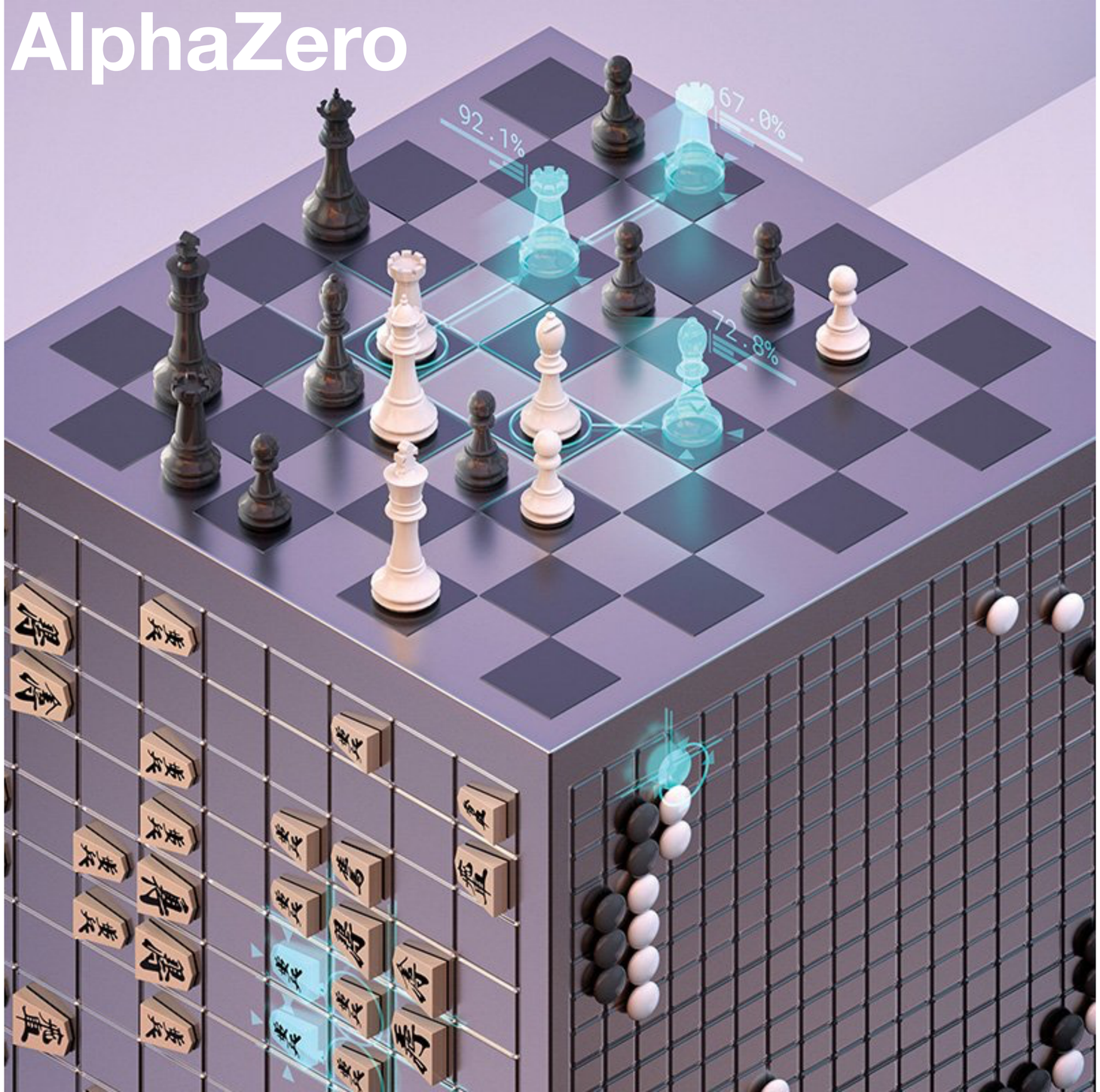
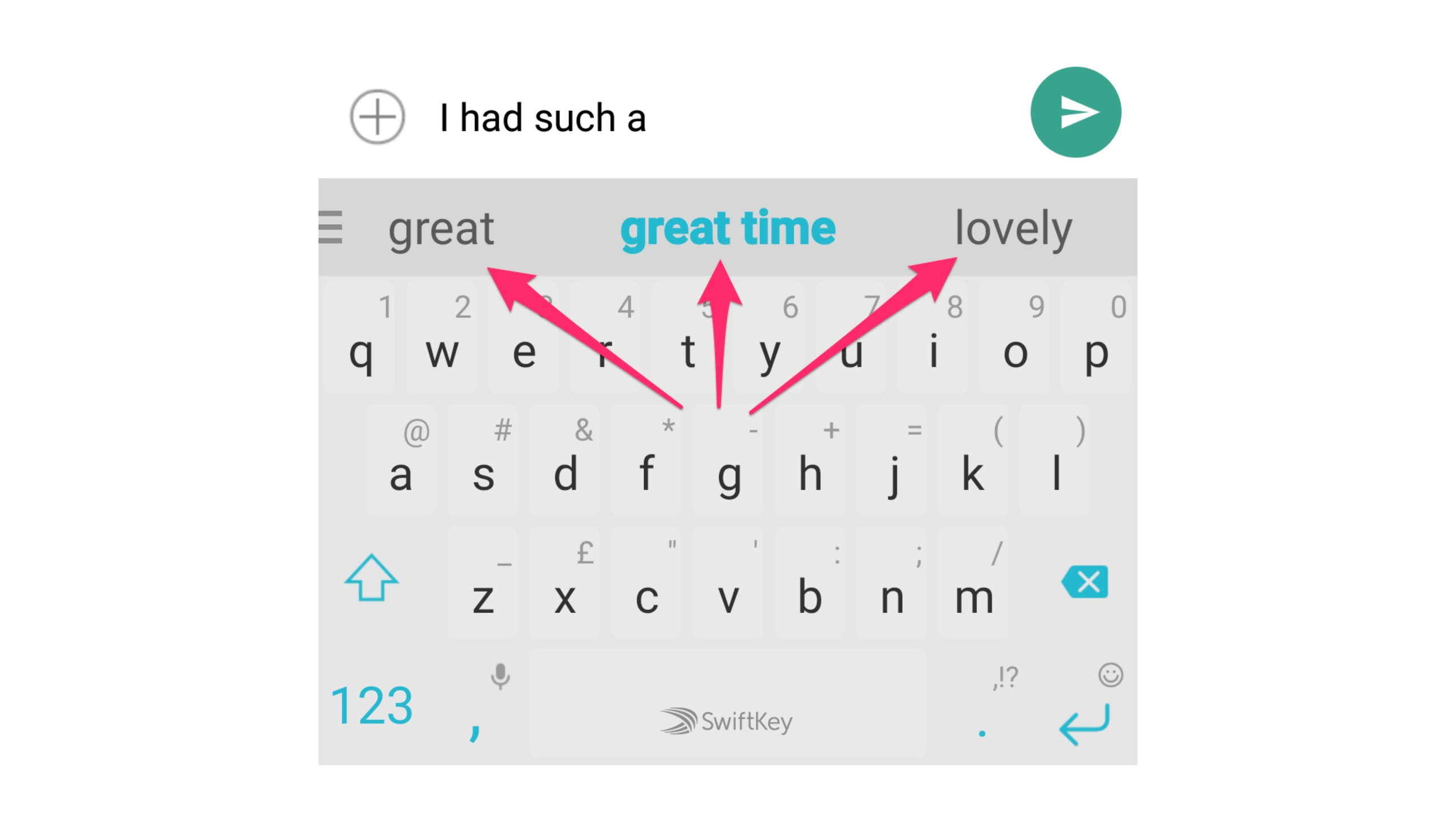
Machine learning for people who already know all the math.

Anton de la Fuente


AlphaZero
92.1%
67.0%
72.8%


I had such a
great
great time
lovely
SwiftKey


Google
How to
how to 意味
how to use
how to 歌詞
how to pronounce
how to build a girl
how to basic
how to be single
how to avoid a climate disaster
Google 検索
I'm Feeling Lucky
不適切な検索候補の報告


Google
どうやって
どうやって
どうやって 英語
どうやって死にてぇ マイキー
どうやって生きていこう
どうやって 韓国語
どうやって戦えばいいんだ
どうやってここまで
どうやって学校に行きますか 英語
Google 検索
I'm Feeling Lucky
不適切な検索候補の報告

Written by AI 人工知能が書いた

The best kind of ramen is the one you can eat while doing your taxes. At least, that's what the government's recently published "Ramen: The Global Ramen Experiment" suggests.

Next, I will tell you about string theory. String theory is a mathematical model of our universe, a theory that describes not only how our spacetime is constructed, but how it is built into the fabric of reality.

OpenAI

an armchair in the shape of an avocado
アボカドの形をした椅子

OpenAI

a tapir made of accordion
アコーディオンで作った獏さん

OpenAI

Outline

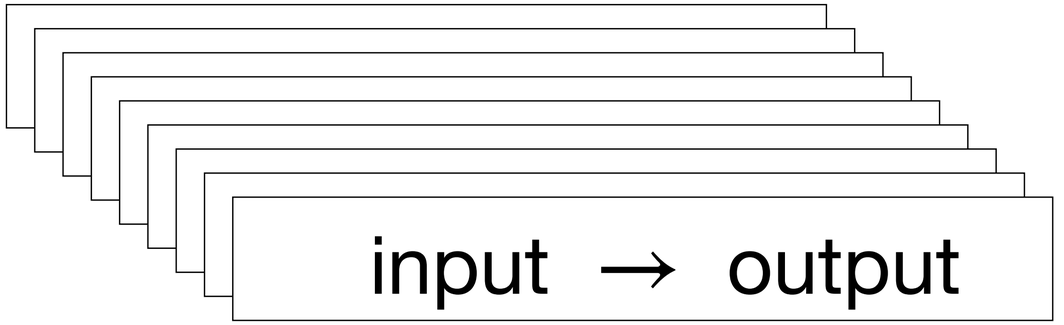
- How is playing chess and writing text curve fitting?
- What do these functions look like?
- Why don’t we overfit or get stuck in a local minimum?

Outline

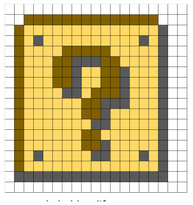
- How is playing chess and writing text curve fitting?
- What do these functions look like?
- Why don’t we overfit or get stuck in a local minimum?

