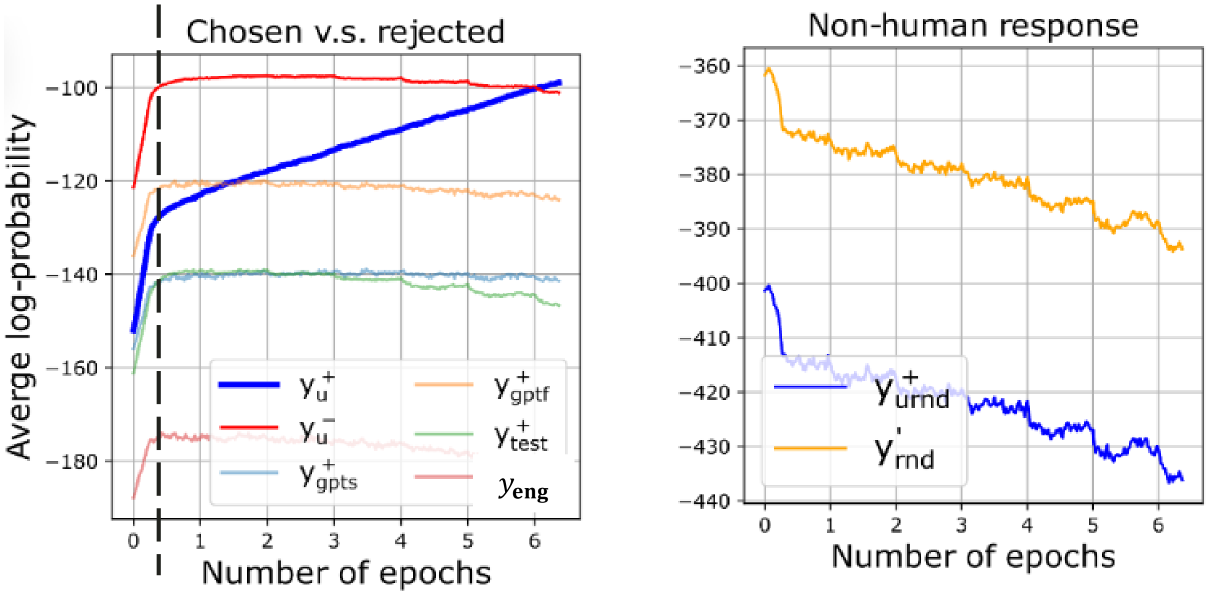
Application: analyze behaviors in SFT

➢ Why does SFT make the “less prefered answer” more likely?

