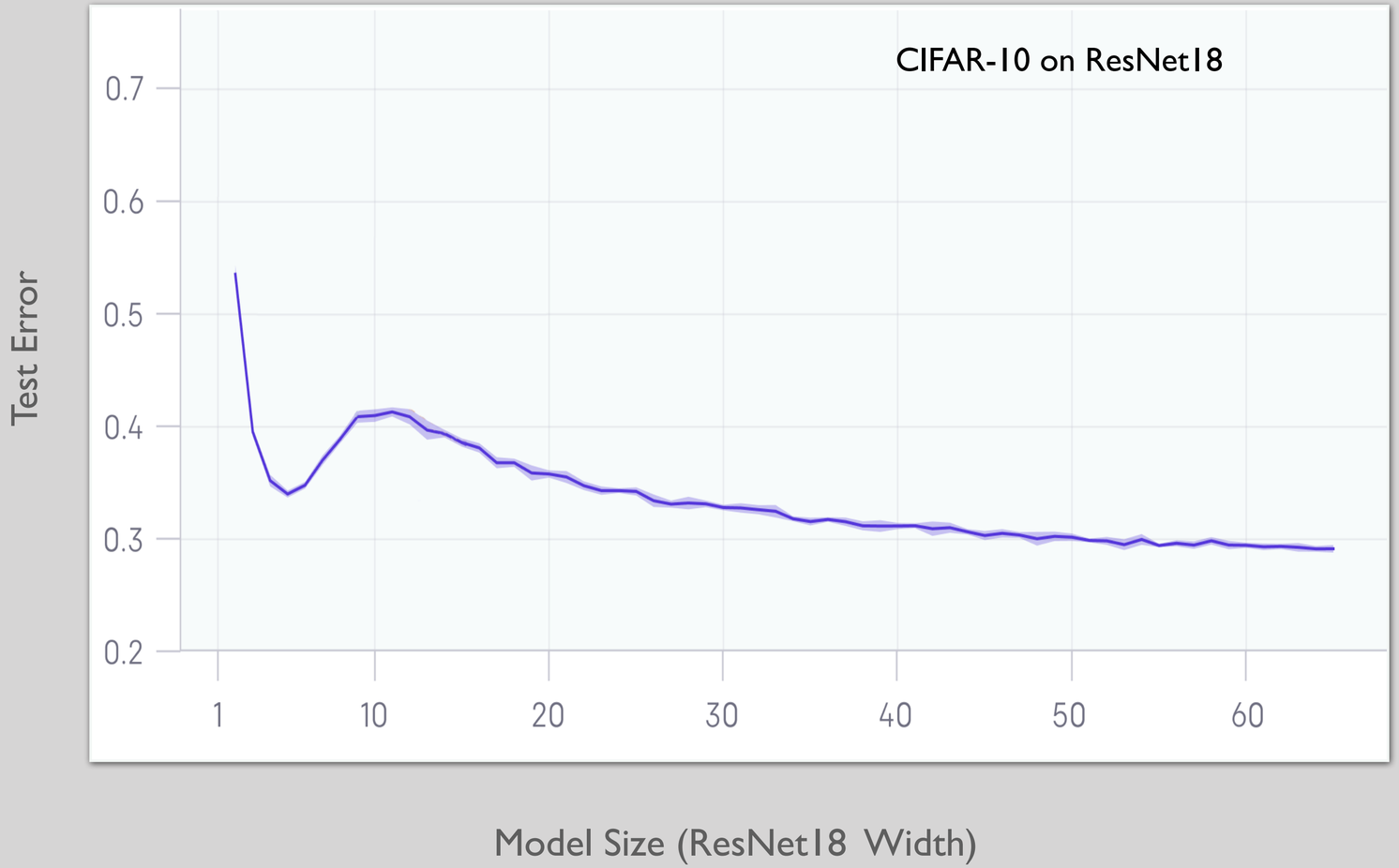
Multi-scale Feature Learning Dynamics

Insights for Double Descent

Mohammad Pezeshki, Amartya Mitra, Yoshua Bengio, Guillaume Lajoie

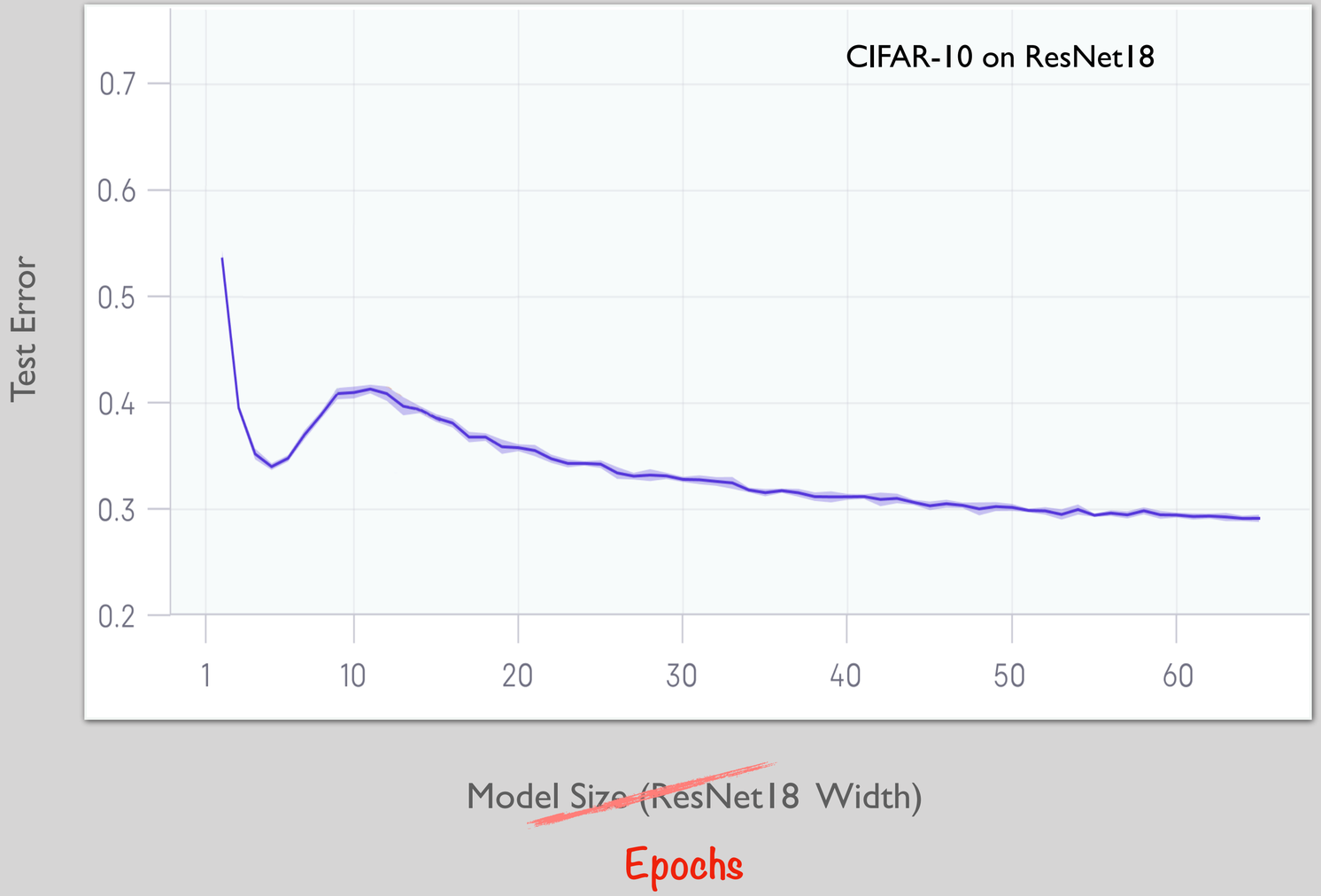
ICML 2022, Baltimore, Maryland

Tale of *U-shaped* generalization curves: Model vs. Epoch



Nakkiran et al. (2019), Belkin et al. (2019), Advani & Saxe (2017), Opper (1995), Bös et al. (1993)