Goal of machine learning

- Training data 学習データ



- New input 新しい入力 ⟶



Function domain and range

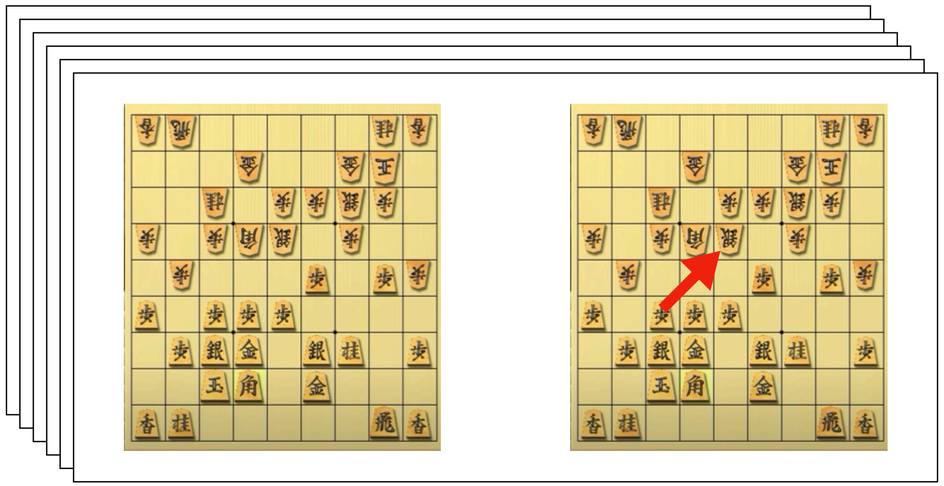
$$\text{Data} = \{(x_\alpha, y_\alpha)\}$$

$$x_\alpha \in \mathbb{R}^{M \times N}$$

$$y_\alpha \in \mathbb{R}^{M' \times N'}$$ one-hot

$$p(x, \theta) \in \mathbb{R}^{M' \times N'}$$ probability distribution
parameters θ

Chess・Shogi 将棋・Go 碁

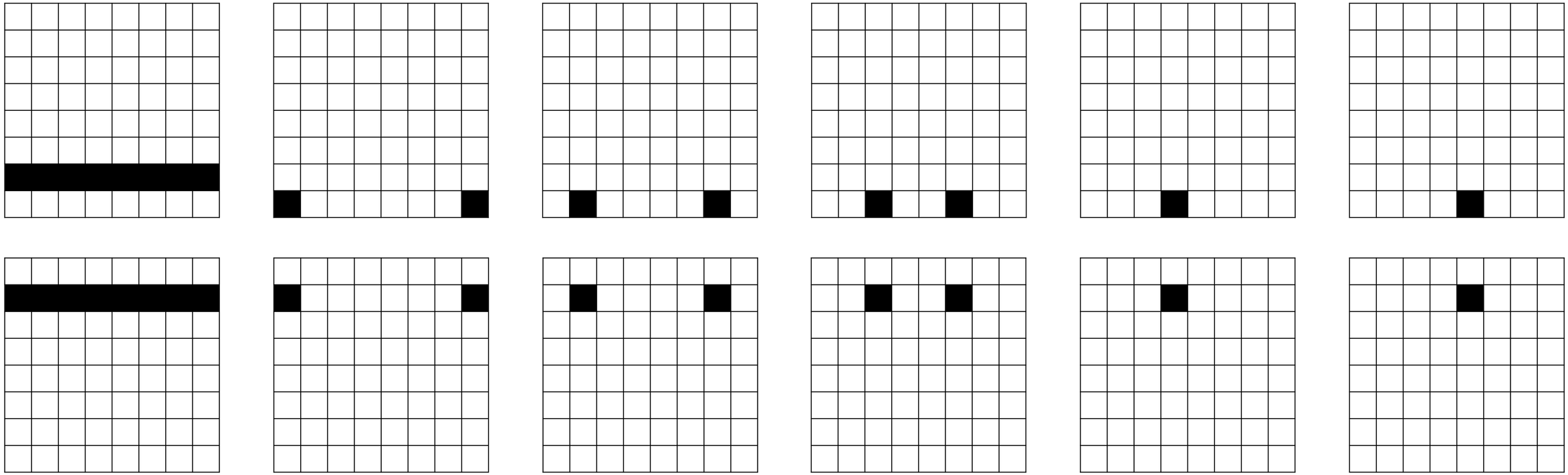


- (Position, move made by winning player)
(駒の位置、勝った棋士の次の一手)

- 44 million games 44百万の試合

Chess input space

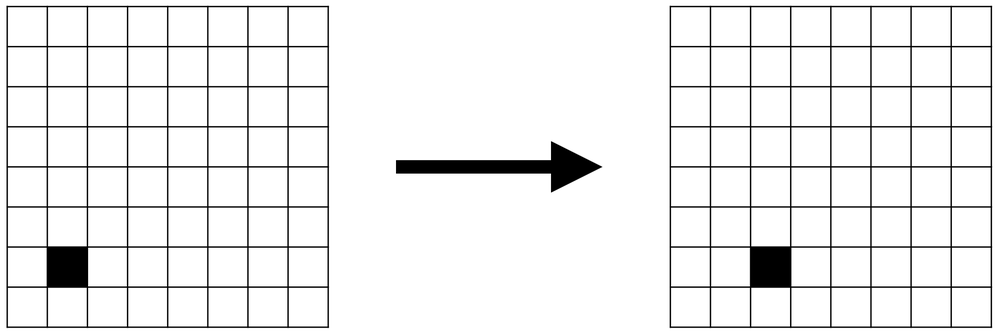
$$x \in \mathbb{R}^{8\times8\times(6+6+1)}$$



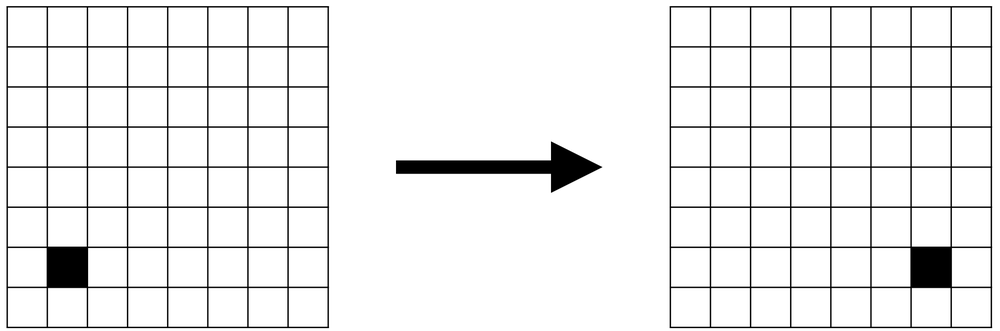
Chess output space

$$y \in \mathbb{R}^{8\times8\times(8\times7+8)}$$

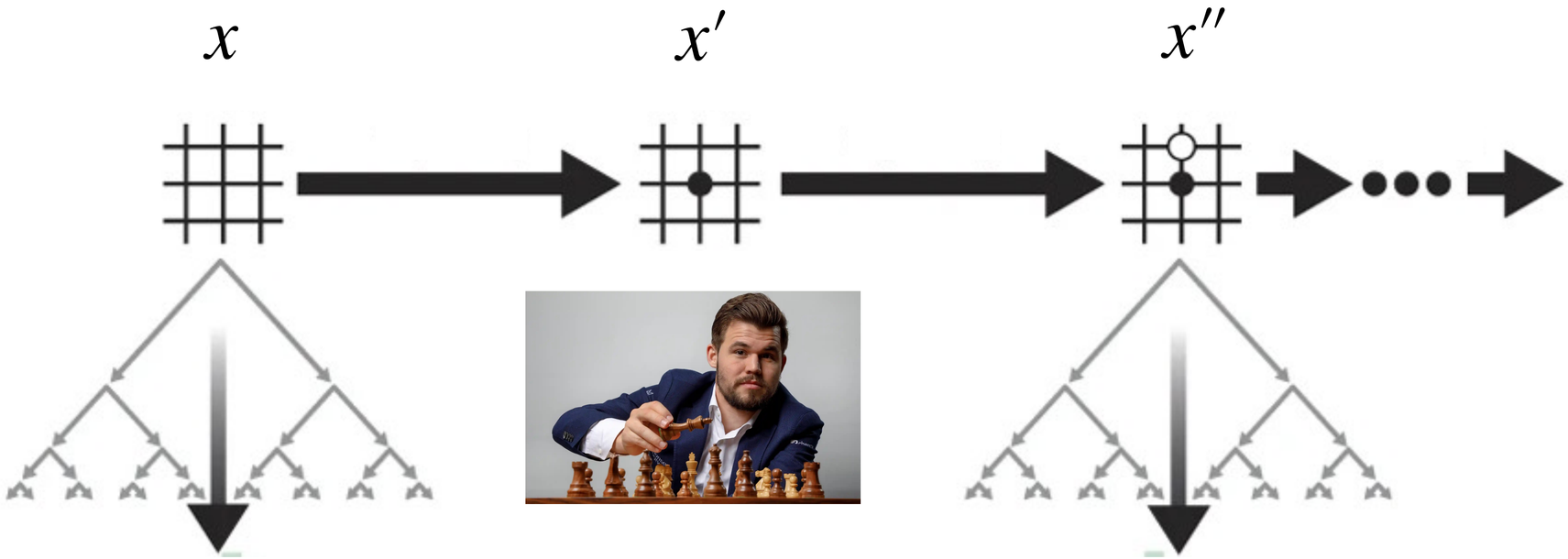
$y_{2\times2\times(3\times1+0)}$



$y_{2\times2\times(3\times5+0)}$



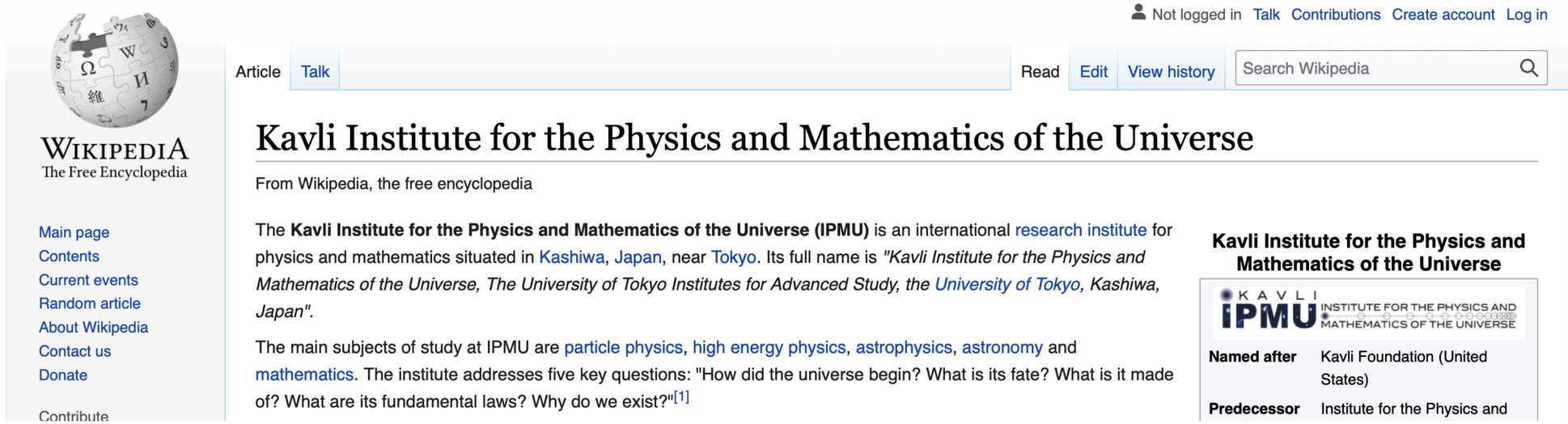
How to play chess



$$y \sim P(x, \theta)$$

$$y'' \sim P(x'', \theta)$$

Text-writing training data



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute

Kavli Institute for the Physics and Mathematics of the Universe

From Wikipedia, the free encyclopedia

The **Kavli Institute for the Physics and Mathematics of the Universe (IPMU)** is an international research institute for physics and mathematics situated in Kashiwa, Japan, near Tokyo. Its full name is *"Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo Institutes for Advanced Study, the University of Tokyo, Kashiwa, Japan"*.

