

# Machine Learning for Molecules

## Lessons and Challenges of Data-Centric Chemistry

---

Feb 5th, 2022

Ichigaku Takigawa

<https://itakigawa.github.io/>

RIKEN Center for Advanced Intelligence Project (AIP)

Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University





TAKIGAWA Ichigaku  
瀧川 一学

<https://itakigawa.github.io>

# Hi, I am a machine-learning researcher

- 1995-2004 Hokkaido Univ (Grad School. Engineering)  
2004 PhD Computer Science
- 2005-2011 Kyoto Univ (Inst. Chemical Research)  
Bioinformatics Center, Assist. Prof.  
Grad School Pharmaceutical Sciences, Assit. Prof.
- 2012-2018 Hokkaido Univ (Grad School. Info Sci & Tech)  
Large-Scale Knowledge Processing Lab, Assoc. Prof.  
2015-2018 JST PRESTO for Materials Informatics
- 2019- RIKEN Center for AI Project @ ATR Kyoto  
2019- Hokkaido Univ  
(Inst. Chemical Reaction Design & Discovery)

RIKEN AIP Kyoto, Kyoto Univ CiRA, Nikon  
jointly working on stem cell biology.



TAKIGAWA Ichigaku  
瀧川 一学

<https://itakigawa.github.io>

## But also, I am a machine-learning user

- 1995-2004 Hokkaido Univ (Grad School. Engineering)  
2004 PhD Computer Science
- 2005-2011 Kyoto Univ (Inst. Chemical Research)  
Bioinformatics Center, Assist. Prof.  
Grad School Pharmaceutical Sciences, Assit. Prof.
- 2012-2018 Hokkaido Univ (Grad School. Info Sci & Tech)  
Large-Scale Knowledge Processing Lab, Assoc. Prof.  
2015-2018 JST PRESTO for Materials Informatics
- 2019- RIKEN Center for AI Project @ ATR Kyoto  
2019- Hokkaido Univ  
(Inst. Chemical Reaction Design & Discovery)

RIKEN AIP Kyoto, Kyoto Univ CiRA, Nikon  
jointly working on stem cell biology.

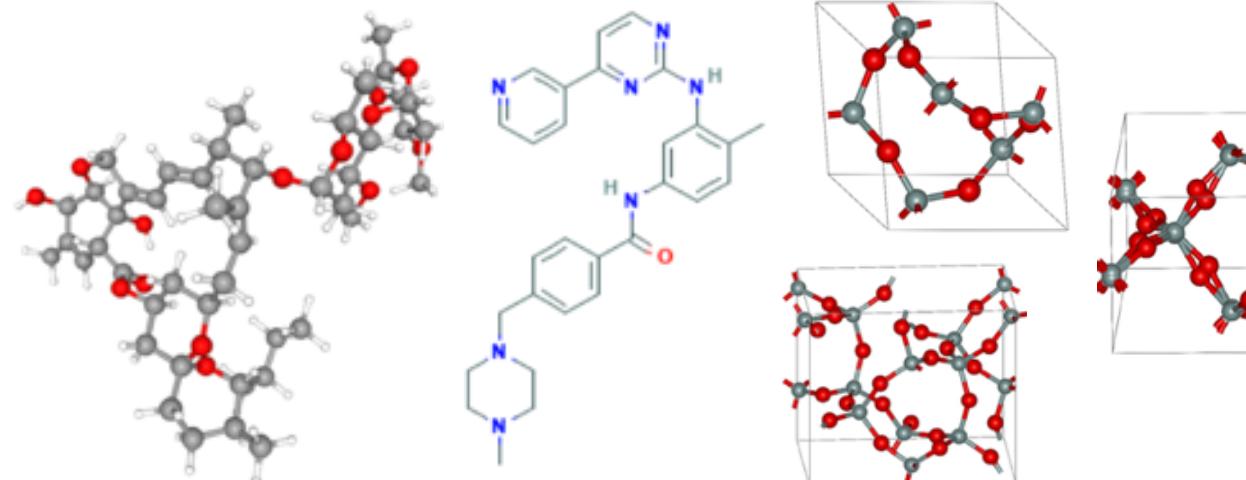
# Inst. Chemical Reaction Design and Discovery



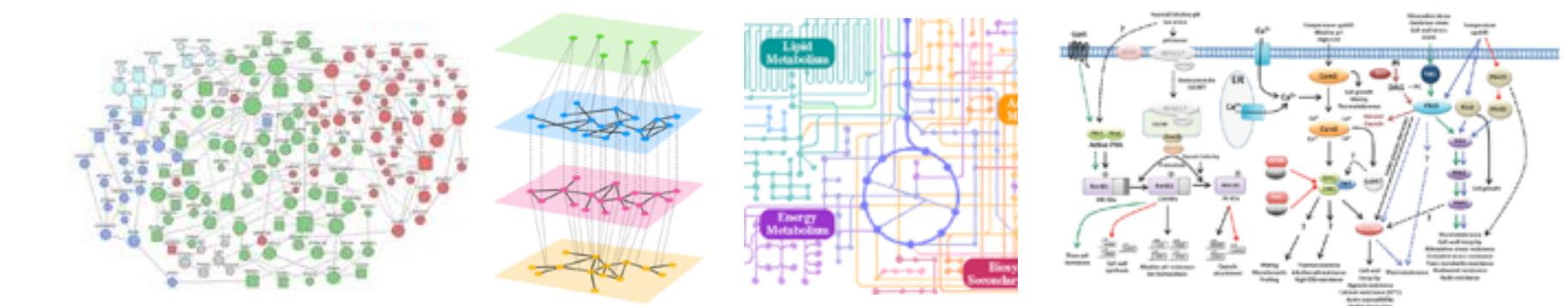
# My interest: Machine learning with discrete structures

*Discrete structures*, i.e., combinatorial or algebraic structures  
such as sets, groups, permutations, combinations, sequences, trees, and graphs

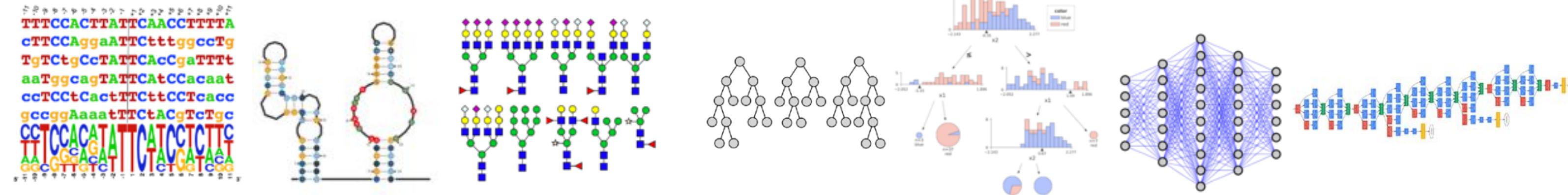
## Target objects



## Relations between target objects



## ML models



# Molecules clearly have a combinatorial aspect

c&en

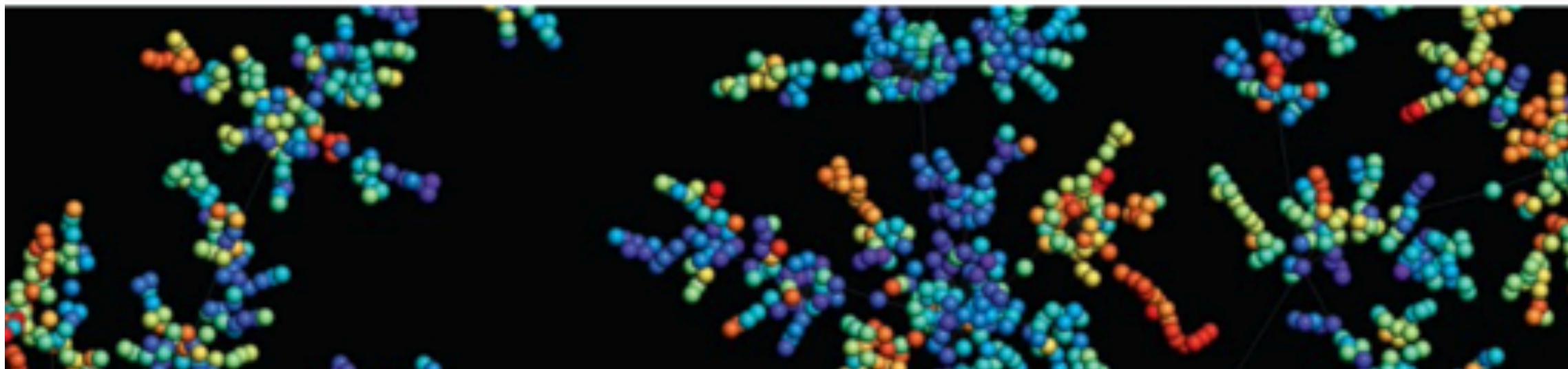
COMPUTATIONAL CHEMISTRY

## Exploring chemical space: Can AI take us where no human has gone before?

Artificial intelligence is helping us find novel, useful molecules. For the field to really take off, though, these tools will need to be accessible to the wider chemistry community

by Sam Lemonick

April 6, 2020 | A version of this story appeared in **Volume 98, Issue 13**



### BY THE NUMBERS

**$10^{180}$**

An upper estimate of the number of possible molecules

**$10^{80}$**

Estimated number of atoms in the universe

**$10^{60}$**

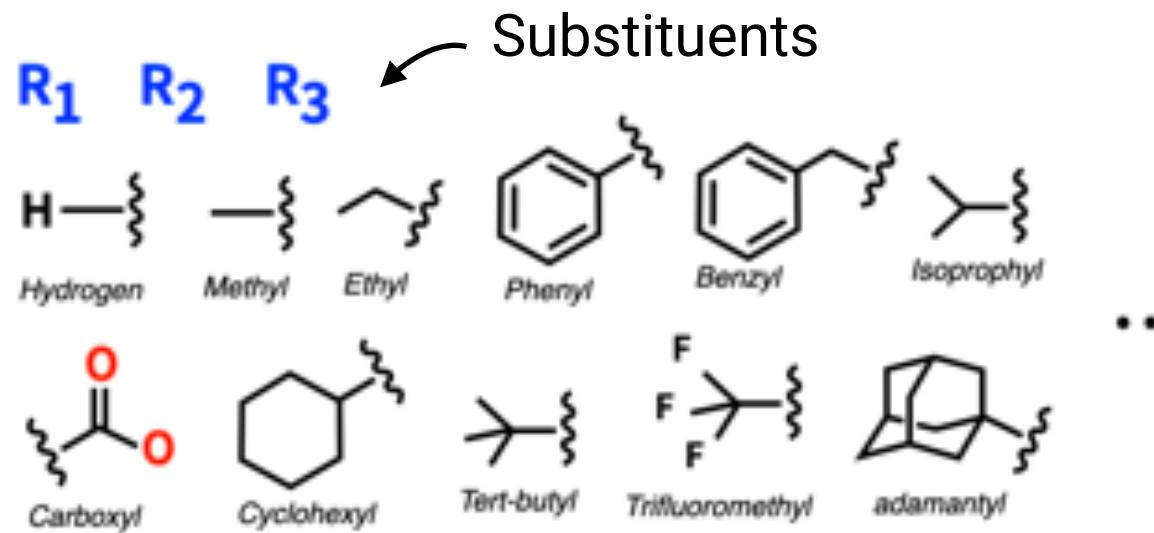
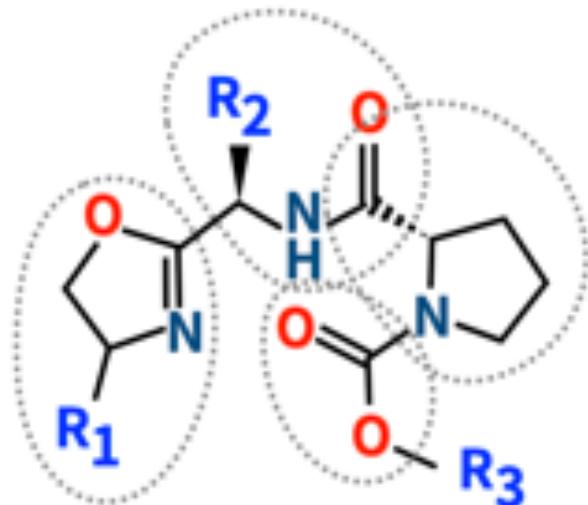
An estimate of the number of possible small organic molecules

**$10^8$**

The number of organic and inorganic substances in the CAS database

# How can ML leverage these combinatorial nature?

- A molecule is a set of atoms and bonds

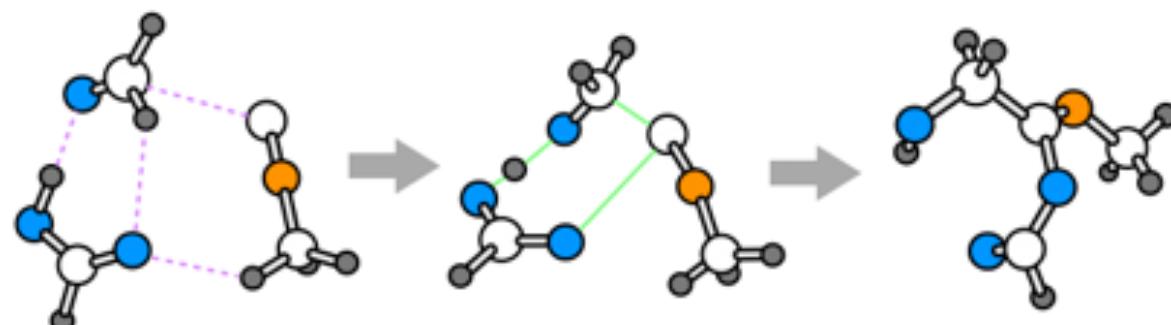


**Compositionality and hierarchy**

Similar to natural language?

We combine words to make any complicated sentences.

- A chemical reaction is a recombination pattern of bonds



The underlying rules are (largely) governed by **many-body quantum chemistry of electrons**

# This talk

**This slide is available at**  
<https://itakigawa.github.io/news.html>

---

A quick review on the **dark side** and **light side** of ML  
from both viewpoints as an ML algorithm researcher and an ML practitioner/user

- 1. What actually ML is?
- 2. The **dark side**: Modern aspects of ML
- 3. The **light side**: Deep learning for molecules

May the ML Force be with you...

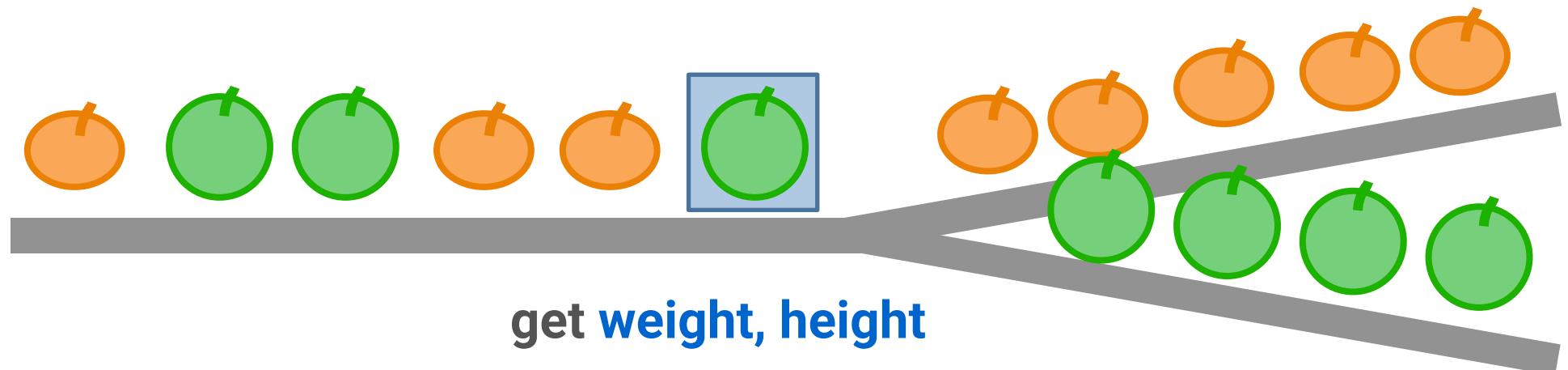
Science is built up of facts, as a house is built of stones;  
but an accumulation of facts is no more a science than  
a heap of stones is a house.

*Henri Poincaré "Science and hypothesis"*



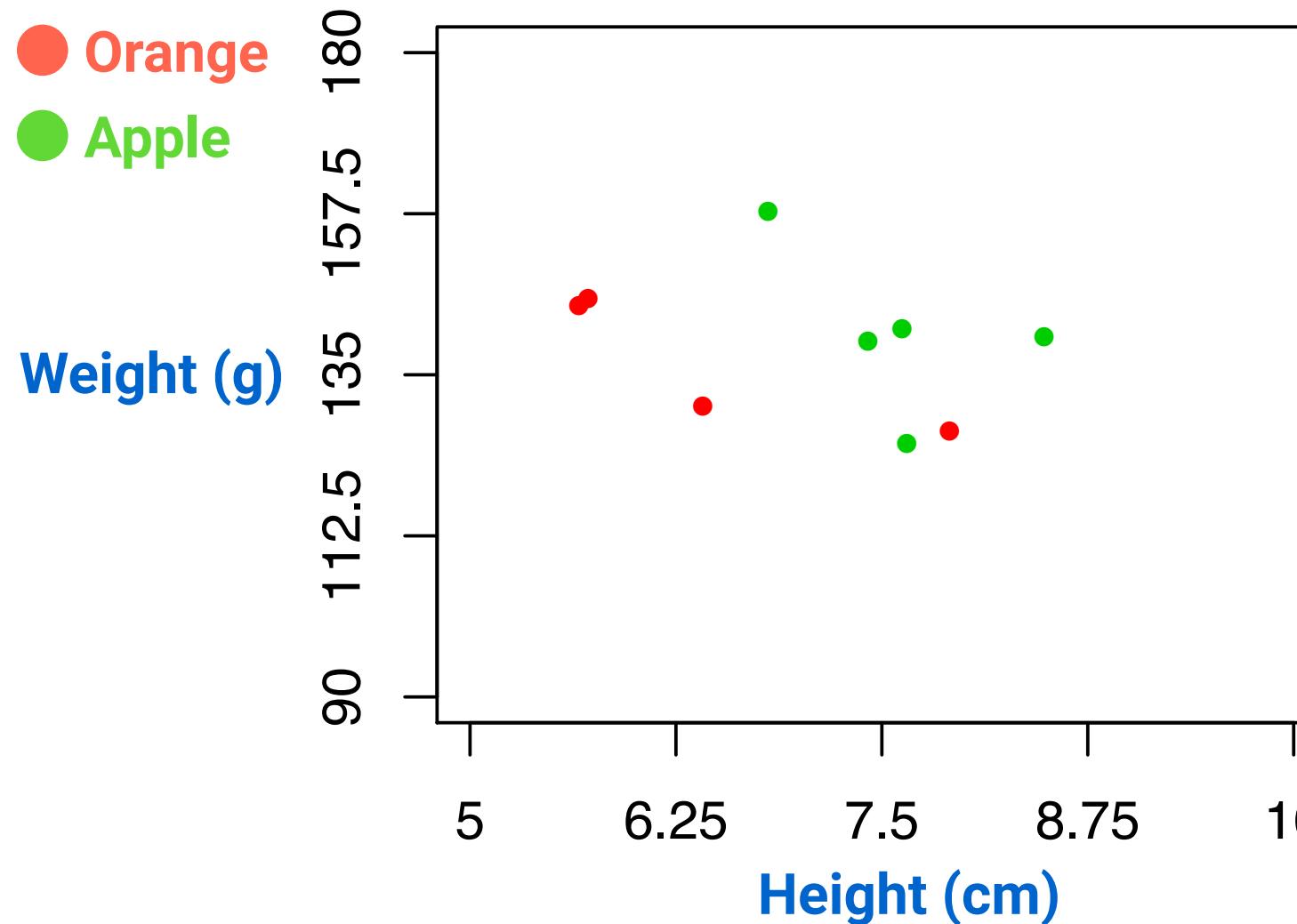
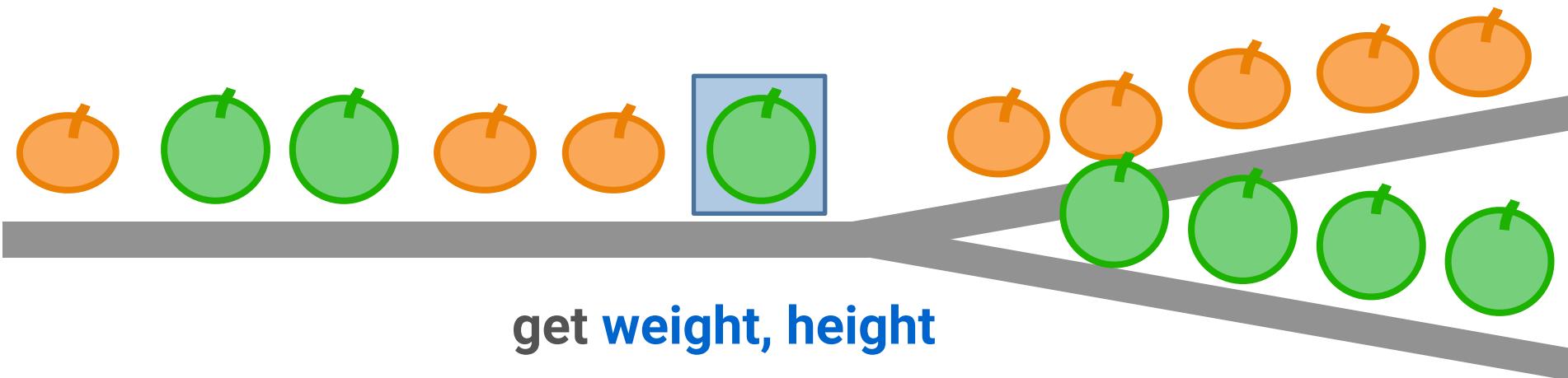
# ML converts data into "prediction"

---



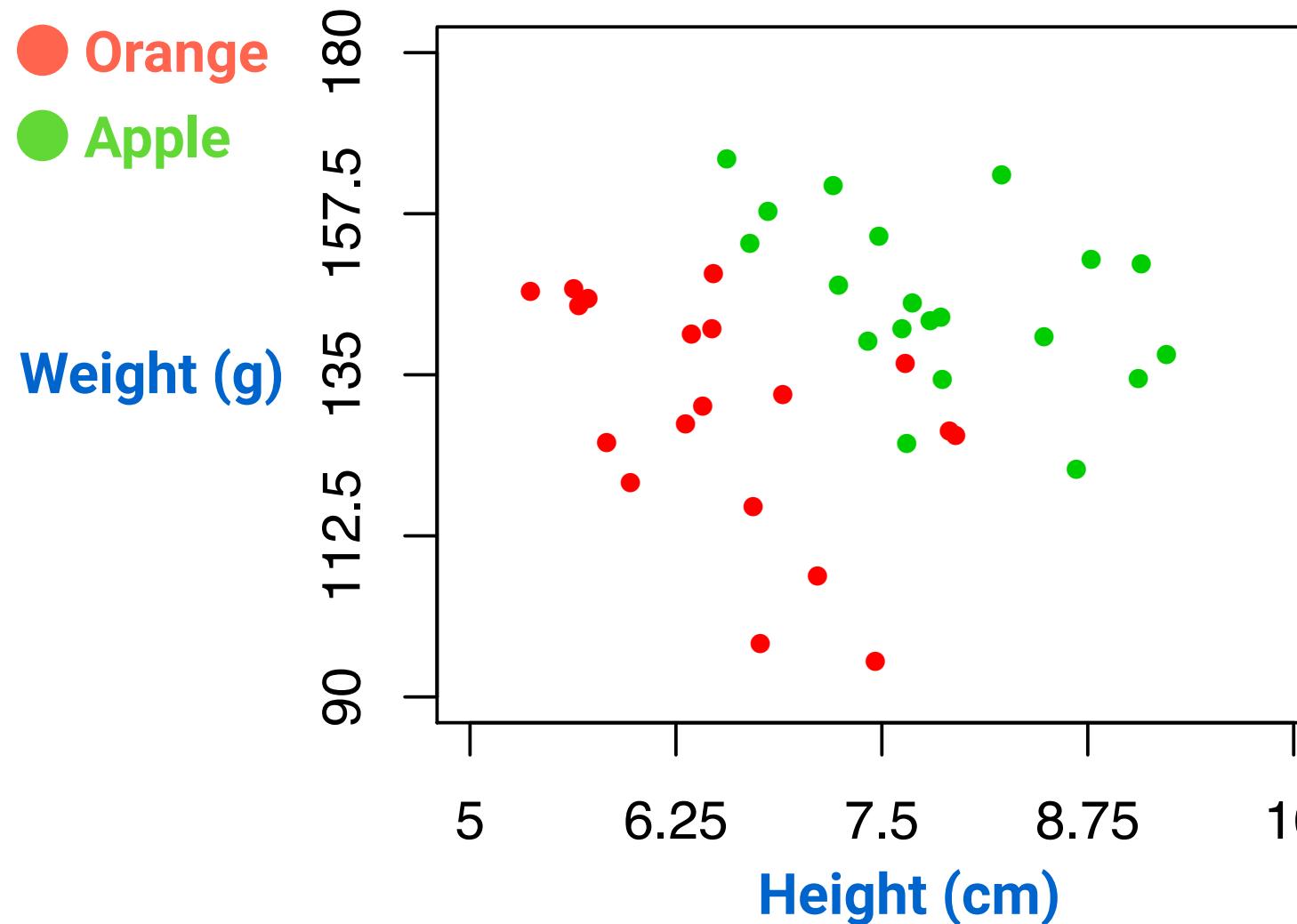
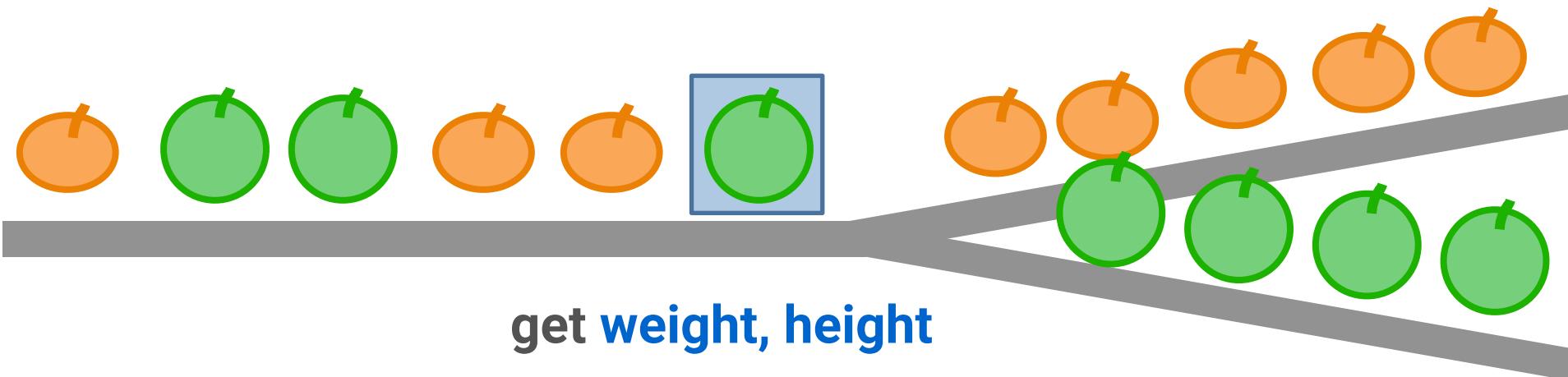
# ML converts data into "prediction"