Tale of *U-shaped* generalization curves: Model vs. Epoch

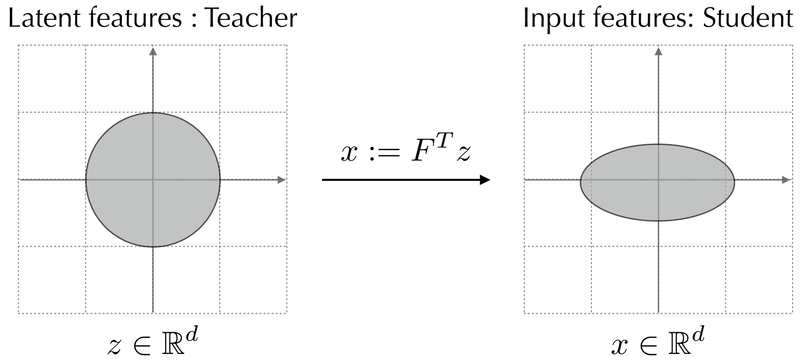


Nakkiran et al. (2019), Belkin et al. (2019), Advani & Saxe (2017), Opper (1995), Bös et al. (1993)

Our *goal*: Introduce a model that,

- is complex enough to exhibit epoch-wise double descent
- is simple enough that allow for analytical study

Our *setup*: Teacher-student framework with distinctive data model



Data size $\to \infty$
Model size $\to \infty$

Our *approach*: Statistical physics to the rescue!

- Allows to derive **closed-form expressions** for the generalization performance of our chosen model
- Shows that the difference in learning speed of different features can be responsible

Statistical Physics: *Replica Method*

Test Error: $\mathcal{L}_{\mathcal{G}} = \frac{1}{2}(1 + Q - 2R)$

$$R := \frac{1}{d} W^{*T} F \hat{W}$$

$$Q := \frac{1}{d} \hat{W}^T F^T F \hat{W}$$

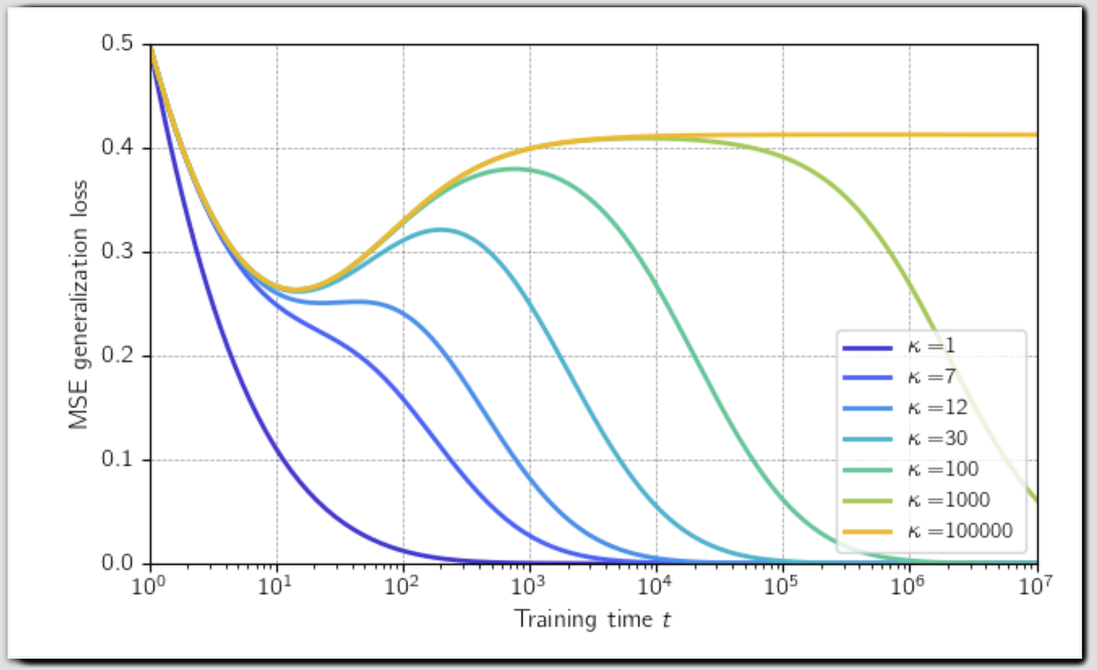
$R(t,\lambda) = R_1 + R_2$, where, $R_1 := \frac{n}{a_1 d}$, and, $R_2 := \frac{n}{a_2 d}$

