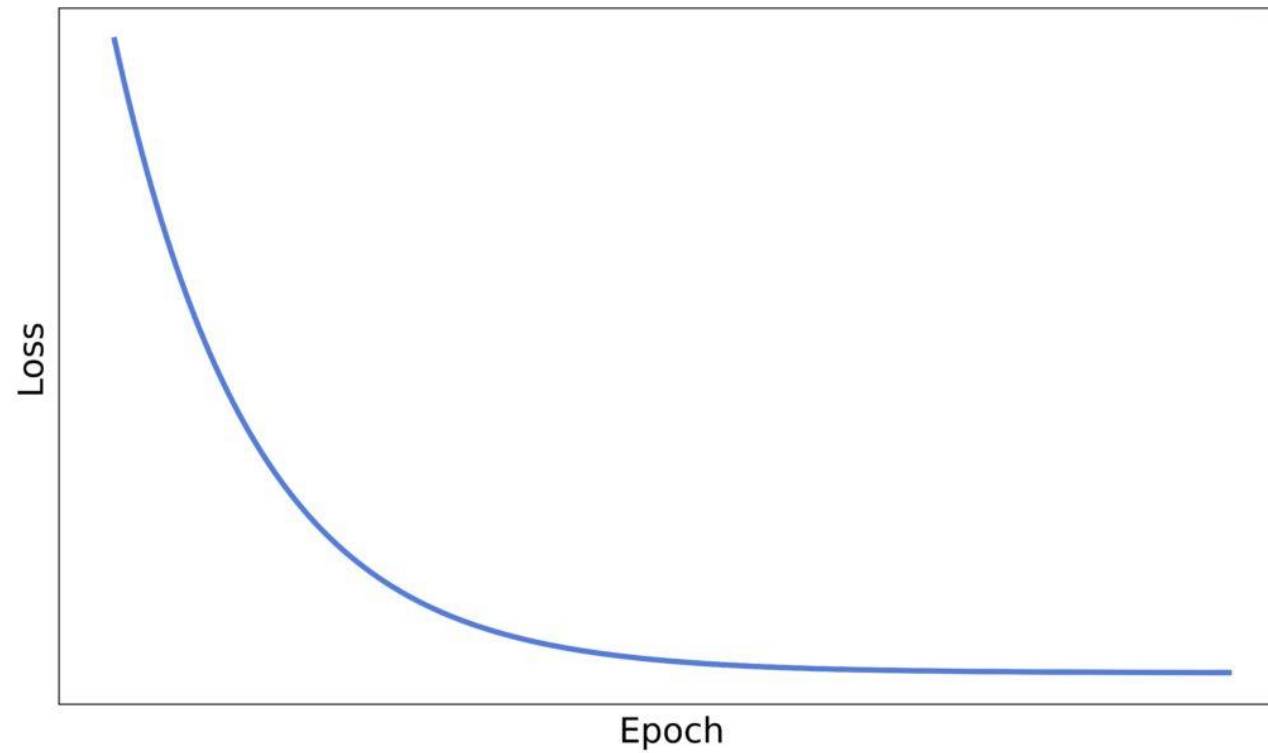
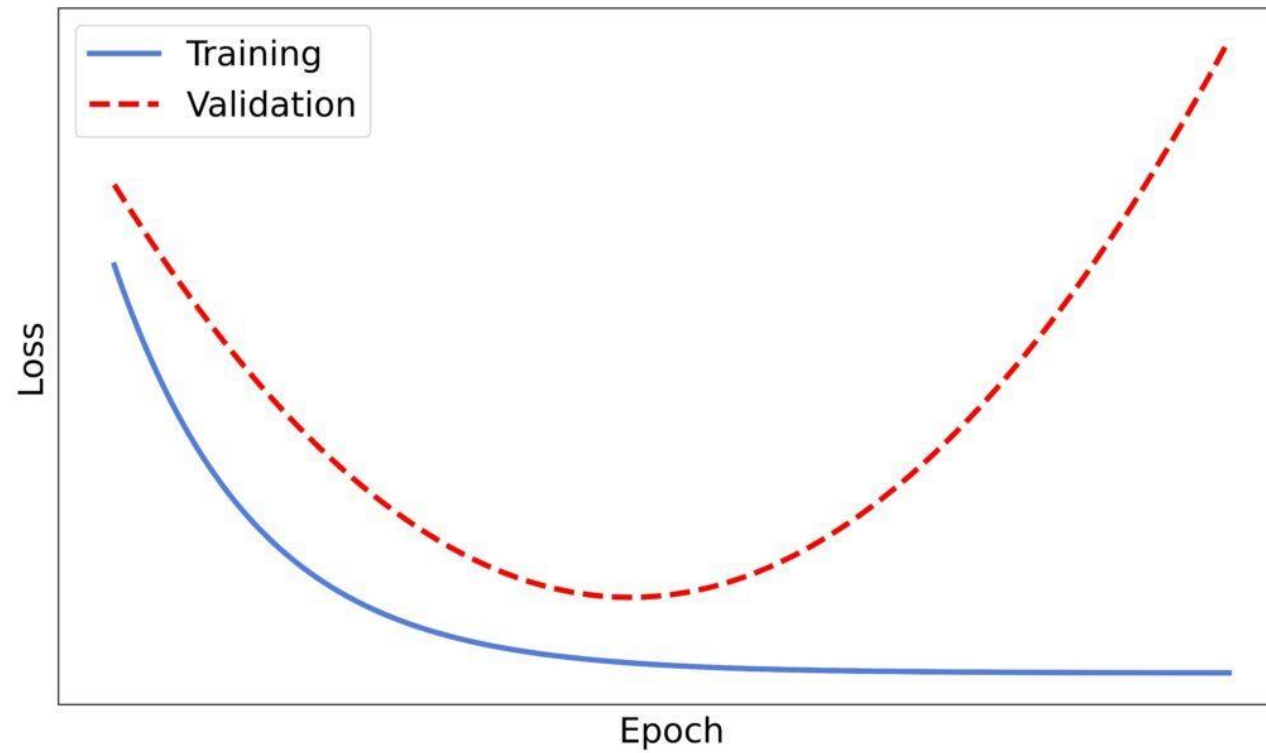