---



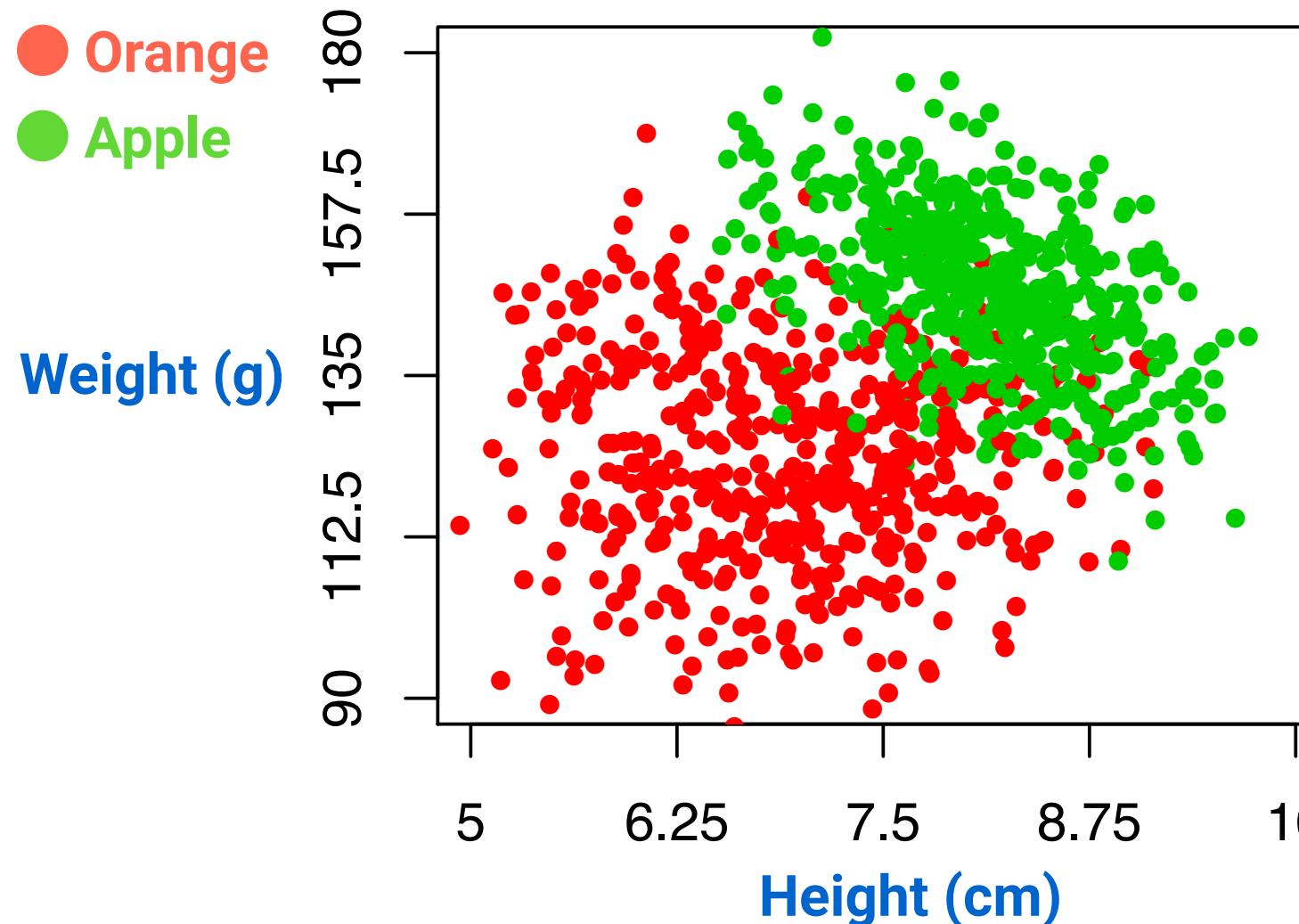
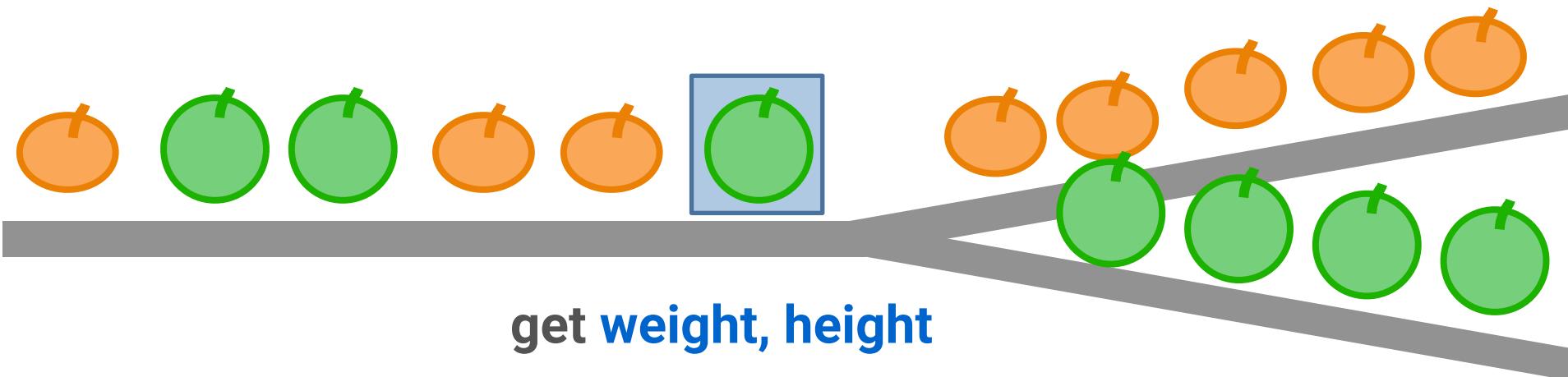
# ML converts data into "prediction"

---



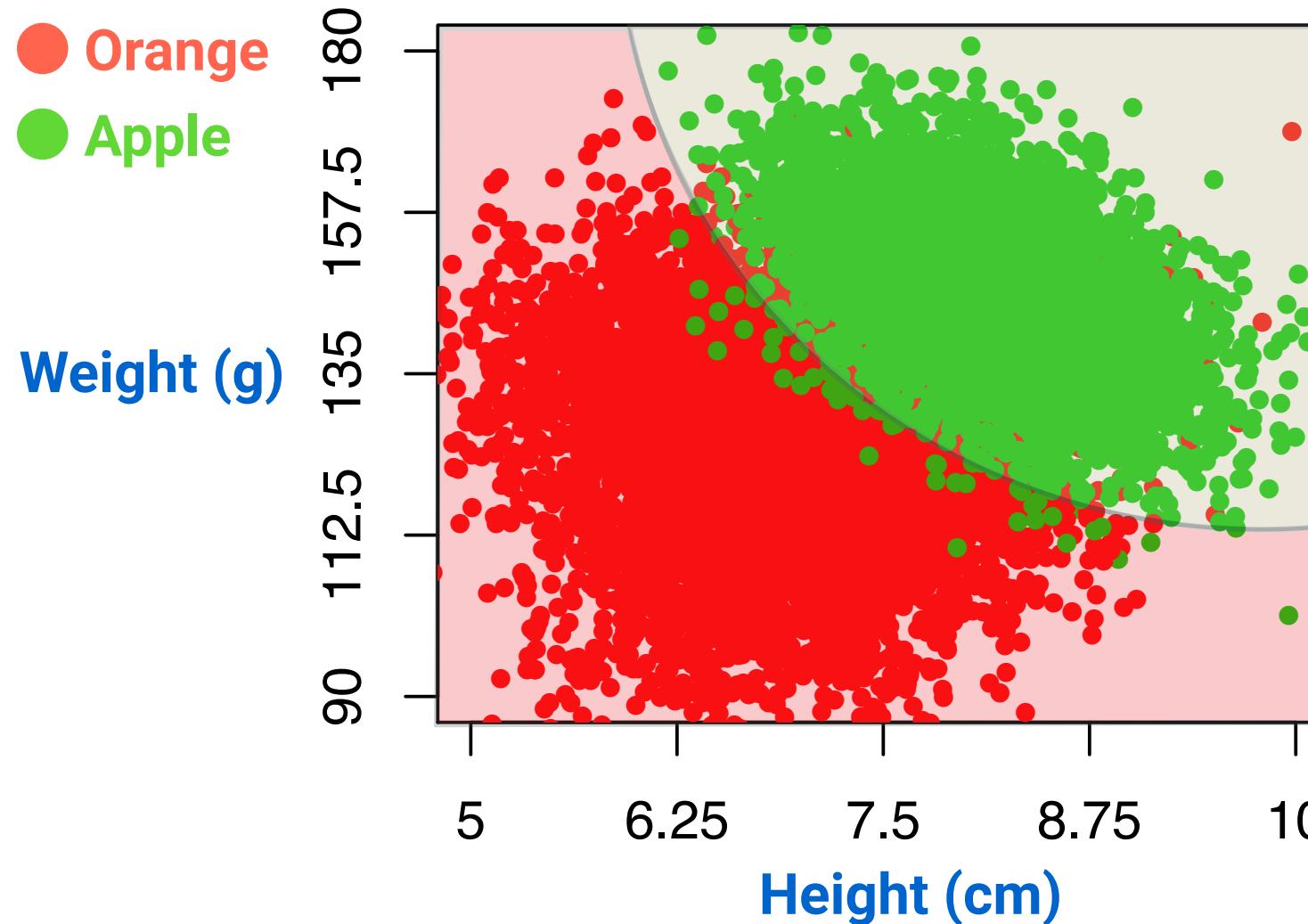
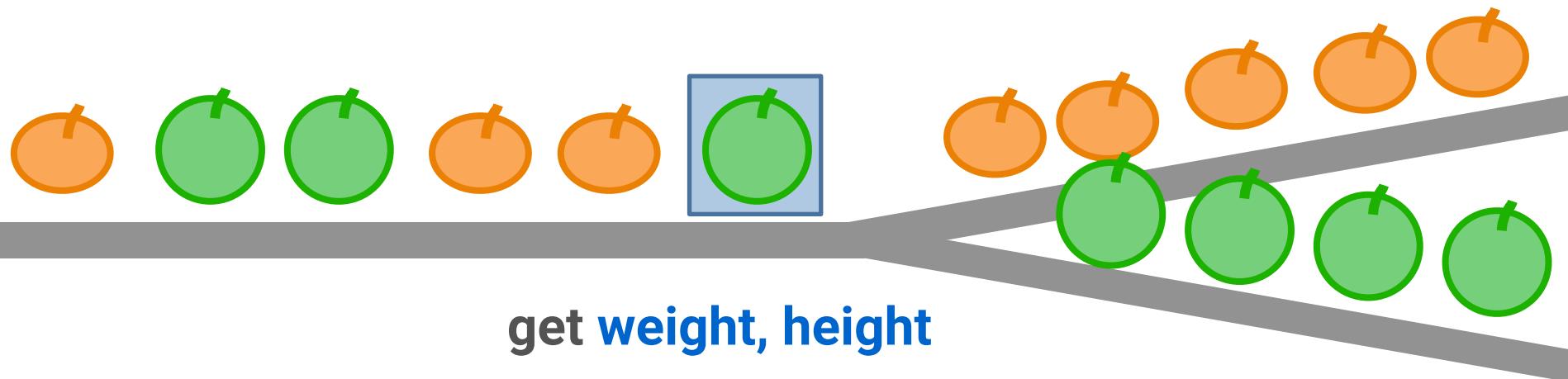
# ML converts data into "prediction"

---



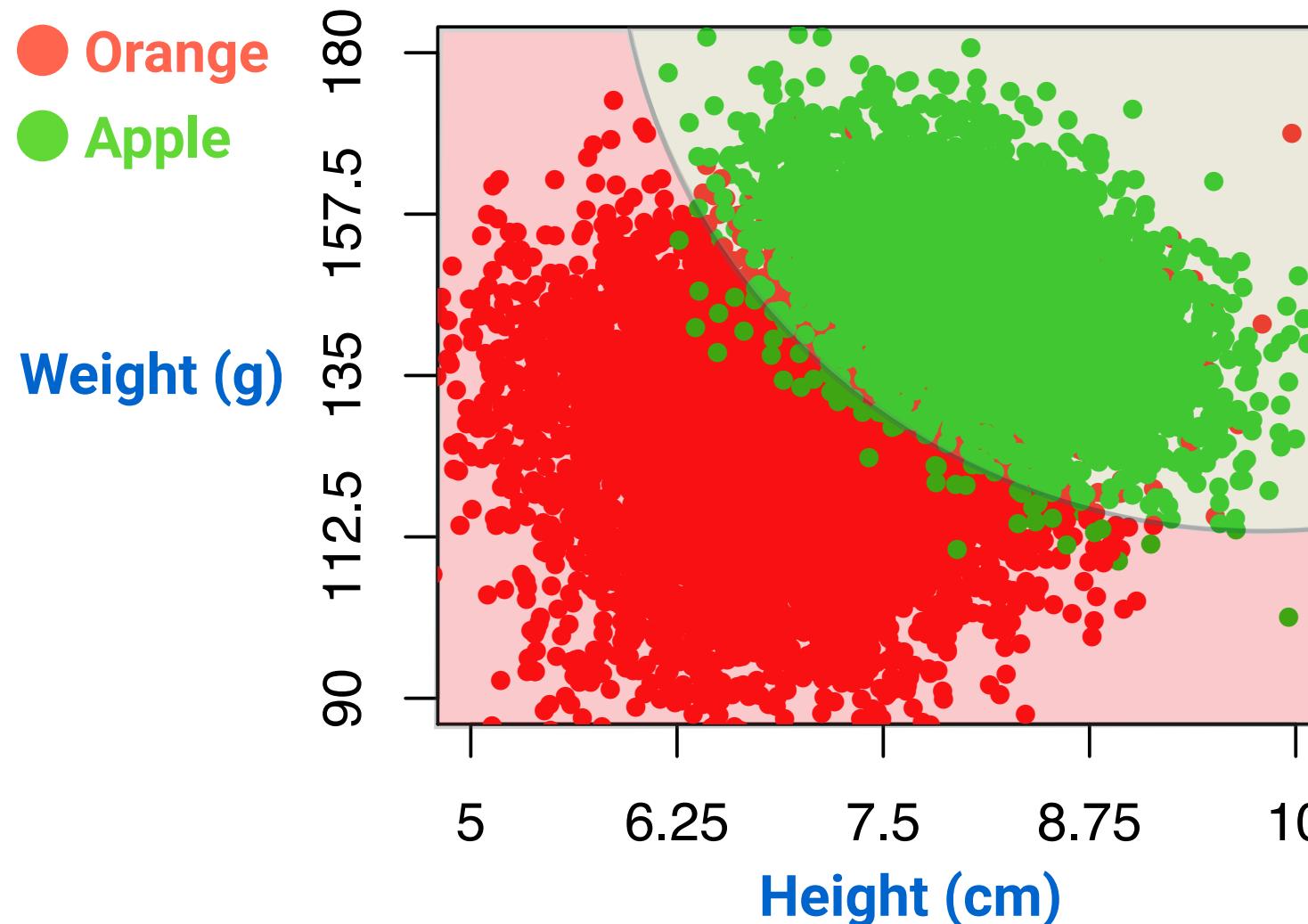
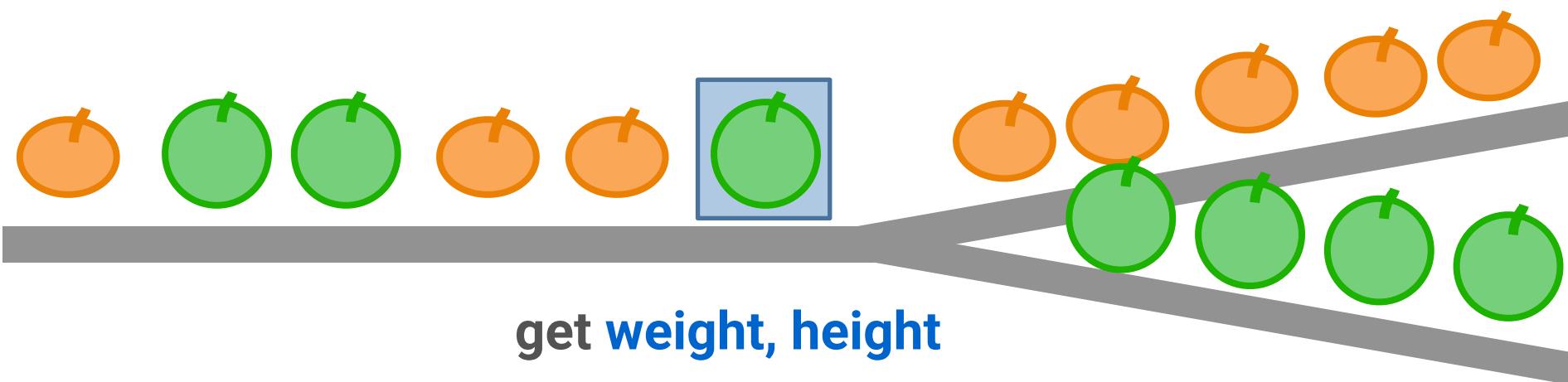
# ML converts data into "prediction"

---

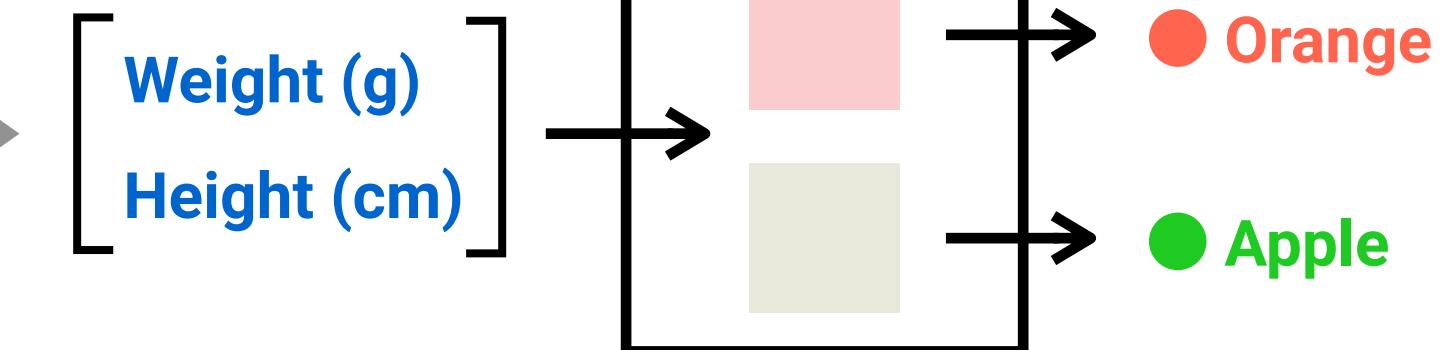


# ML converts data into "prediction"

---

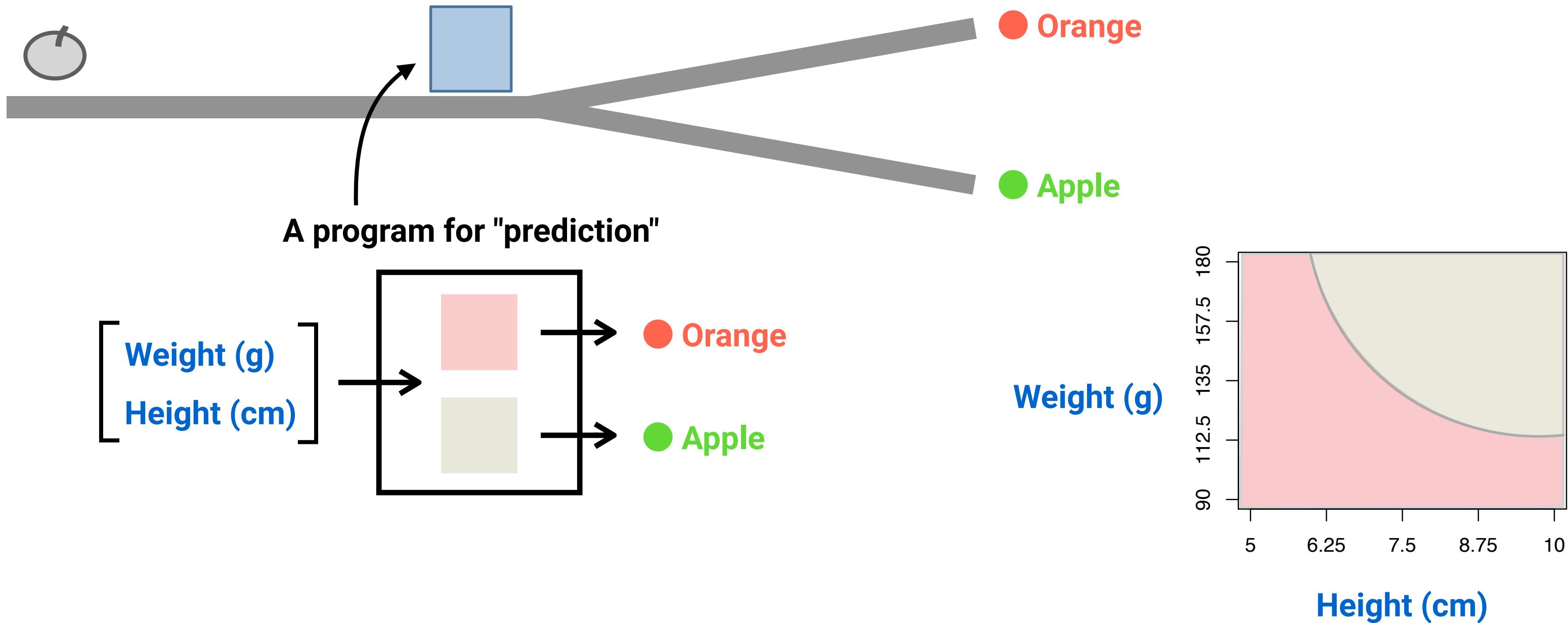


A program for "prediction"



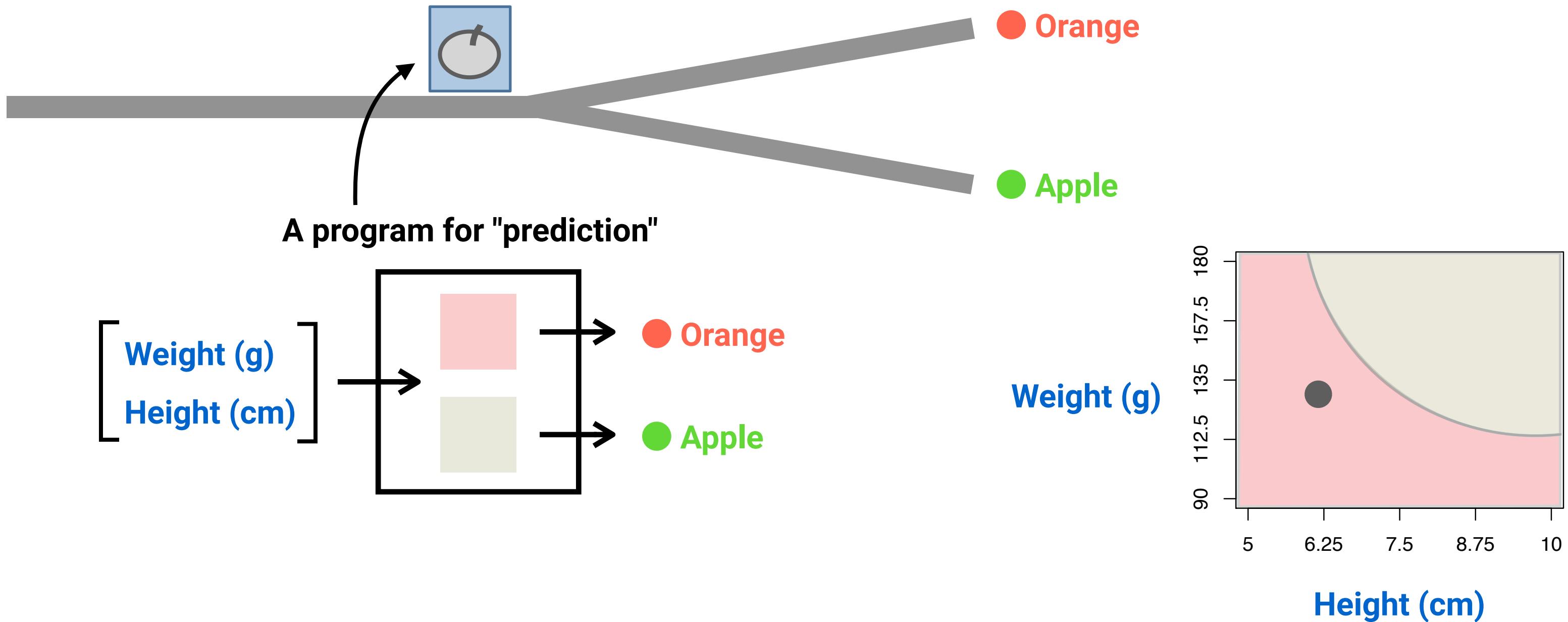
# ML converts data into "prediction"

---



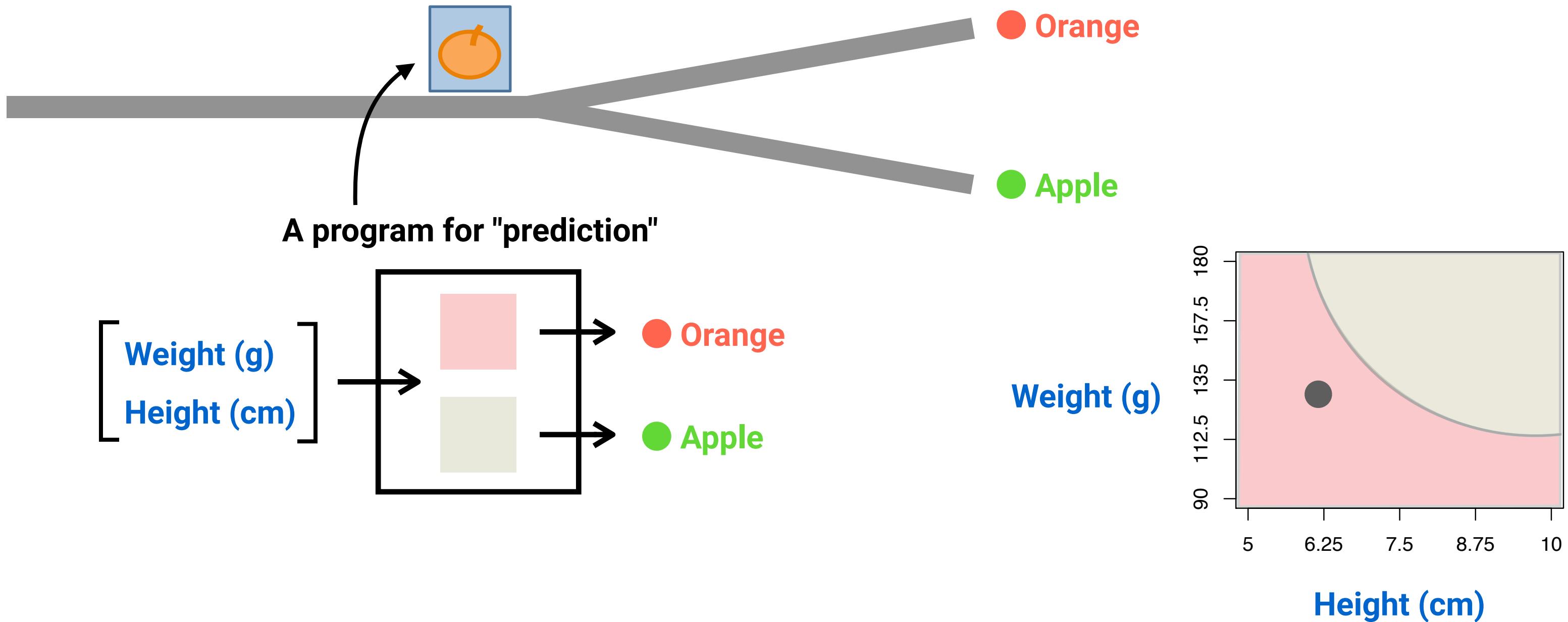
# ML converts data into "prediction"

---

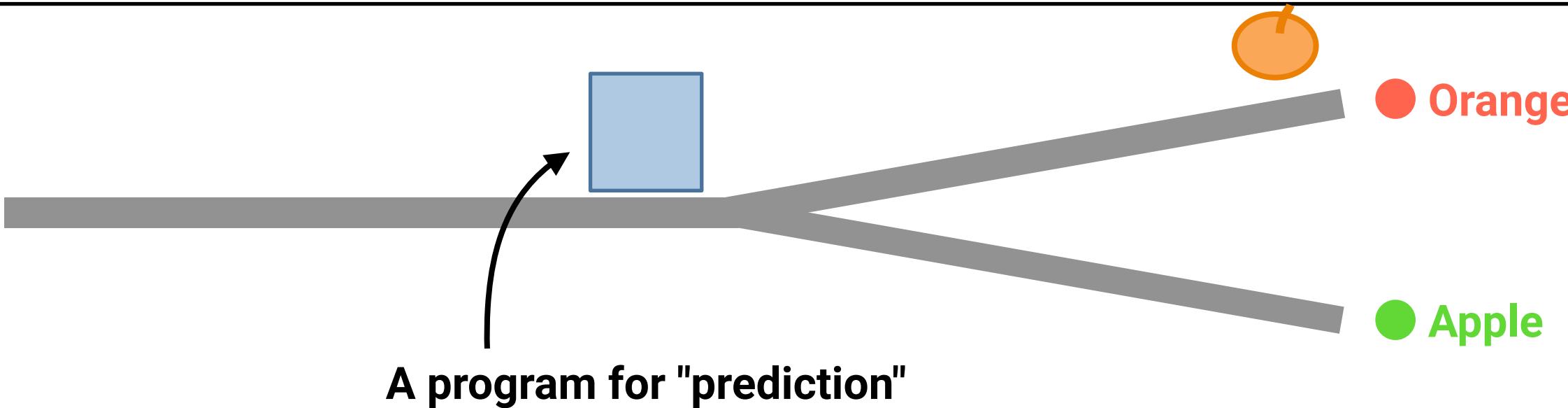


# ML converts data into "prediction"

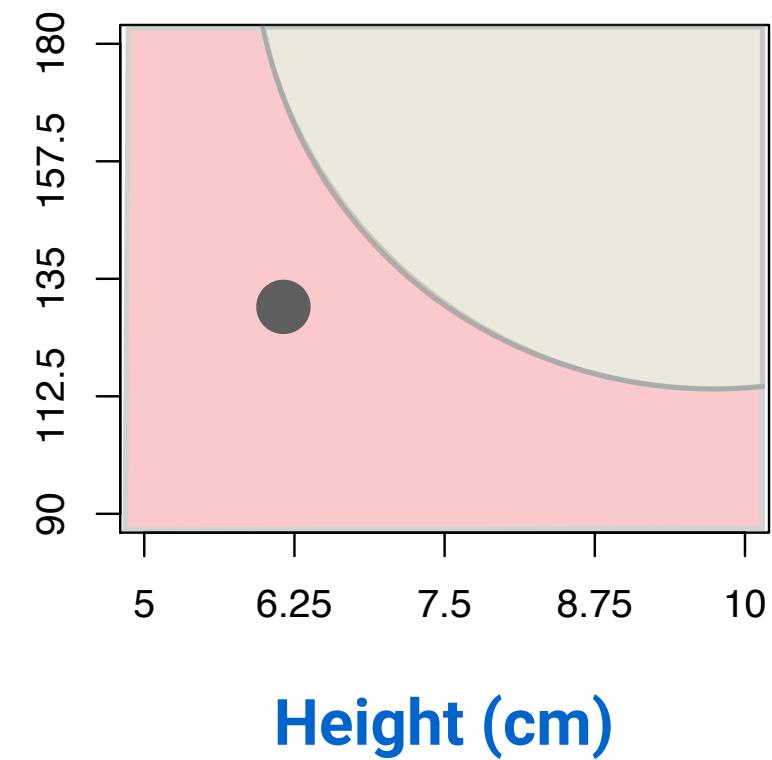
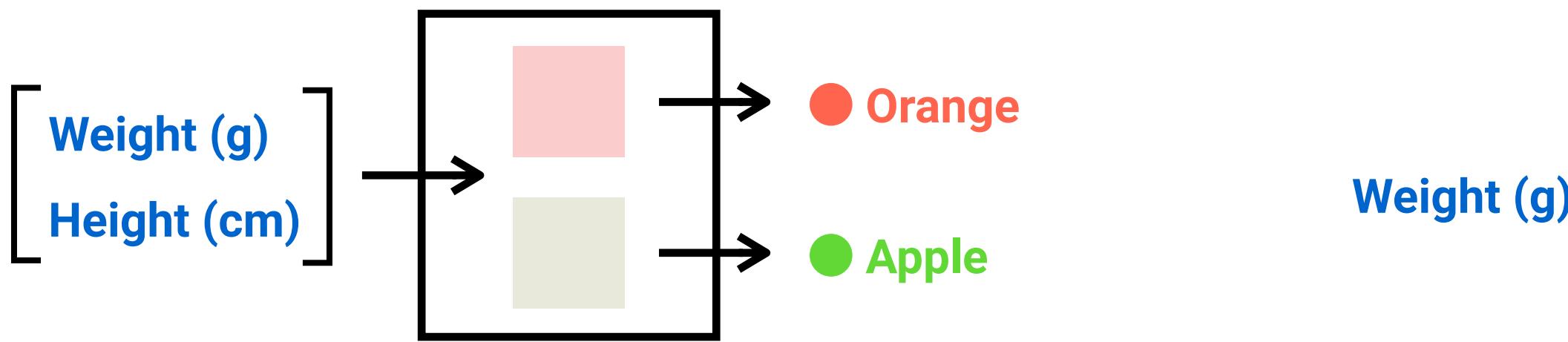
---