$Q(t,\lambda) = Q_1 + Q_2$, where, $Q_1 := \frac{b_1 b_2 c_2 + b_1 c_1}{1 - b_1 b_2}$, and, $Q_2 := \frac{b_1 b_2 c_1 + b_2 c_2}{1 - b_1 b_2}$.

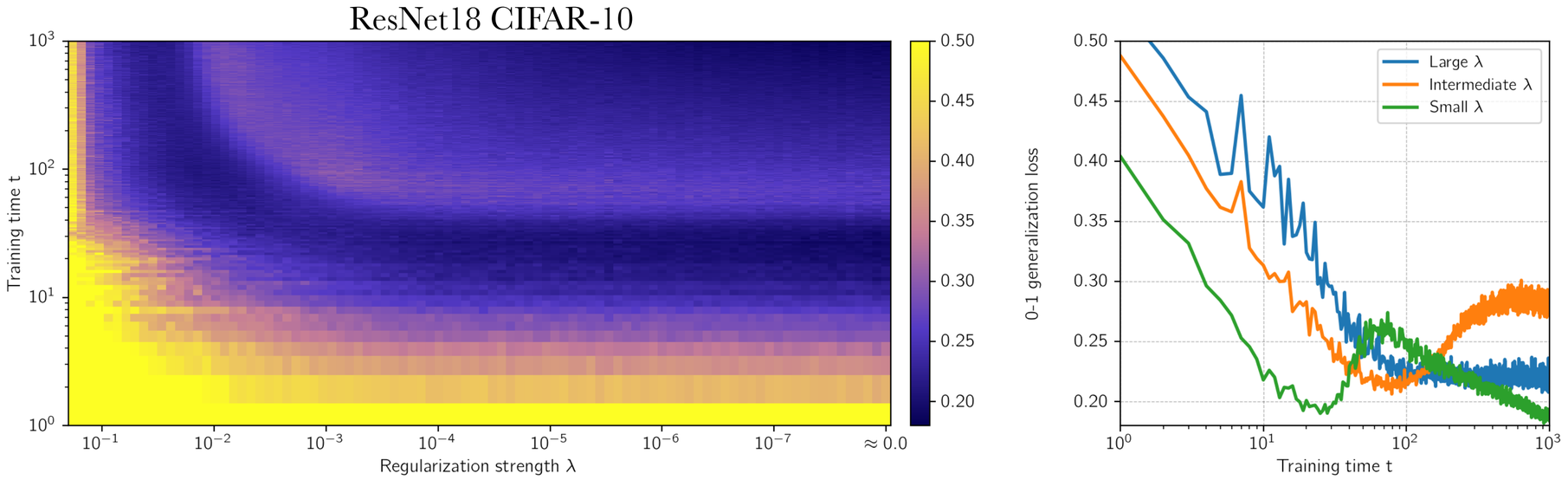
$\alpha_1 = \frac{n}{p}$, $\alpha_2 = \frac{n}{d-p}$, $\gamma_1 = \frac{1}{\sigma_1^2 \eta t}$ $\gamma_2 = \frac{1}{\sigma_2^2 \eta t}$