grokking

# neural networks: loss graph



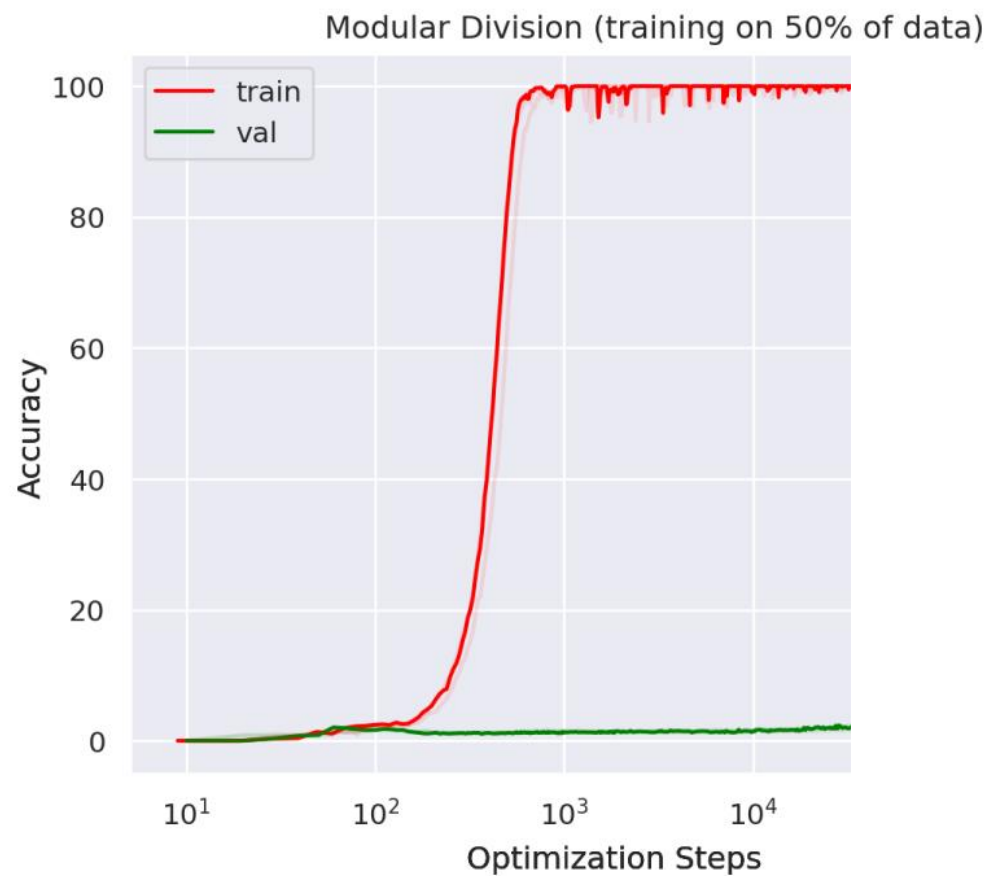
# neural networks: overfitting



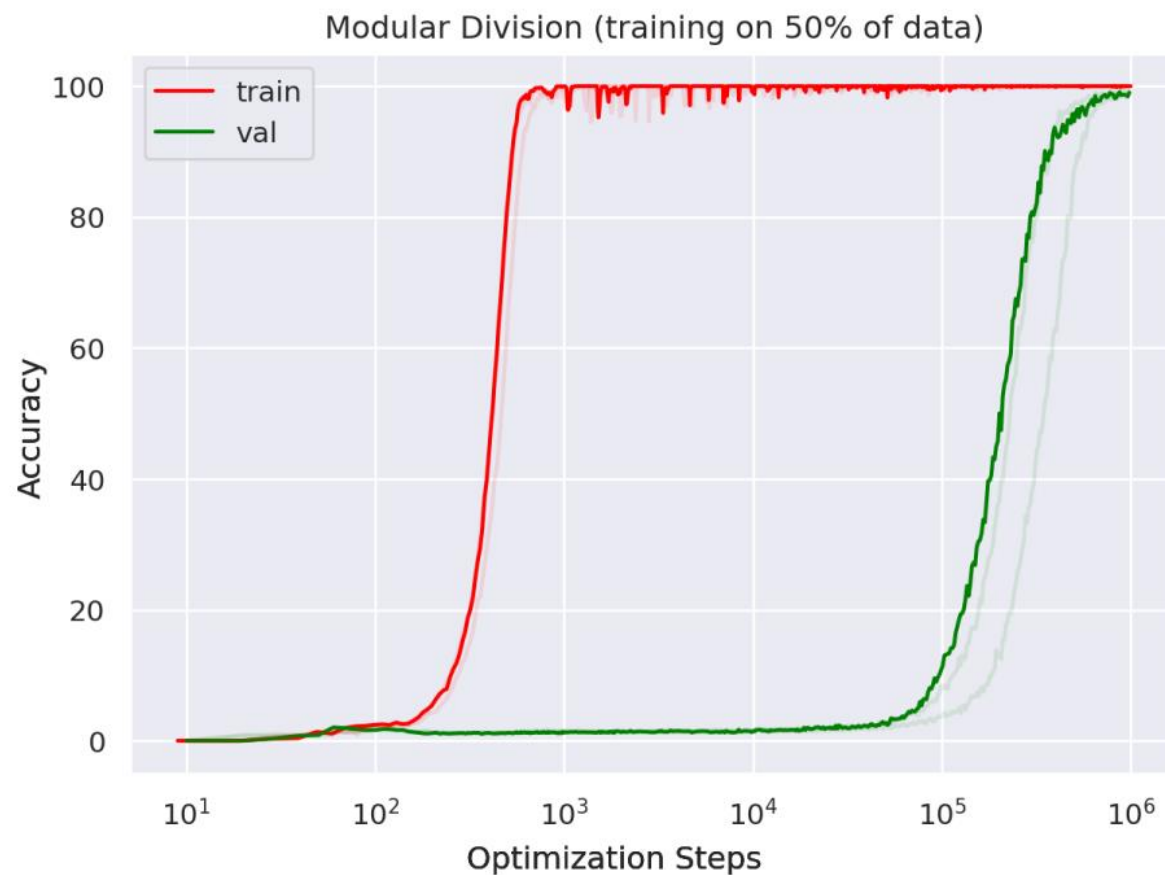
neural networks: weight decay

$$\lambda \|w\|_2$$

# grokking: epoch-loss graph



# grokking: epoch-loss graph



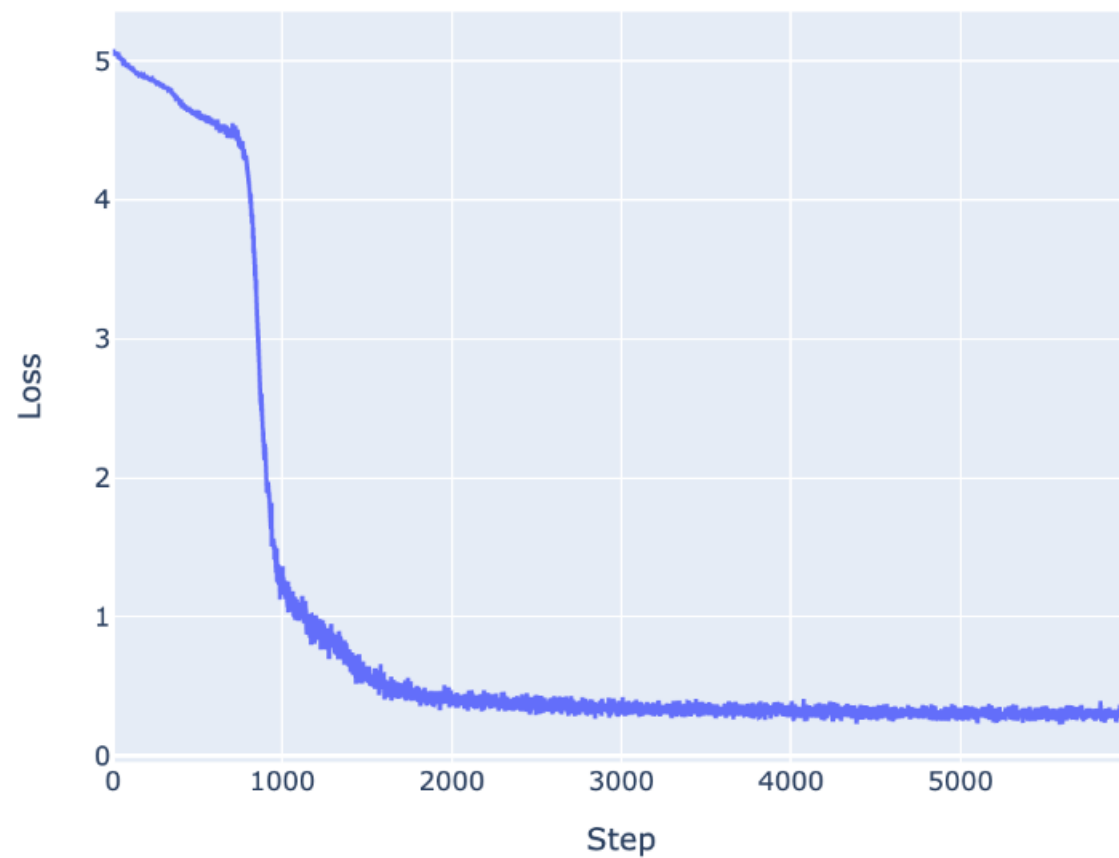
# grokking: example

★	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

ColpafoDeHjJpxmcyloxytjwJnZvixkplαQnnmuvLbηiροZqATgvESoRphsnwkrγldOIbXhW·μcNc|TXβKnoZUGYVθierξn  
CnOHyCIGtmgmNrvNrlεVηC TJtYiW o n m a W Y p Q D l i c x a η Z  
l y i p e r N v u K i r n J c m X O o n a l O h c o X n o U l ξ  
p c i f k j J P α D F n c v γ π c l z n g i l o r θ c n y t r o π μ n n c k γ r t D  
a f u k ε t o v u j V h n Q i p d O i β c n y t r o π μ n n c k γ r t D  
D n m n N T ε n λ γ u z p d c K γ m v n k c γ r W U t o i O n A d δ q o Y j Z w t  
e H μ γ η v p B a r p n m w l n d c K γ m v n k c γ r W U t o i O n A d δ q o Y j Z w t  
J R P e N a r p n m w l n d c K γ m v n k c γ r W U t o i O n A d δ q o Y j Z w t  
P j F P e r γ l O n r d d k ε J J Q c D K A p c x d z μ v w t h ν β j λ  
x p M r a n r M X v O i t ξ Y G X k e E n a ν F w i l x k m p r t  
Y c n p B o M x w o P α G B t θ n o W η μ T β c ζ m w i β s q T n a X n i f K o η  
l d n δ c x w o M x w o P α G B t θ n o W η μ T β c ζ m w i β s q T n a X n i f K o η  
x t Q r W x n i R W j ζ η w j n v γ z p h y O Y α n e a ν P U c i r s q v i a X n n f K o η  