# ML converts data into "prediction"

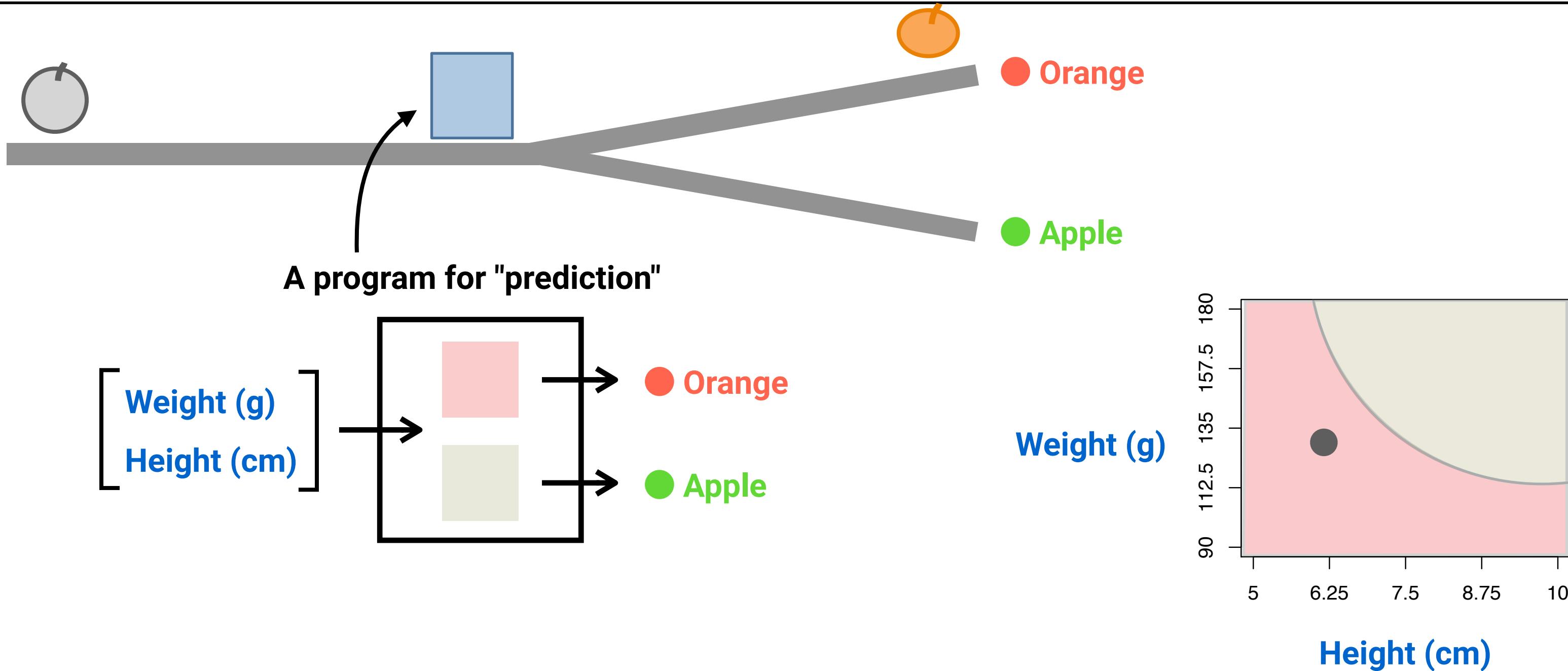


# A program for "prediction"



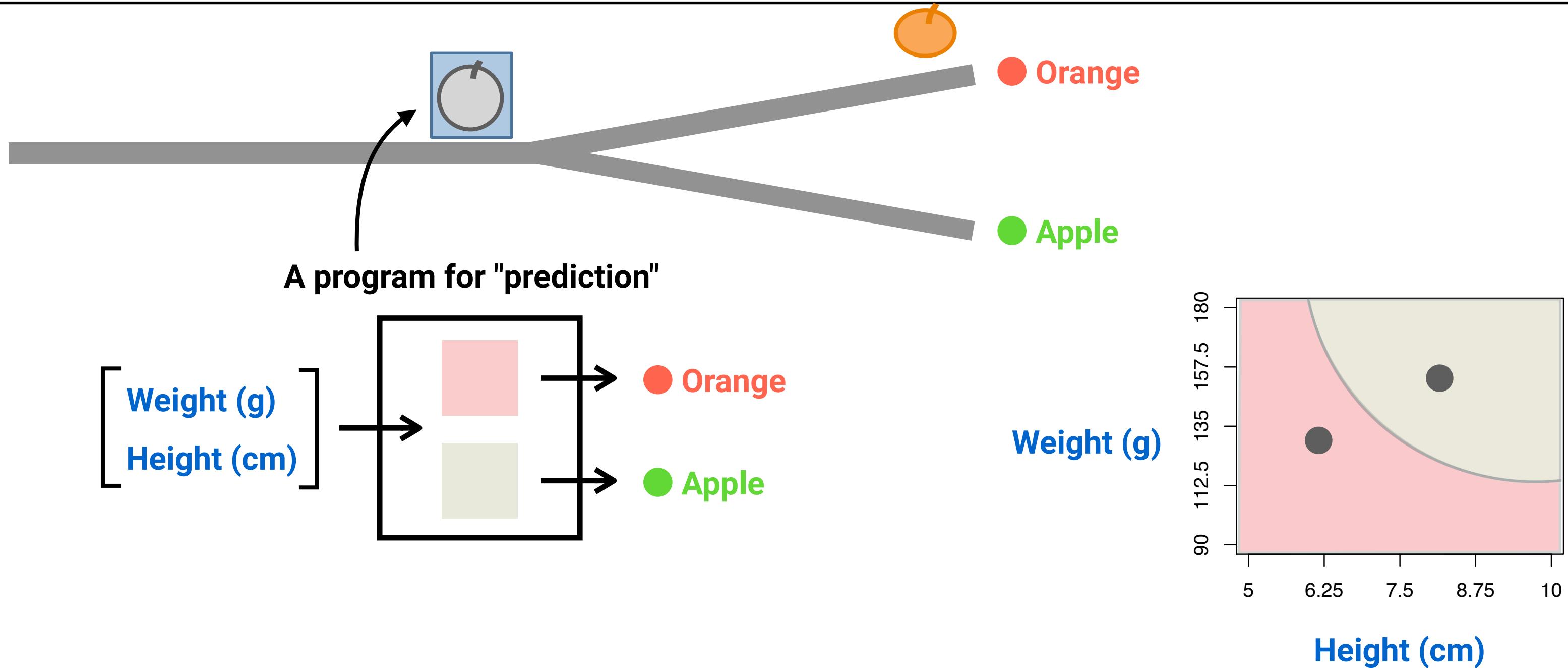
# ML converts data into "prediction"

---



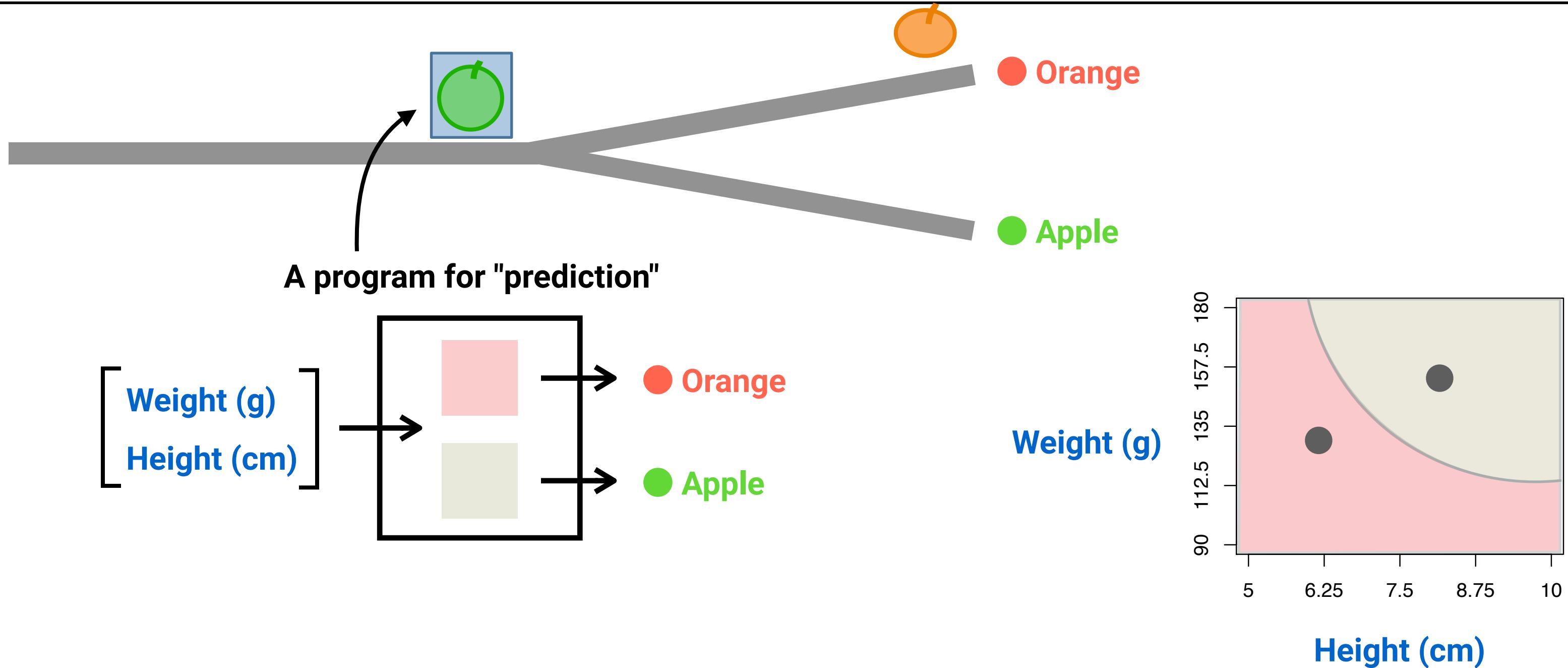
# ML converts data into "prediction"

---



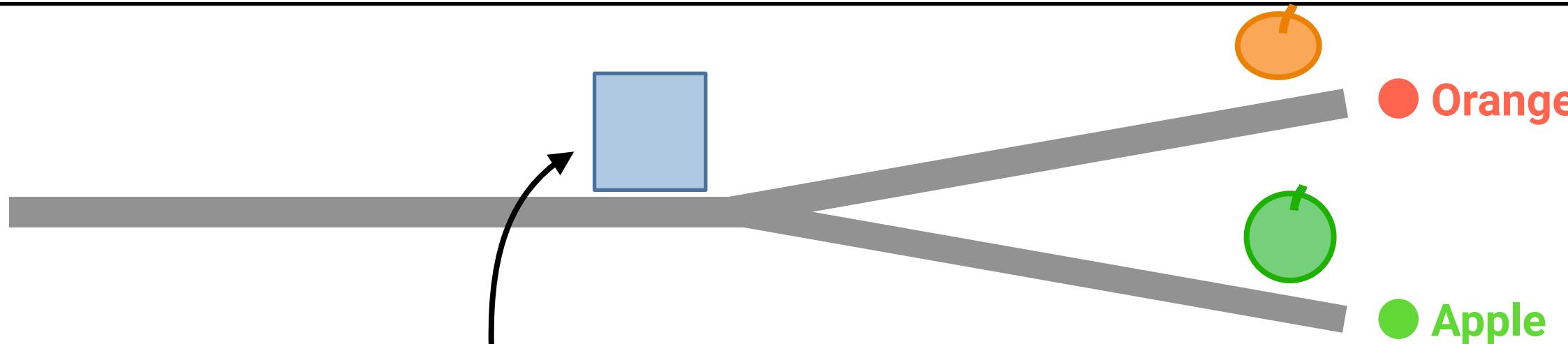
# ML converts data into "prediction"

---

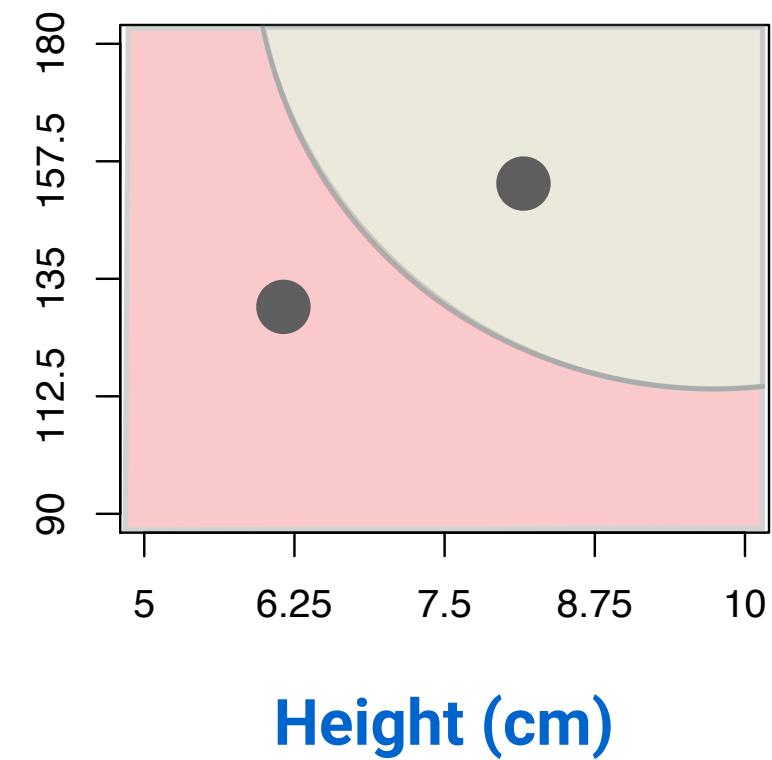
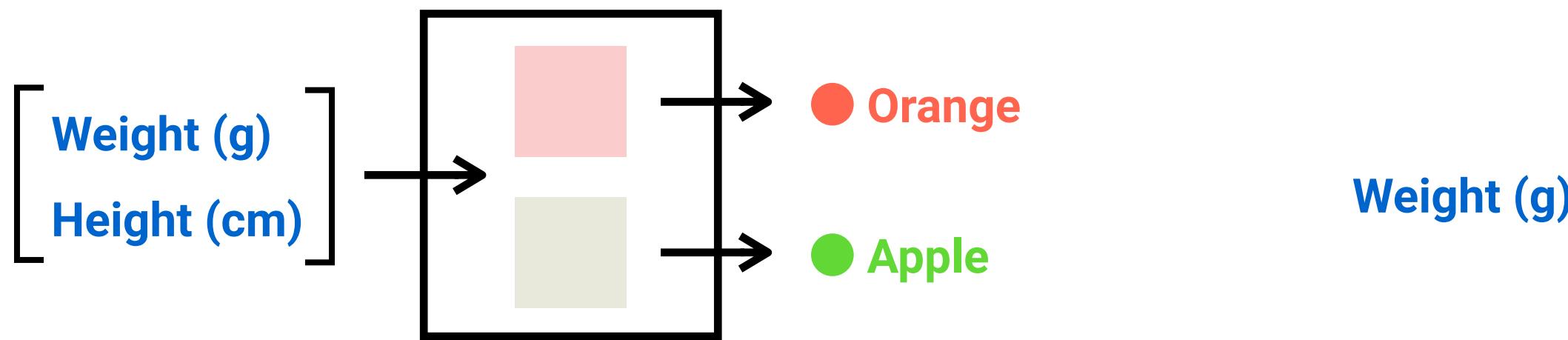


# ML converts data into "prediction"

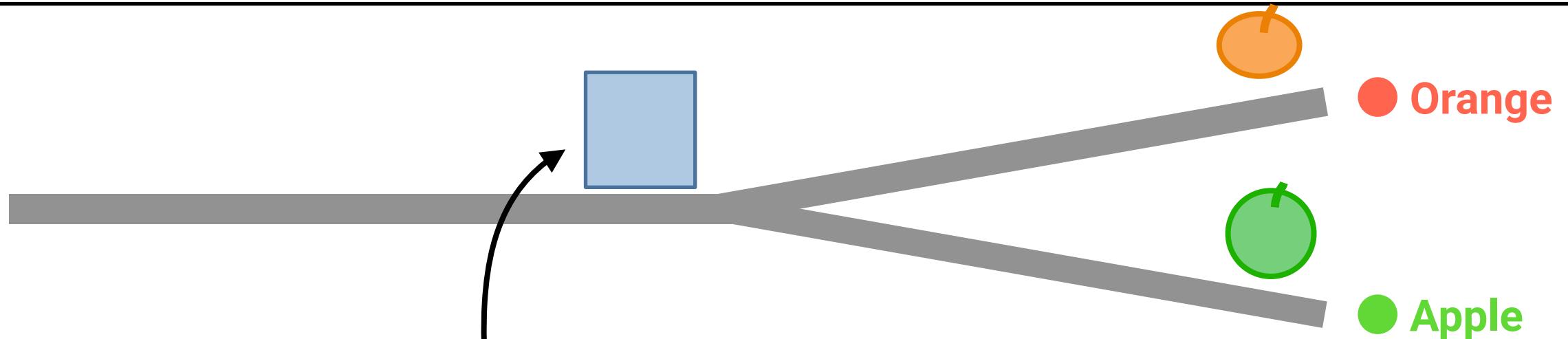
---



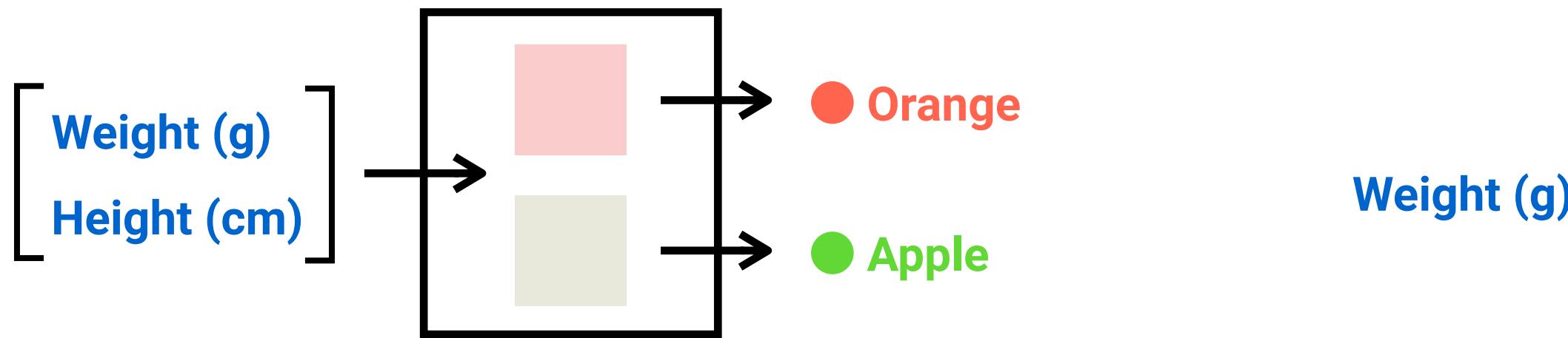
**A program for "prediction"**



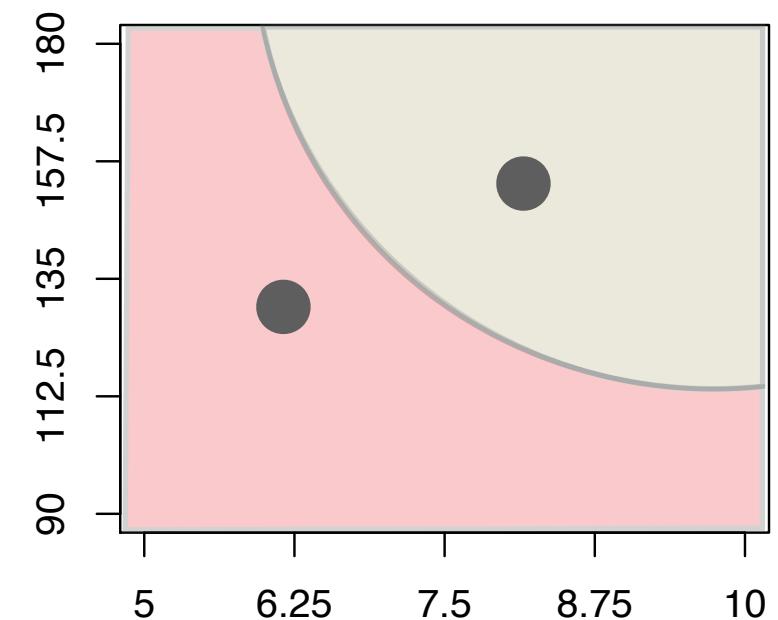
# ML converts data into "prediction"



A program for "prediction"



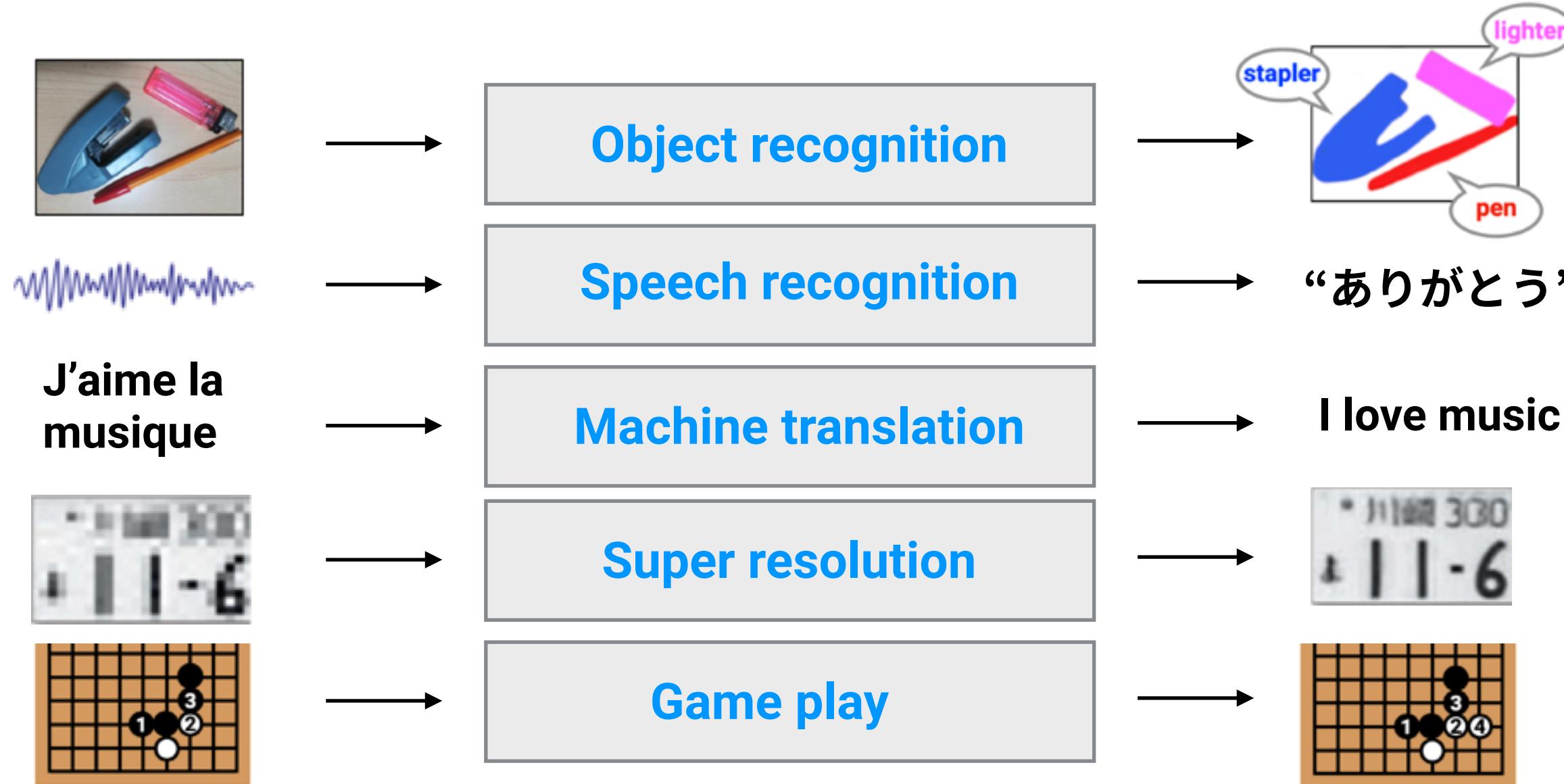
Now we got a computer program to predict "orange or apple" for any **unseen** ones **directly from collected data**



**Height (cm)**

# ML is a new (lazy) way of programming

ML generates a computer program just **by giving many input-output examples** even when we **don't know** the underlying mechanism between inputs and outputs.



# This simple idea is more powerful than you may think

---

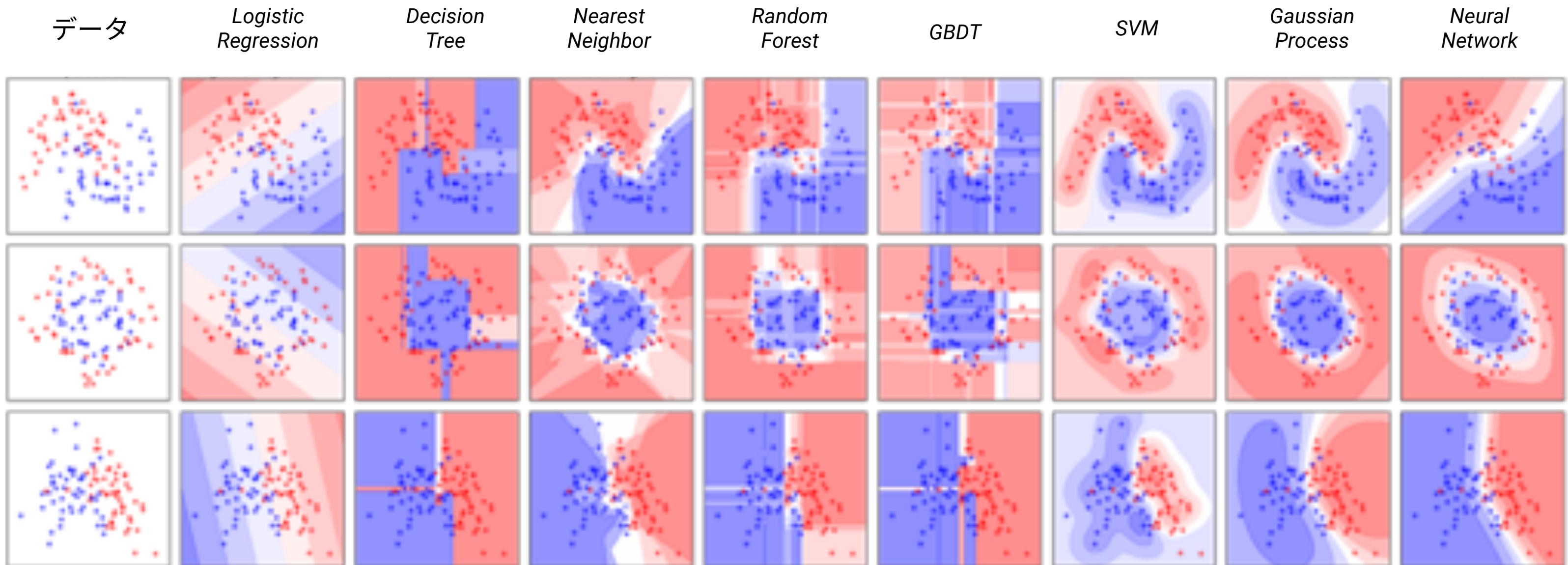
Remarkably powerful when we have relevant input-output examples (**it's useless if we don't**)



# Many ways to mathematically represent the boundary

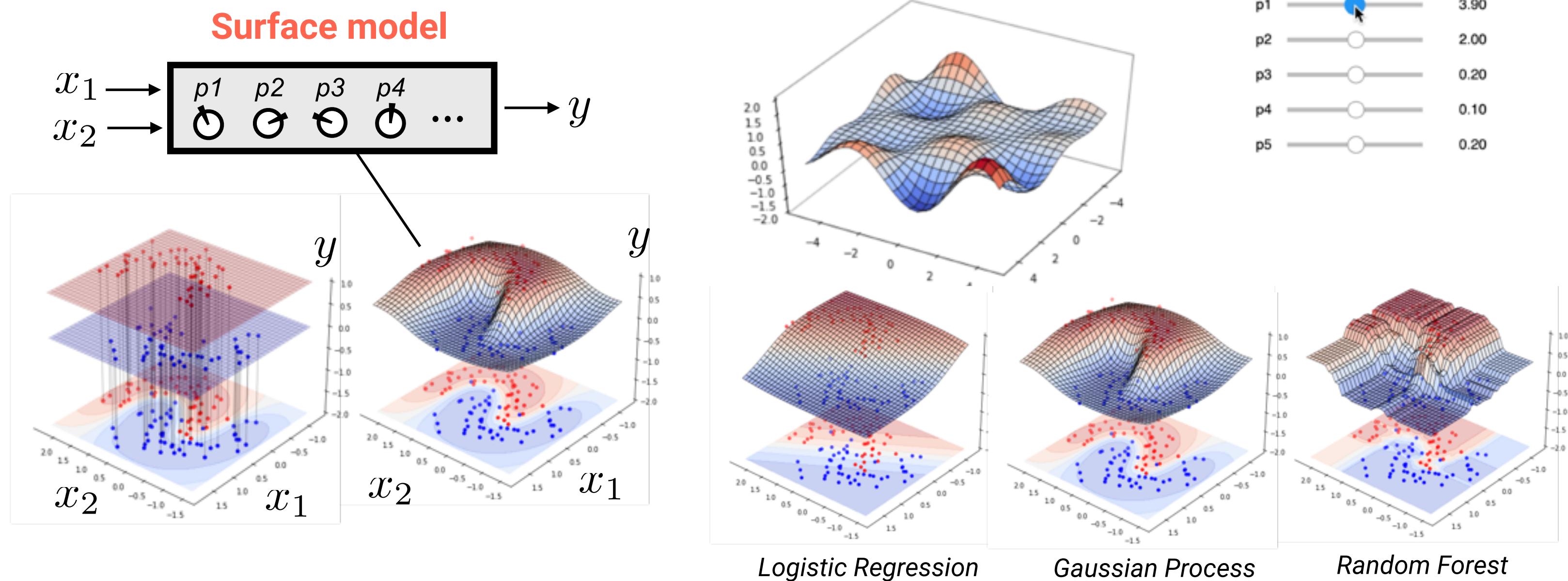
---

This is why you see too many algorithms when you start to learn ML...