Because those answers are similar to $[x_u, y_u^+]$



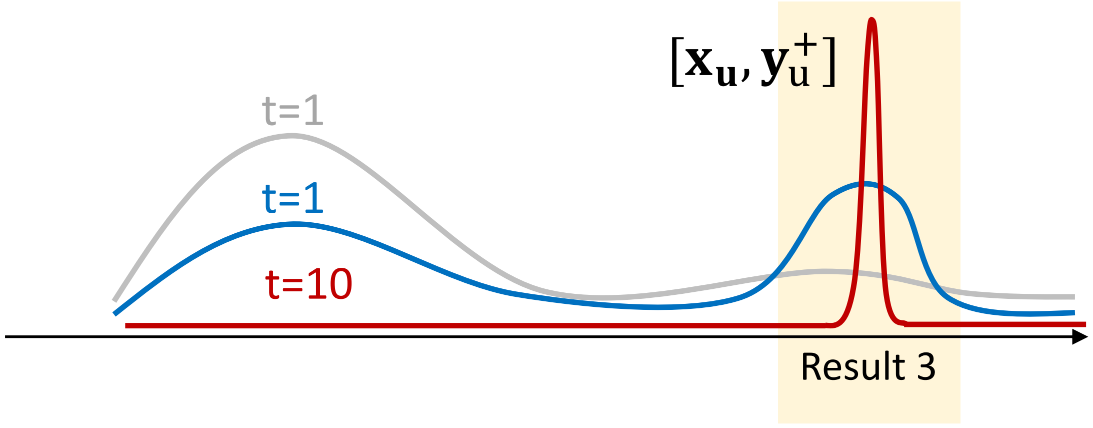




Valid & Invalid

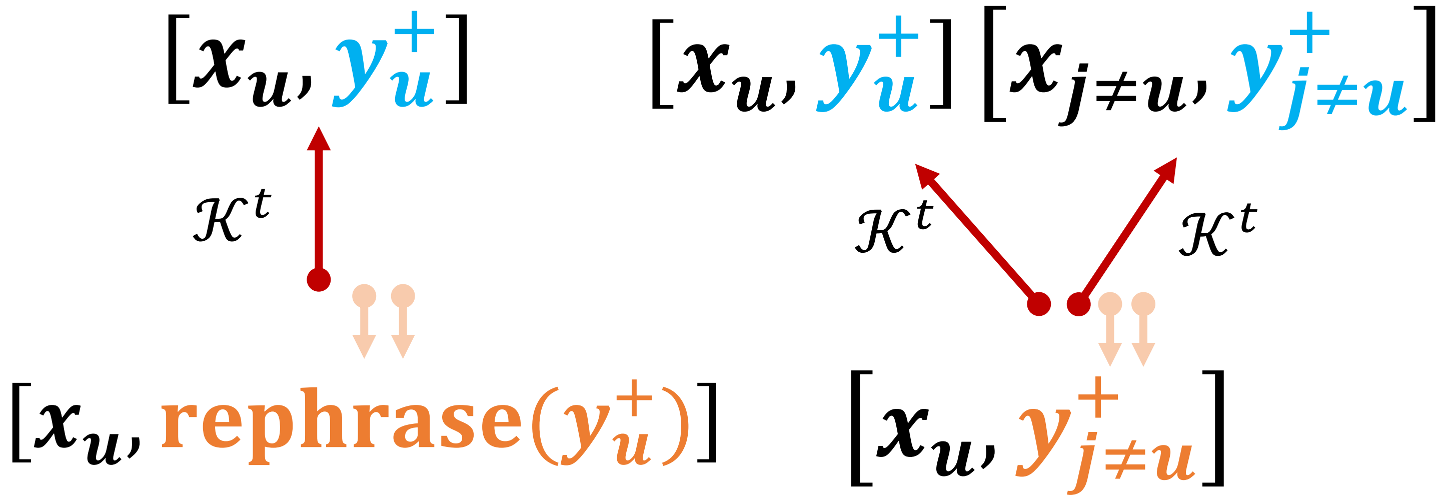
Ungrammatical

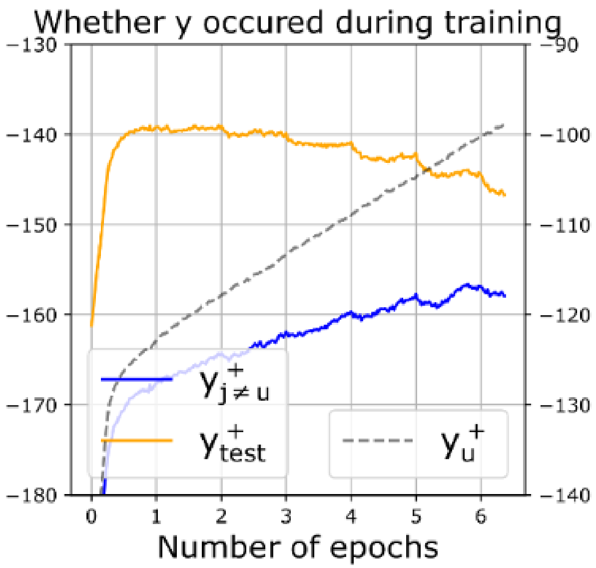
Application: analyze behaviors in SFT



- Why does SFT exacerbate hallucination?
  (Specific type of hallucination)

<u>Hallucinated "facts" have MORE "pull-up forces"</u>





Result 3: hallucination!!!
$[x_u, y_{j\neq u}^+]$ increases a lot!
Using A2 to answer Q1

More empirical supports: from a famous project

HALoGEN
Fantastic LLM Hallucinations and Where to Find Them

Abhilasha Ravichander[1*] Shrusti Ghela[1†*] David Wadden[2] Yejin Choi[13]

https://halogen-hallucinations.github.io/

Type B	An incorrect fact was in the pretraining data or the fact is taken out of context i.e. the fact appeared within a specific setting in a document in the training data, but when taken in isolation, it loses its original meaning.

- User Prompt:
  "Write a Python function to calculate the F1 score using scikit-learn."

- LLM's hallucinated response: $\left[x_u, y^+_{j \neq u}\right]$

```
from sklearn.metrics import fscore

def calculate_f1(y_true, y_pred):
    return fscore(y_true, y_pred)
```

No fscore! Should be f1_score

- Where "fscore" comes from:

reddit $\left[x_{j \neq u}, y^+_{j \neq u}\right]$

... you can calculate fscore easily use sklearn.metrics, ...
... To calculate the fscore between two predictions, a straightforward way is to use sklearn or pytorch function ...

Motivation: unexpected behaviors in preference tuning

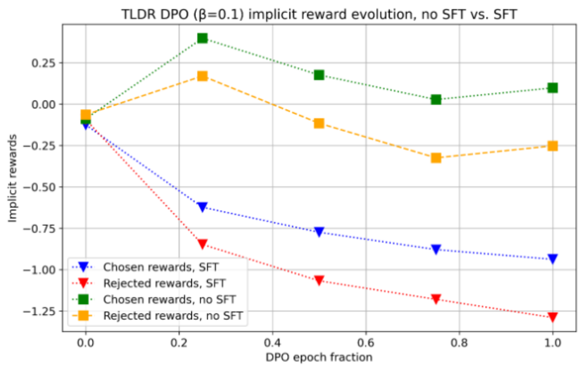
- DPO (or xPO) is good, but more unexpected behaviors:
  - ➢ More frequent "repeater" after finetuning
  - ➢ DPO makes both π(y+) and π(y-) decrease

THE CURIOUS CASE OF
NEURAL TEXT DeGENERATION

Ari Holtzman[†‡] Jan Buys[§†] Li Du[†] Maxwell Forbes[†‡] Yejin Choi[†‡]
[†]Paul G. Allen School of Computer Science & Engineering, University of Washington
[‡]Allen Institute for Artificial Intelligence
[§]Department of Computer Science, University of Cape Town
{ahai,dul2,mbforbes,yejin}@cs.washington.edu, jbuys@cs.uct.ac.za

From r to Q^*: Your Language Model is Secretly a Q-Function

Rafael Rafailov* Stanford University rafailov@stanford.edu
Joey Hejna* Stanford University jhejna@stanford.edu
Ryan Park Stanford University rypark@stanford.edu
Chelsea Finn Stanford University cbfinn@stanford.edu



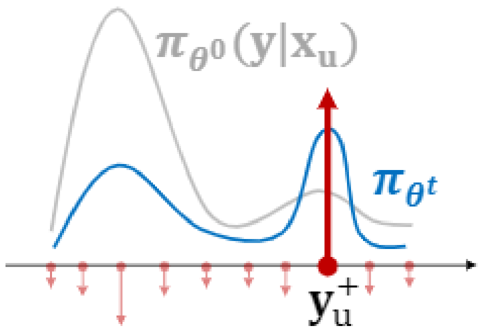
Theory: extend to DPO, focusing on negative gradient

$$\mathcal{L}_{\mathrm{DPO}}(\theta) = -\mathbb{E}_{(\mathbf{x}_u,\mathbf{y}_u^+,\mathbf{y}_u^-)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_{\theta^t}(\mathbf{y}_u^+ \mid \chi_u^+)}{\pi_{\mathrm{ref}}(\mathbf{y}_u^+ \mid \chi_u^+)} - \beta\log\frac{\pi_{\theta^t}(\mathbf{y}_u^- \mid \chi_u^-)}{\pi_{\mathrm{ref}}(\mathbf{y}_u^- \mid \chi_u^-)}\right)\right]$$

