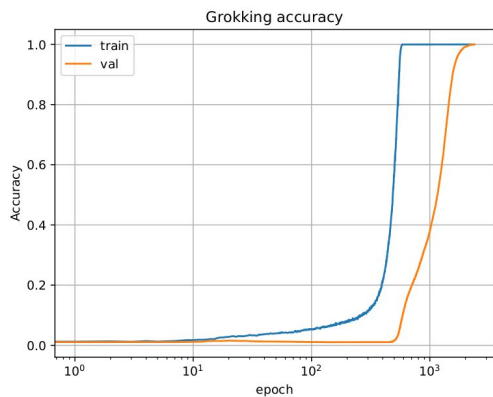


Периодическое поведение при обучении глубоких нейронных сетей на примере эффекта гроккинга

Мельник Юрий Максимович

Кафедра ММП, факультет ВМК
МГУ им. Ломоносова

Научный руководитель: кандидат ф.-м. наук Ветров Д.П
Научный консультант: Южаков Т.А.



★	a	b	c	d	e
a	a	d	?	c	d
b	c	d	d	a	c
c	?	e	d	b	d
d	a	?	?	b	c
e	b	b	c	?	a

Визуализация свойства масштабной инвариантности

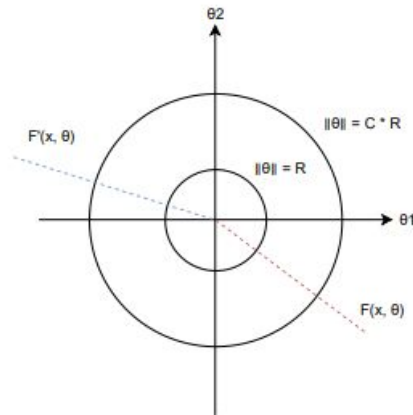
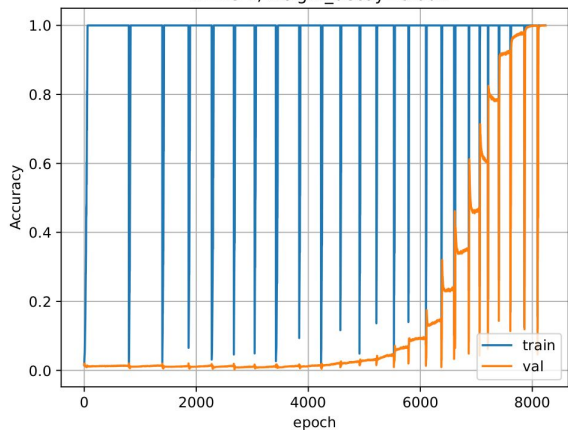
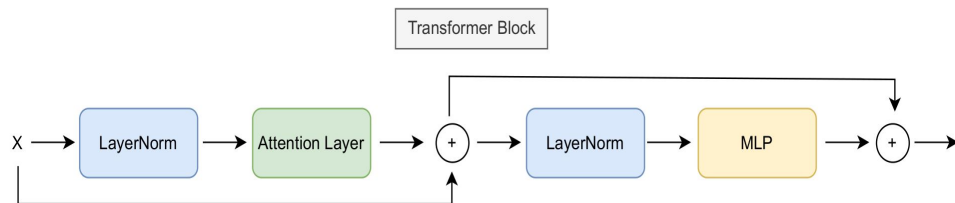


График точности характерный для эффекта “гроккинга” (левый рисунок)
Синтетические данные для задачи “гроккинга” (правый рисунок)

2L Transformer with LayerNorm on Modular Addition (p=97)
lr=1e-1, weight_decay=0.001



$$\log p(y|x, \theta) = \log p(y|x, C\theta), \quad \forall C > 0$$



Трансформерная архитектура, использовавшаяся в экспериментах

Периодическое поведение при обучении модели в задаче “гроккинга”