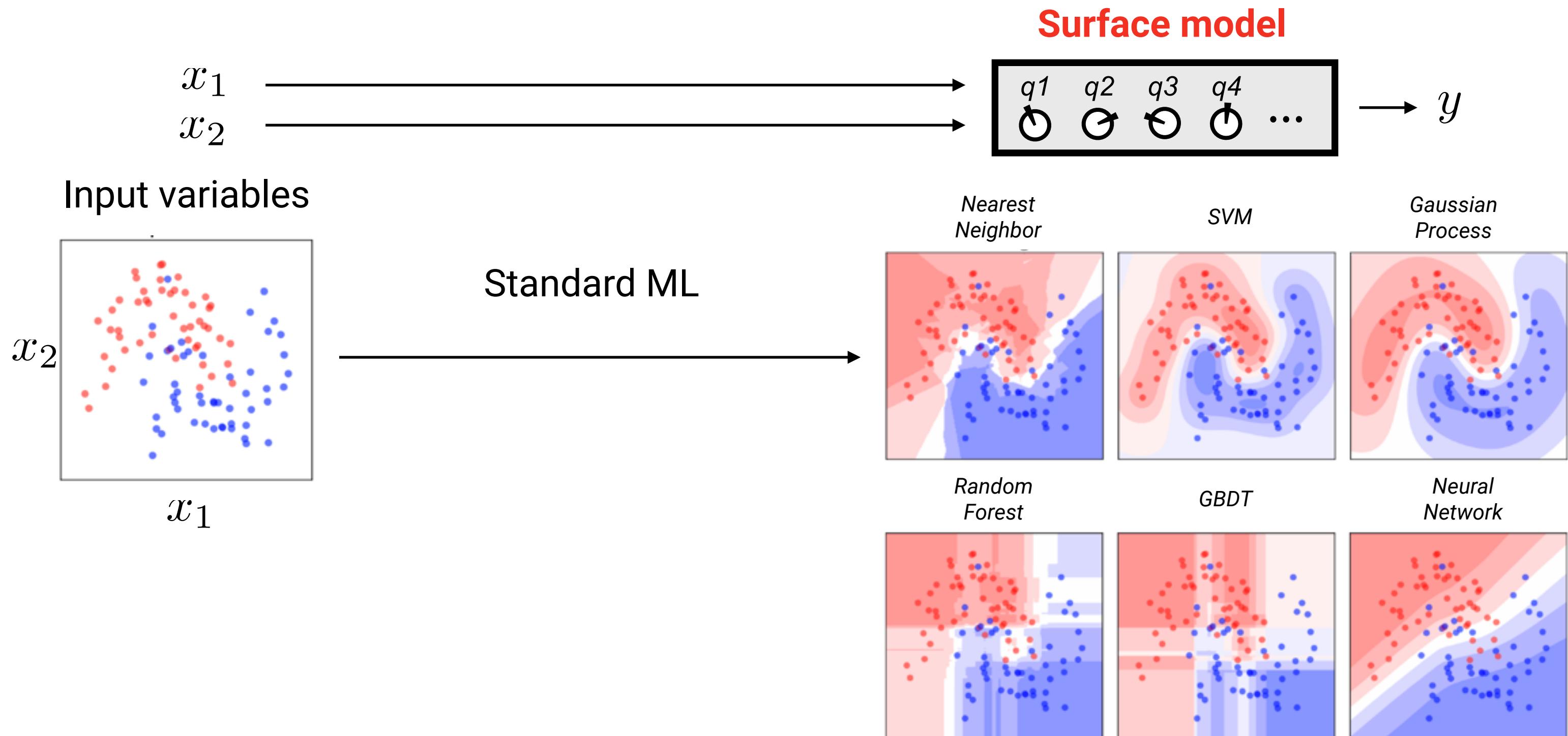


# But anyway, we're just tweaking parameters for a good fit

Internally, we're just **fitting a surface** to given points by adjusting its parameter values.

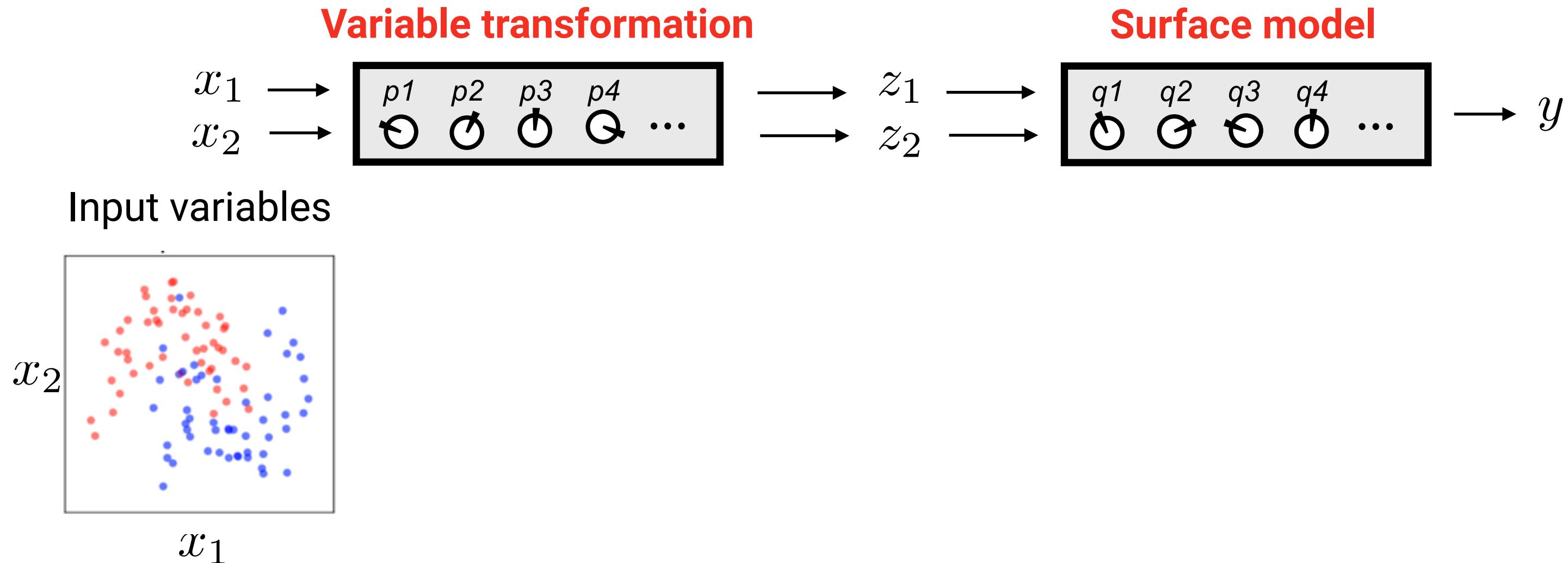


# Deep Learning (Representation Learning)



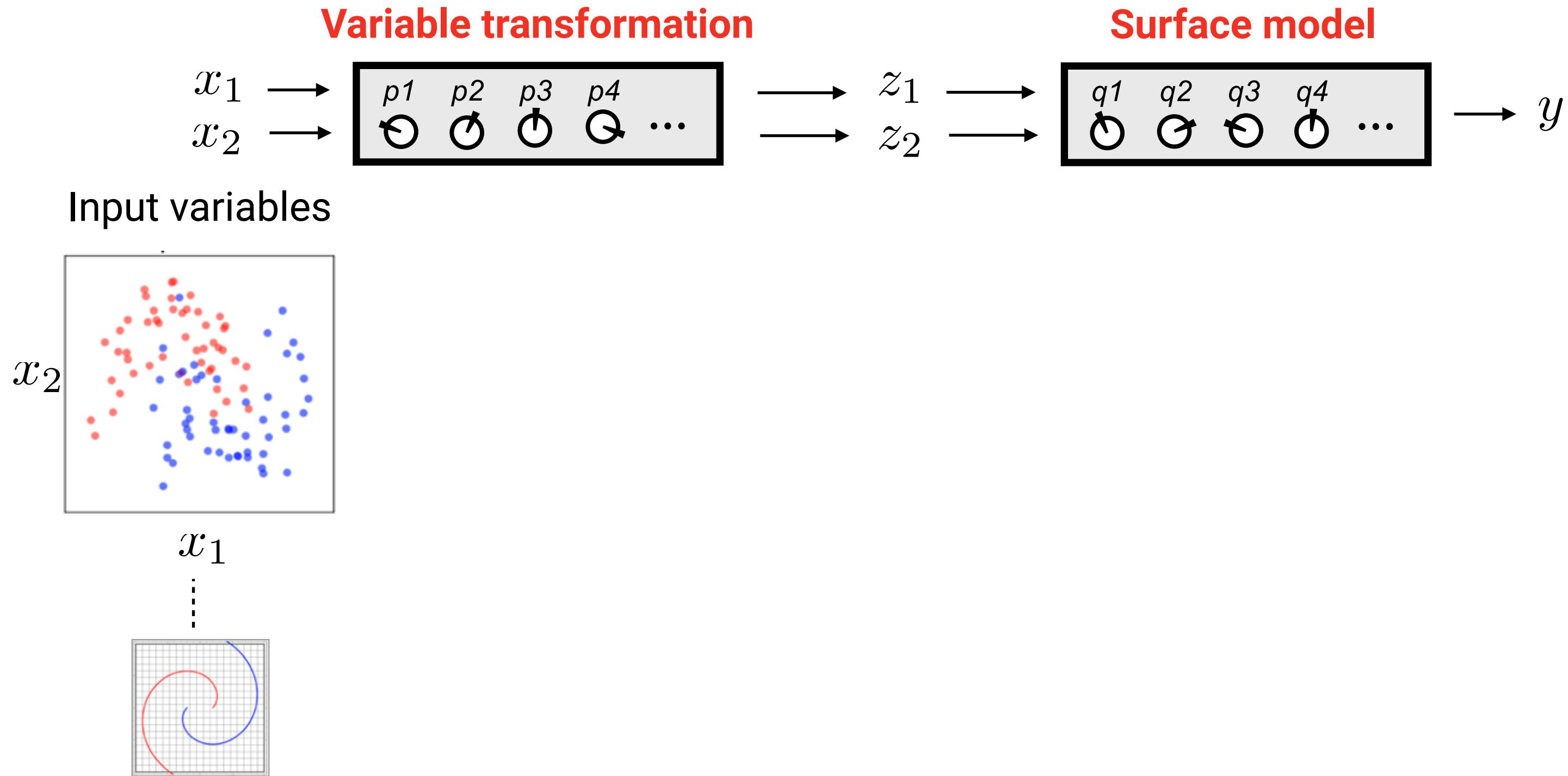
# Deep Learning (Representation Learning)

---

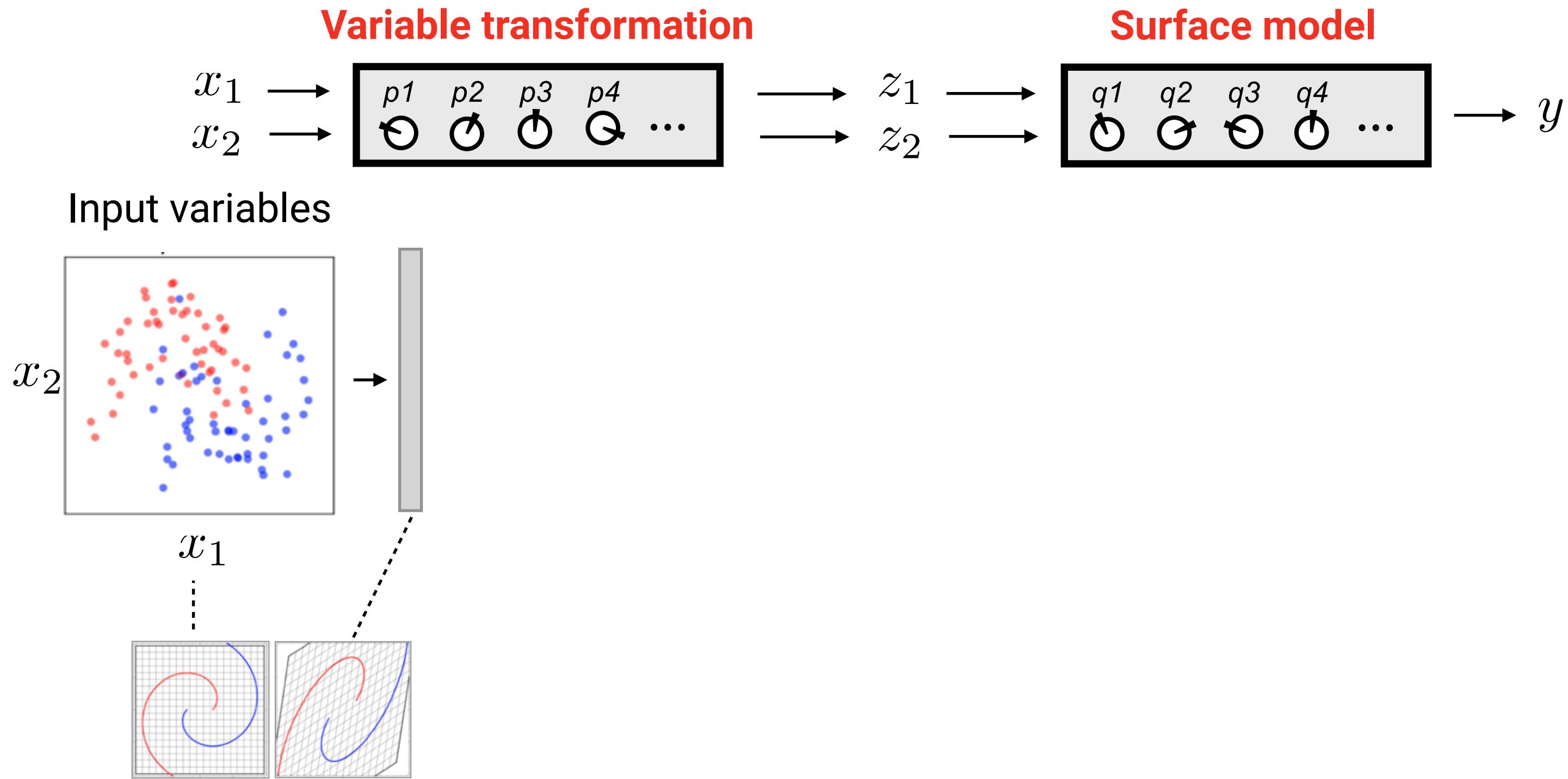


# Deep Learning (Representation Learning)

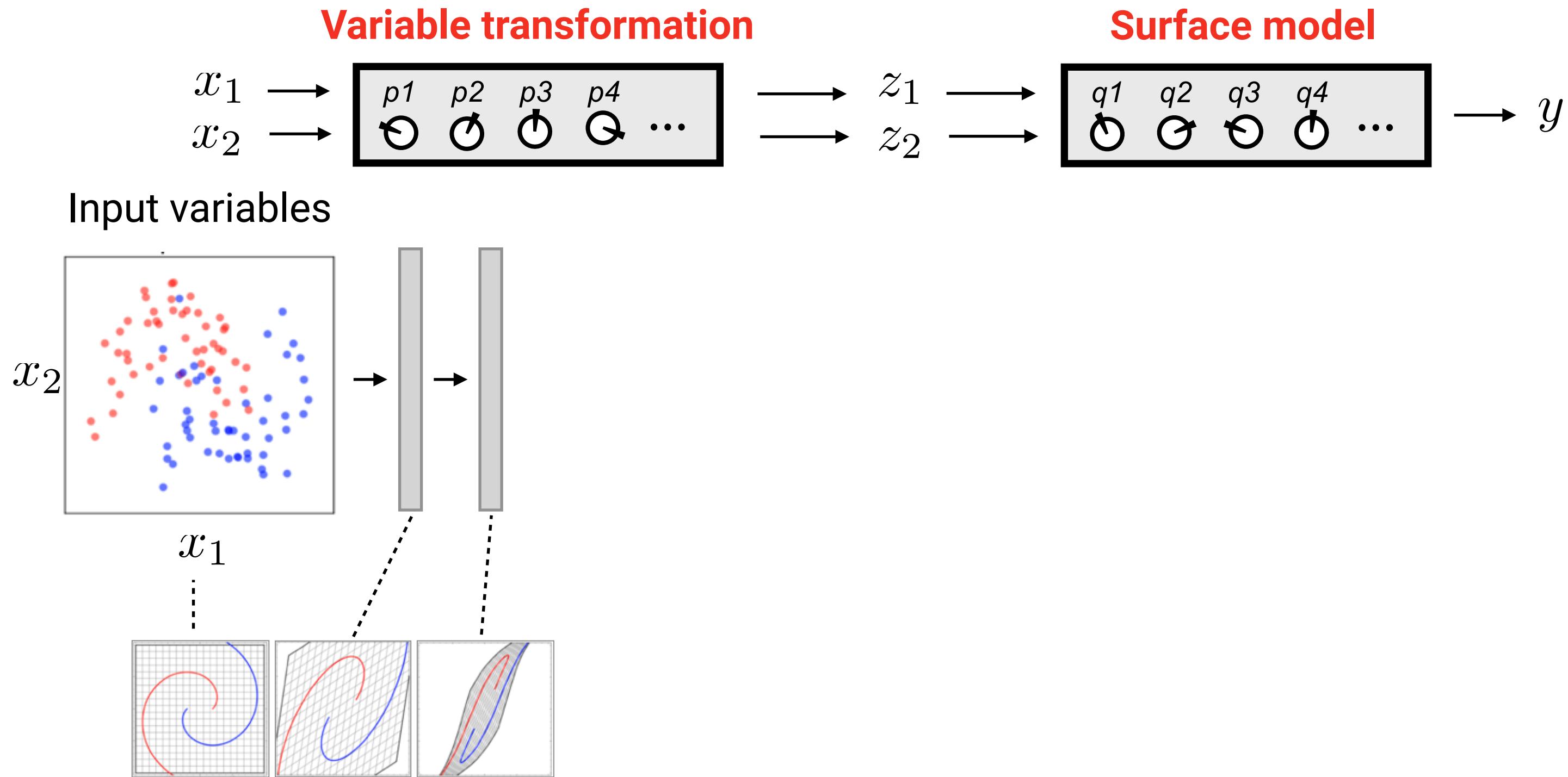
---



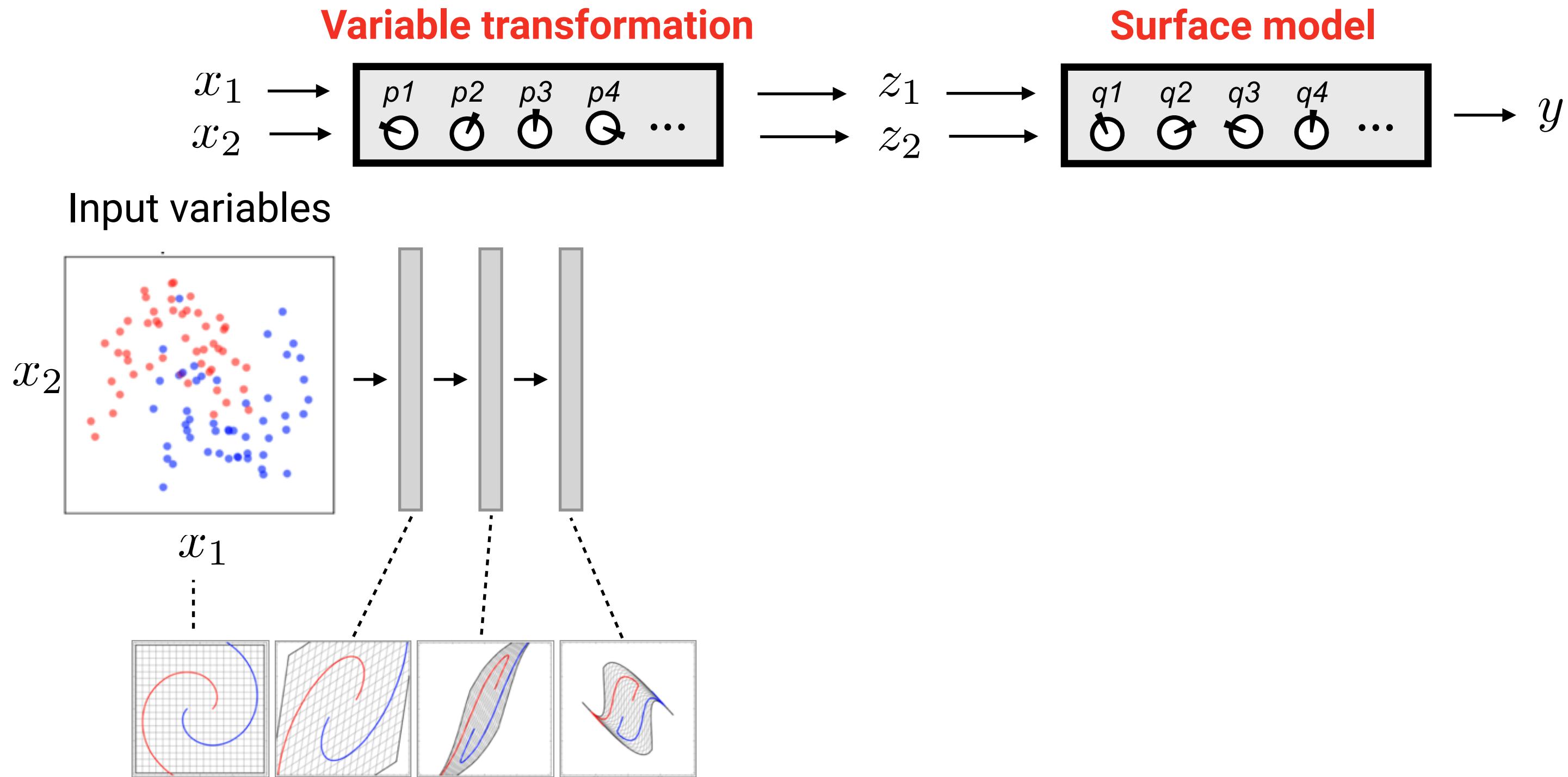
# Deep Learning (Representation Learning)



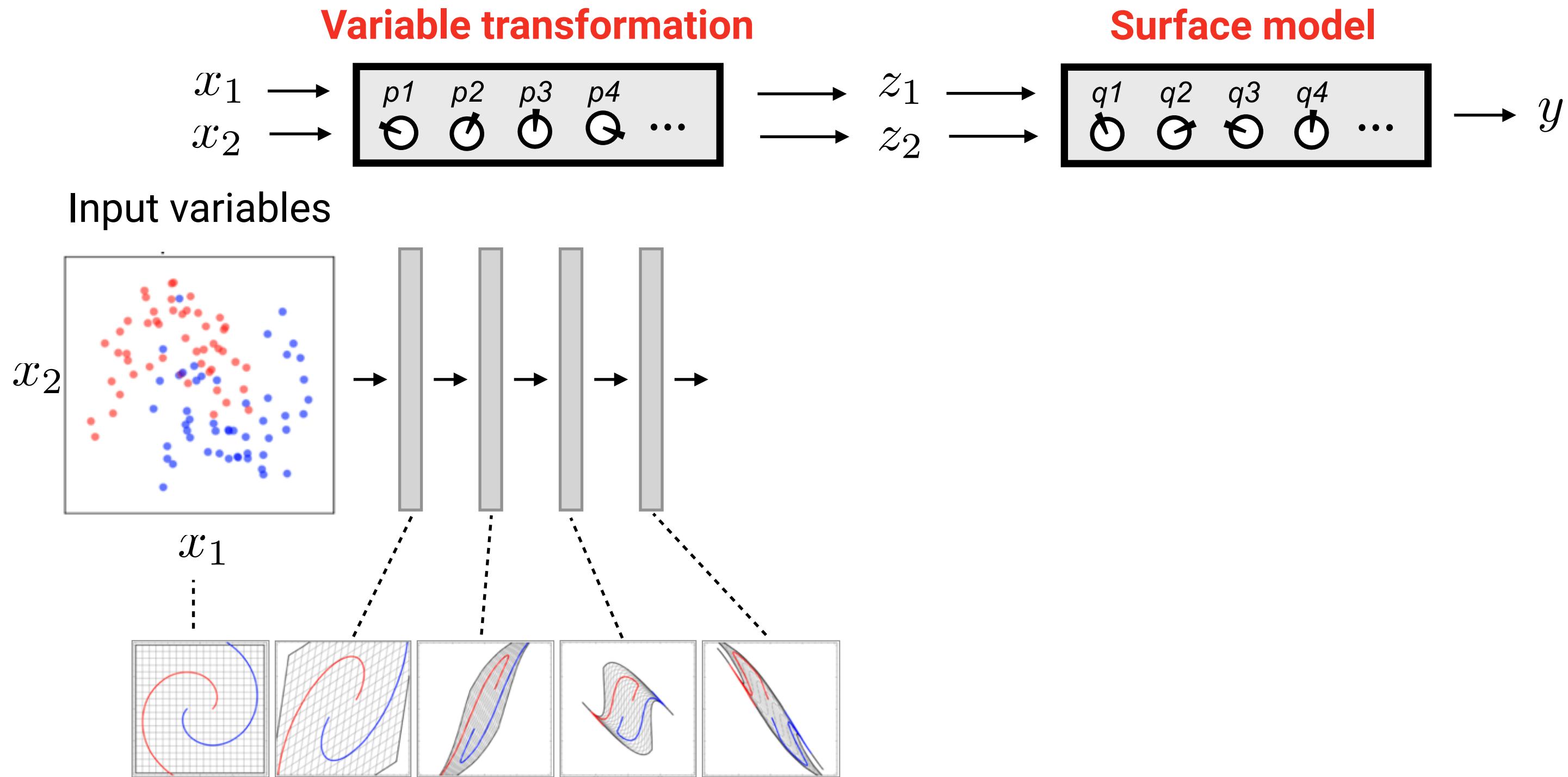
# Deep Learning (Representation Learning)