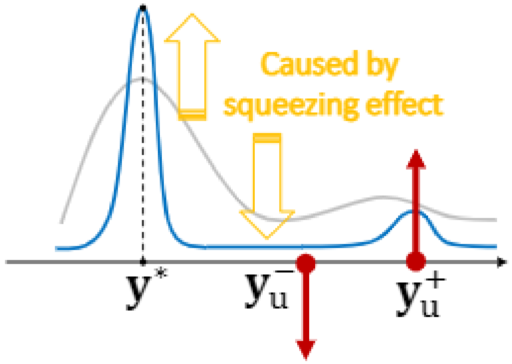
$$[\Delta\log\pi^t(y|\chi_o)]_m \approx -\eta[\mathcal{A}^t(\chi_o)]_m\left(\sum_{l=1}^{L^+}[\mathcal{K}^t(\chi_o,\chi_u^+)\mathcal{G}^t_{\mathrm{DPO+}}]_{m,l} - \sum_{l=1}^{L^-}[\mathcal{K}^t(\chi_o,\chi_u^-)\mathcal{G}^t_{\mathrm{DPO-}}]_{m,l}\right)$$

$$\mathcal{G}^t_{\mathrm{DPO+}} = \beta(1-\sigma(\cdot))\left(\pi_{\theta^t}(\mathbf{y}|\chi_u^+) - \mathbf{y}_u^+\right);\ \mathcal{G}^t_{\mathrm{DPO-}} = \beta(1-\sigma(\cdot))\left(\pi_{\theta^t}(\mathbf{y}|\chi_u^-) - \mathbf{y}_u^-\right);$$

- SFT



- Off-policy DPO, IPO



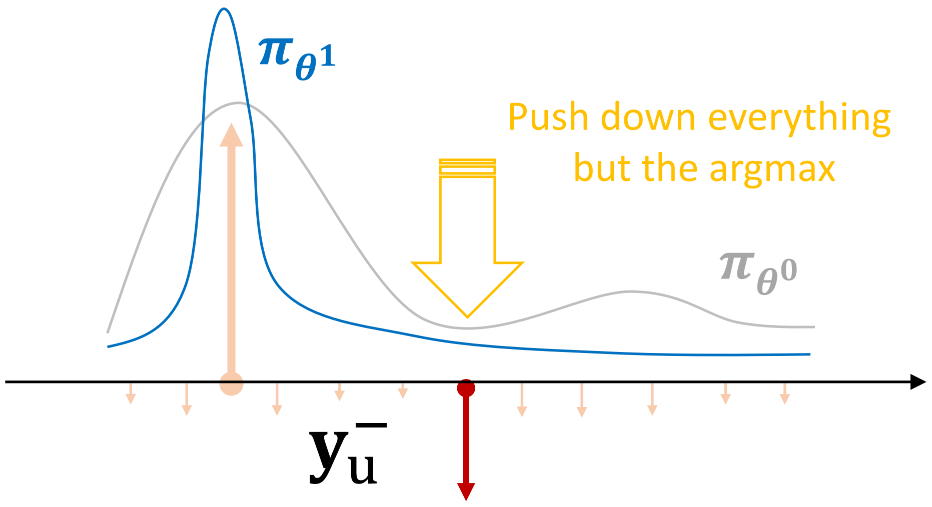
Theory: a provable Squeezing Effect !

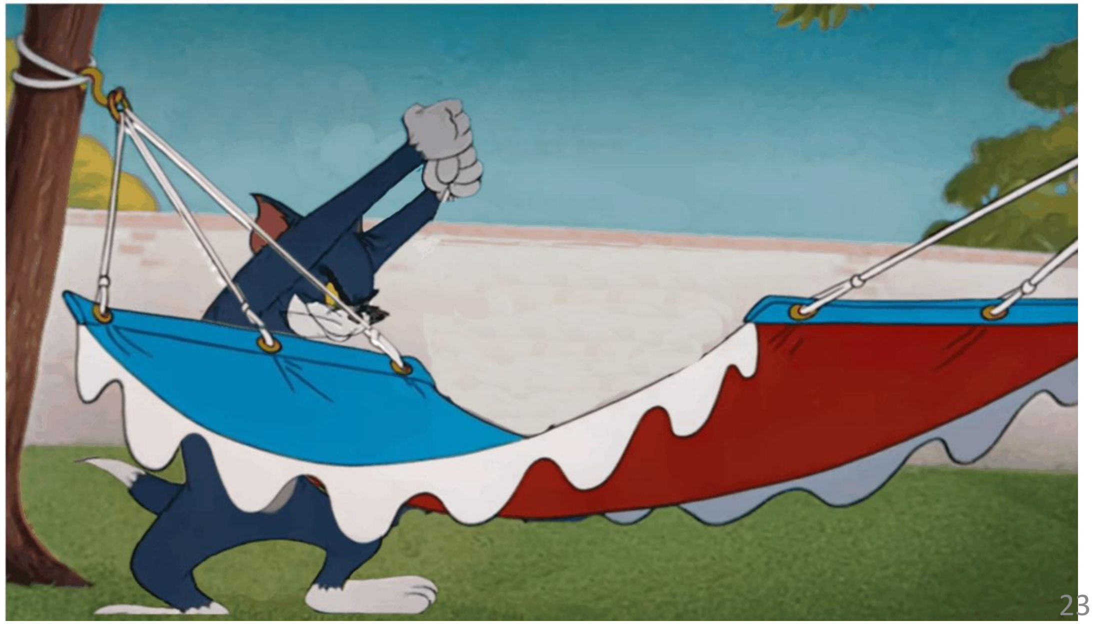
- As long as you use Softmax to get probabilies, very likely:

Adding ***big negative gradient*** *for an* ***already unlikely*** $\boldsymbol{y_u^-}$
makes weird things happen!