y λ w ζ θ s k n c o W j ζ η w j n v γ z p h y O Y α n e a ν P U c i r s q v i a X n n f K o η  
b F FK n Q d o S G y U v H j n D U x k e d θ n r Z c o c t p r B n k g y e b  
J L y l η n v F n L U g T x η w P α β i H Z q n W s c p c l D v g j k a G t n y L a v u h y D c n  
ζ n v D v y z a η p T x η w P α β i H Z q n W s c p c l D v g j k a G t n y L a v u h y D c n  
n k R c o h i t y n d n F a β i H Z q n W s c p c l D v g j k a G t n y L a v u h y D c n  
p λ a S p l β T G G n k m U β i P r v n W A z i e l T w n y L a v u h y D c n  
Q v k c n G δ u K p μ s u c i G v n Y n p B g v t o u T L o i y l U E O t y j j r  
n i o l m x b t R n G v f z d p c r e t i o v i c u j j n θ n l a c d y v l p n f o s c c o k  
v L o J p v f K Q n W f k i n c l x θ n v W s c p c l D v g j k a G t n y L a v u h y D c n  
b n c q p h n Q n W f k i n c l x θ n v W s c p c l D v g j k a G t n y L a v u h y D c n  
l i s x R s p L Q n c v j a i n k i p z c G r n W η r c b j i n m p l b a f u  
o z q a l y F X n k β b e l E y z d n β c p n W Q a U i r c i K η a b x t o y ζ v v λ O φ  
T g η X n k β b e l E y z d n β c p n W Q a U i r c i K η a b x t o y ζ v v λ O φ  
g n p x ξ o x X n e r v G μ k l J S r K d k p n G F n v K a w Y v K ξ o V b M  
E S o L h f h ξ δ n u i c X n E a η λ A M s W x o v q η F n v K a w Y v K ξ o V b M  
R p β x u p w s j n u i c U n η E a μ w V i z θ n c D N b A J S z s r v e v x m q k n E i y  