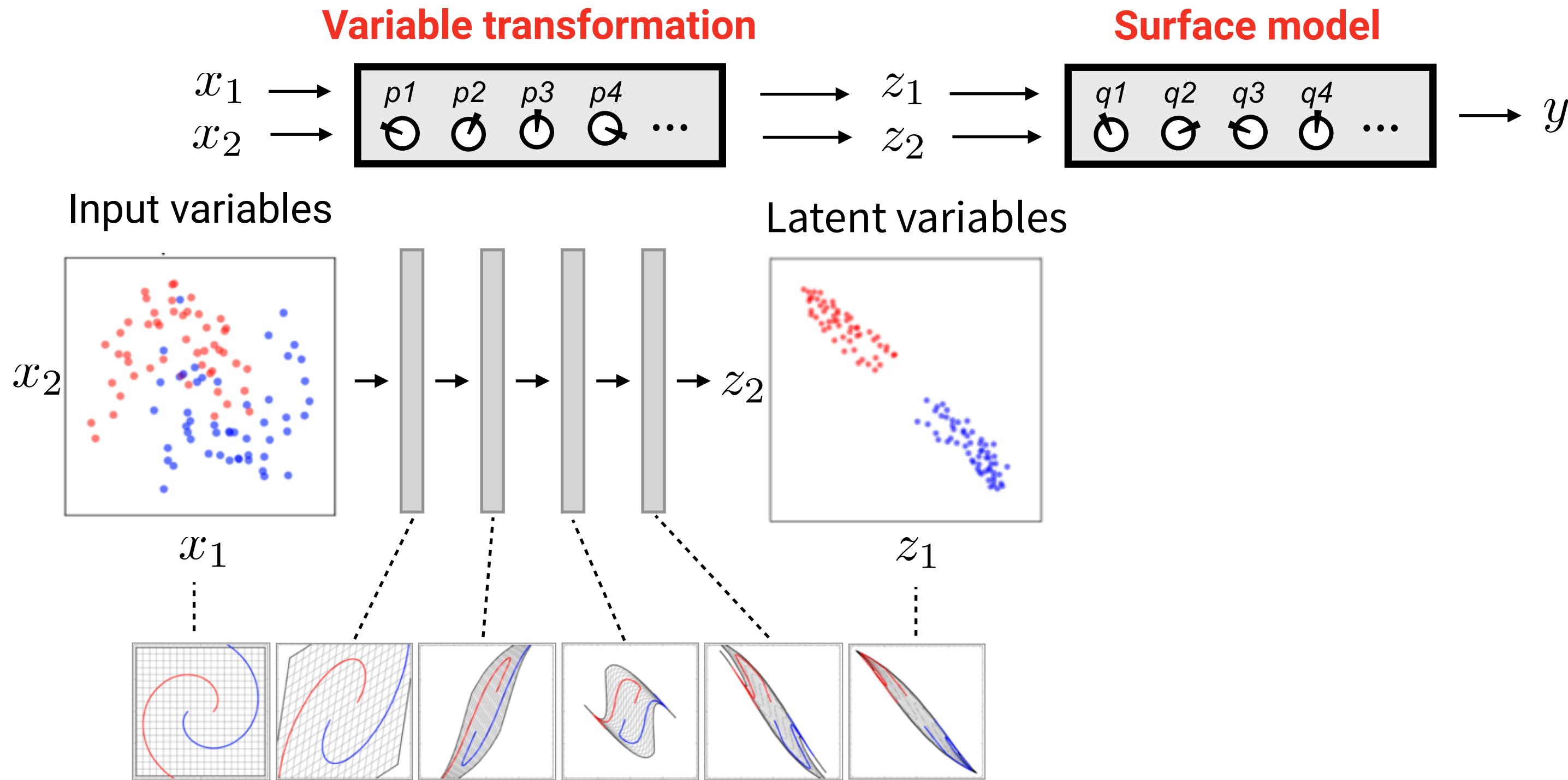
# Deep Learning (Representation Learning)



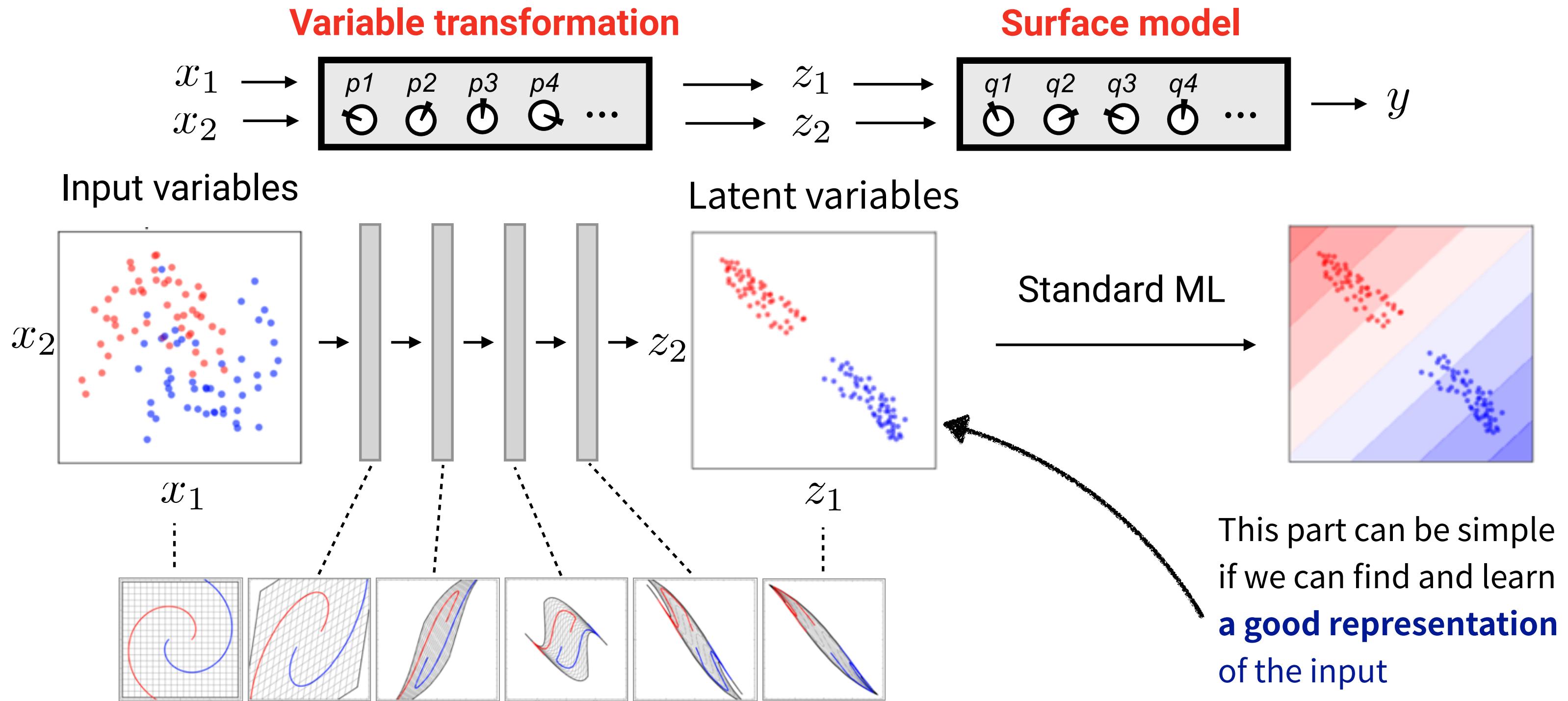
# Deep Learning (Representation Learning)



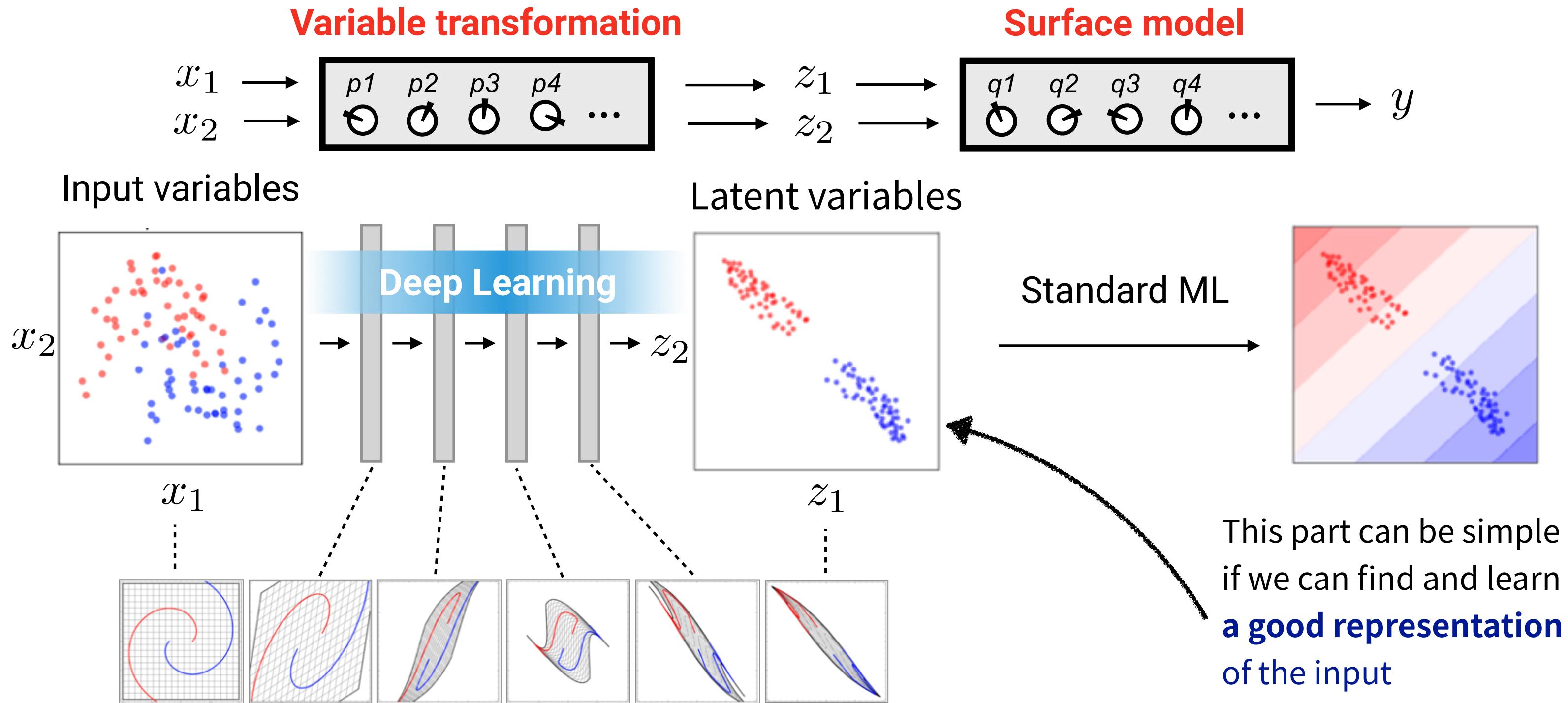
# Deep Learning (Representation Learning)



# Deep Learning (Representation Learning)



# Deep Learning (Representation Learning)



# Pitfall: The lure of wishful wordings

---

The current ML is stunningly powerful but it's very different from our sci-fi image of "AI".  
Be careful about these **"wishful"** wordings that needlessly distract and mislead us!

*"Artificial intelligence" doesn't mean that we have something artificial also intelligent like us.  
"Machine learning" doesn't mean that machines actually learn things like us.*

The cover of IEEE Spectrum magazine features a large, bold title 'IEEE Spectrum' at the top. Below the title is a photograph of several white, cylindrical objects, possibly AI hardware or circuit boards. A red banner across the middle contains the text 'Why Is AI So Dumb?' in white, followed by 'A SPECIAL REPORT' in smaller white text. At the bottom of the cover, there are four small article thumbnails with titles and page numbers: 'What's Next for Deep Learning' (P. 26), 'Inside DeepMind's Robot Lab' (P. 34), 'The 7 Biggest Weaknesses of Neural Nets' (P. 42), and 'Surprise! One of them is math' (P. 42). The publication information 'FOR THE TECHNOLOGY INSIDER OCTOBER 2021' is at the bottom right.

SIGART Newsletter No. 57 April 1976

---

## ARTIFICIAL INTELLIGENCE MEETS NATURAL STUPIDITY

Drew McDermott

MIT AI Lab Cambridge, Mass 02139

As a field, artificial intelligence has always been on the border of respectability, and therefore on the border of crackpottery. Many critics <Dreyfus, 1972>, <Lighthill, 1973> have urged that we are over the border. We have been very defensive toward this charge, drawing ourselves up with dignity when it is made and folding the cloak of Science about us. On the other hand, in private, we have been justifiably proud of our willingness to explore weird ideas, because pursuing them is the only way to make progress.

# This talk

**This slide is available at**  
<https://itakigawa.github.io/news.html>

---

A quick review on the **dark side** and **light side** of ML  
from both viewpoints as an ML algorithm researcher and an ML practitioner/user

1. What actually ML is?
2. The **dark side**: Modern aspects of ML
3. The **light side**: Deep learning for molecules

May the ML Force be with you...

Science is built up of facts, as a house is built of stones;  
but an accumulation of facts is no more a science than  
a heap of stones is a house.

*Henri Poincaré "Science and hypothesis"*



# The dark side: Modern aspect of ML

---

- **High dimensionality:** Too many input variables

We tend to use **many input variables** because ML is completely unaware of any information **not** in the input variables. Missing relevant factors results in spurious correlation.

e.g.) 100 x 100 RGB image = **30 thousand** variables

1000 x 1000 RGB image = **3 million** variables

# The dark side: Modern aspect of ML

---

- **High dimensionality:** Too many input variables

We tend to use **many input variables** because ML is completely unaware of any information **not** in the input variables. Missing relevant factors results in spurious correlation.

e.g.) 100 x 100 RGB image = **30 thousand** variables

1000 x 1000 RGB image = **3 million** variables

- **Overrepresentation:** Too many parameters

Remember that we're fitting a surface with *hundreds million* parameters in a *several million* dimensional space!

e.g.) ResNet50: **26 million** params

ResNet101: **45 million** params

EfficientNet-B7: **66 million** params

VGG19: **144 million** params

12-layer, 12-heads BERT: **110 million** params

24-layer, 16-heads BERT: **336 million** params

GPT-2 XL: **1558 million** params

GPT-3: **175 billion** params

# The dark side: Modern aspect of ML

---

- **Data hungriness:** Big data is big for human, but can be too small for ML models...

As a result, it **requires huge good data** to make current ML models work.

# The dark side: Modern aspect of ML

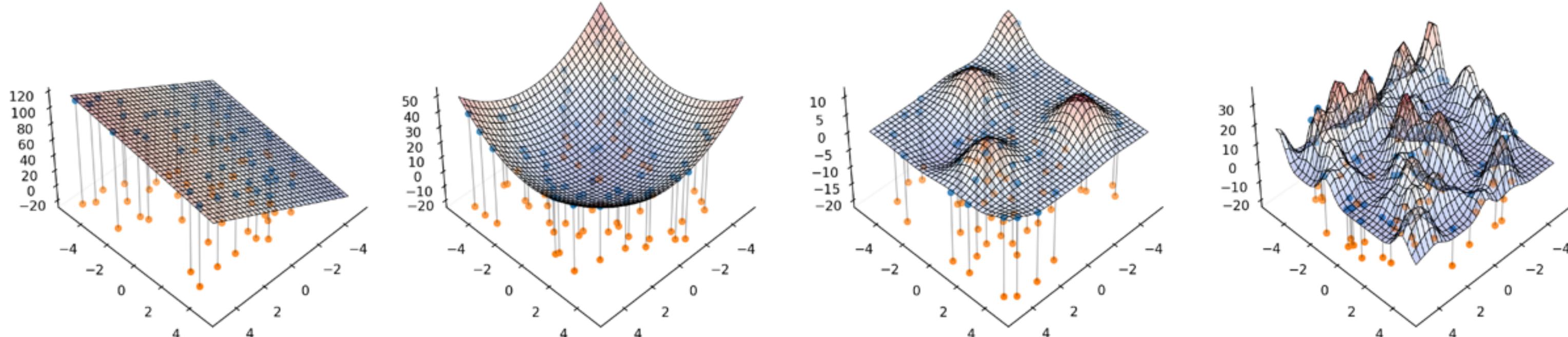
---

- **Data hungeriness:** Big data is big for human, but can be too small for ML models...

As a result, it **requires huge good data** to make current ML models work.

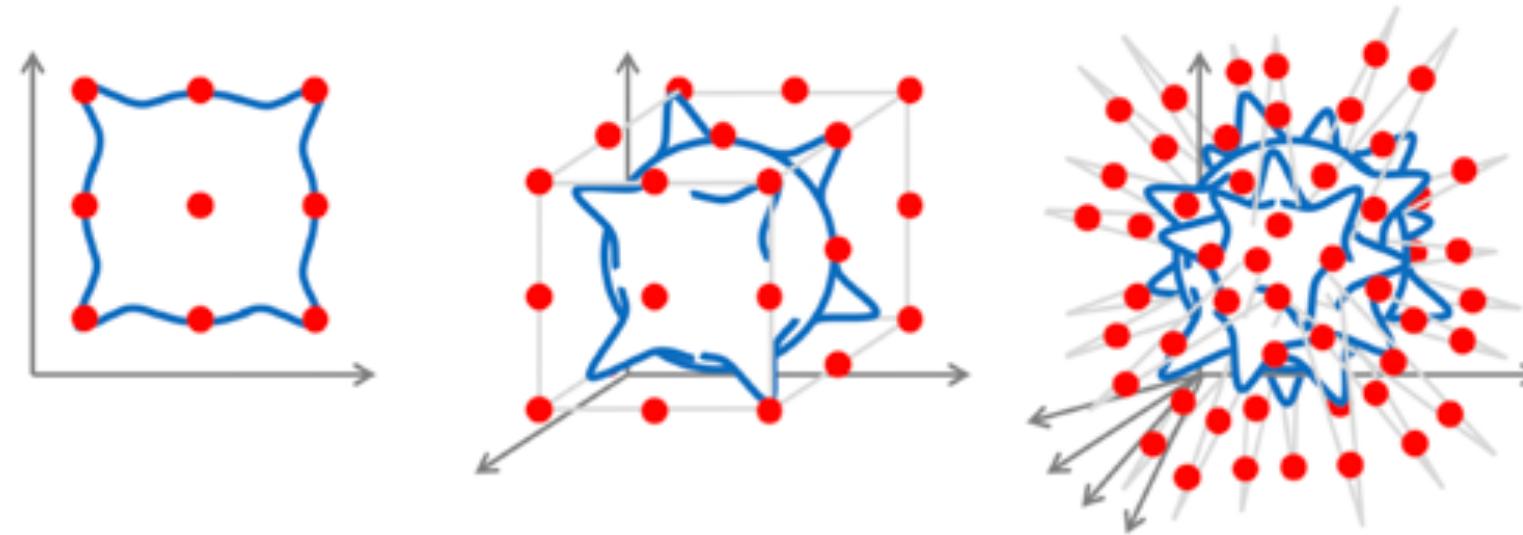
Think twice about how complex the input-output relationship you are trying to find by ML is.

How many samples will be ***statistically sufficient*** to estimate **2-variable** functions like these?  
What if you're fitting a **100-variable** function, or **10-thousand-variable** function?



# The curse of dimensionality

---



$L$ -Lipschitz  $f : \mathbb{R}^d \rightarrow \mathbb{R}$   
gives similar outputs for similar inputs  
 $|f(x) - f(x')| \leq L\|x - x'\|$   
for all  $x, x' \in \mathbb{R}^d$

A classic result in function approximation: If we must approximate a function of  $d$  variables and we know only that it is Lipschitz, say, then we need order  $(1/\varepsilon)^d$  observations on a grid in order to obtain an approximation scheme with uniform approximation error  $\varepsilon$ .

Donoho DL,

High-dimensional data analysis: The curses and blessings of dimensionality.

Plenary Lecture, AMS National Meeting on Mathematical Challenges of the 21st Century. 2000.

Bronstein MM, Bruna J, Cohen T, Veličković P.

Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.

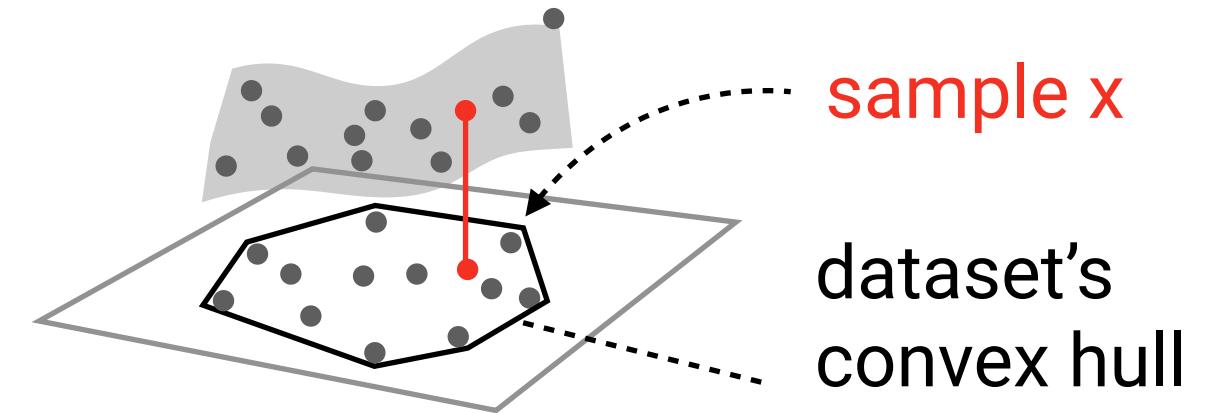
arXiv [cs.LG]. 2021. <http://arxiv.org/abs/2104.13478>

# The interpolation vs extrapolation argument

Balestrieri R, Pesenti J, LeCun Y.

Learning in High Dimension Always Amounts to Extrapolation.  
arXiv [cs.LG]. 2021. <http://arxiv.org/abs/2110.09485>

- If we define 'interpolation' as  
*Interpolation occurs for a sample  $x$  whenever this sample falls inside the given dataset's convex hull.*
- Then, the paper empirically and theoretically demonstrate that "**on any high-dimensional ( $>100$ ) dataset, interpolation almost surely never happens.**"
- "Those results challenge the validity of our current interpolation/extrapolation definition as an indicator of generalization performances."
- In high dimension, even just determining interpolation or extrapolation is already counter-intuitive...



Over 3 hours interview with authors  
@ MLStreetTalk  
<https://youtu.be/86ib0sfdfTw>

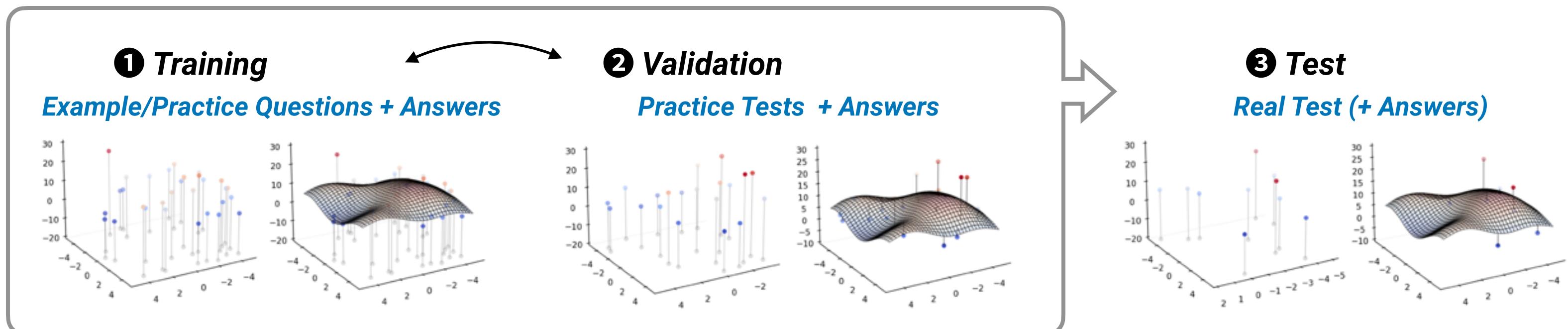


# Assessing ML predictions is intrinsically very hard!

It is quite **obvious** that ML correctly makes prediction for **the given training data**.

ML predictions **must** be evaluated by using **some data except the given training data**,  
but they can mean **any data we haven't seen yet**.

- ✓ We have no other choice but to **use some finite data** for assessing any ML predictions.
- ✓ They are already in hand, and **unintended cheating (data leakage) is very likely to happen**



# Big challenge: Rashomon effect and underspecification

---

## Rashomon Effect: The multiplicity of good ML models

In general, we can have **many equally good but very different ML models** that give equally accurate predictions for the given data.

# Big challenge: Rashomon effect and underspecification

---

## Rashomon Effect: The multiplicity of good ML models

In general, we can have **many equally good but very different ML models** that give equally accurate predictions for the given data.

- Many explanations can exist for a single set of **finite** observations in general .  
(whether they are given by ML or by human experts.)

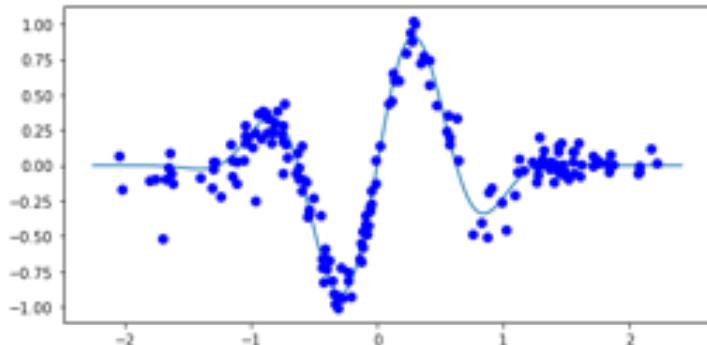
# Big challenge: Rashomon effect and underspecification

## Rashomon Effect: The multiplicity of good ML models

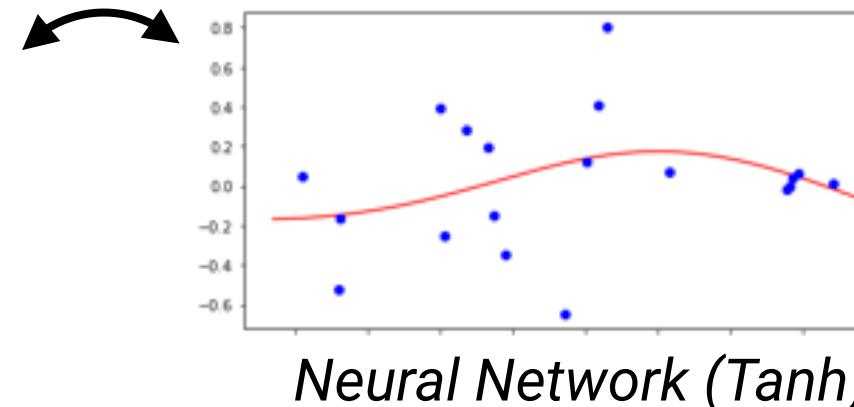
In general, we can have **many equally good but very different ML models** that give equally accurate predictions for the given data.

- Many explanations can exist for a single set of **finite** observations in general .  
(whether they are given by ML or by human experts.)
- They can **largely disagree in a underspecified situation** where data is statistically insufficient.

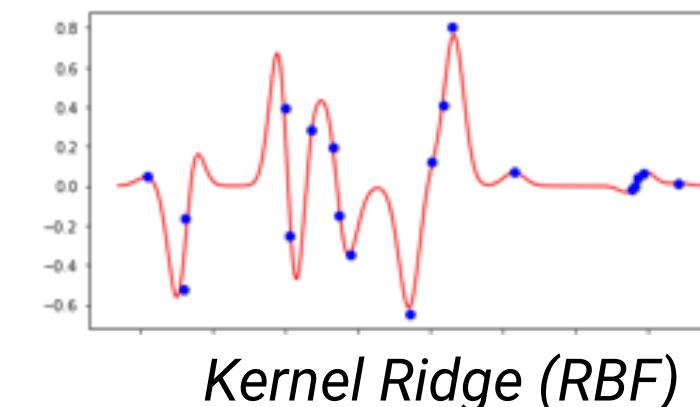
Any ML model will work



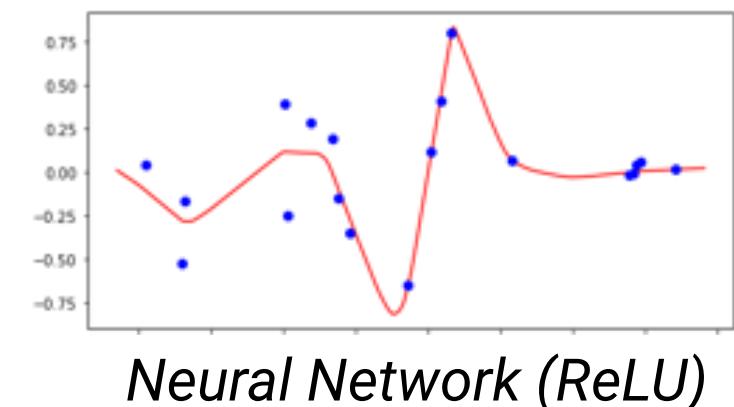
Different models can give very different predictions for out-of-sample cases



Neural Network (Tanh)



Kernel Ridge (RBF)



Neural Network (ReLU)

# Underspecification appears in many practical ML systems?

---

- We often see **ML failures** in real-world domains **even when** we trained them on well-curated data that structurally matched with the application domain.
- "While ML models are validated on held-out data, this validation is **often insufficient to guarantee** that the models will have well-defined behavior when they are used **in a new setting.**"

[https://ai.googleblog.com/2021/10/  
how-underspecification-presents.html](https://ai.googleblog.com/2021/10/how-underspecification-presents.html)



The latest from Google Research

## How Underspecification Presents Challenges for Machine Learning

Monday, October 18, 2021

Posted by Alex D'Amour and Katherine Heller, Research Scientists, Google Research

Machine learning (ML) models are being used more widely today than ever before and are becoming increasingly impactful. However, they often exhibit unexpected behavior when they are used in real-world domains. For example, computer vision models can exhibit surprising sensitivity to irrelevant features, while natural language processing models can depend unpredictably on demographic correlations not directly indicated by the text. Some reasons for these failures are

<https://arxiv.org/abs/2011.03395>

arXiv.org > cs > arXiv:2011.03395

Search...