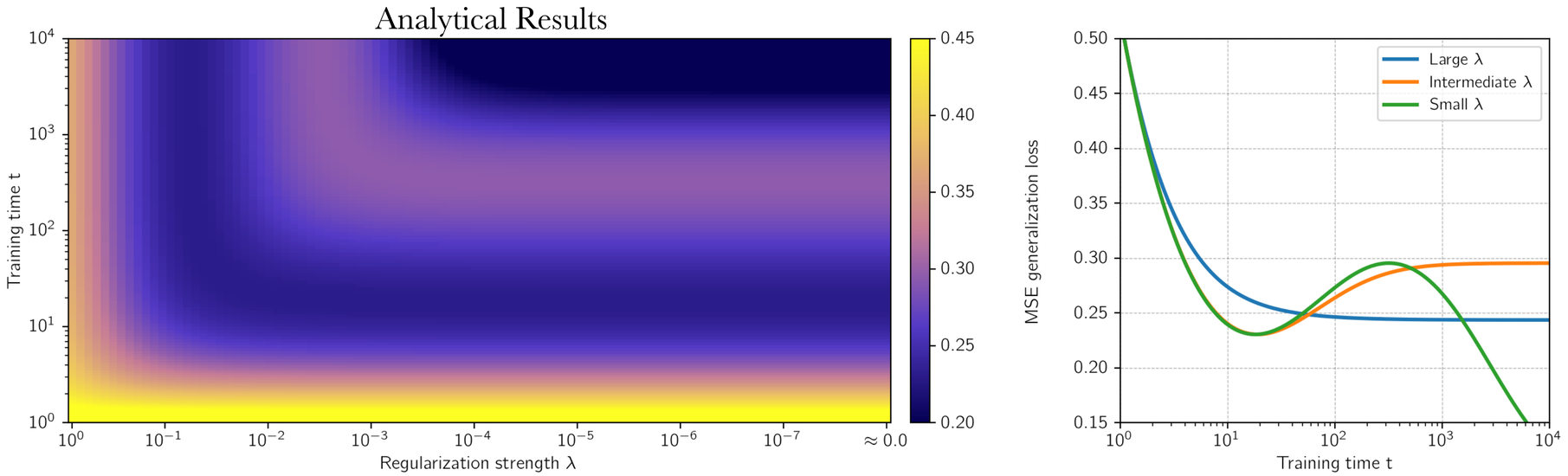
$a_i = 1 + \frac{2\gamma_i}{(1 - \alpha_i - \gamma_i) + \sqrt{(1 - \alpha_i - \gamma_i)^2 + 4\gamma_i}}$, for $i \in \{1, 2\}$.

$b_i = \frac{\alpha_i}{a_i^2 - \alpha_i}$, $c_i = 1 - 2R_i - \frac{n}{d}\frac{2 - a_i}{a_i}$ for $i \in \{1, 2\}$,