n q η n Z β n i y d X f a k y Q e v H v θ m ζ l k C b o n θ n F i n n x X ζ y  
s n w i w Y R Z o y i R j δ l p x K t n p v o p m ζ l k C b o n θ n F i n n x X ζ y  
k i θ n S n r D o M A p f p k u γ l y A p o i P t w j n v m e c v o n l v ξ i n d c z V Y F t F S β Y V η  
γ l d n M T M i M i n g o o μ v G δ o i P t w j n v m e c v o n l v ξ i n d c z V Y F t F S β Y V η  
O i B z M i M i n g o o μ v G δ o i P t w j n v m e c v o n l v ξ i n d c z V Y F t F S β Y V η  
h β u h p y i o α μ D r ζ c N ε r i n l Y O o η i v ξ i n d c z V Y F t F S β Y V η  
W λ n c ζ ζ M z β P o ε μ R 9 γ J e o k p i β a y U a l p m n j n a j w ξ η p z i  
μ C L h j β o λ r ξ E r i w v d o E v y K c N n y o O η T δ B b R r o r a g W Z A Zh T r z M p S s  
N c l p o η u v o J i p R i o B b R r o r a g W Z A Zh T r z M p S s  
X r n β η a e m j o K c W w L p w p κ l d x n n i n x A b p u η x r γ H D h t G i n c o a p y  
β K S n β η a e m j o K c W w L p w p κ l d x n n i n x A b p u η x r γ H D h t G i n c o a p y  
n U e z a D o h c η n η D h a z θ y A n e r T κ i θ e i O n n c β z i c l d n i o j B  
o z a H z a D o h c η n η D h a z θ y A n e r T κ i θ e i O n n c β z i c l d n i o j B  
U o E γ n β r o r G n d R z W i o a i c ξ f λ V E t e b v n v n c s E R θ U q t  
G V t θ s β r o r G n d R z W i o a i c ξ f λ V E t e b v n v n c s E R θ U q t  
V z k i o λ h r o a h W A e Z j n l n o n v c n B Y j V E n i Y X V v q A j  
ε n n k H r t η w i W c n B Y j V E n i Y X V v q A j