Help | Advanced S

Computer Science > Machine Learning

(Submitted on 6 Nov 2020 ([v1](#)), last revised 24 Nov 2020 (this version, v2))

## Underspecification Presents Challenges for Credibility in Modern Machine Learning

Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D. Hoffman, Farhad Hormozdiari, Neil Houlsby, Shaobo Hou, Ghassen Jerfel, Alan Karthikesalingam, Mario Lucic, Yian Ma, Cory McLean, Diana Mincu, Akinori Mitani, Andrea Montanari, Zachary Nado, Vivek Natarajan, Christopher Nielson, Thomas F. Osborne, Rajiv Raman, Kim Ramasamy, Rory Sayres, Jessica Schrouff, Martin Seneviratne, Shannon Sequeira, Harini Suresh, Victor Veitch, Max Vladymyrov, Xuezhi Wang, Kellie Webster, Steve Yadlowsky, Taedong Yun, Xiaohua Zhai, D. Sculley

ML models often exhibit unexpectedly poor behavior when they are deployed in real-world domains. We identify underspecification as a key reason for these failures. An ML pipeline is underspecified when it can

# Many points in ML are still open problems!

---

Definitely **we still need to rethink our traditional understanding** on concepts like generalization, overfitting, bias-variance tradeoff, interpolation/extrapolation, the curse of dimensionality, etc.

- **Too high expressive power**
  - Zero training error on random labels. (DL can represent any function / memorize entire data)
- **Benign overfitting (when interpolating noisy training data)**
  - Low test error even when we have zero training error on *noisy* training data.
- **Implicit regularization**
  - Stochastic optimization like SGD prefers low-complexity solutions.

Berner J, Grohs P, Kutyniok G, Petersen P.

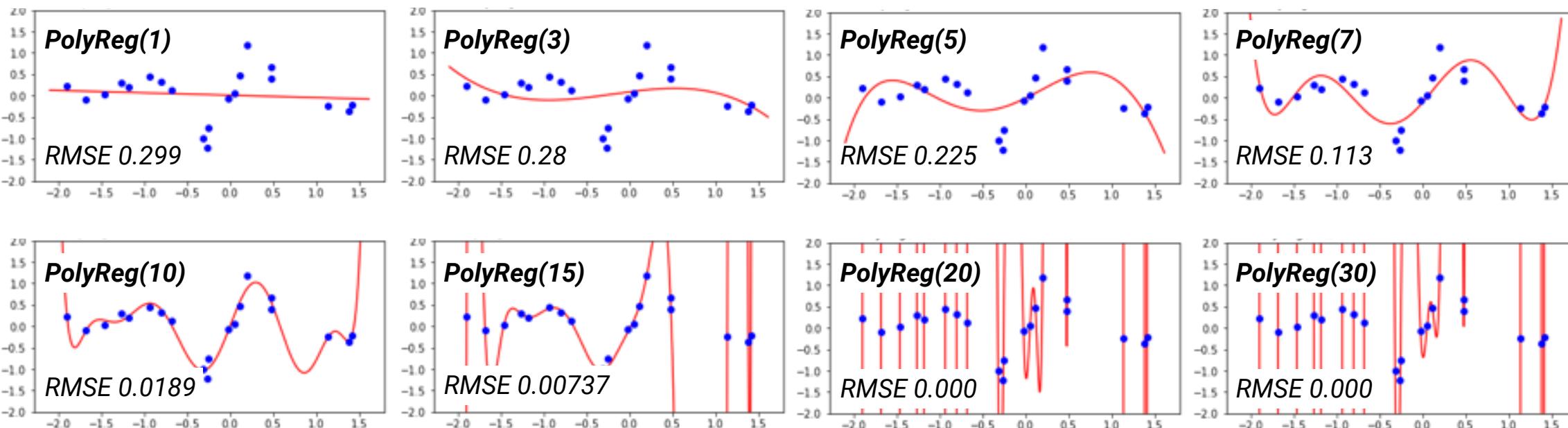
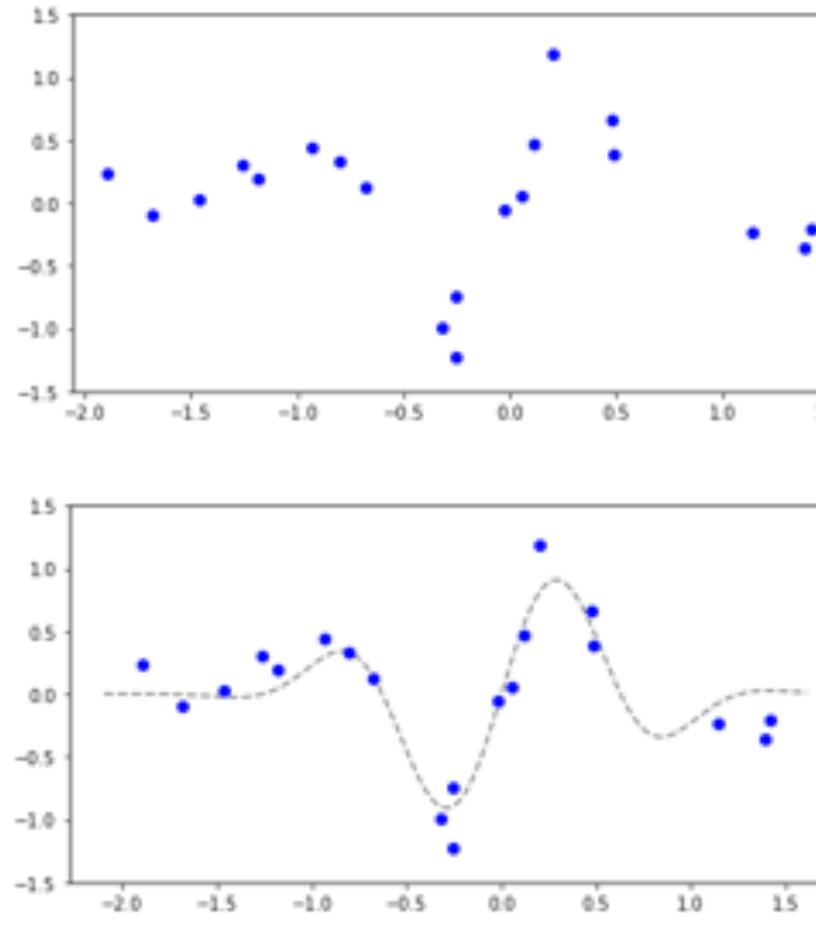
[The Modern Mathematics of Deep Learning](https://arxiv.org/abs/2105.04026). arXiv [cs.LG]. 2021. <http://arxiv.org/abs/2105.04026>

Zhang C, Bengio S, Hardt M, Recht B, Vinyals O.

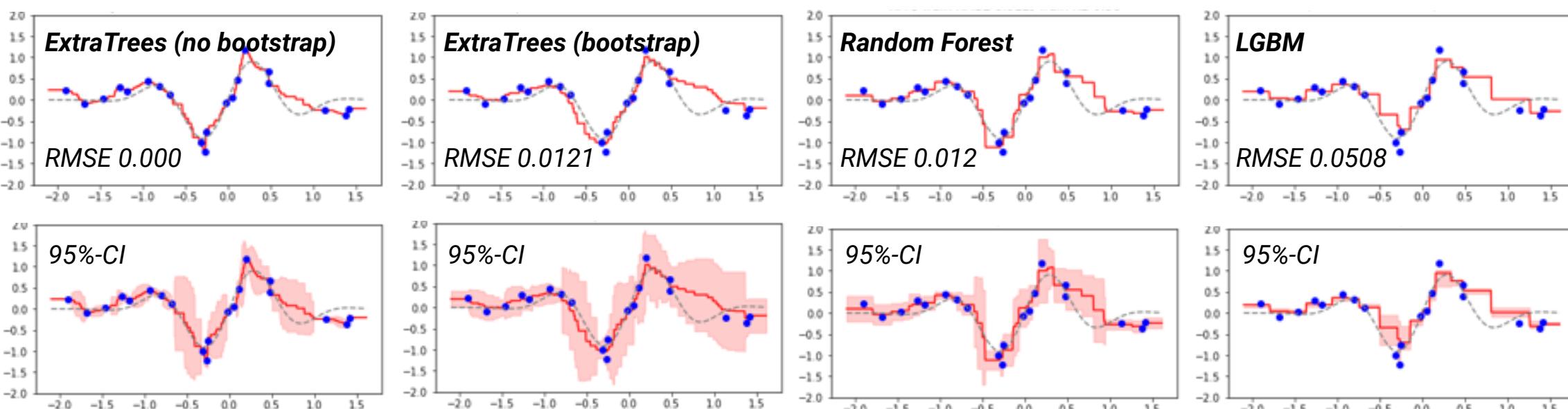
[Understanding deep learning \(still\) requires rethinking generalization](https://doi.org/10.1145/3468221.3468254). Commun ACM. 2021;64: 107–115.

# A toy case: 'Benign' zero training error on noisy data

Problematic overfitting by polynomial regression of order k



clearly overfitted but harmless (still informative)



also we can assess  
the uncertainty

# This talk

**This slide is available at**  
<https://itakigawa.github.io/news.html>

---

A quick review on the **dark side** and **light side** of ML  
from both viewpoints as an ML algorithm researcher and an ML practitioner/user

1. What actually ML is?
2. The **dark side**: Modern aspects of ML
3. The **light side**: Deep learning for molecules

May the ML Force be with you...

Science is built up of facts, as a house is built of stones;  
but an accumulation of facts is no more a science than  
a heap of stones is a house.

*Henri Poincaré "Science and hypothesis"*



# ML's interests: How to tame the high dimensionality?

---

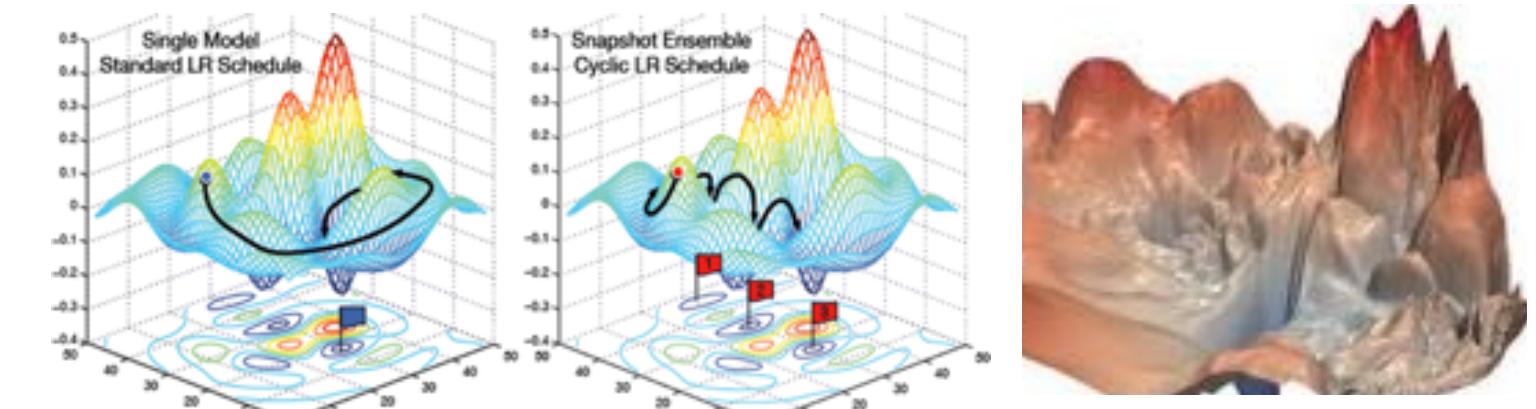
## 1. Regularization (inc. implicit regularization by SGD)

Control, restrict, or stabilize the solution space or optimization

Trying to find the global minimum of  
very bumpy loss landscape...

## 2. Good initial value (warm start) of general use

Large-scale pretraining and its transfer



# ML's interests: How to tame the high dimensionality?

## 1. Regularization (inc. implicit regularization by SGD)

Control, restrict, or stabilize the solution space or optimization

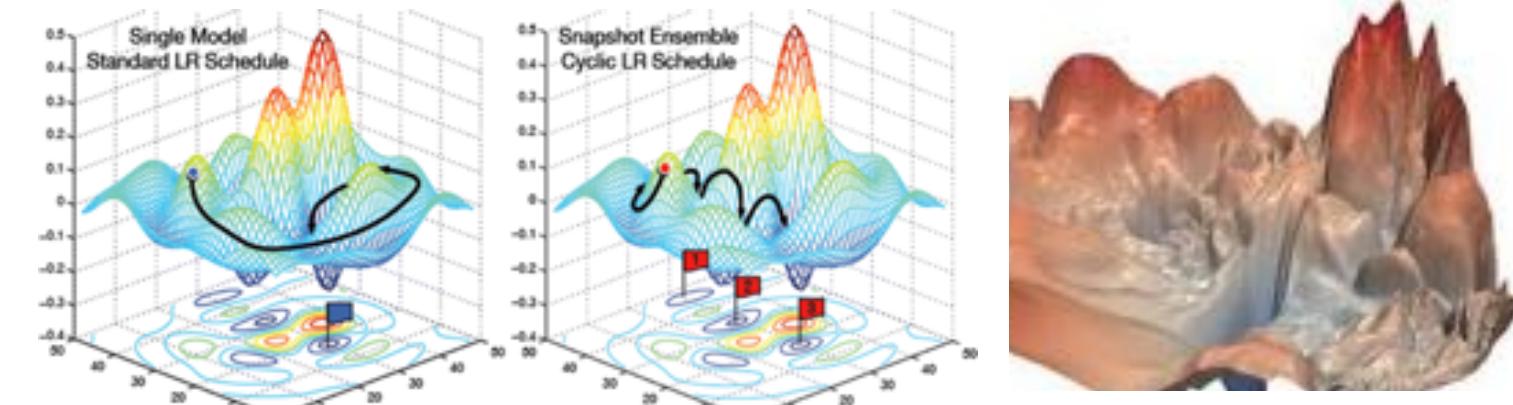
Trying to find the global minimum of very bumpy loss landscape...

## 2. Good initial value (warm start) of general use

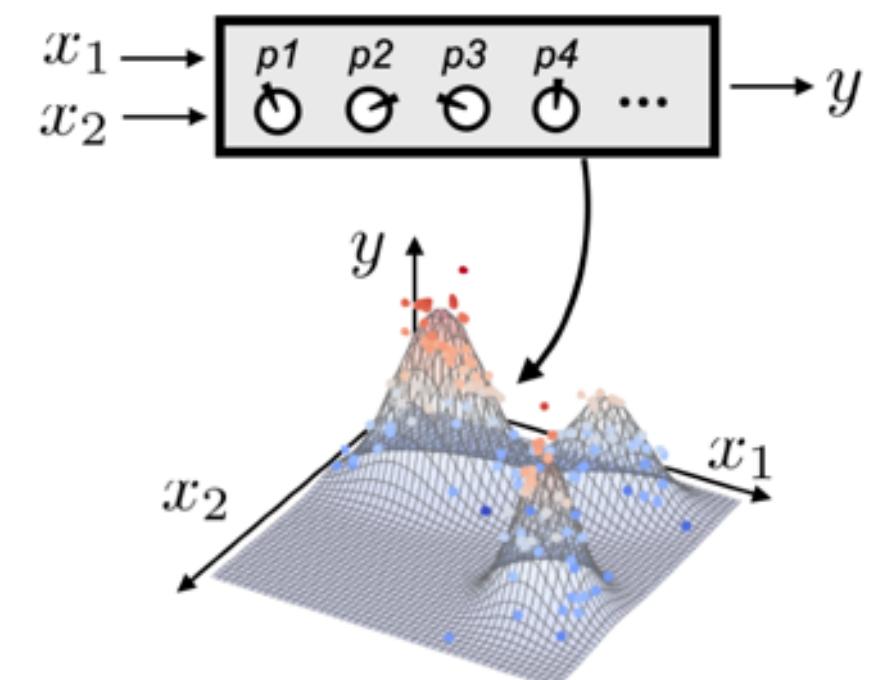
Large-scale pretraining and its transfer

## 3. Relevant "inductive bias"

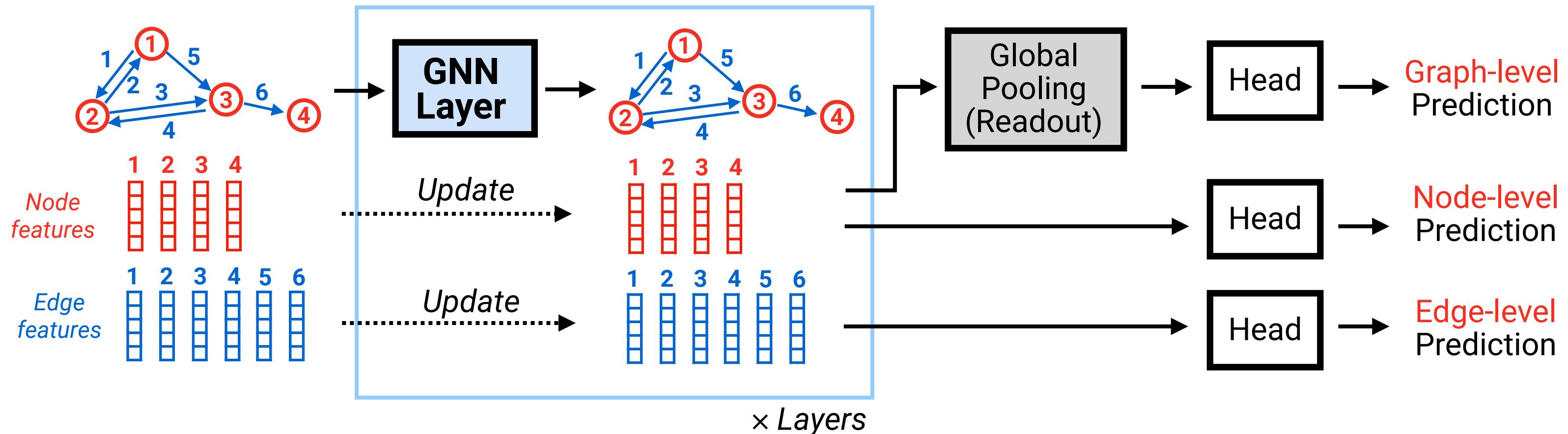
ML models that can represent any function **make worse** the risk of Rashomon effect, underspecification, and spurious correlation.



**Design inductive biases (features, architectures, models, etc) by chemical knowledge, expert intuition, and theory**  
so that ML models don't unintentionally represent a function that lacks any chemical validity.



# Graph Neural Networks (GNNs) in general



## Travel Time Estimation (Google Maps, Baidu Maps)

Derrow-Pinion A, She J, Wong D, Lange O, Hester T, Perez L, et al. [ETA Prediction with Graph Neural Networks in Google Maps](#). CIKM 2021

Fang X, Huang J, Wang F, Zeng L, Liang H, Wang H. [ConSTGAT: Contextual Spatial-Temporal Graph Attention Network for Travel Time Estimation at Baidu Maps](#). KDD 2020

## Siri Triggering (Apple)

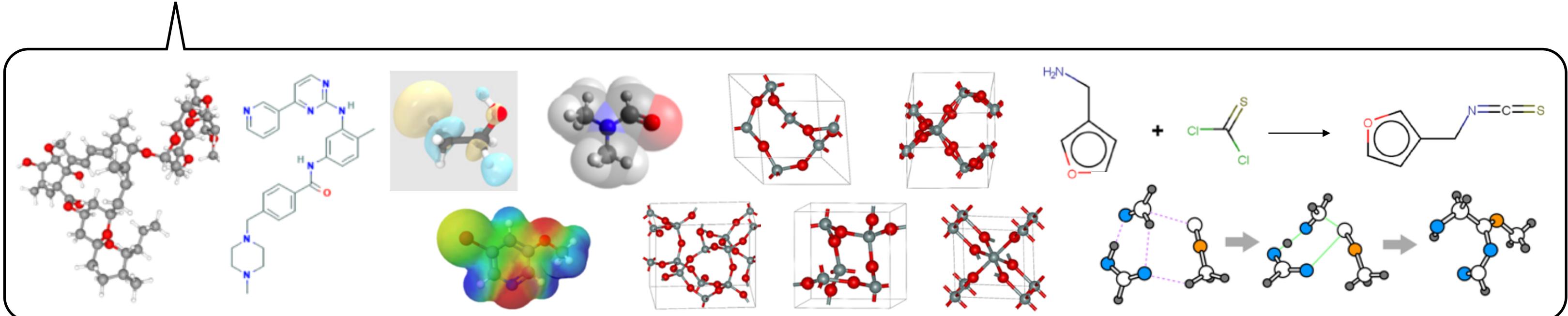
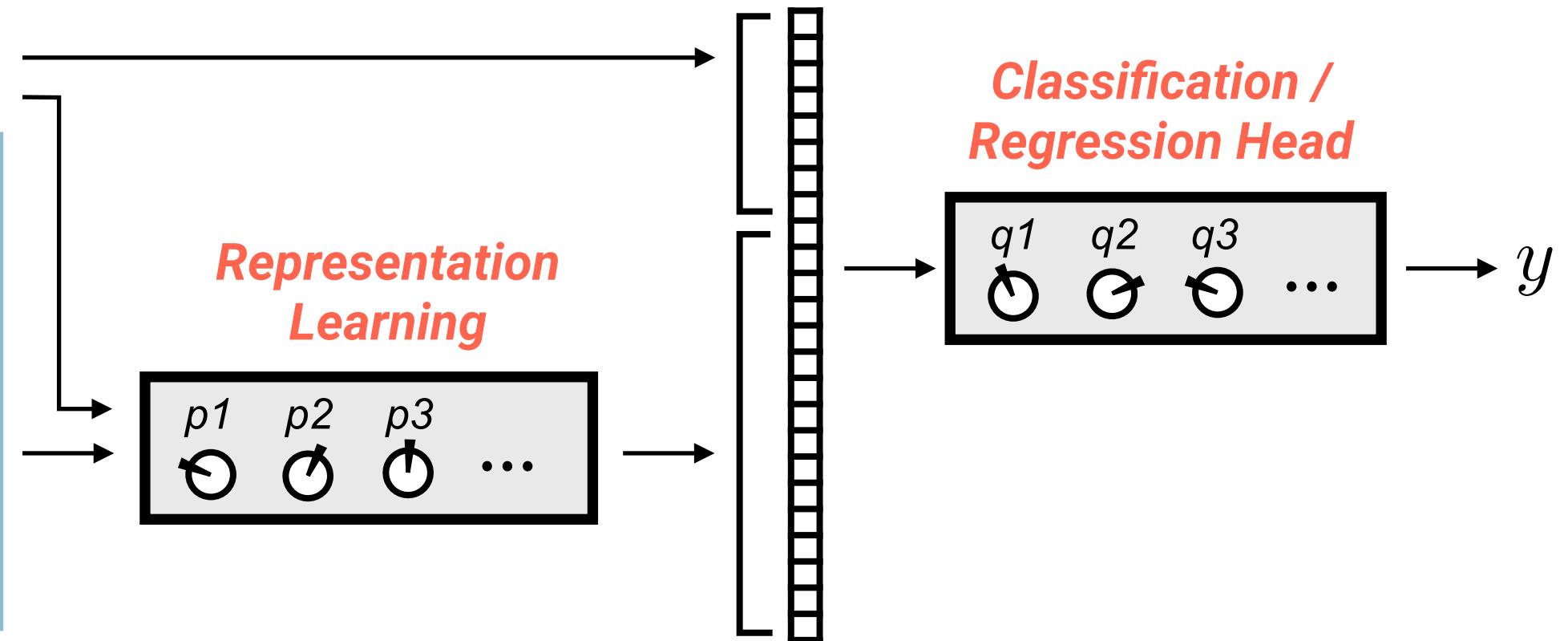
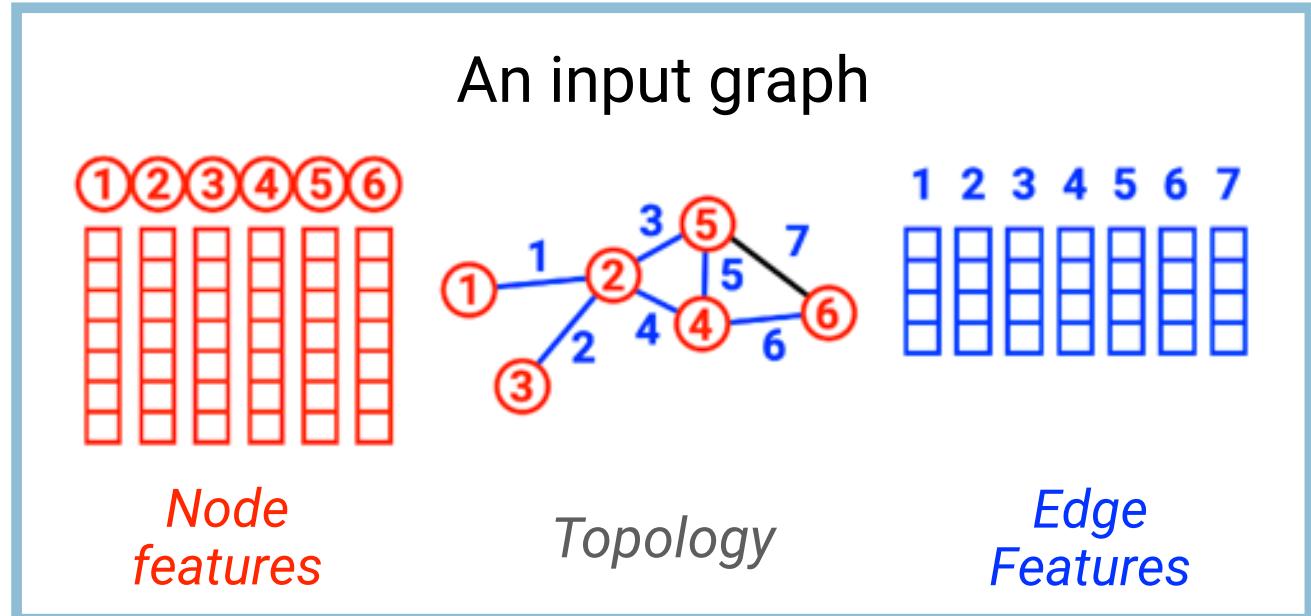
Dighe P, Adya S, Li N, Vishnubhotla S, Naik D, Sagar A, et al. [Lattice-Based Improvements for Voice Triggering Using Graph Neural Networks](#). ICASSP 2020

## Knowledge Collection (Amazon)

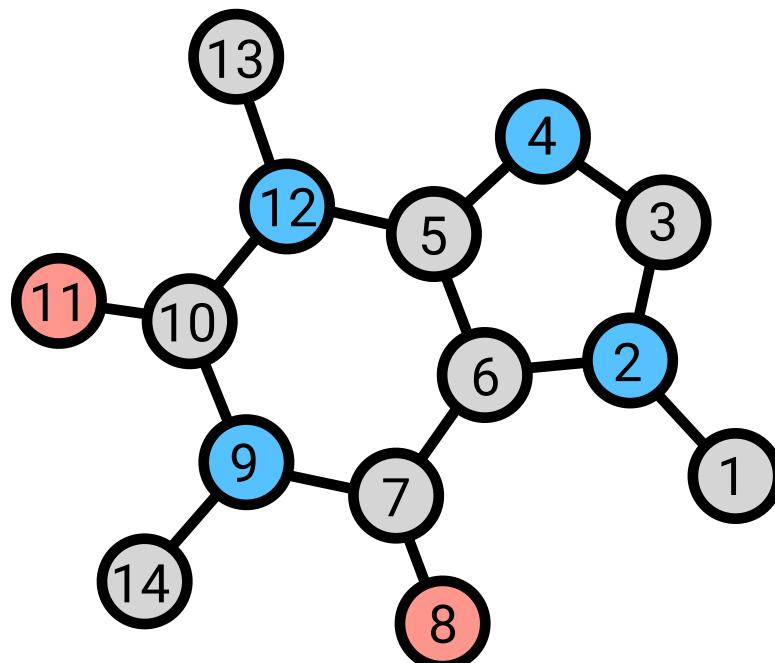
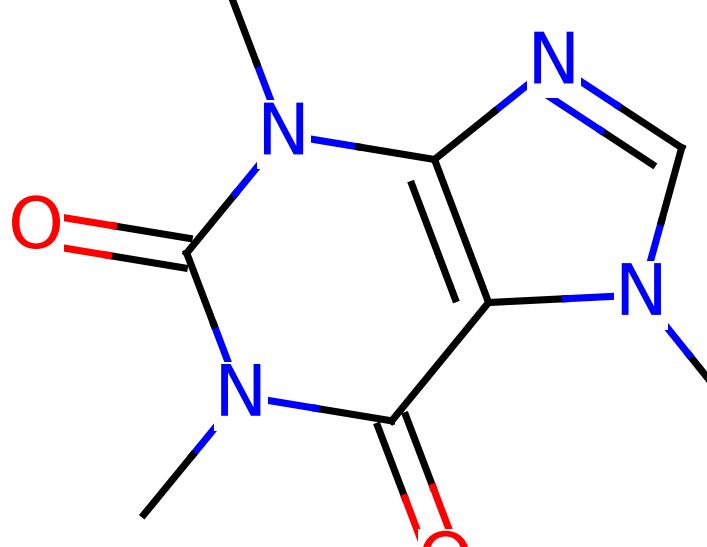
Dong XL, He X, Kan A, Li X, Liang Y, Ma J, et al. [AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types](#). KDD 2020

# Graph neural networks for chemical problems

Other Info (Conditions, Environment, ...)