The main subjects of study at IPMU are particle physics, high energy physics, astrophysics, astronomy and mathematics. The institute addresses five key questions: "How did the universe begin? What is its fate? What is it made of? What are its fundamental laws? Why do we exist?"[1]

Kavli Institute for the Physics and Mathematics of the Universe

KAVLI IPMU INSTITUTE FOR THE PHYSICS AND MATHEMATICS OF THE UNIVERSE

Named after	Kavli Foundation (United States)
Predecessor	Institute for the Physics and

- (The Kavli Institute for, the)
- (The Kavli Institute for the, Physics)
- (The Kavli Institute for the Physics, and)
- Next word prediction, 600 Billion words 次の単語を予測、6千億の単語

Encode words as vectors

$$\mathbb{R}^{50000}$$

$$\begin{aligned} \text{The} &= (1,0,0,0,\cdots) \\ \text{is} &= (0,1,0,0,\cdots) \\ \text{apple} &= (0,0,1,0,\cdots) \\ \text{Japan} &= (0,0,0,1,\cdots) \end{aligned}$$

presentations

Sentences are a concatenation of words

$$x = \left[x_{\text{Life}}, x_{\text{is}}, x_{\text{great}}, x_{!}, x_{\text{pad}}, x_{\text{pad}}, \cdots, x_{\text{pad}} \right]$$

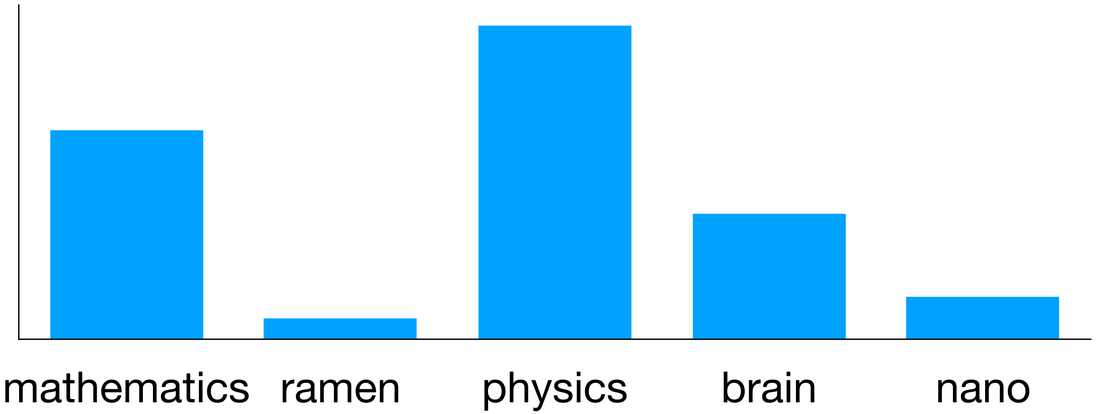
$$\text{input} \in \mathbb{R}^{50000 \times 512}$$

$$\text{output} \in \mathbb{R}^{50000} \quad \text{probability distribution}$$

How to write text

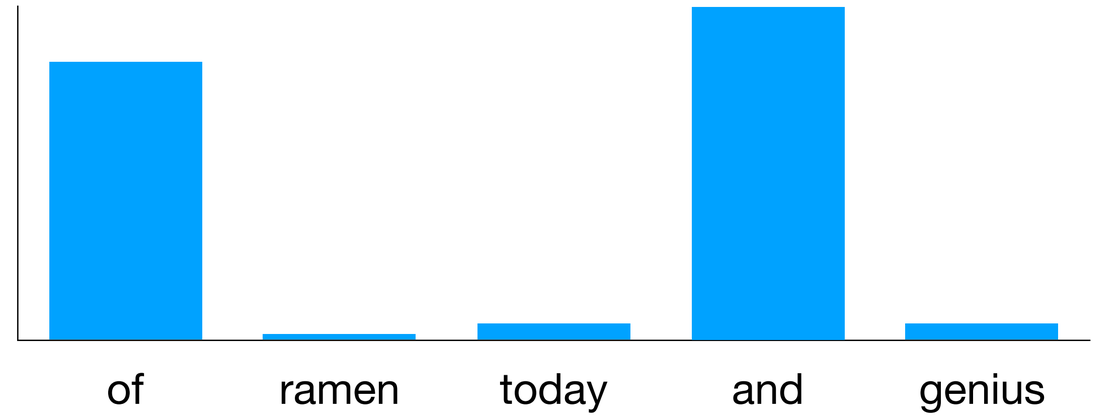
$x =$ The Kavli Institute for the

$P(x, \theta)$



$x' =$ The Kavli Institute for the physics

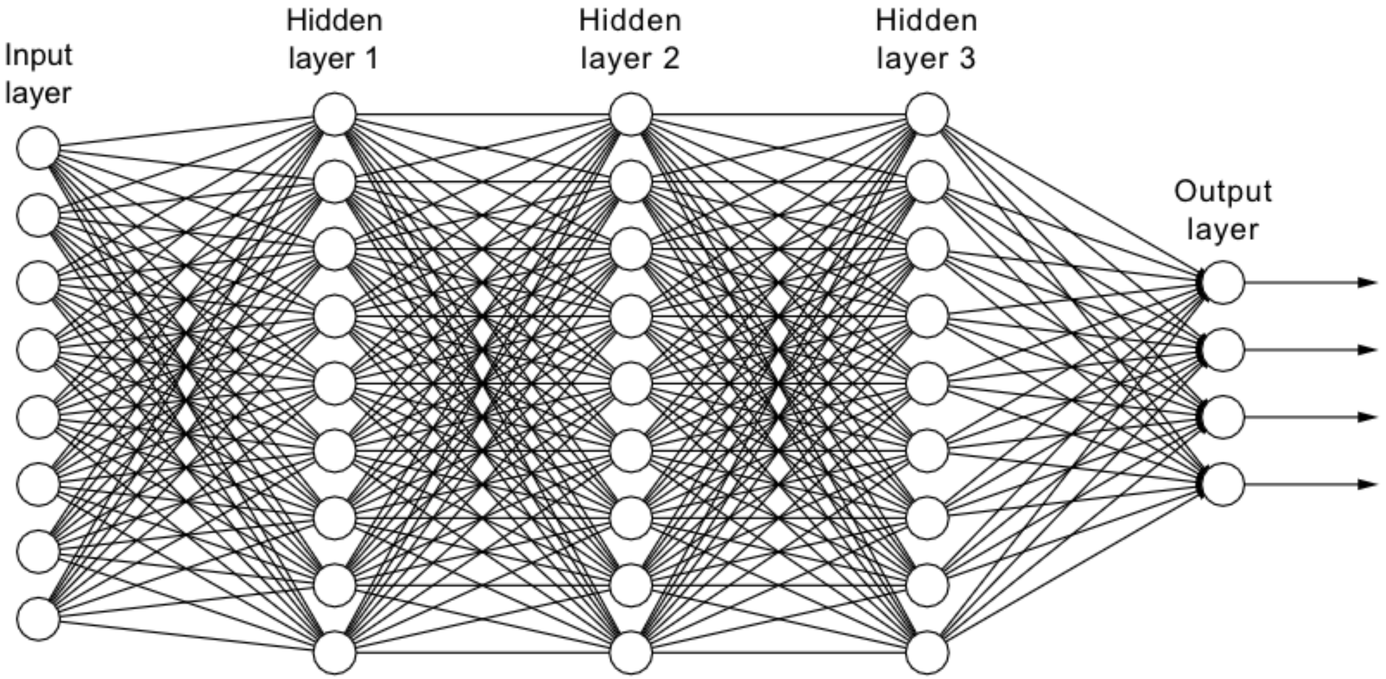
$P(x', \theta)$



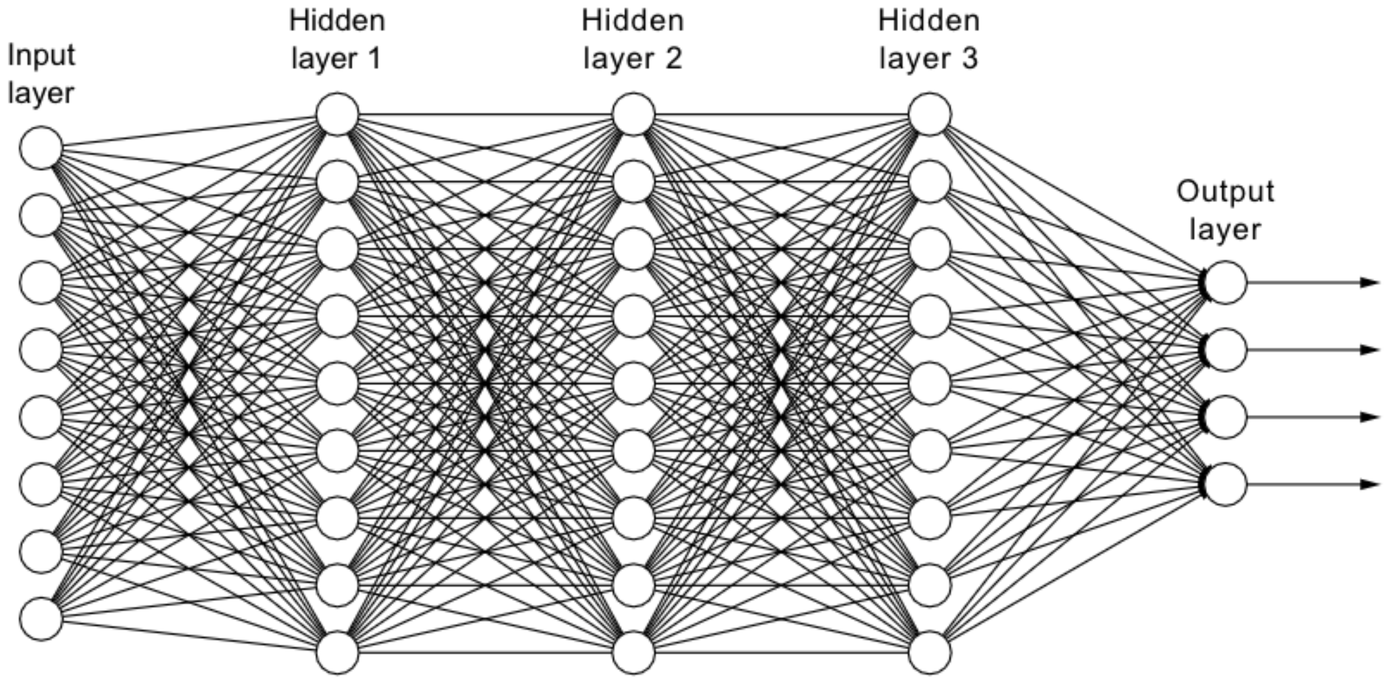
Outline

- How is playing chess and writing text curve fitting?
- What do these functions look like?
- Why don’t we overfit or get stuck in a local minimum?