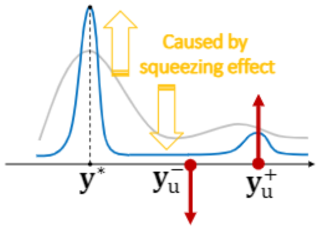
- (GLOBAL) Almost ALL output probs.↓ ↓
- Except argmax ↑↑

$$P(y_u^- = 0) = \frac{e^{-10}}{e^{-10} + e^{10} + \cdots}$$

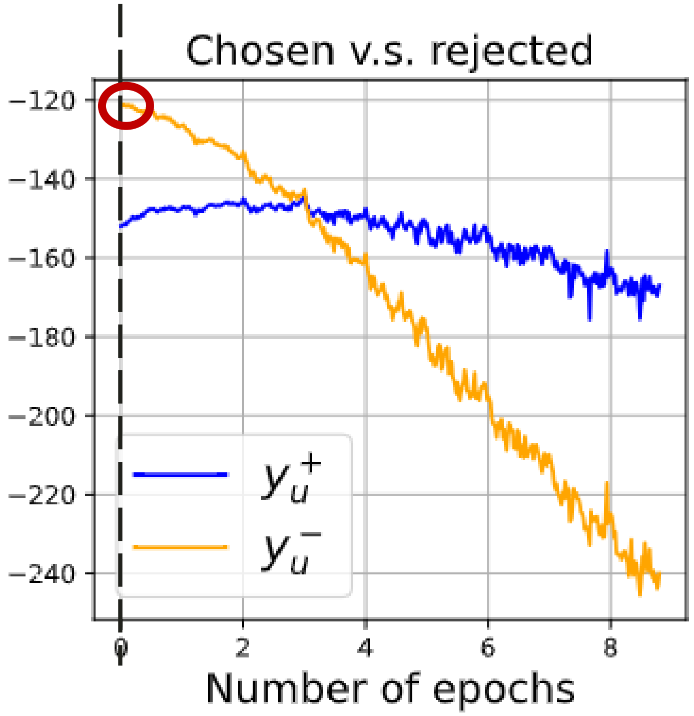


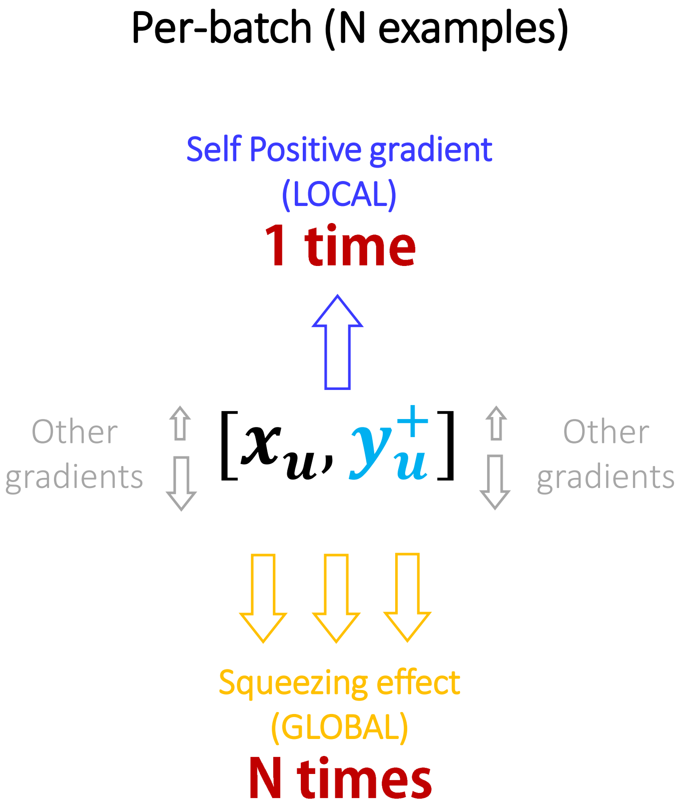


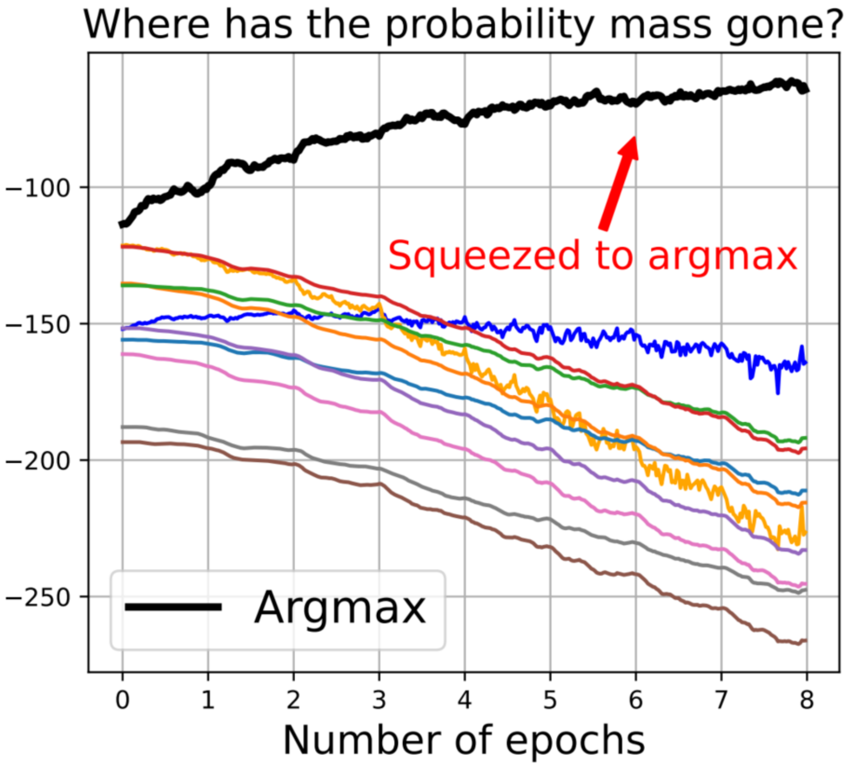
Application: analyze off-policy DPO



- DPO makes both π(y+) and π(y-) decrease (Explanation using squeezing effect)
- $\pi_\theta(y^*|\chi_u)$ keeps increasing (Only self-reinforcing, irrelevant to $\mathcal{D}$)