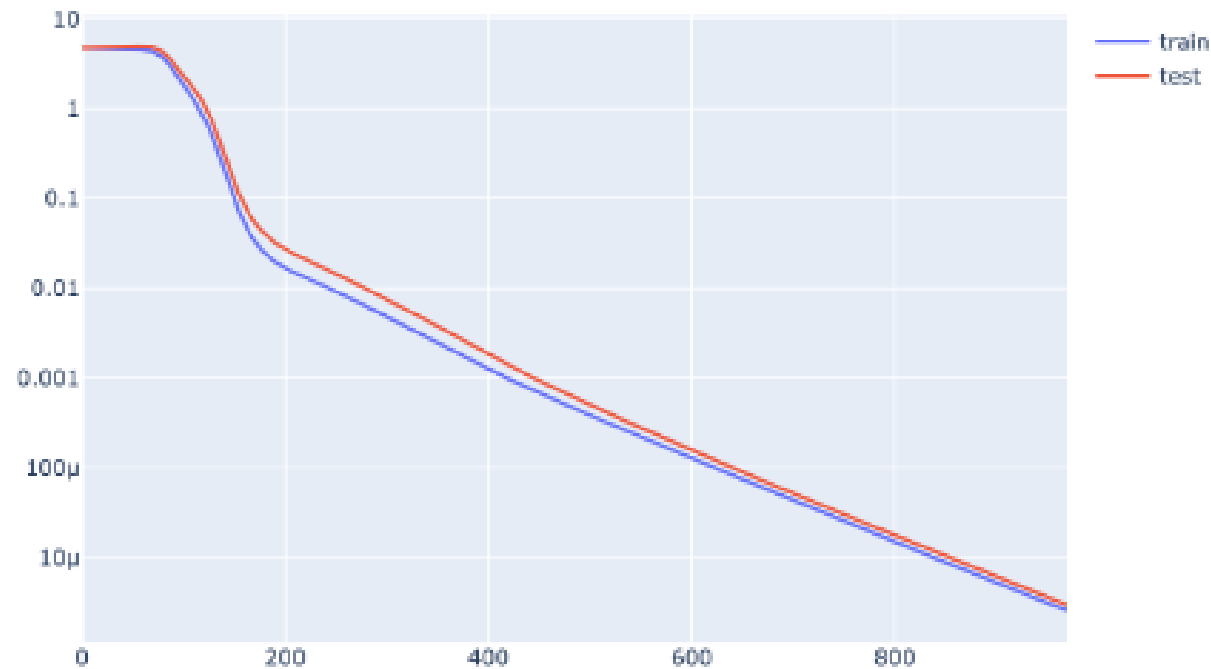


# phase change: epoch-loss graph



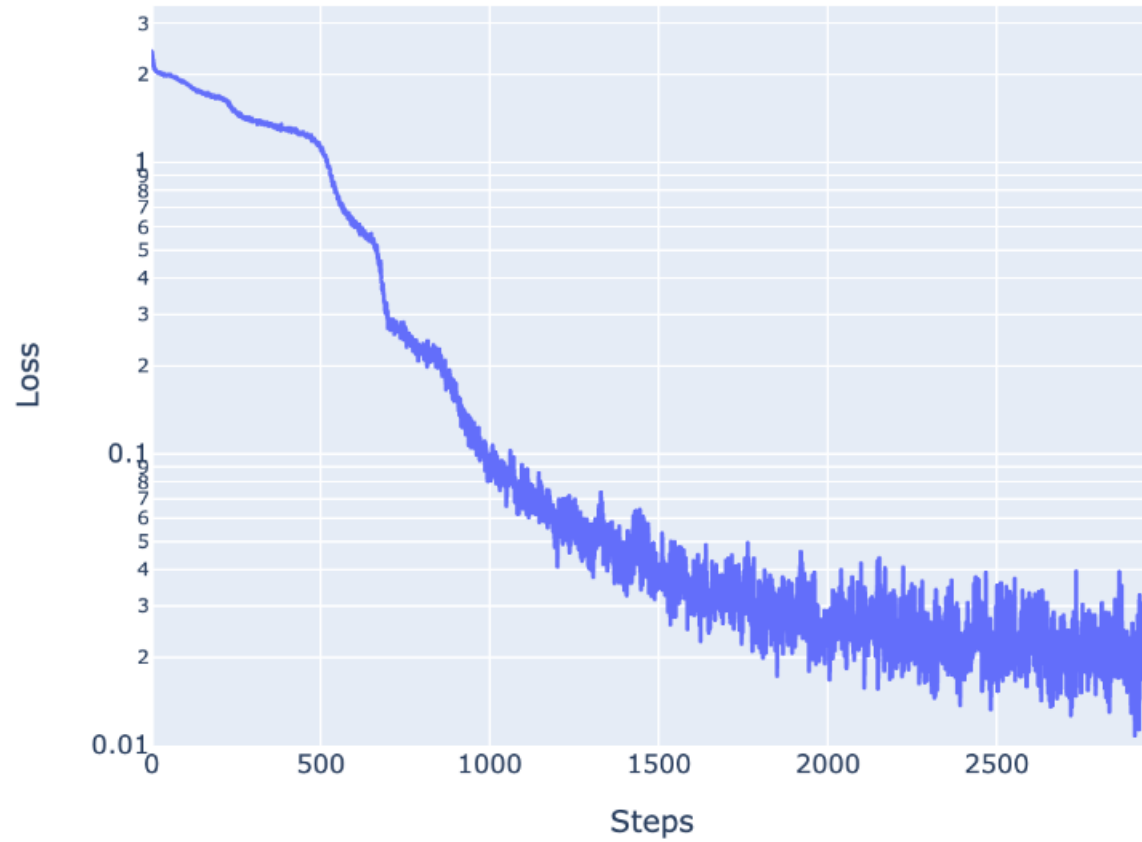
# example: addition mod 113

Train + Test Loss curves for modular addition trained on 95% of the data

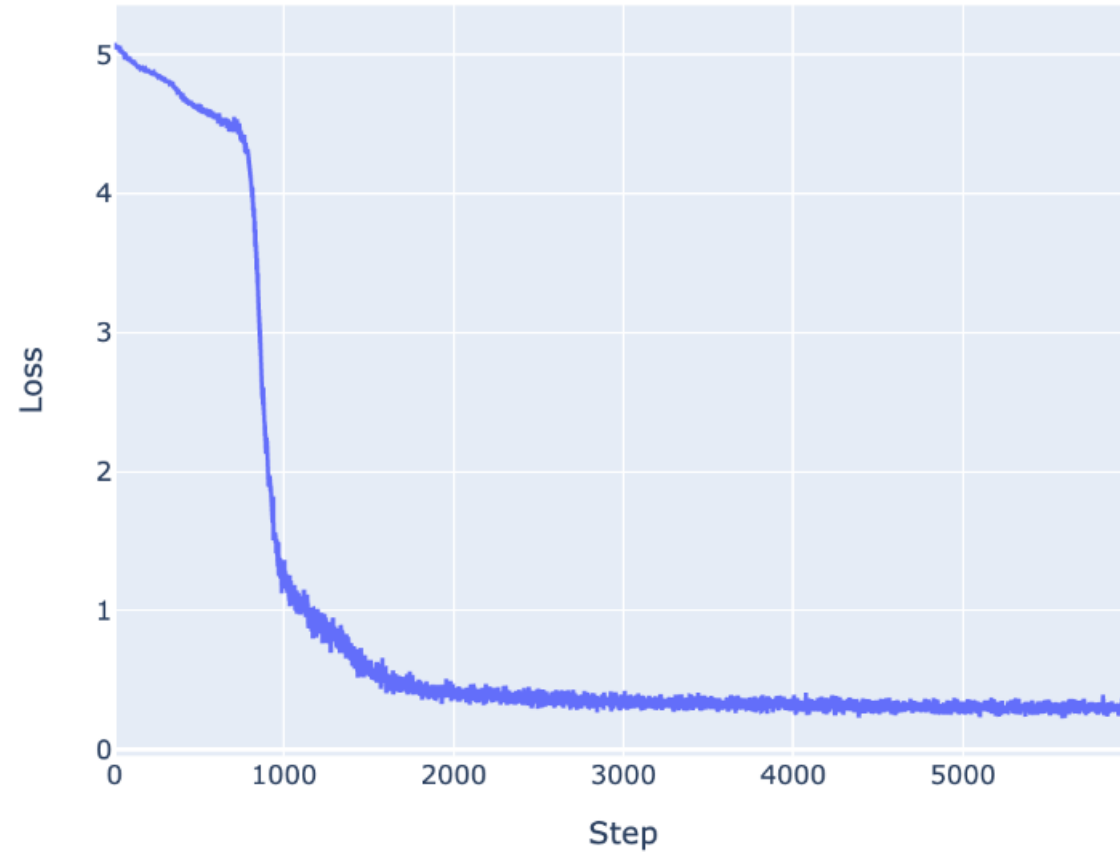