Neural Network



Neural Network



$$h^1 = \max(0, W^1 x + b^1)$$

$$h^2 = \max(0, W^2 h^1 + b^2)$$

$$h^3 = \max(0, W^3 h^2 + b^3)$$

$$h^4 = W^4 h^3 + b^4$$

$$p_i = \frac{e^{h_i^4}}{\sum_j e^{h_j^4}}$$

Transformers

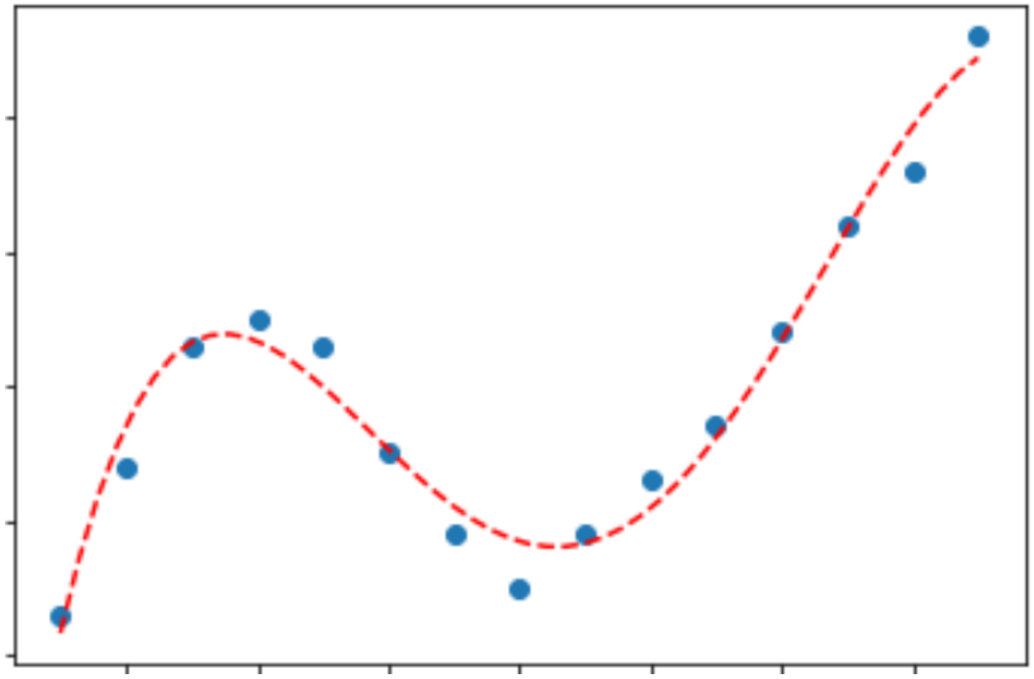
$$[h_1, h_2, h_3, \cdots, h_{512}] \qquad h_i \in \mathbb{R}^N$$

$$h_i' = V\alpha_{ij}h_j \qquad \alpha_{ij} = \frac{e^{h_i M h_j}}{\sum_k e^{h_i M h_k}}$$

Outline

- How is playing chess and writing text curve fitting?
- What do these functions look like?
- Why don’t we overfit or get stuck in a local minimum?

Loss



$$\mathrm{Loss}(\theta) = -\sum_{\alpha=1}^{N_{\mathrm{data}}} \left(y_\alpha - p(x_\alpha, \theta)\right)^2$$

Multiple class loss

$$\text{Data} = \{(x_\alpha, y_\alpha)\}$$

$$x_\alpha \in \mathbb{R}^{M \times N}$$

$$y_\alpha \in \mathbb{R}^{M' \times N'}$$ one-hot

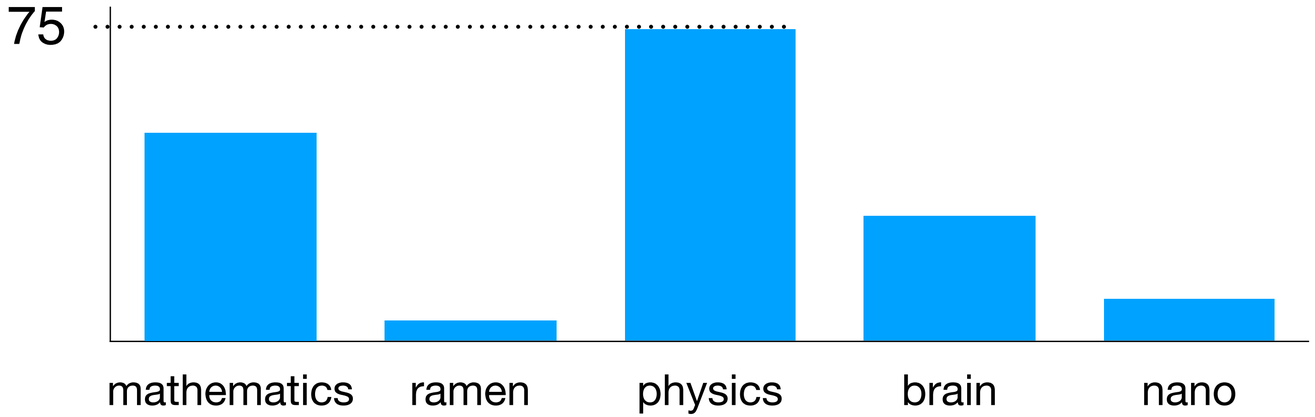
$$\text{label}_\alpha =$$ index of non-zero element

Multiple class loss

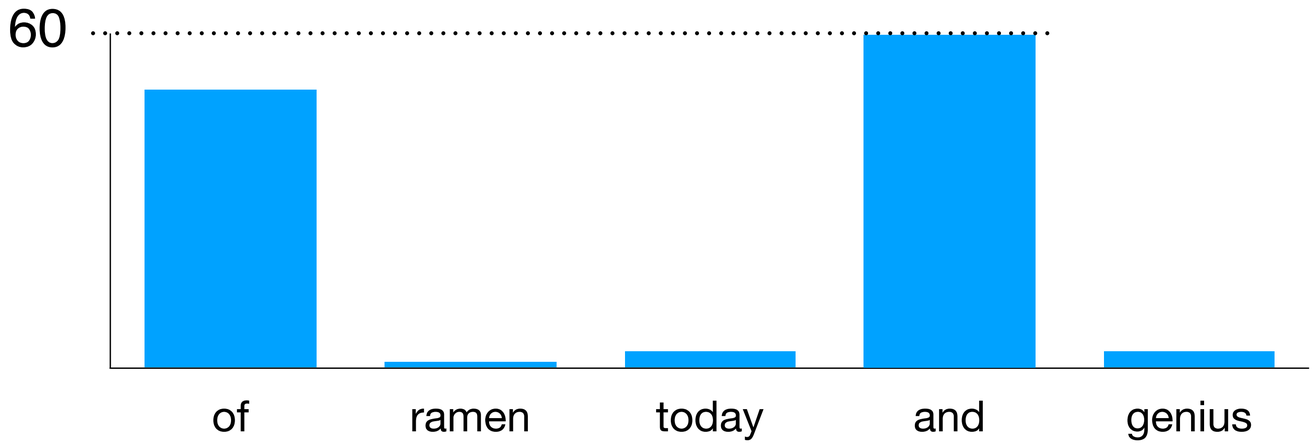
$$\mathrm{Loss}(\theta) = -\sum_{\alpha=1}^{N_{\mathrm{data}}} \ln p_{\mathrm{label}_\alpha}(x_\alpha, \theta)$$