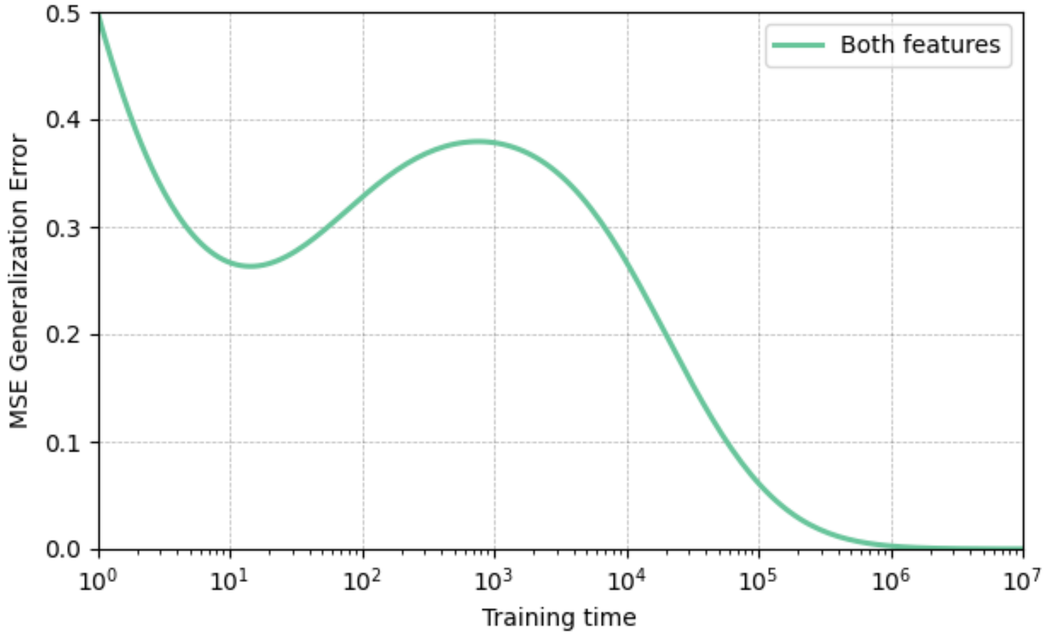


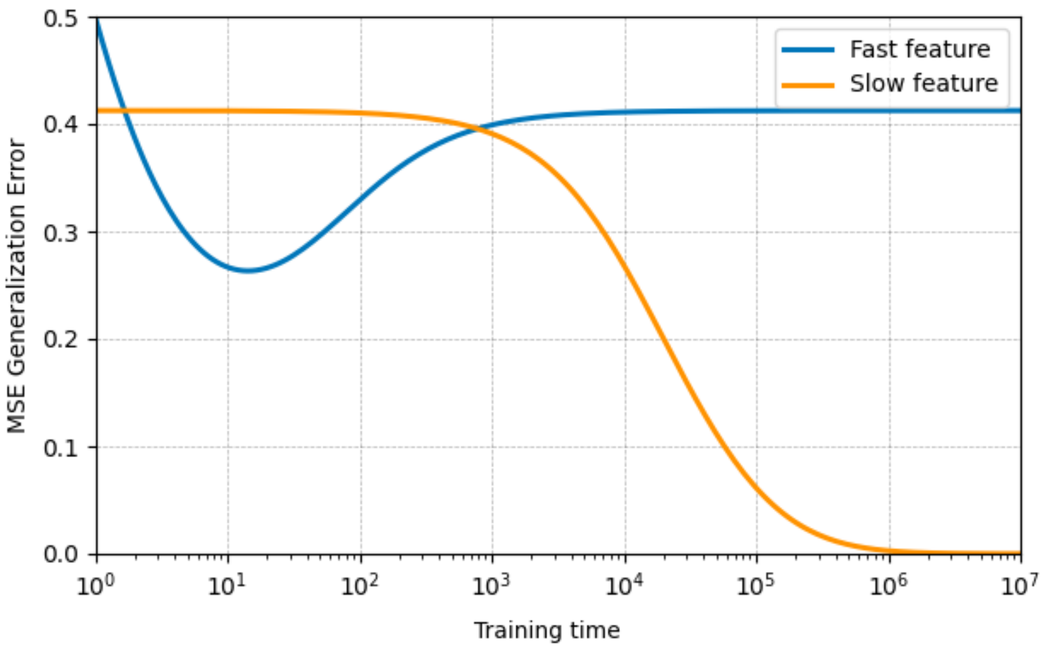
Qualitative results:





Key understanding:





At the same time a fast-learning feature overfits, a slow-learning feature starts to fit.

Thank you!

Check out our paper for more amazing details!