Modular addition mod 113 loss curve, trained on 95% of the data

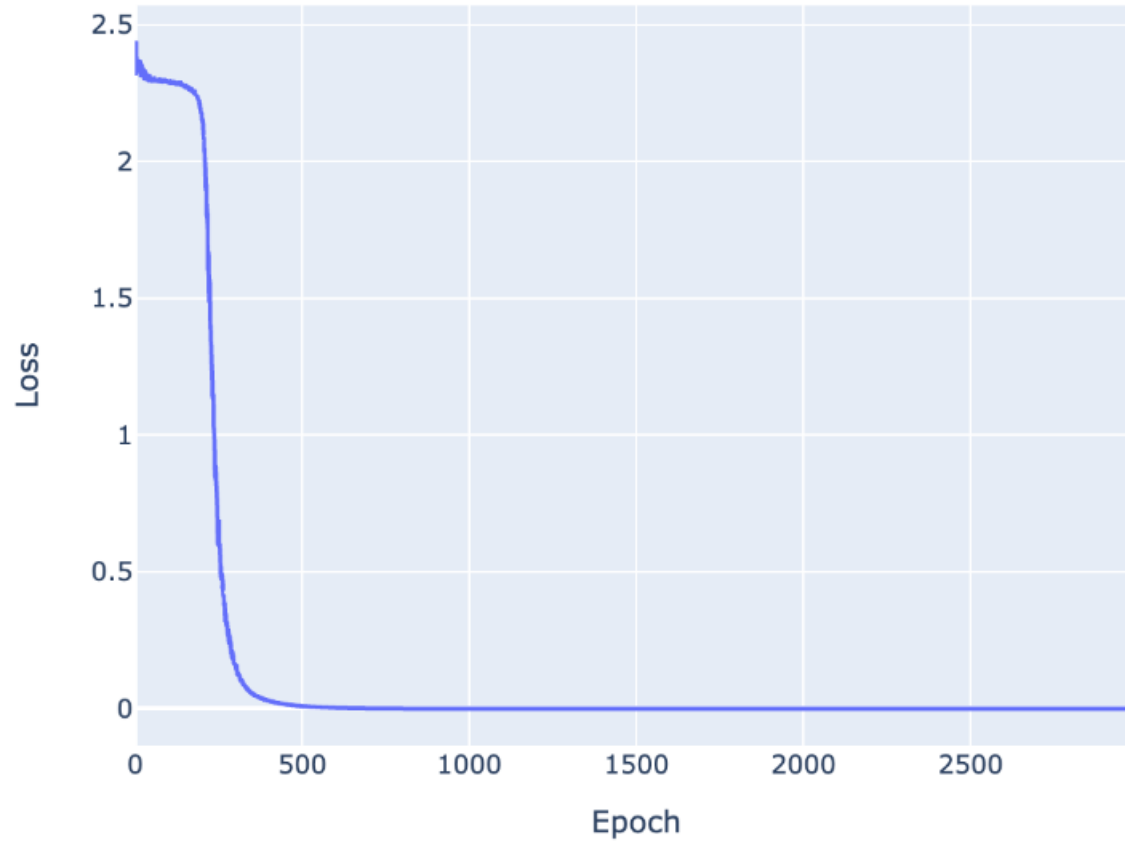
example: 5 digit addition



example: predicting repeated subsequences



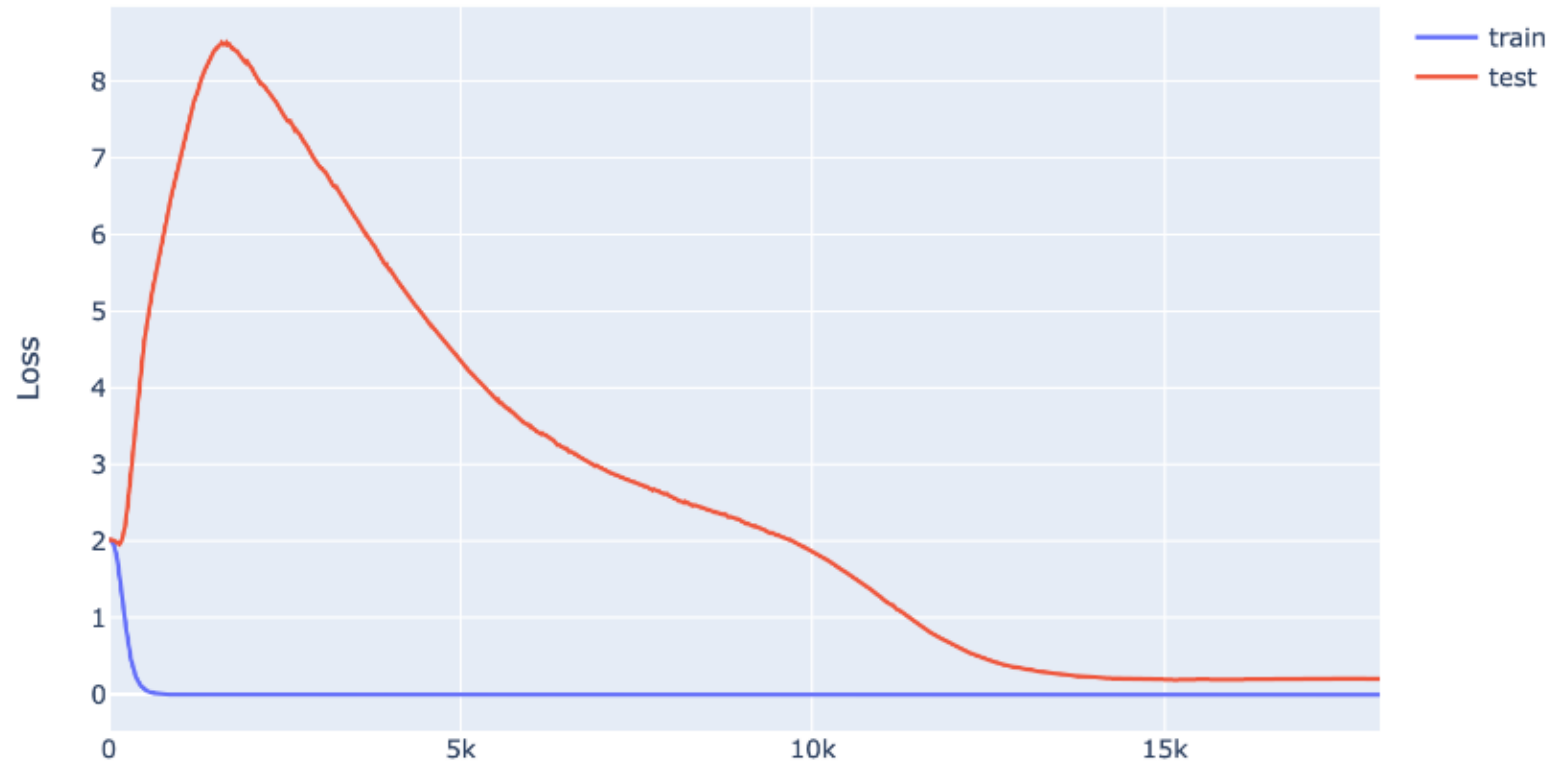