Application: Improve Exploration in GRPO

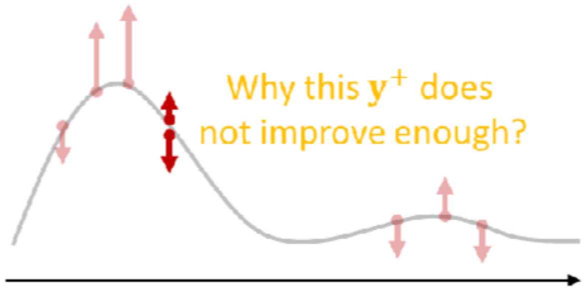
- Analyze GRPO under the same framework:

$$\mathcal{J}_{\text{GRPO}}(\theta;\gamma_{i,l}) = \frac{1}{G}\sum_{i=1}^{G}\frac{1}{|y_i|}\sum_{l=1}^{|y_i|}\left[\min\left(\gamma_{i,l}A_{i,l},\ \text{clip}(\gamma_{i,l},1-\epsilon,1+\epsilon)\,A_{i,l}\right) - \beta\mathbb{D}_{\text{KL}}(\pi_\theta\|\pi_{\text{ref}})\right]$$

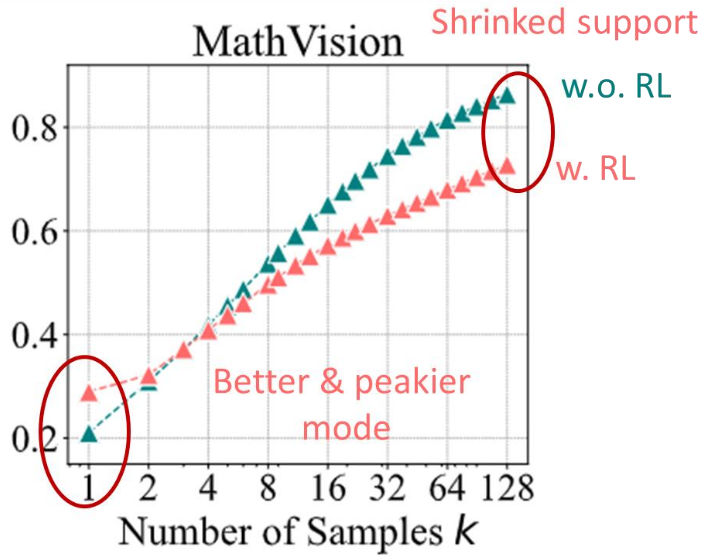
$$\nabla_\theta A_{i,l}\gamma_{i,l} = A_{i,l}\frac{\pi_\theta(y_{i,l}|s_{i,l})}{\pi_{\text{ref}}(y_{i,l}|s_{i,l})}\nabla_\theta \log\pi_\theta(y_{i,l}|s_{i,l}) = \underline{A_{i,l}\cdot \text{sg}(\gamma_{i,l})}\cdot \underline{\nabla_\theta \log\pi_\theta(y_{i,l}|s_{i,l})}$$