Example

The Kavli Institute for the



The Kavli Institute for the physics

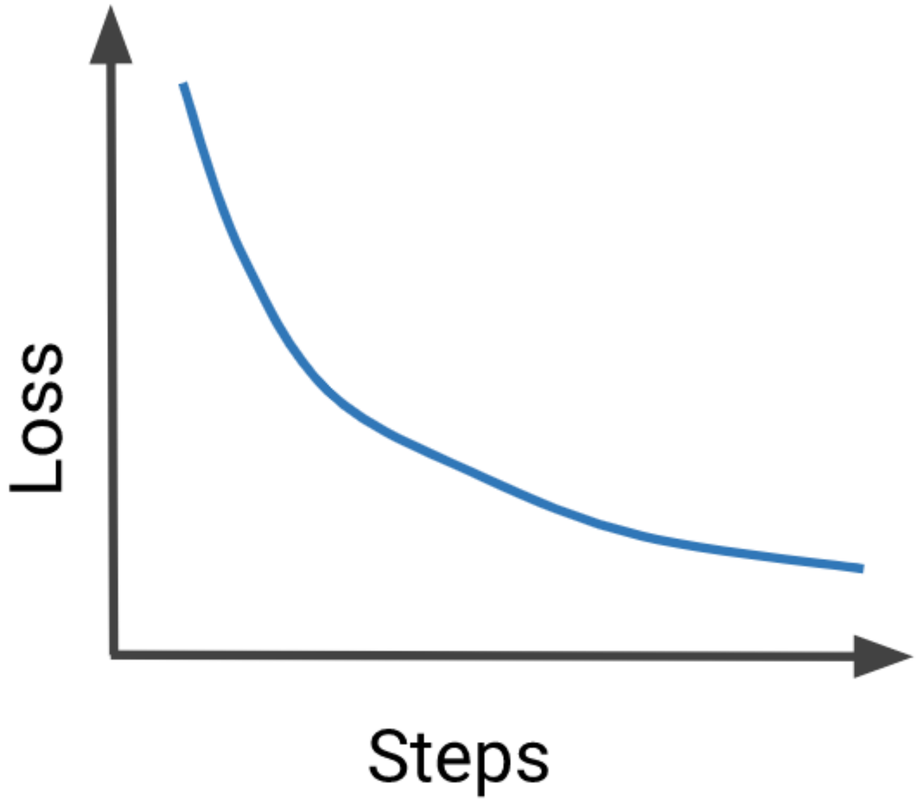


$$\text{Loss}(\theta) = -\log 0.75 - \log 0.6 - \cdots$$

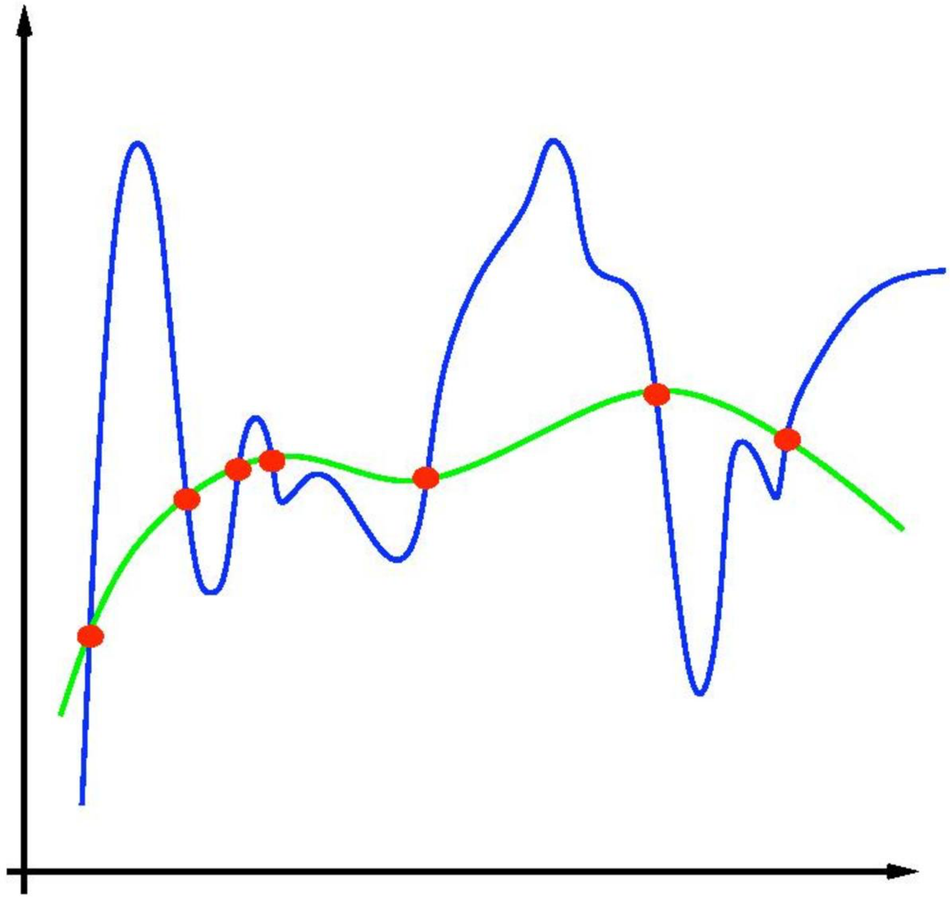
Gradient Descent

$$\theta = \theta - \epsilon \partial_\theta \text{Loss}$$

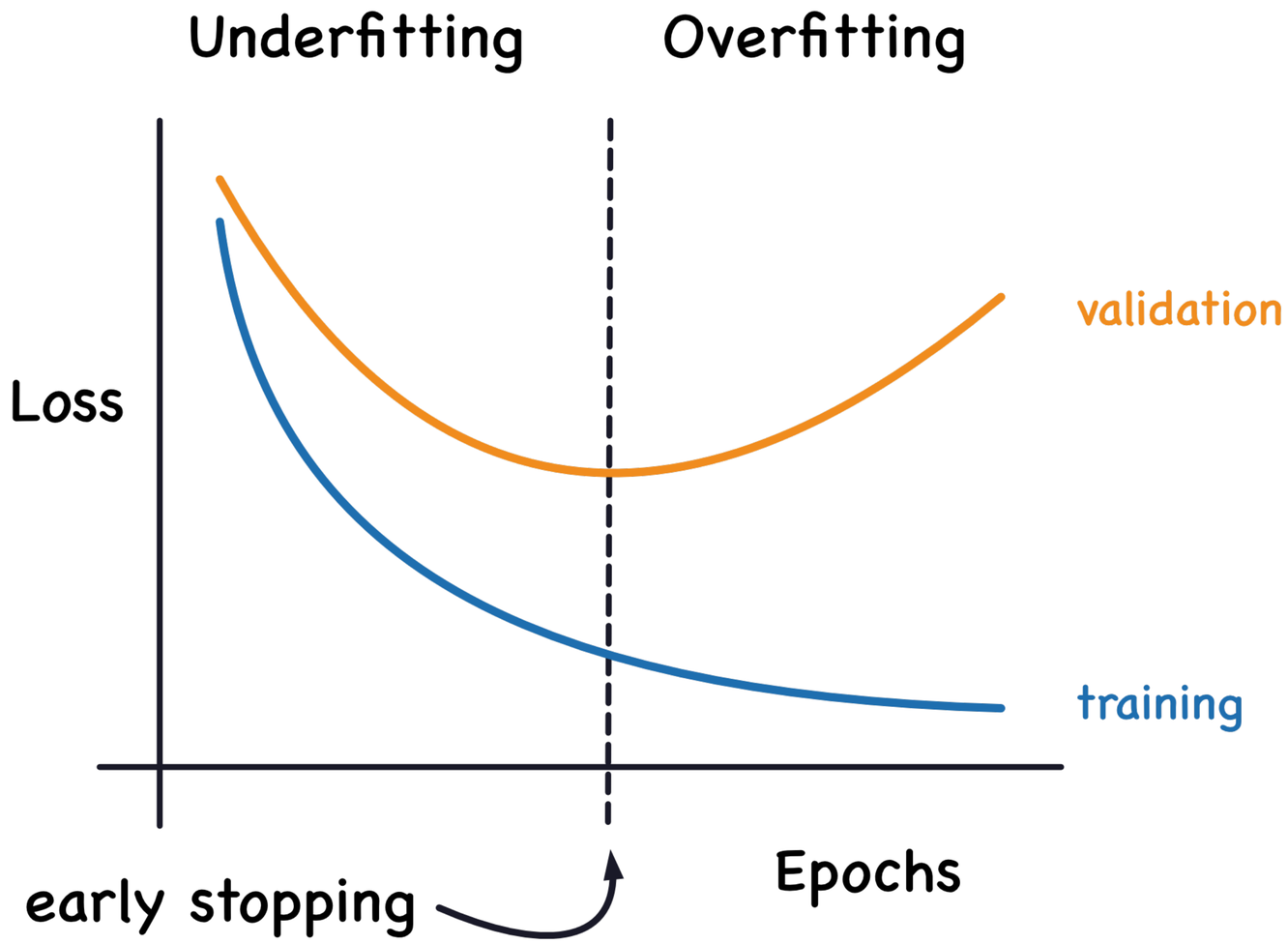
Loss decreases during learning (optimization)



Overfitting



Preventing Overfitting



Local Minimum?

$$\theta = \theta - \epsilon \partial_\theta \text{Loss}$$

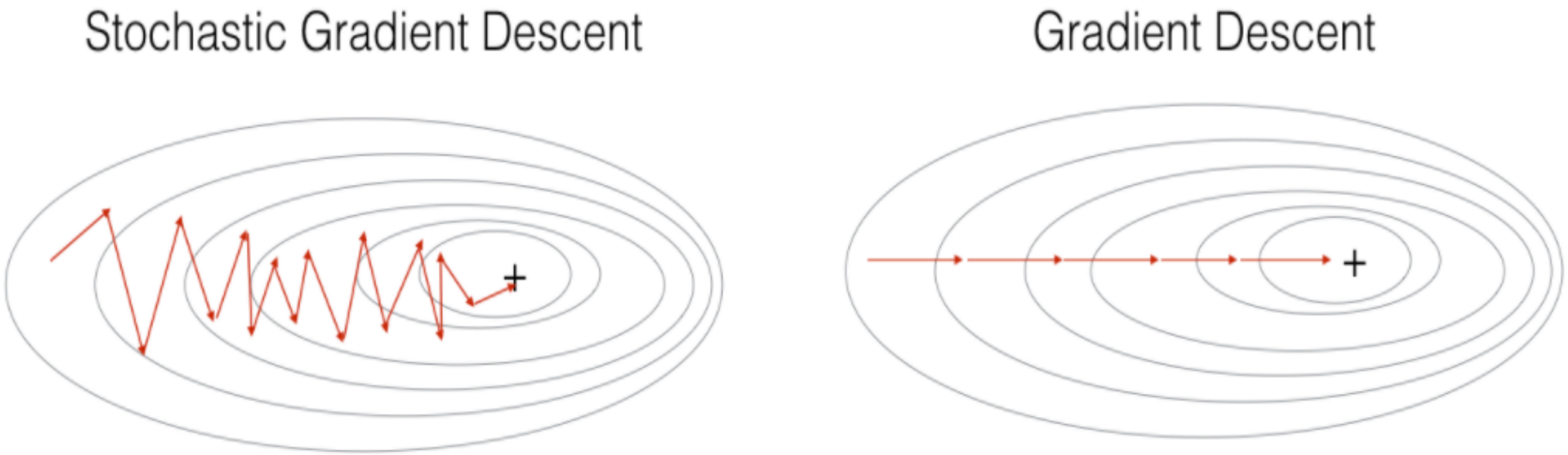
$$= \theta + \epsilon \partial_\theta \sum_{\alpha=1}^{N_{\text{data}}} \ln p_{\text{label}_\alpha}$$

Stochastic Gradient Descent

$$\theta = \theta - \epsilon \partial_\theta \text{Loss}$$

$$= \theta + \epsilon \partial_\theta \sum_{\alpha=1}^{N_{\text{batch}}} \ln p_{\text{label}_\alpha}$$

Stochastic Gradient Descent

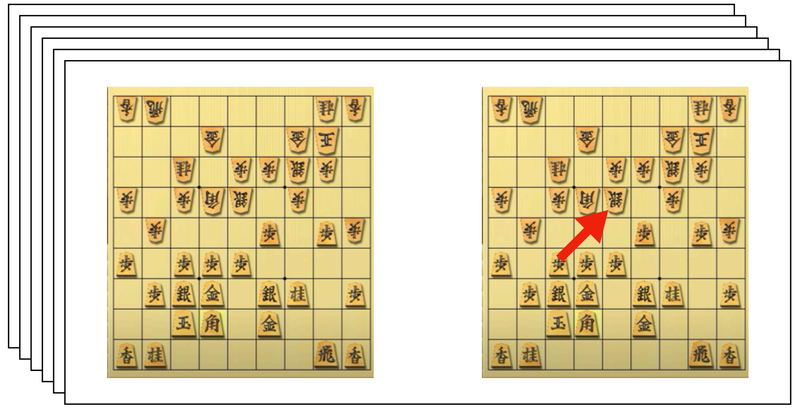


Outline

- How is playing chess and writing text curve fitting?
- What do these functions look like?
- Why don’t we overfit or get stuck in a local minimum?