example: finding the max element



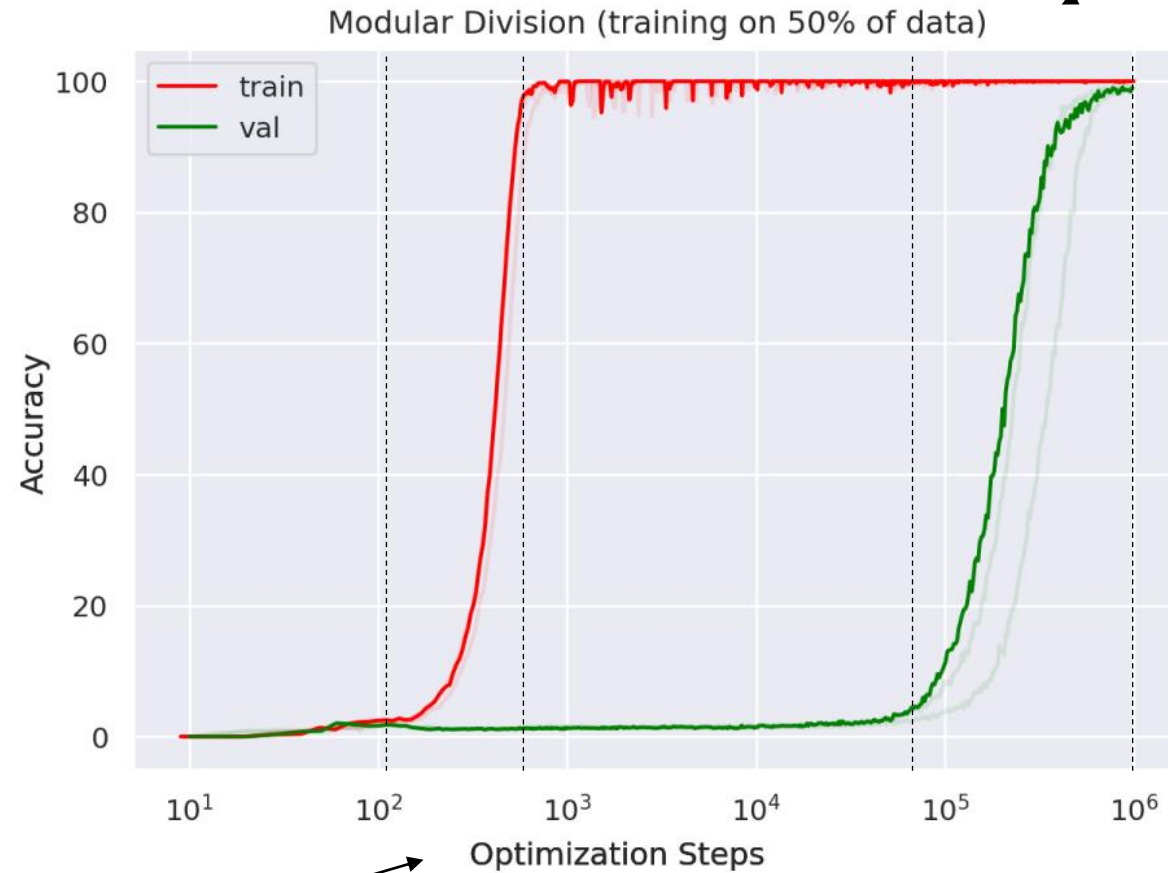
# phase change: possible explanation

- A lottery ticket hypothesis-inspired explanation
- A random walk explanation
- An evolutionary explanation

phase change + weight decay + small sample



# grokking: explanation



phase change

memorization



# bonus: $x + y \bmod p$ algorithm

1.  $x, y$
2.  $\cos kx, \sin kx, \cos ky, \sin ky$
3.  $\cos k(x + y), \sin k(x + y)$
4.  $\cos k(x + y - z)$

$$k = \frac{2\pi}{p}$$