# Molecular graphs



## Node features

1. Atomic number
2. # of directly-bonded neighbors
3. # of hydrogens
4. Formal charge
5. Atomic mass
6. Is in a ring?

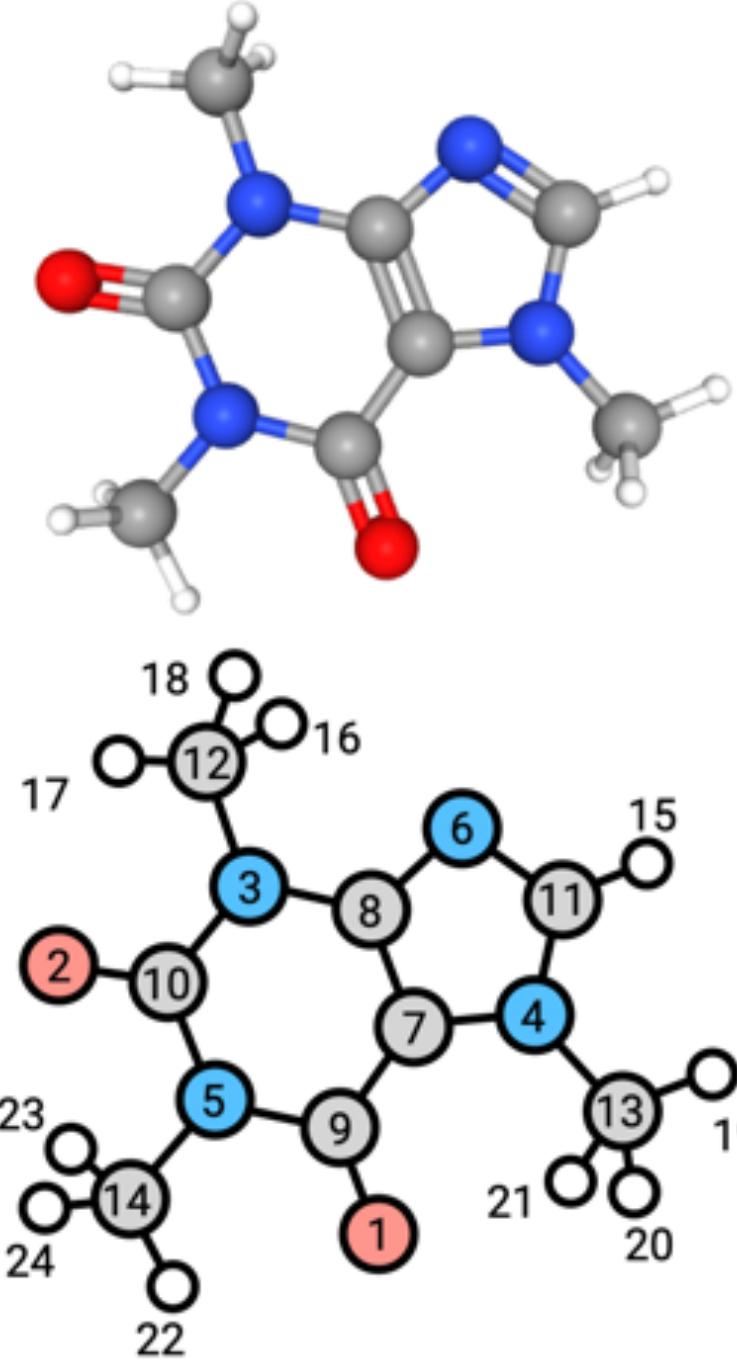
		1.	2.	3.	4.	5.	6.
1	C	6	4	3	0	12.011	0
2	N	7	3	0	0	14.007	1
3	C	6	3	1	0	12.011	1
4	N	7	2	0	0	14.007	1
5	C	6	3	0	0	12.011	1
6	C	6	3	0	0	12.011	1
7	C	6	3	0	0	12.011	1
8	O	8	1	0	0	15.999	0
9	N	7	3	0	0	14.007	1
10	C	6	3	0	0	12.011	1
11	O	8	1	0	0	15.999	0
12	N	7	3	0	0	14.007	1
13	C	6	4	3	0	12.011	0
14	C	6	4	3	0	12.011	0

## Edge features

1. Bond type
2. Stereochemistry

		1.	2.
1	2	1	0
2	3	12	0
3	4	12	0
4	5	12	0
5	6	12	0
6	7	12	0
7	8	2	0
7	9	12	0
9	10	12	0
10	11	2	0
10	12	12	0
12	13	1	0
9	14	1	0
6	2	12	0
12	5	12	0

# Geometric graphs



Nodes

		Z	x	y	z
1	O	8	0.470	2.569	0.001
2	O	8	-3.127	-0.444	-0.000
3	N	7	-0.969	-1.313	0.000
4	N	7	2.218	0.141	-0.000
5	N	7	-1.348	1.080	-0.000
6	N	7	1.412	-1.937	0.000
7	C	6	0.858	0.259	-0.001
8	C	6	0.390	-1.026	-0.000
9	C	6	0.031	1.422	-0.001
10	C	6	-1.906	-0.250	-0.000
11	C	6	2.503	-1.200	0.000
12	C	6	-1.428	-2.696	0.001
13	C	6	3.193	1.206	0.000
14	C	6	-2.297	2.188	0.001
15	H	1	3.516	-1.579	0.001
16	H	1	-1.045	-3.197	-0.894
17	H	1	-2.519	-2.760	0.001
18	H	1	-1.045	-3.196	0.896
19	H	1	4.199	0.780	0.000
20	H	1	3.047	1.809	-0.899
21	H	1	3.047	1.808	0.900
22	H	1	-1.809	3.165	-0.000
23	H	1	-2.932	2.103	0.888
24	H	1	-2.935	2.102	-0.885

Edges

		type	distance
1	9	2	1.228
2	10	2	1.236
3	8	1	1.388
3	10	1	1.417
3	12	1	1.458
4	7	1	1.365
4	11	1	1.371
4	13	1	1.443
5	9	1	1.420
5	10	1	1.442
5	14	1	1.459
6	8	1	1.369
6	11	2	1.317
7	8	2	1.368
7	9	1	1.427
11	15	1	1.082
12	16	1	1.094
12	17	1	1.093
12	18	1	1.094
13	19	1	1.093
13	20	1	1.093
13	21	1	1.093
14	22	1	1.092
14	23	1	1.095
14	24	1	1.095

+ Can be added

Non-geometric node features

Non-geometric edge features

# Geometric Deep Learning (GDL)

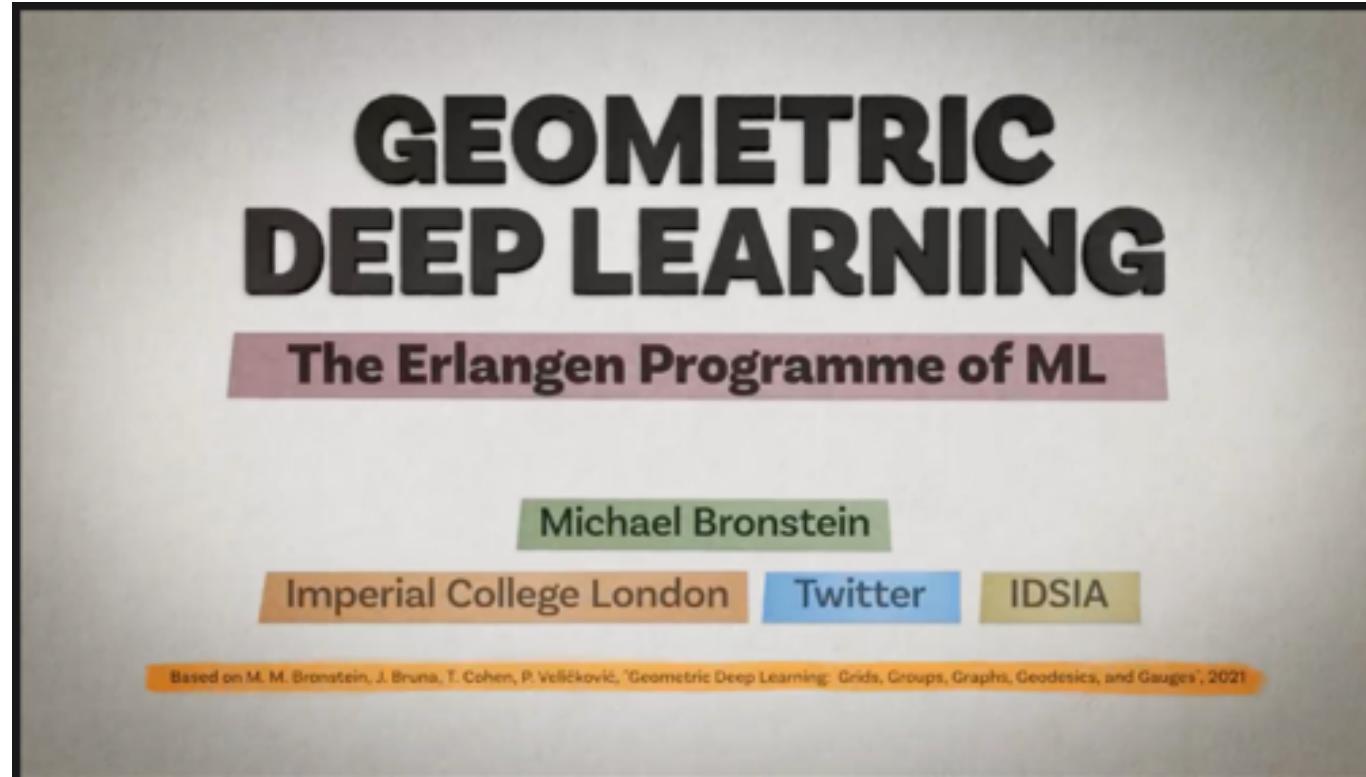
<https://geometricdeeplearning.com>

Bronstein MM, Bruna J, Cohen T, Veličković P.

[Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges.](#)

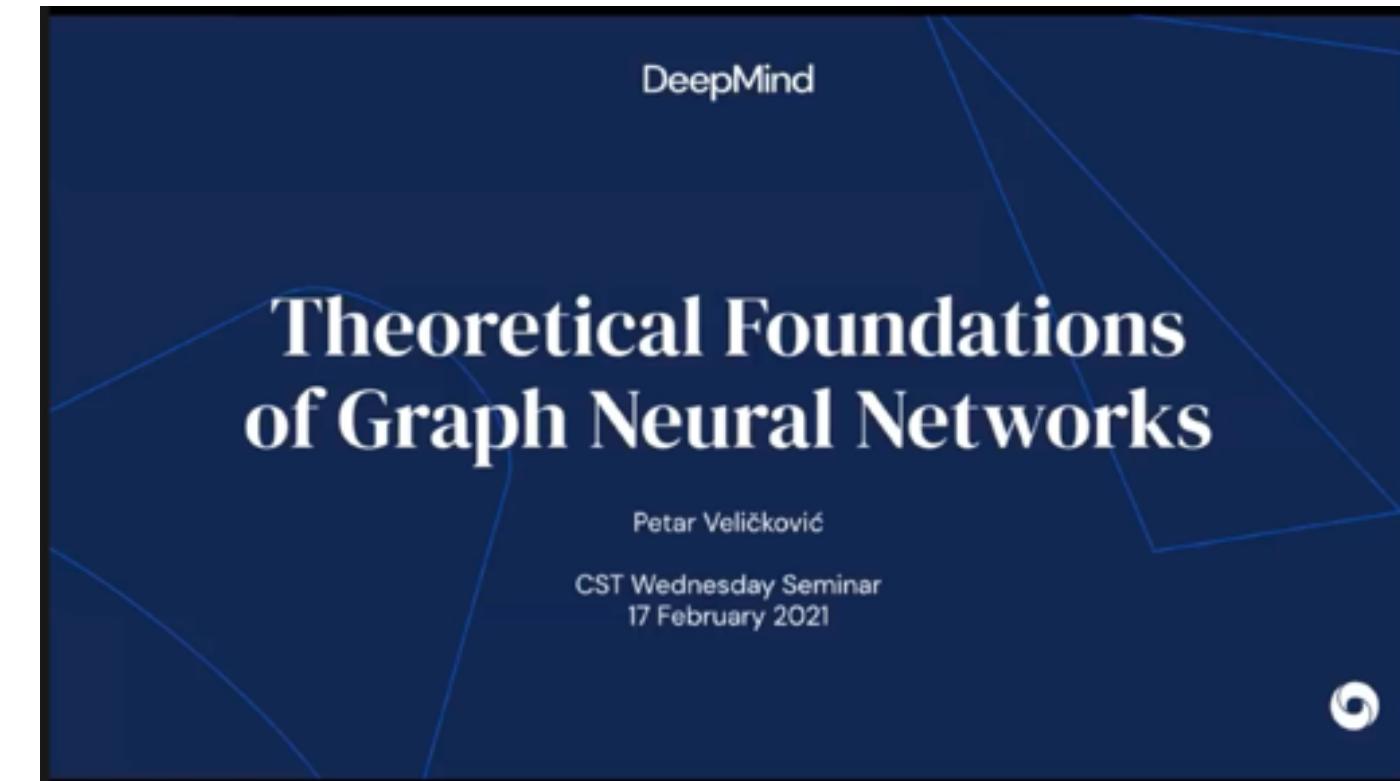
arXiv [cs.LG]. 2021. <http://arxiv.org/abs/2104.13478>

*ICLR 2021 Keynote (Michael Bronstein)*

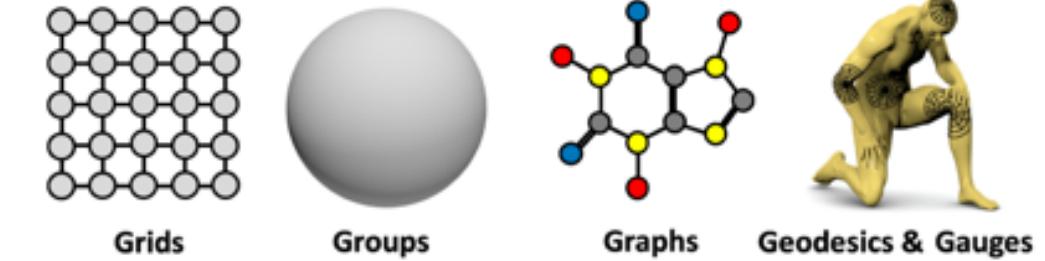


<https://youtu.be/w6Pw4MOzMuo>

*Seminar Talk (Petar Veličković)*



<https://youtu.be/uF53xsT7mjc>

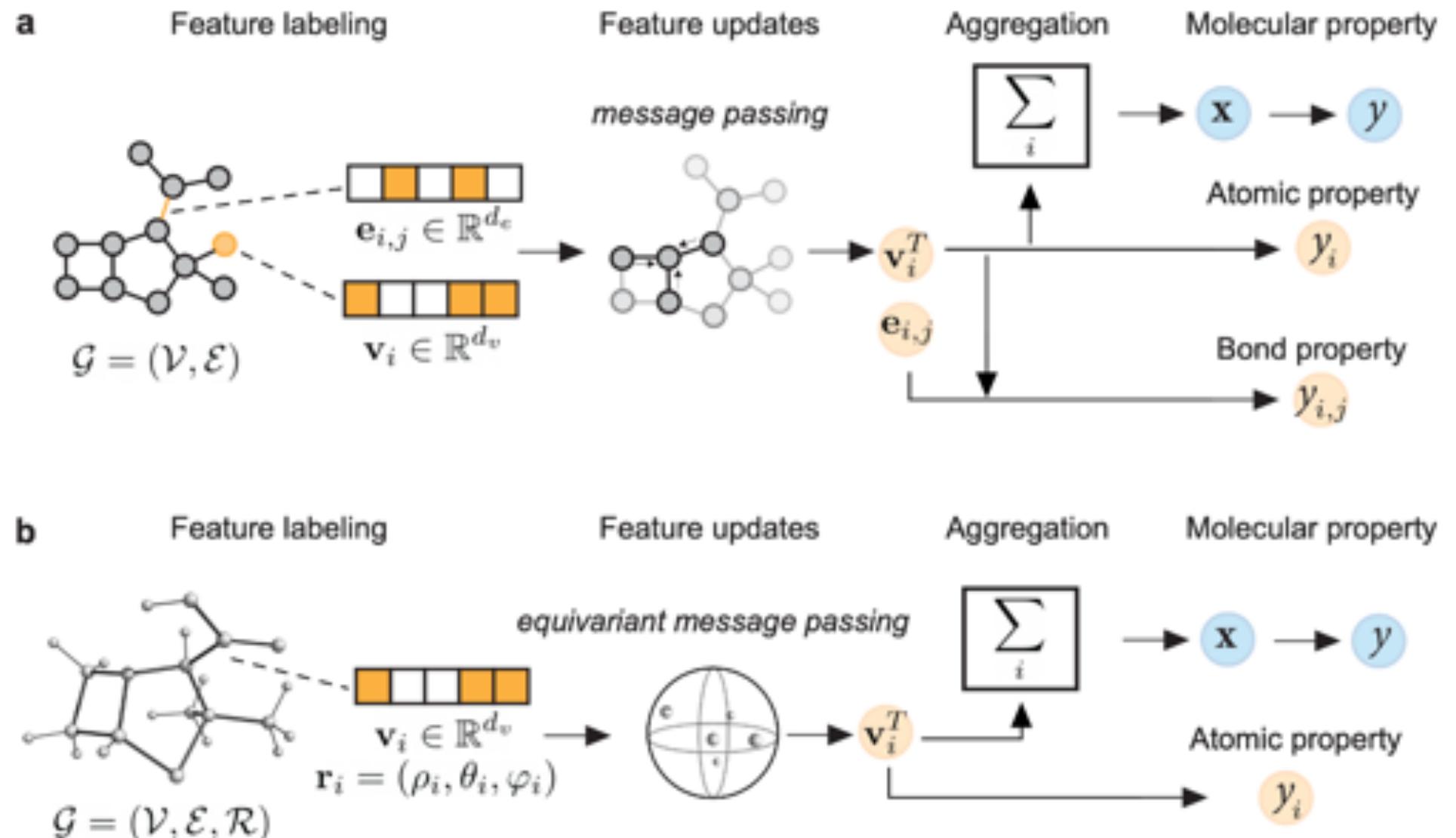


# GDL for molecular representations

Atz K, Grisoni F, Schneider G.

Geometric deep learning on molecular representations.

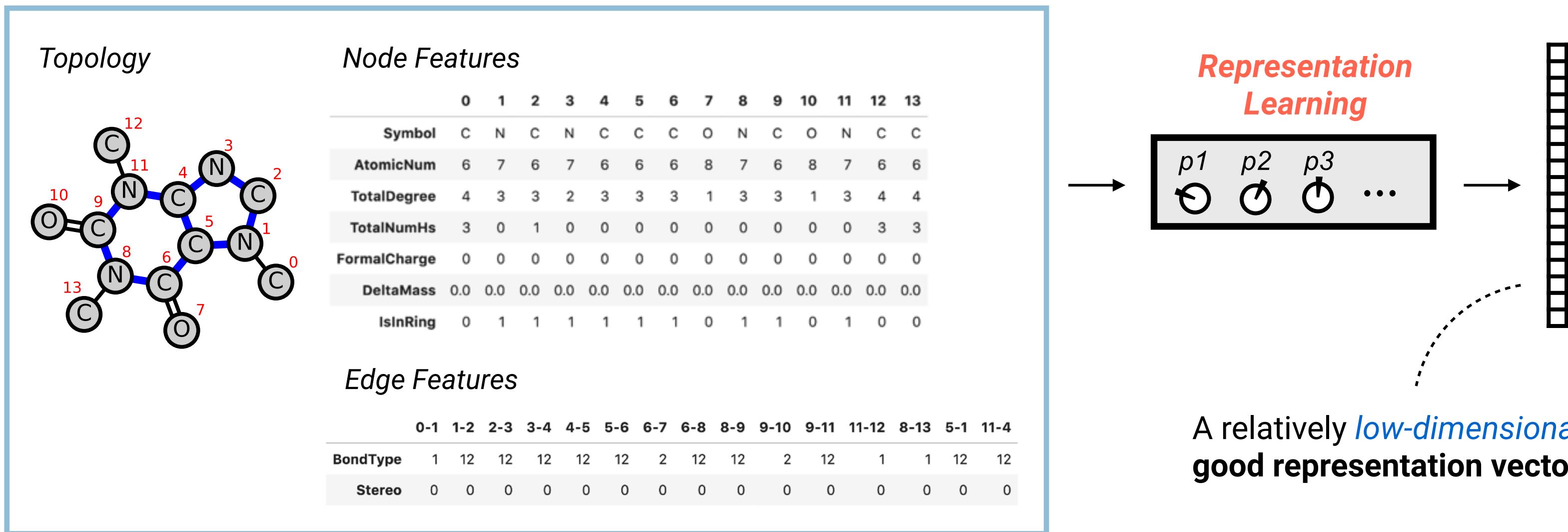
Nature Machine Intelligence. 2021;3: 1023–1032. <https://arxiv.org/abs/2107.12375>



We need to consider equivariance under E(3) or SE(3) for some geometric features (coordinates, forces, vector field, etc)

# Graph Representation Learning

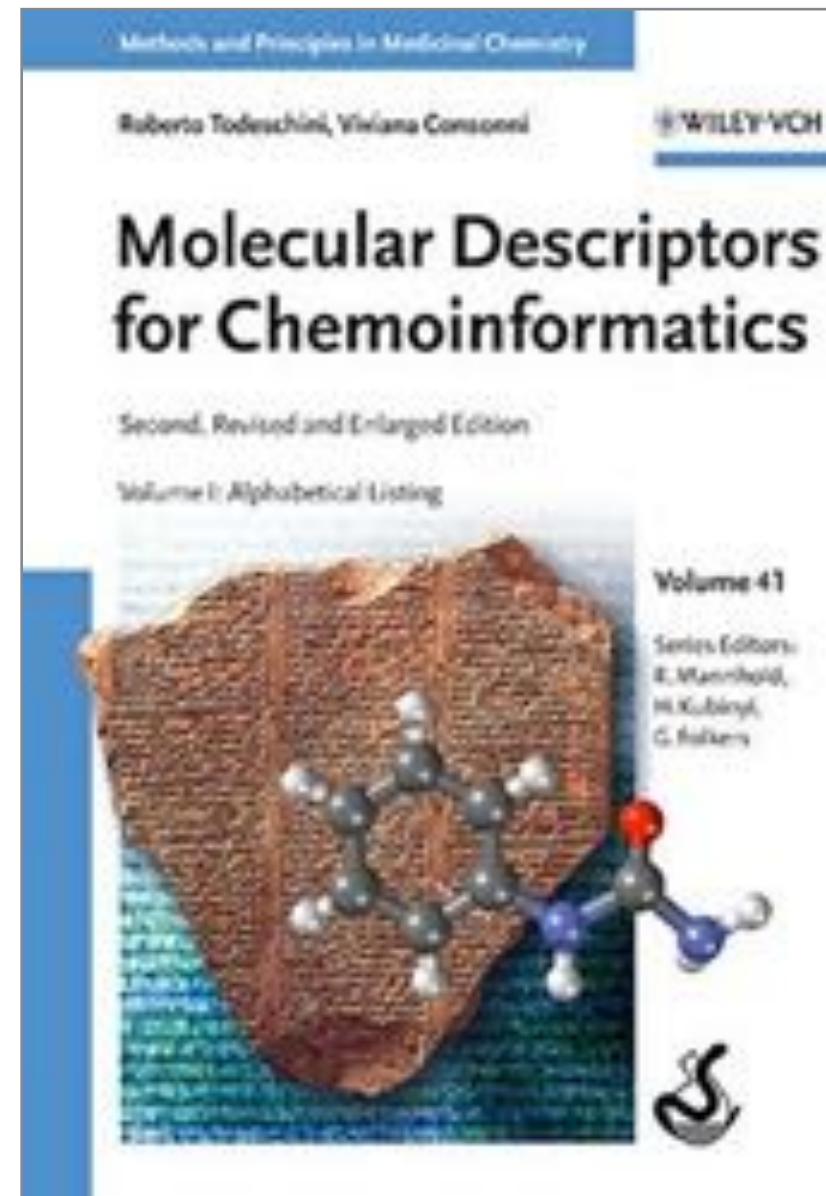
We can seek for **a good representation vector** that can be computed from a molecular graph, which is expected to be superior to any man-made descriptors!



# Beyond man-made descriptors

- **0-Dimensional Descriptors**
  - Constitutional descriptors
  - Count descriptors
- **1-Dimensional Descriptors**
  - List of structural fragments
  - Fingerprints
- **2-Dimensional Descriptors**
  - Graph invariants
- **3-Dimensional Descriptors**
  - 3D MoRSE, WHIM, GETAWAY, ...
  - Quantum-chemical descriptors
  - Size, steric, surface, volume, ...
- **4-Dimensional Descriptors**
  - GRID, CoMFA, Volsurf, ...

"Vol 1 contains an alphabetical listing of **more than 3,300 descriptors**"

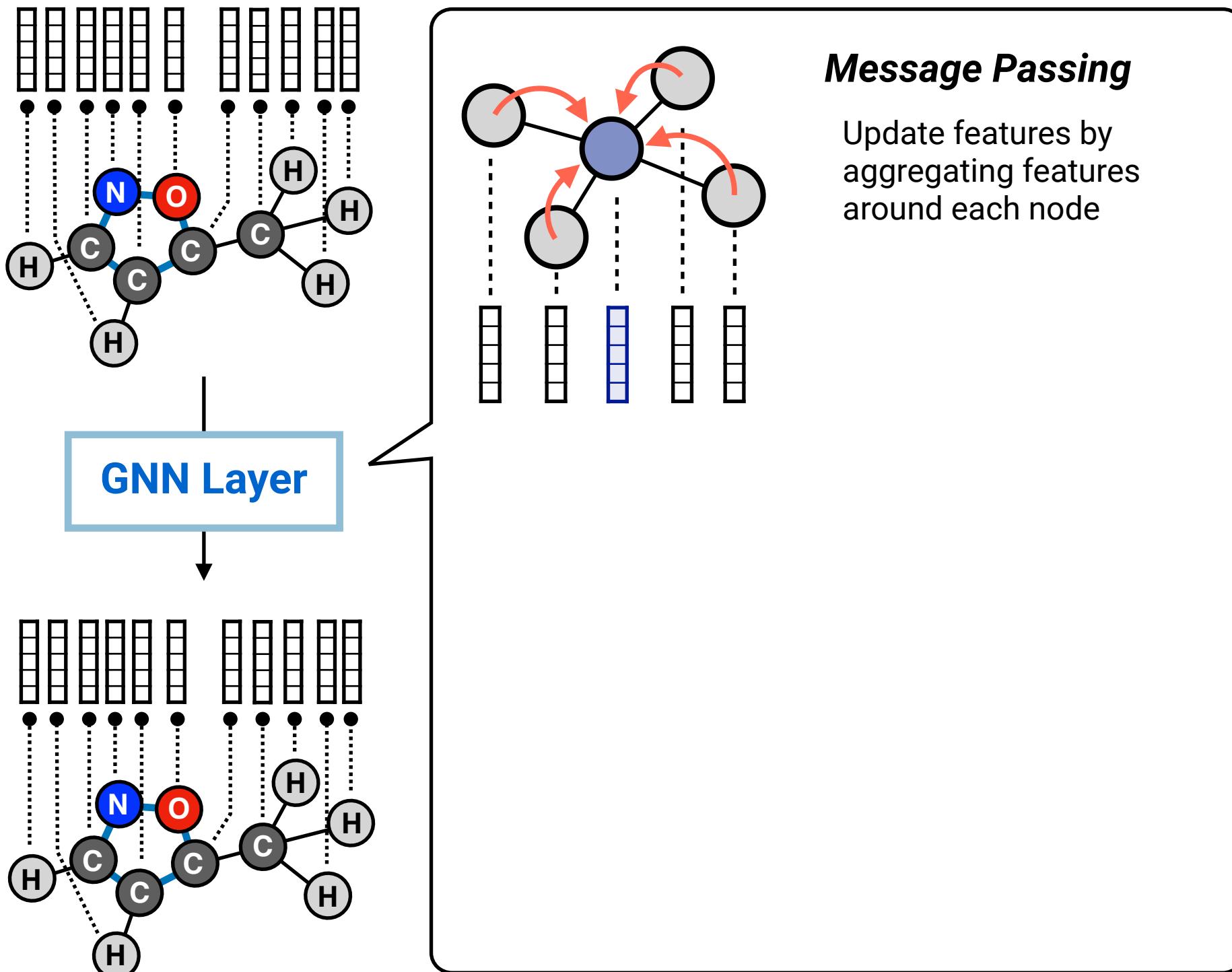


"Dragon calculates **5,270** molecular descriptors"