Reinforcement learning 強化学習

Self play 自分と対戦 →



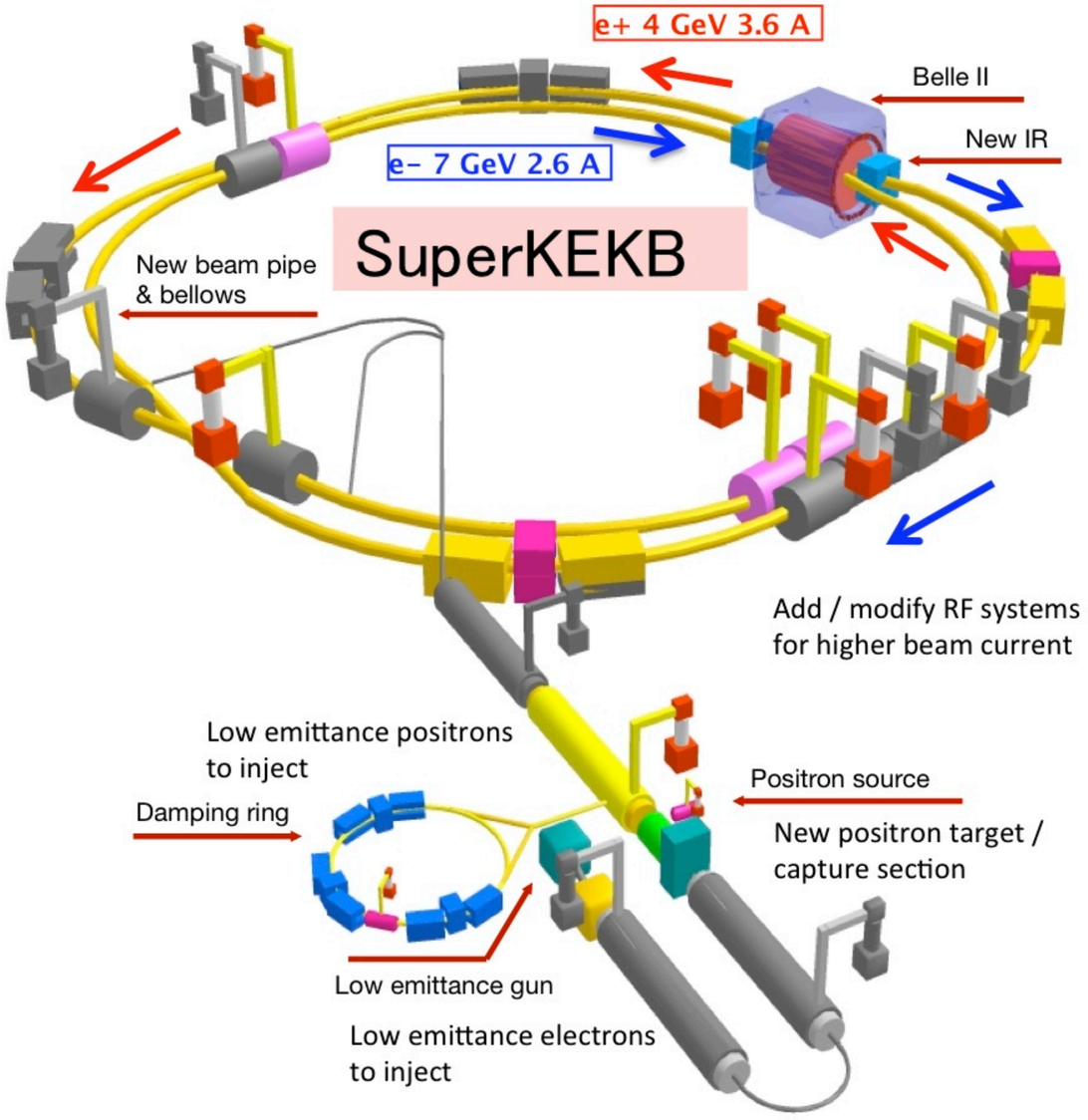
→ Learn 学び

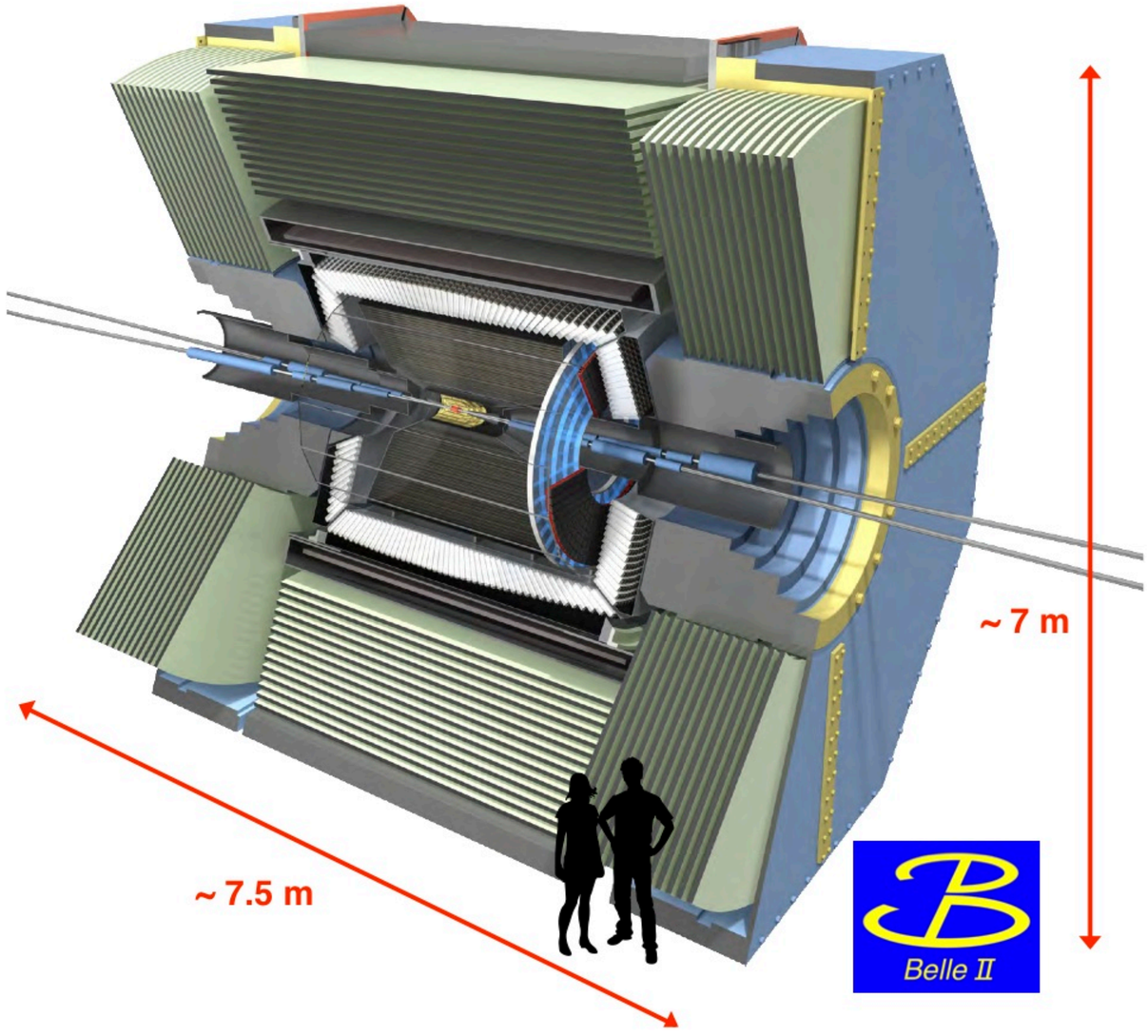
(Position, move made by winning player)
(駒の位置、勝った棋士の次の一手)