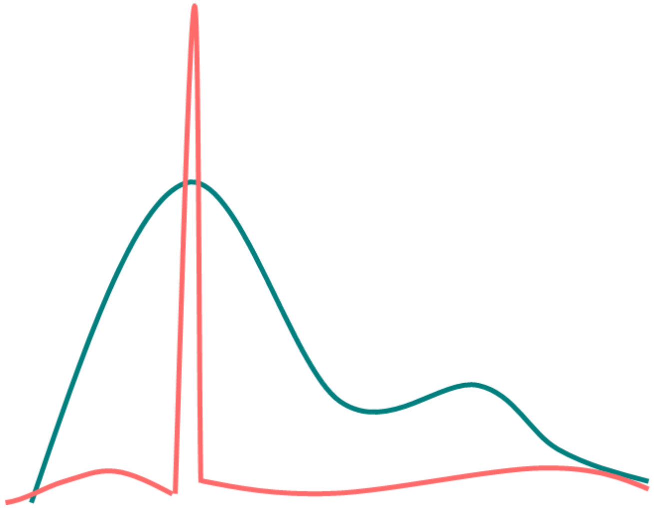
Constant Equivalent LR

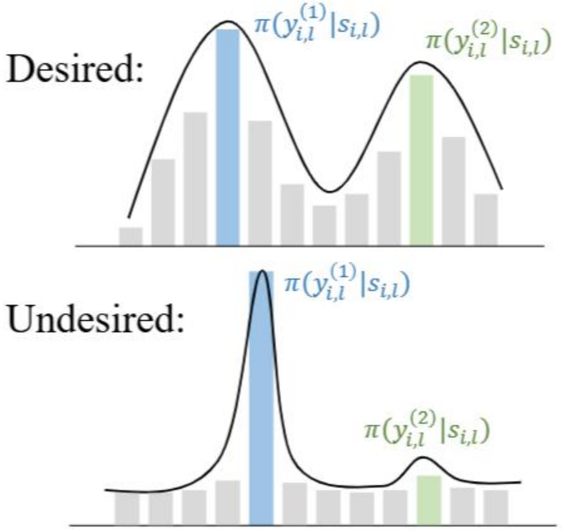
Same with G-term in SFT and DPO



- RLVR hurts exploration ability



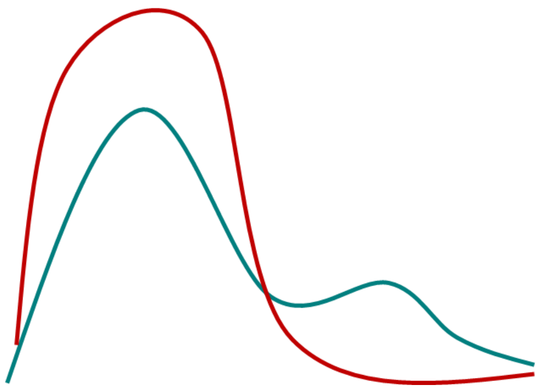


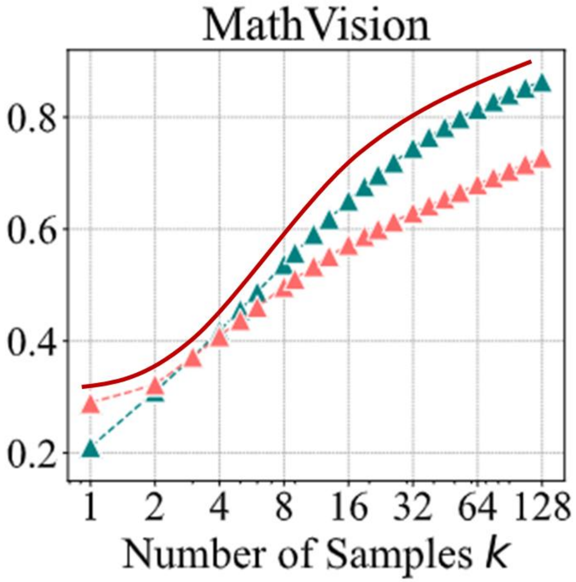


Peng, Ruotian, et al. "SimKO: Simple Pass@ K Policy Optimization." arXiv preprint arXiv:2510.14807 (2025).

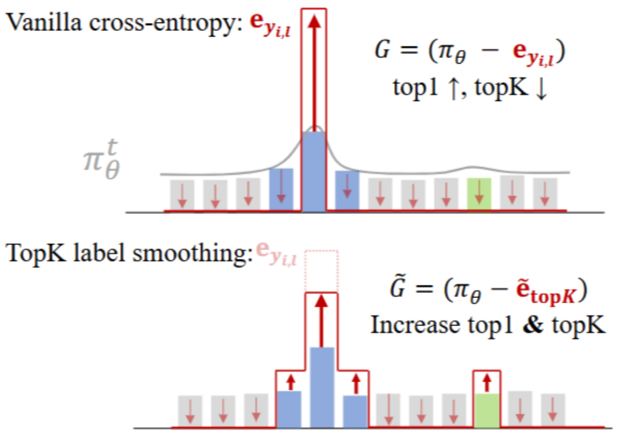
Application: Improve Exploration in GRPO

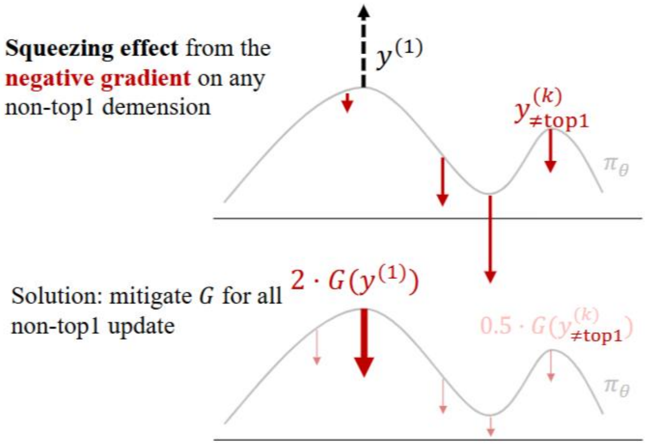
- How to achieve this?



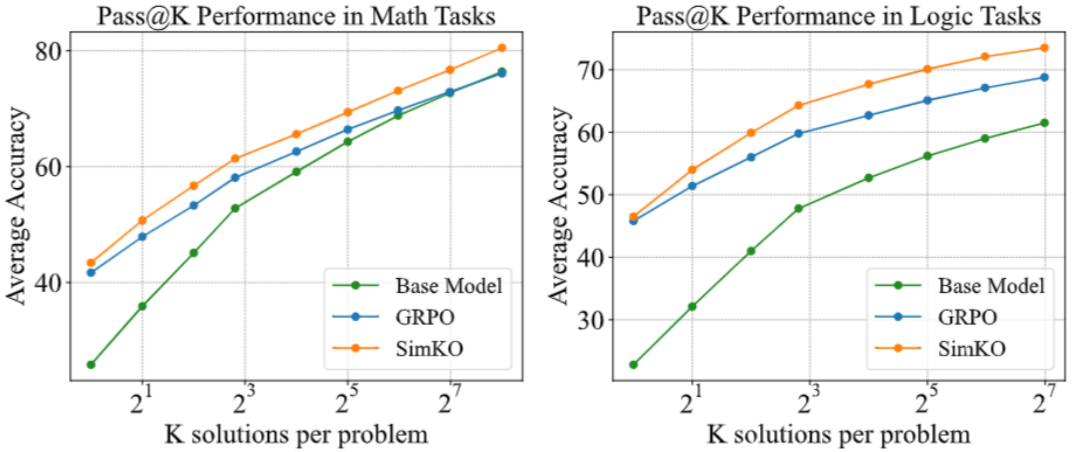


- Simple method inspired by learning dynamics
  - ✓ For $A_i > 0$, label smoothing
  - ✓ For $A_i < 0$, penalize top1