Repeat 繰り返す

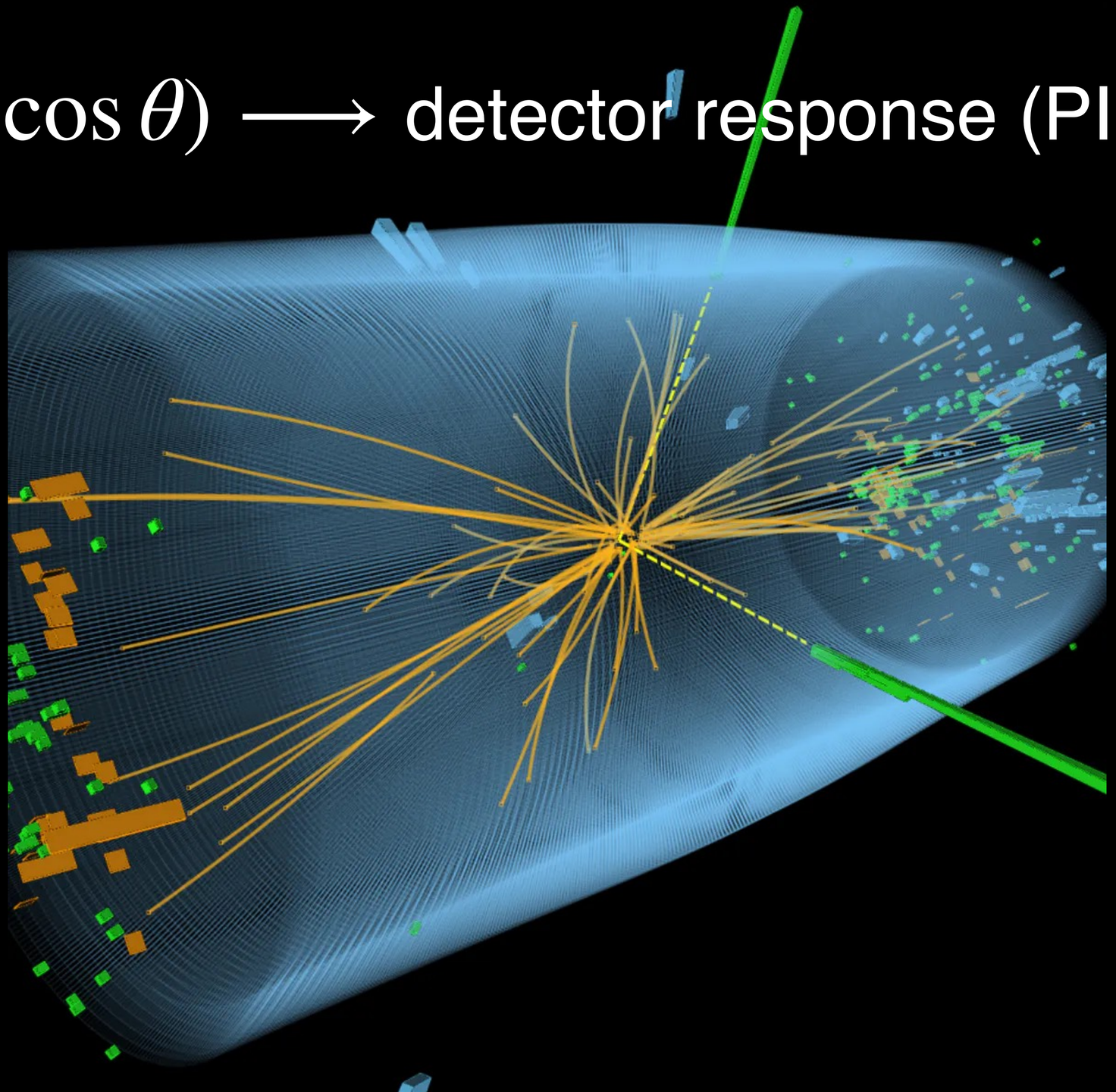
Curve fitting inside a loop

generate data → curve fit → generate data → curve fit → ⋯


e+ 4 GeV 3.6 A
Belle II
New IR
e− 7 GeV 2.6 A
SuperKEKB
New beam pipe
& bellows
Add / modify RF systems
for higher beam current
Low emittance positrons
to inject
Positron source
Damping ring
New positron target /
capture section
Low emittance gun
Low emittance electrons
to inject

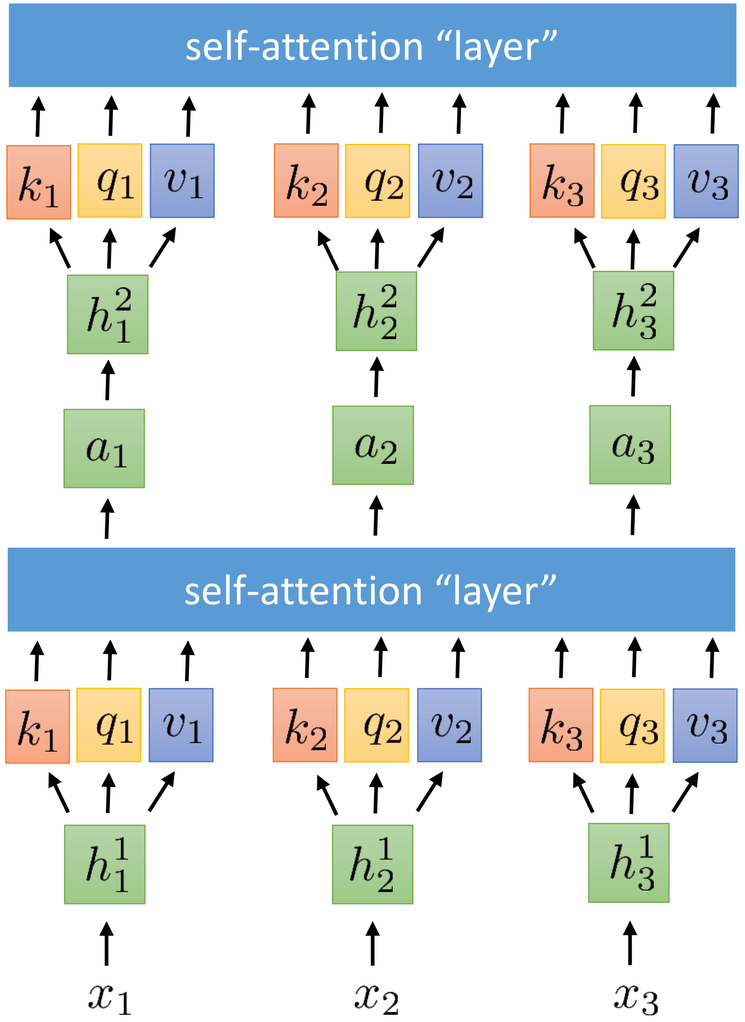

~ 7 m
~ 7.5 m
Belle II

(particle, p, $\cos\theta$) $\longrightarrow$ detector response (PID variables)



Appendix

Transformer



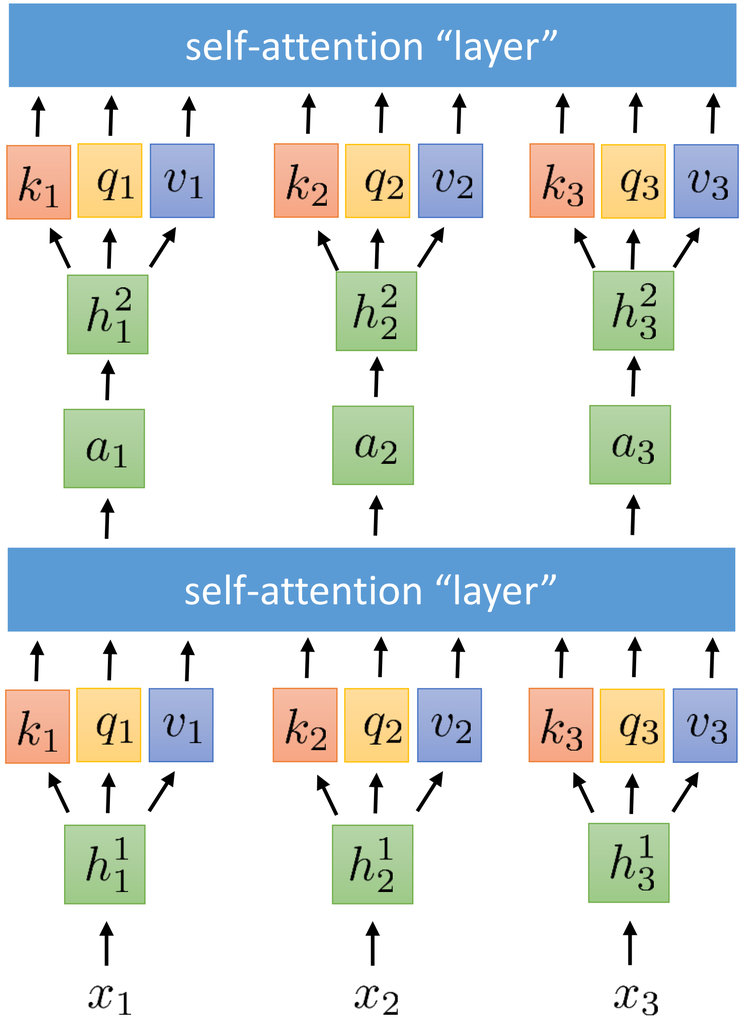
$$k_i = Kh_i$$

$$q_i = Qh_i$$

$$v_i = Vh_i$$

$$h_i^1 = Ex_i$$

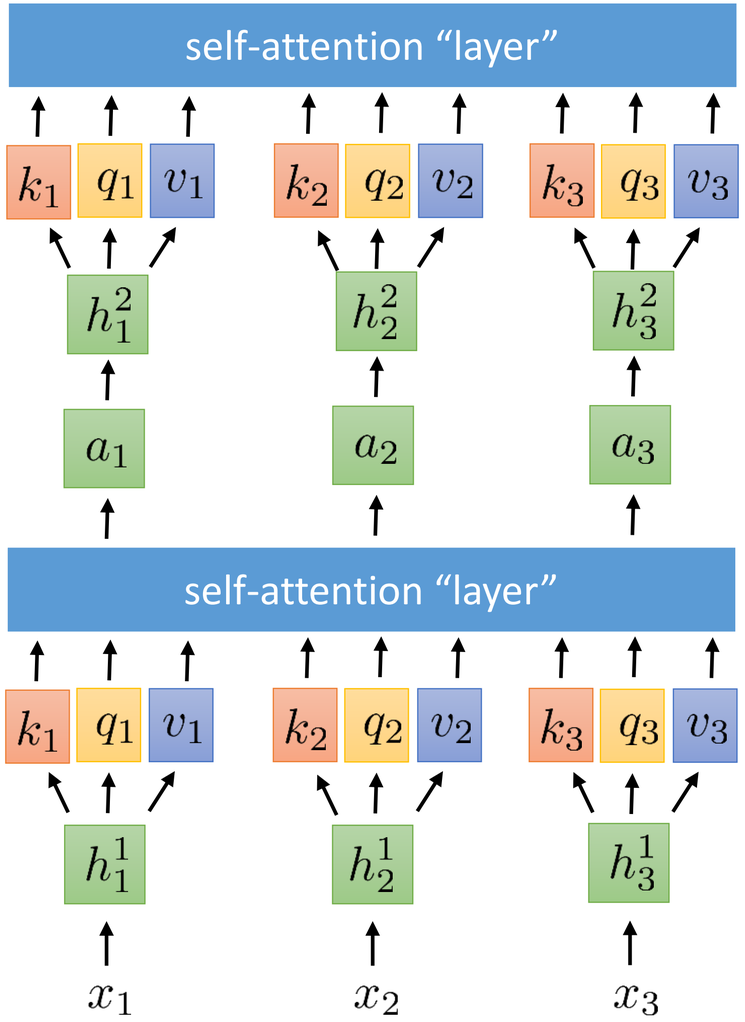
Transformer



$$a_i = A_{ij} h_j^1$$

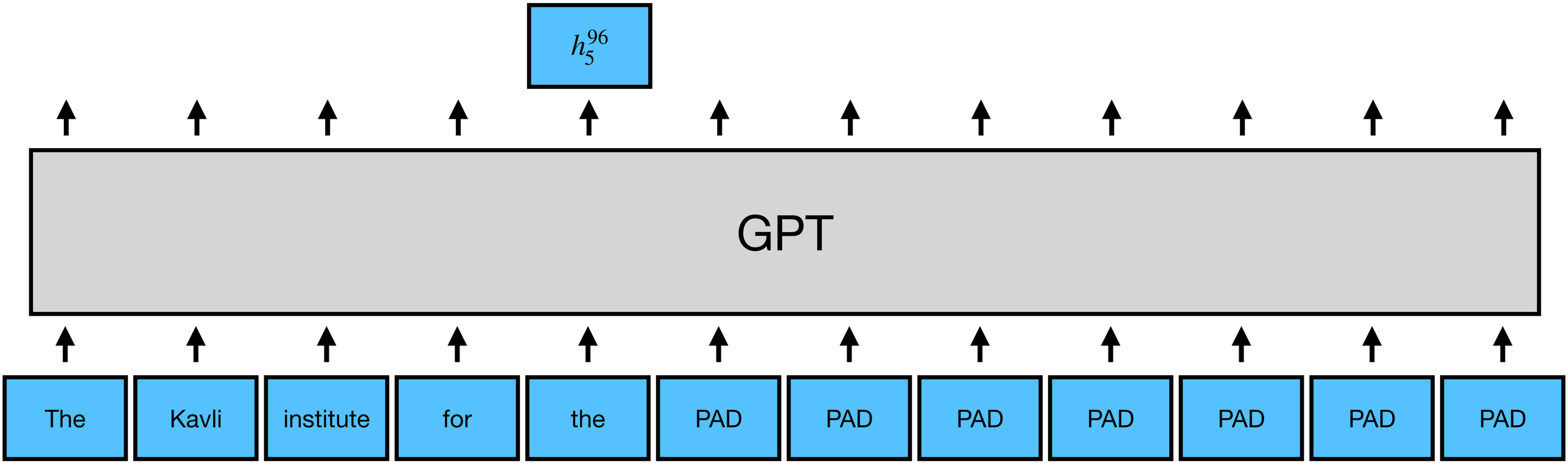
$$A_{ij} = \frac{e^{q_i \cdot k_j}}{\sum_l e^{q_i \cdot k_l}}$$