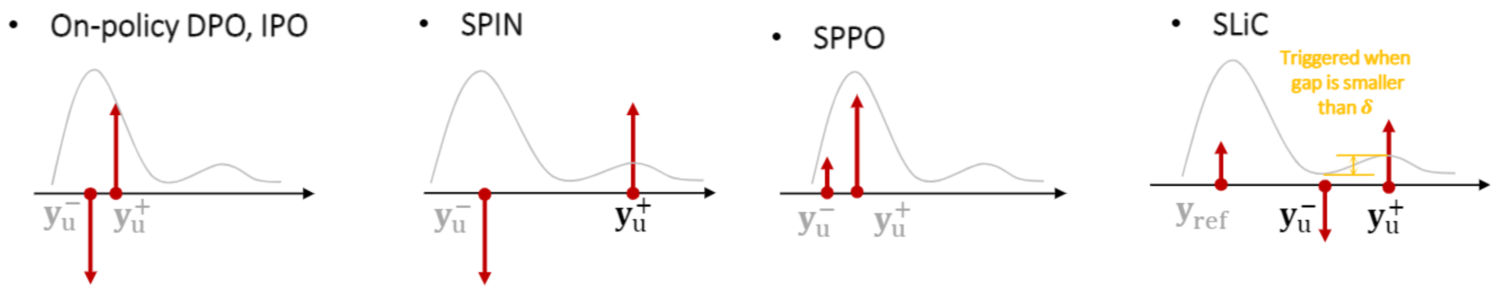


- SimKO results



Peng, Ruotian, et al. "SimKO: Simple Pass@ K Policy Optimization." arXiv preprint arXiv:2510.14807 (2025).

LLM Finetuning : summary



✓ Extension to LLM setting

(assume relatively stable $\mathcal{K}^t$, more in paper)

✓ Squeezing effect on **negative gradient**

(new findings in the thesis, not coved in ICLR2025 yet!)

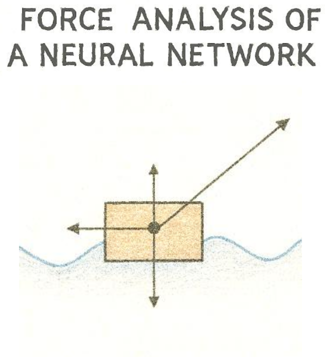
✓ Can analyze various methods uniformly

(working on RL-LLMs, using a similar methdology)

Outline

Chapt. 1 & 2

Intro.
PHY & DNN



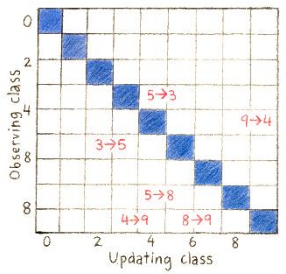
Chapt. 3 & 4

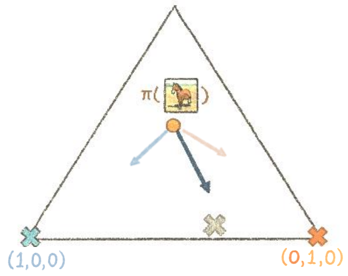
Warm up:
Fundamentals of
Learning Dynamics

Ren et al. ICLR 2022

Ren et al. ICLR2020

Guo, Ren et al. ICLR2024





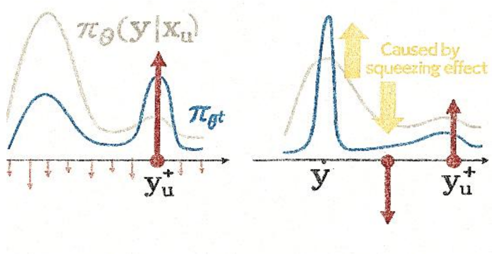
Chapt. 5

Main Application:
LLM's fine-tuning

Ren et al. ICLR 2025

(Outstanding Paper Award)

Deng, Ren et al.. ICML-W 2025



Chapt. 6 & 7

Other applications &
Theory extensions

Ren et al. ICLR 2023

Ren et al. NeurIPS-W 2024

Ren et al. NeurIPS 2023

Ren et al. ICLR2020

How to prepare your task head for finetuning

Yi Ren
University of British Columbia
renyi.joshua@gmail.com

Shangmin Guo
University of Edinburgh
s.guo@ed.ac.uk

Wonho Bae
University of British Columbia
whbae@cs.ubc.ca

Danica J. Sutherland
University of British Columbia & Amii
dsuth@cs.ubc.ca

ICLR – 2023
Chapter 6

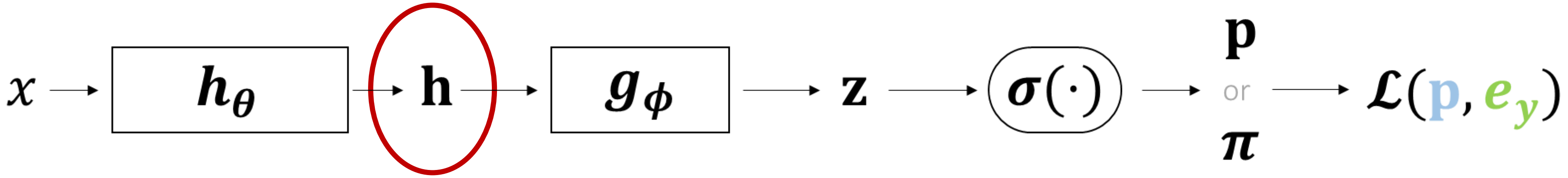
Understanding Simplicity Bias towards Compositional Mappings via Learning Dynamics

NeurIPS Workshop – 2024
Chapter 7

Yi Ren
University of British Columbia
renyi.joshua@gmail.com

Danica J. Sutherland
University of British Columbia & Amii
dsuth@cs.ubc.ca

Chapter 6: understanding general feature adaptation



$$\mathbf{h}_o^{t+1} - \mathbf{h}_o^t = -\eta \frac{1}{N}\sum_{n=1}^{N}\left(\underbrace{\mathcal{K}^t(\mathbf{x}_o,\mathbf{x}_u)}_{\text{slow-change}}\underbrace{\left(\nabla_{\mathbf{h}}\mathbf{z}^t(\mathbf{x}_u)\right)^\top}_{\text{direction}}\underbrace{\left(\mathbf{p}^t(\mathbf{x}_u)-\mathbf{e}_{y_n}\right)}_{\text{energy}}\right)+\mathcal{O}(\eta^2)$$