Transformer

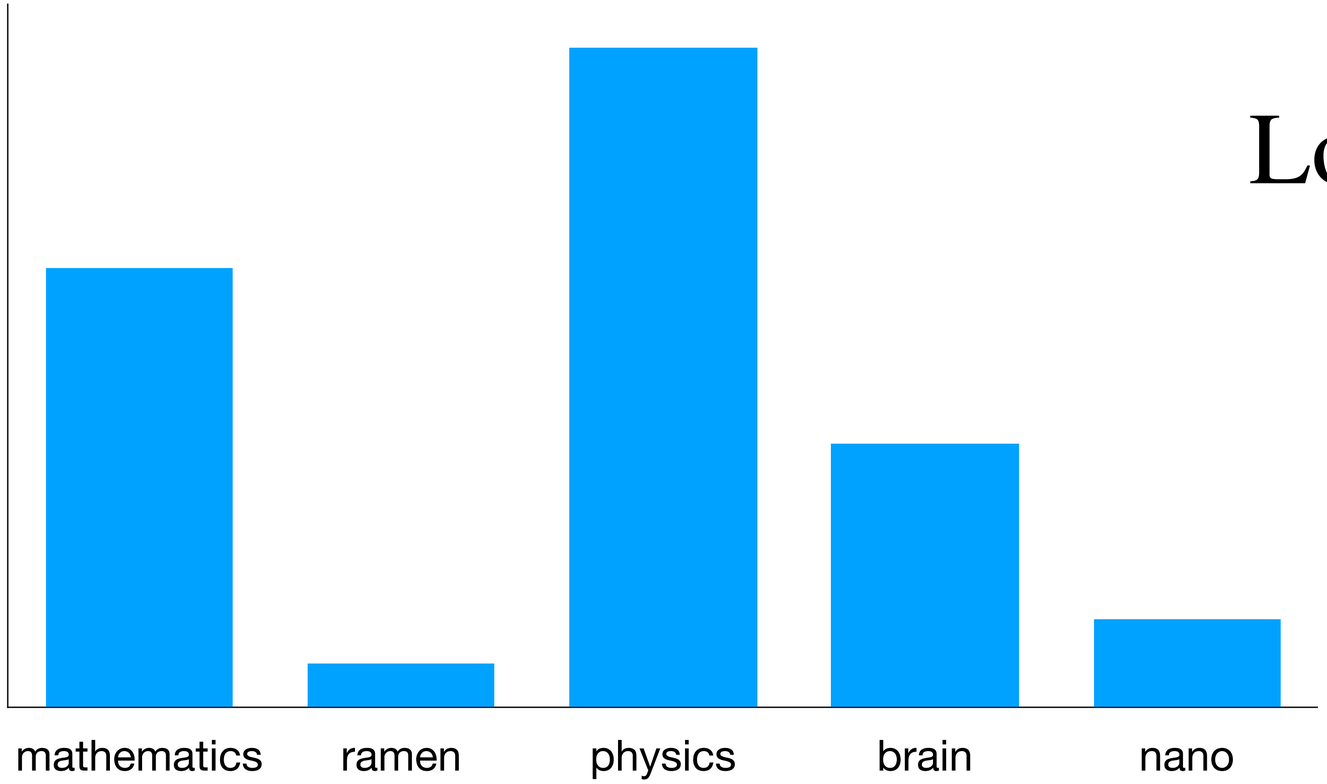


$$h_i^2 = Wa_i$$

Training (next word prediction)

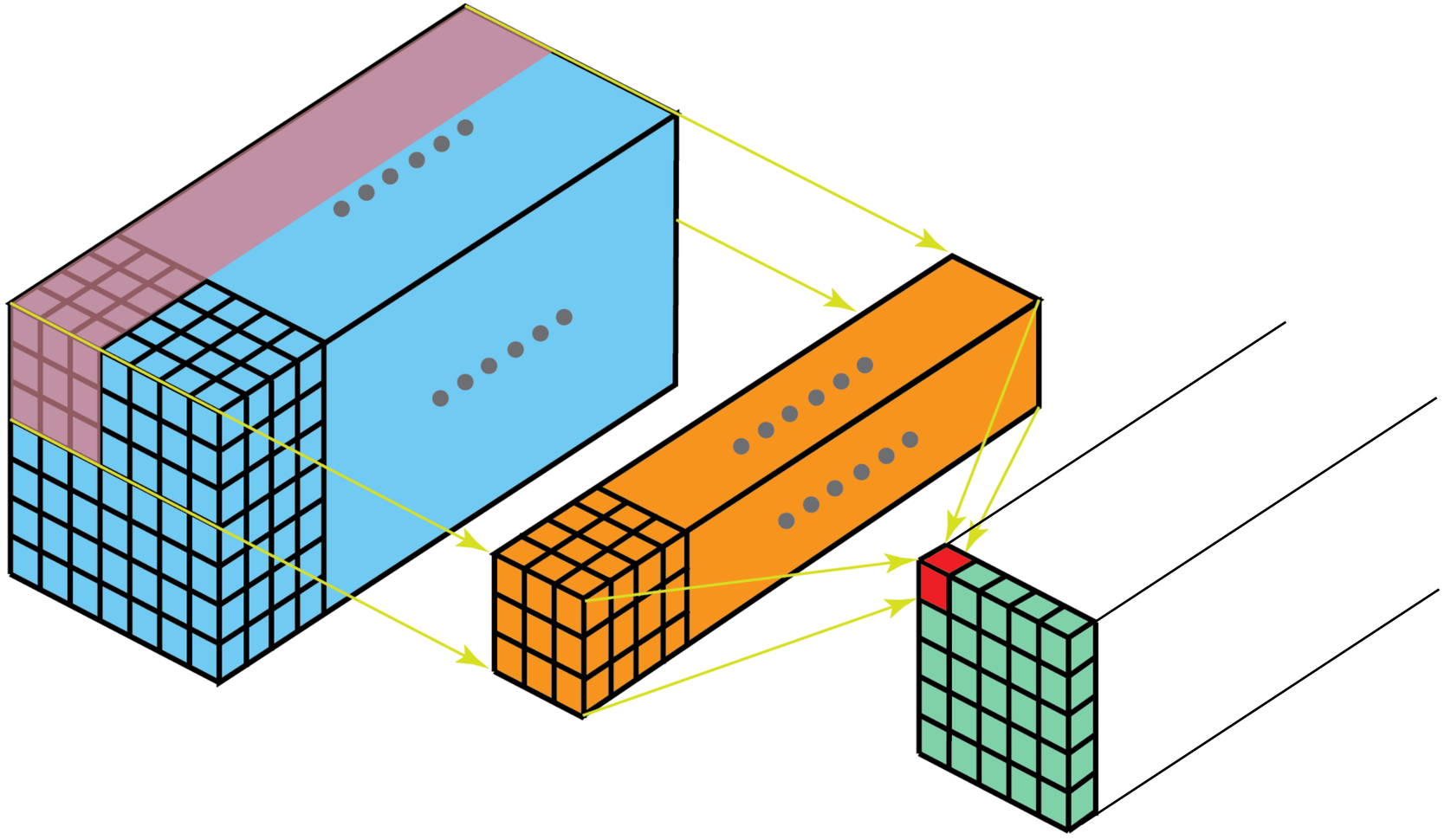


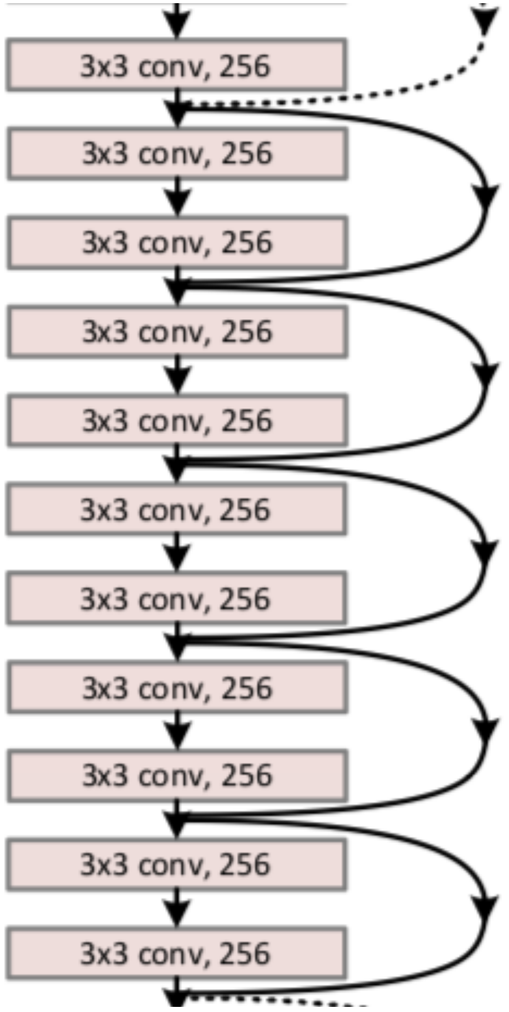
$$P_a = \frac{e^{x'_{5a}}}{\sum_b e^{x'_{5b}}} \qquad x'_5 = E^T h_5^{96}$$



$$\text{Loss} = \text{Loss} - \ln P_{i_{\text{Physics}}}$$

Convolutional Neural Networks





Thank you