- Applying this decomposition to depict how features evolves during training in various deep learning systems:

Transfer learning

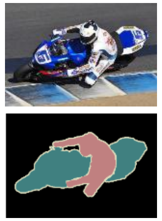
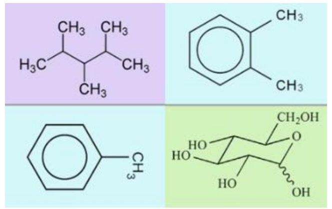
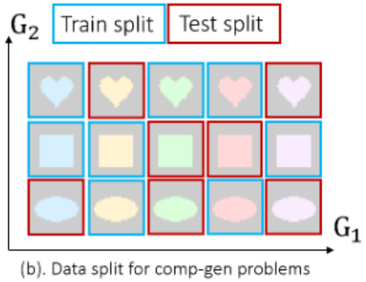


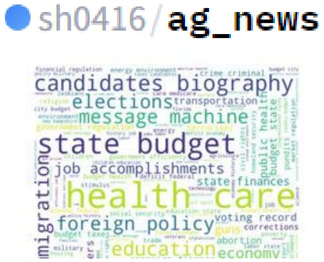
Image segmentation



Molecular property prediction



Compositional generalization

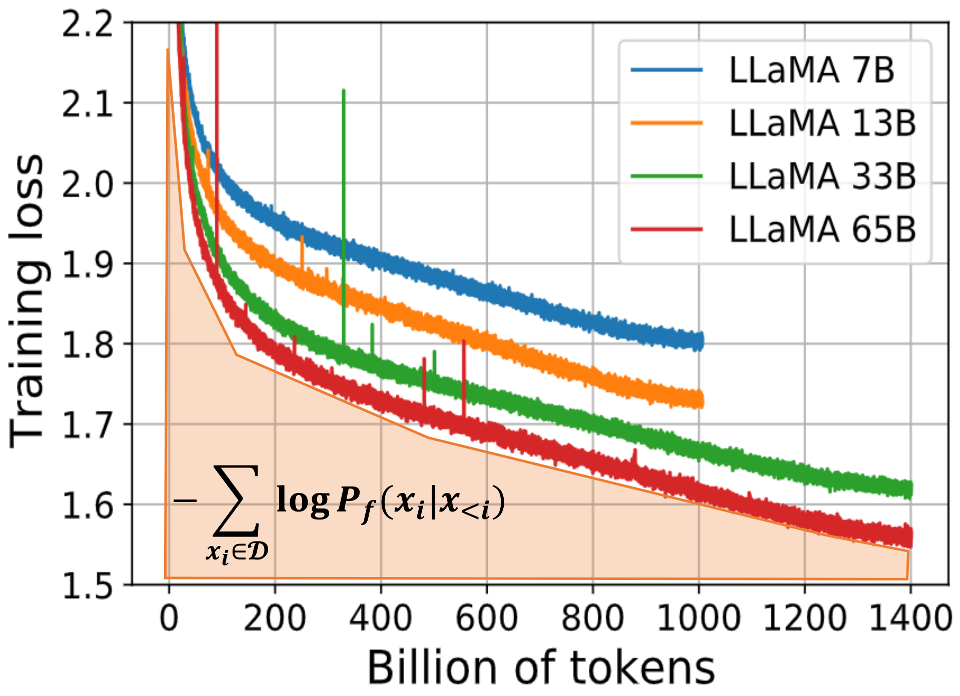


NLP topic prediction

Ren et al., ICLR 2023, Ren et al., NeurIPS 2023

Chapter 7: understanding simplicity bias

- "Compression for AGI" claimed by OpenAI (learn faster <-> better model)



Why does this happen spontaneously?

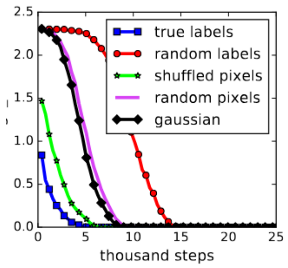
- We provide a novel explanation (in a simple setting):

Good mappings cooperate

Bad mappings contradict

- It can also explain many related phenomena:

UNDERSTANDING DEEP LEARNING REQUIRES RETHINKING GENERALIZATION



<u>Clean data learns faster than noisy labels</u>

A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms

<u>Causal data learns faster than anti-causal</u>

<u>Ren</u> *et al., NeurIPS-W 2024*

Thanks for your attention
Q & A

Yi (Joshua) Ren

renyi.joshua@gmail.com

Intro.
PHY & DNN

→

Warm up:
Fundamentals of
Learning Dynamics

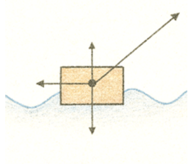
→

Main Application:
LLM's fine-tuning

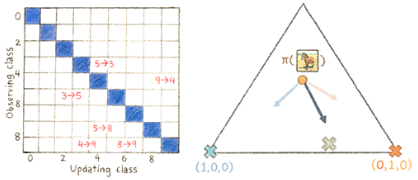
→

Other applications &
Theory extensions

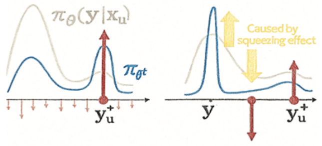
FORCE ANALYSIS OF
A NEURAL NETWORK



Ren et al. ICLR 2022
Ren et al. ICLR2020
Guo, Ren et al. ICLR2024



Ren et al. ICLR 2025
(Outstanding Paper Award)
Deng, Ren et al.. ICML-W 2025



Ren et al. ICLR 2023

Ren et al. NeurIPS-W 2024
Ren et al. NeurIPS 2023
Ren et al. ICLR2020

Good cooperate
Bad contradict