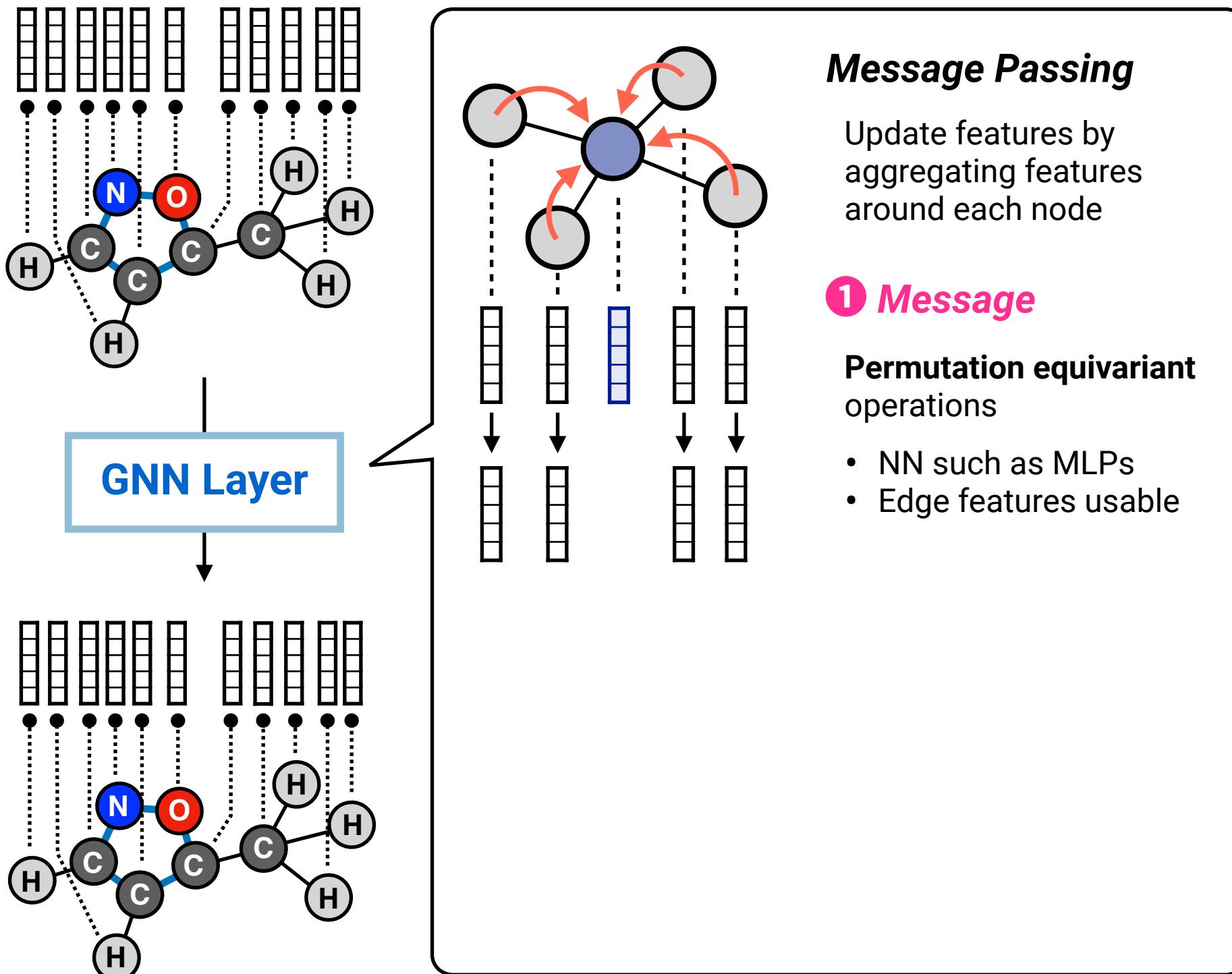
The screenshot shows the DRAGON 7.0 software interface. At the top, there's a logo for "kode chemoinformatics" and the product name "DRAGON 7.0". Below the logo is a navigation bar with tabs for "DESCRIPTORS", "FINGERPRINTS", and "PROJECT AND SCIENTIFIC WORKS". The "DESCRIPTORS" tab is selected. The main area displays a table of molecular descriptor blocks. The table has columns for "BLOCK NO", "BLOCK NAME", and "DESCRIPTORS". The data is as follows:

BLOCK NO	BLOCK NAME	DESCRIPTORS
1	Constitutional	47
2	Ring descriptors	32
3	Topological indices	75
4	Walk and path counts	46
5	Connectivity indices	37
6	Information indices	50
7	2D matrix-based descriptors	607
8	2D autocorrelations	213
9	Burden eigenvalues	96
10	P-VSA-like descriptors	55

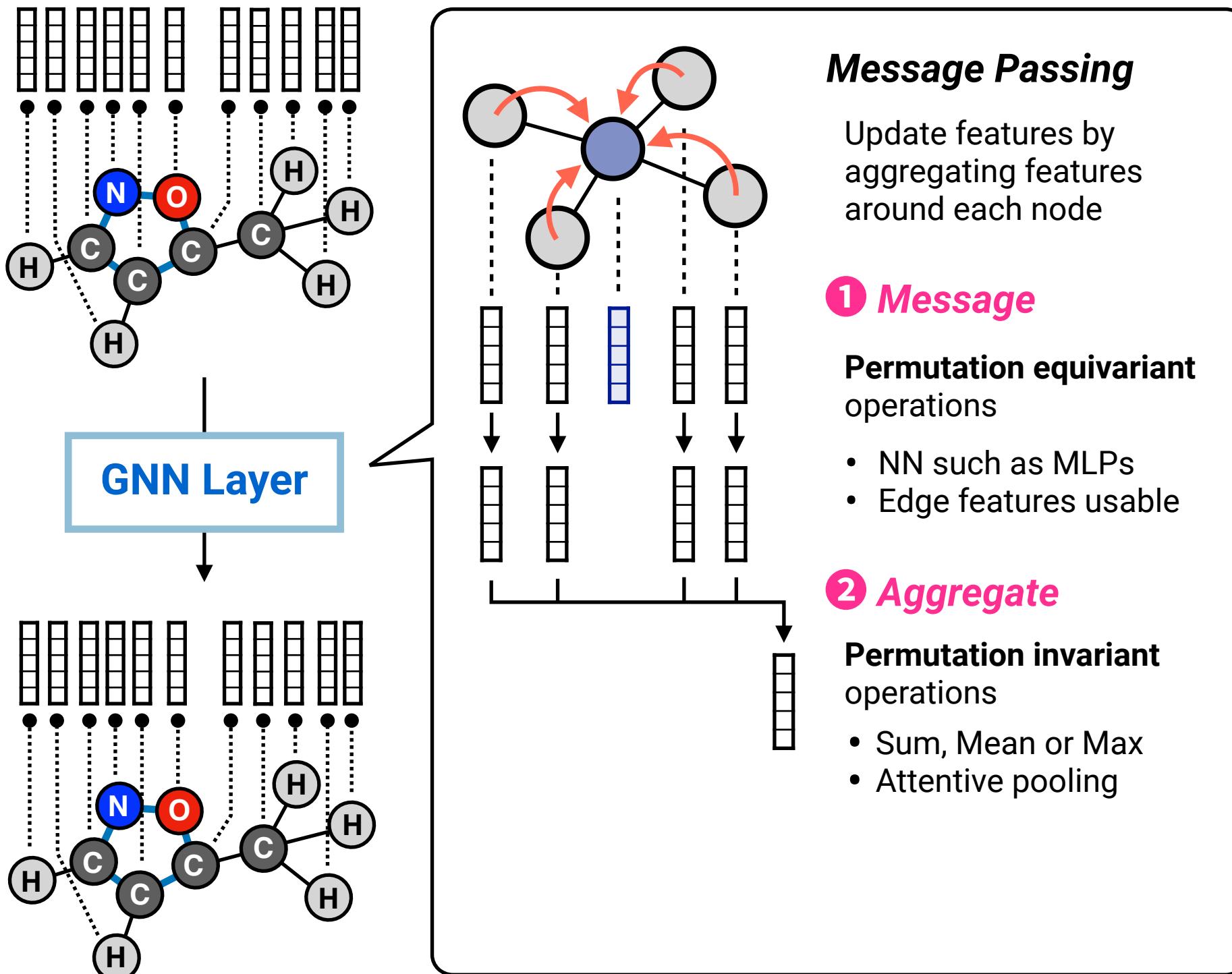
# Message Passing: The inner workings of GNNs



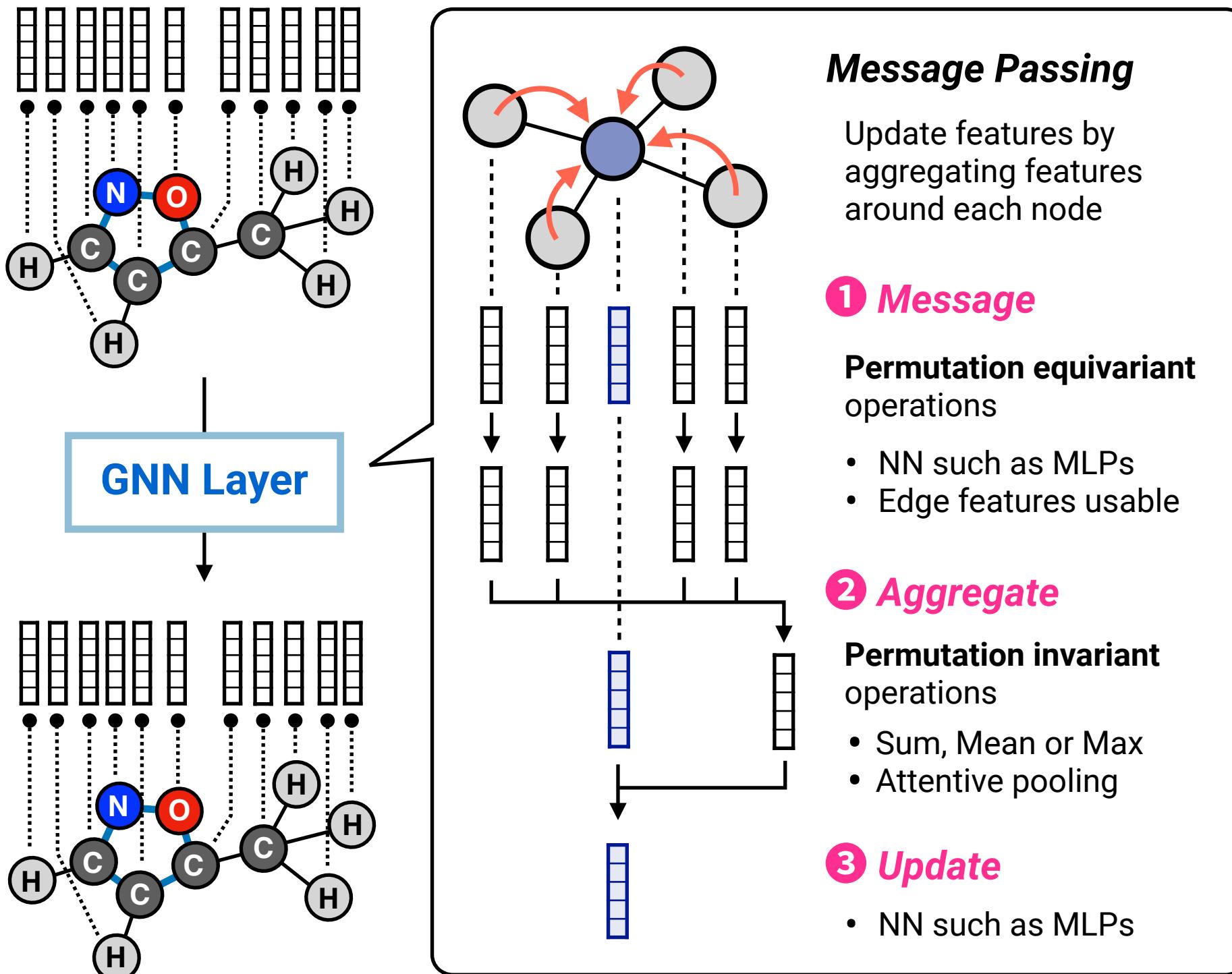
# Message Passing: The inner workings of GNNs



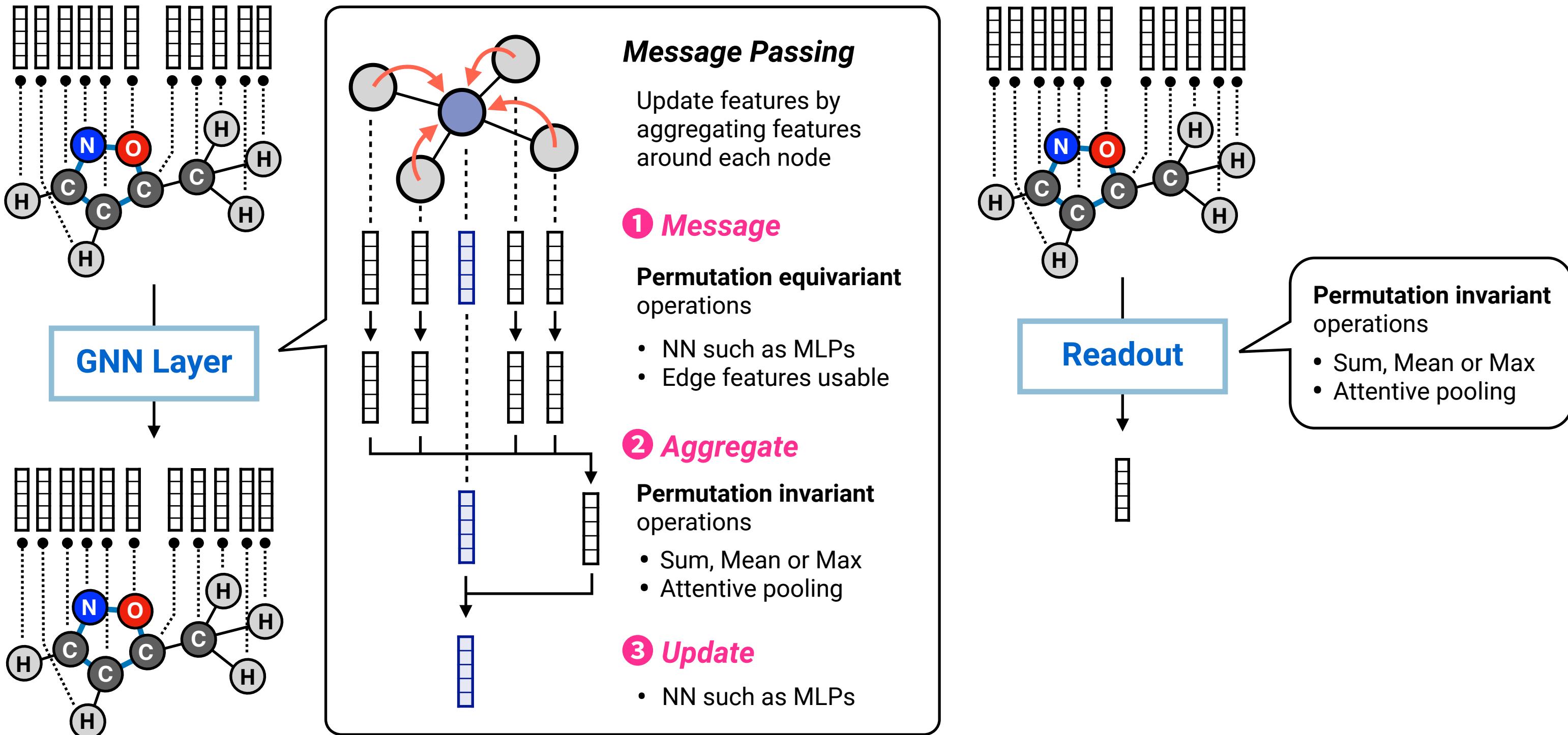
# Message Passing: The inner workings of GNNs



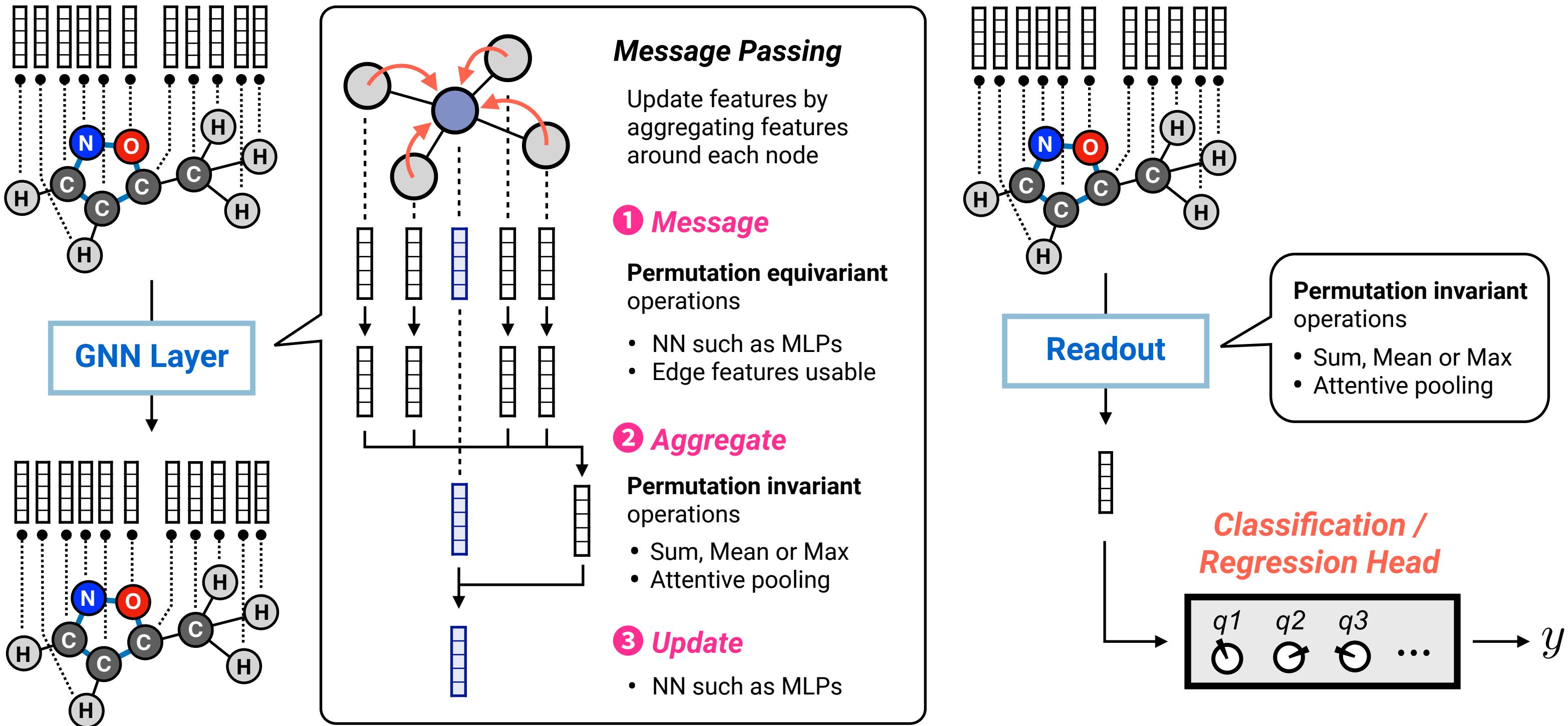
# Message Passing: The inner workings of GNNs



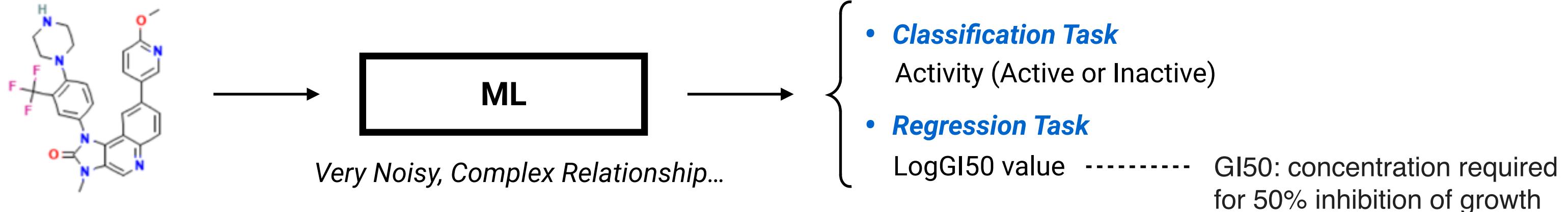
# Message Passing: The inner workings of GNNs



# Message Passing: The inner workings of GNNs



# Use Case 1: Virtual Screening (QSAR/QSPR)



**NCI Human Tumor Cell Line Growth Inhibition Assay (PubChem AID 1)**

Active (2,814)			Inactive (48,922)		
Structure	CID	SID	Activity	Score	LogGI50_M
	5298	121832	Active	67	-8
	363173	493713	Active	43	-6.5871
	399631	530868	Active	51	-7.0678
	390324	521601	Inactive	0	-4
	390311	521588	Inactive	0	-4
	390312	521589	Inactive	4	-4.214

# Use Case 1: Virtual Screening (QSAR/QSPR)

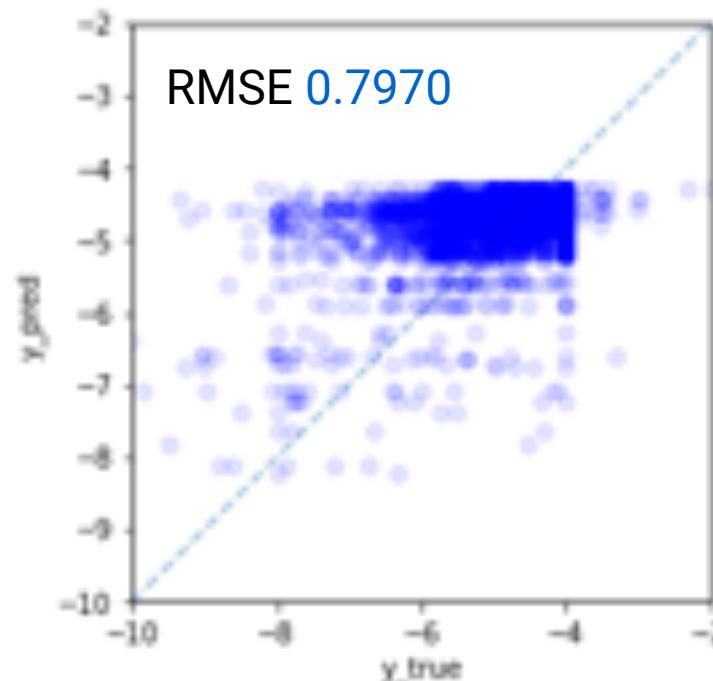
## Standard ML

ExtraTrees  
w/ ECFP6(1024)

- **Classification Task** Activity (Active or Inactive)

95.079%

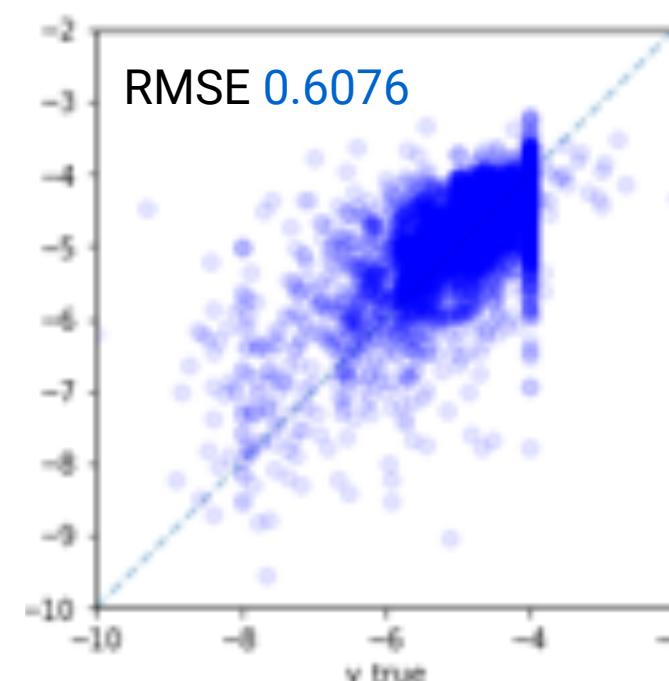
- **Regression Task** LogGI50 value



## GNN

ChemProp  
(Directed MPNN)

95.604%



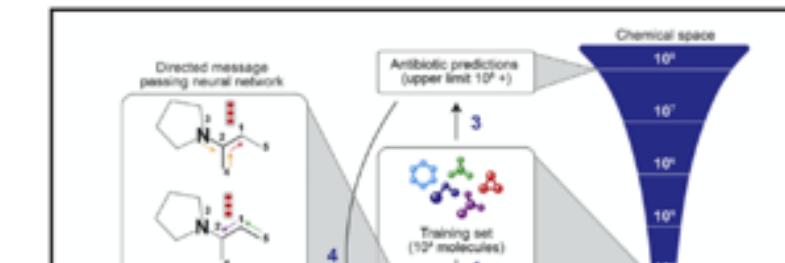
## ChemProp (Yang et al, 2019)

from MIT MLPDS (Machine Learning for Pharmaceutical Discovery and Synthesis) Consortium

## Cell

### A Deep Learning Approach to Antibiotic Discovery

#### Graphical Abstract



#### Authors

Jonathan M. Stokes, Kevin Yang,  
Kyle Swanson, ..., Tommi S. Jaakkola,  
Regina Barzilay, James J. Collins

#### Correspondence

regina@csail.mit.edu (R.B.),  
jimjc@mit.edu (J.J.C.)

## nature

NEWS | 20 February 2020

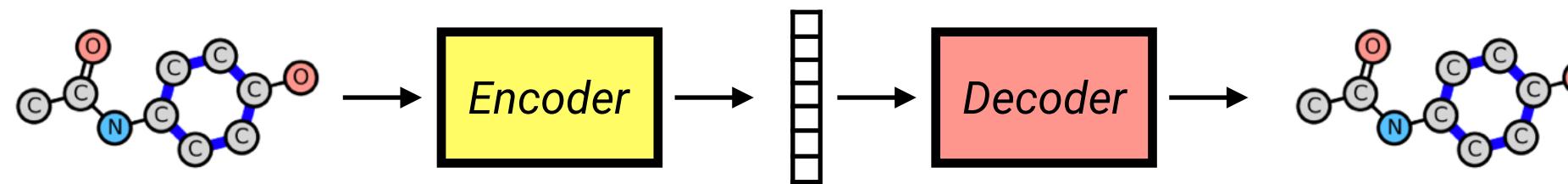
### Powerful antibiotics discovered using AI

Machine learning spots molecules that work even against 'untreatable' strains of bacteria.

Jo Marchant

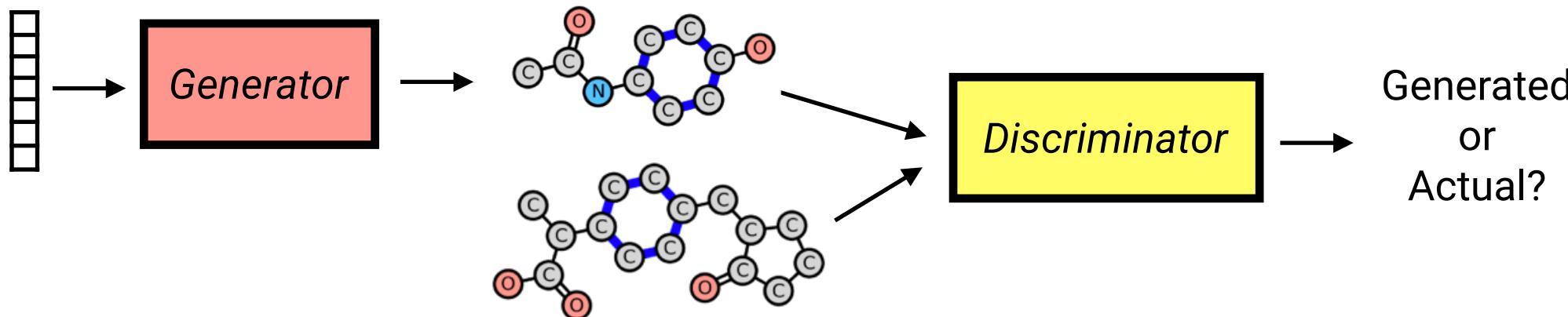
# Use Case 2: Molecule Generation

- Autoencoder-based / Flow-based



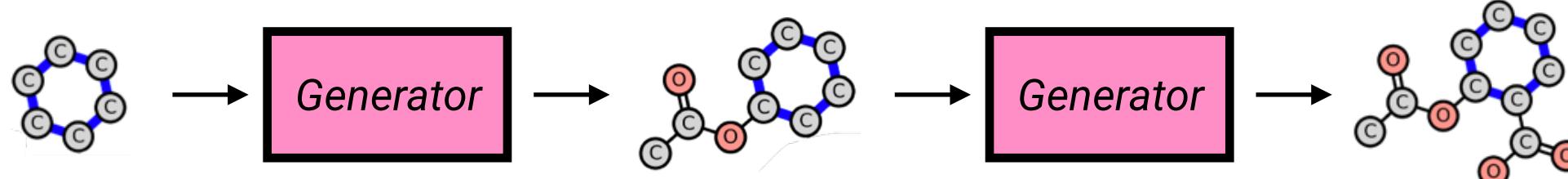
Faez F, Ommi Y, Baghshah MS, Rabiee HR.  
[Deep Graph Generators: A Survey.](#)  
*IEEE Access.* 2021;9: 106675–106702.

- Adversarial



GNNs are important building block to design generation modules.

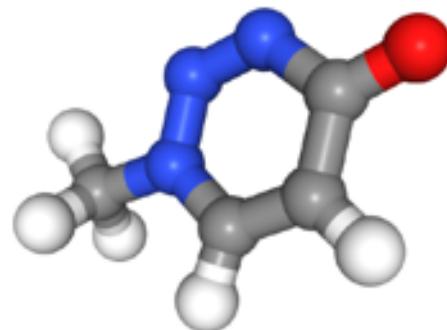
- Autoregressive
- Reinforcement-learning-based



The design patterns to generate images, sounds, texts, are quite useful!

# Use Case 3: Fast Approximation for QM Calculations

input



gdb\_21014

	x	y	z
O	0.314096	-0.129589	-0.389150
C	0.111219	2.102676	-0.051749
C	2.331344	3.941075	0.212303
O	4.667017	2.677399	0.437948
C	6.152491	3.062553	-1.780599
C	4.732264	5.009654	-3.282819
C	2.562527	5.549427	-2.143825
H	-1.771427	3.048695	0.071772
H	1.977918	5.086871	1.919865
H	8.050245	3.696867	-1.222422
H	6.372399	1.276980	-2.825015
H	5.428656	5.805758	-5.033531
H	1.118529	6.857080	-2.763050

Quantum mechanical calculations

~ 1000 sec

by solving a one-electron Schrödinger equation (Kohn–Sham equation)

$$\hat{H}\Psi = E\Psi$$

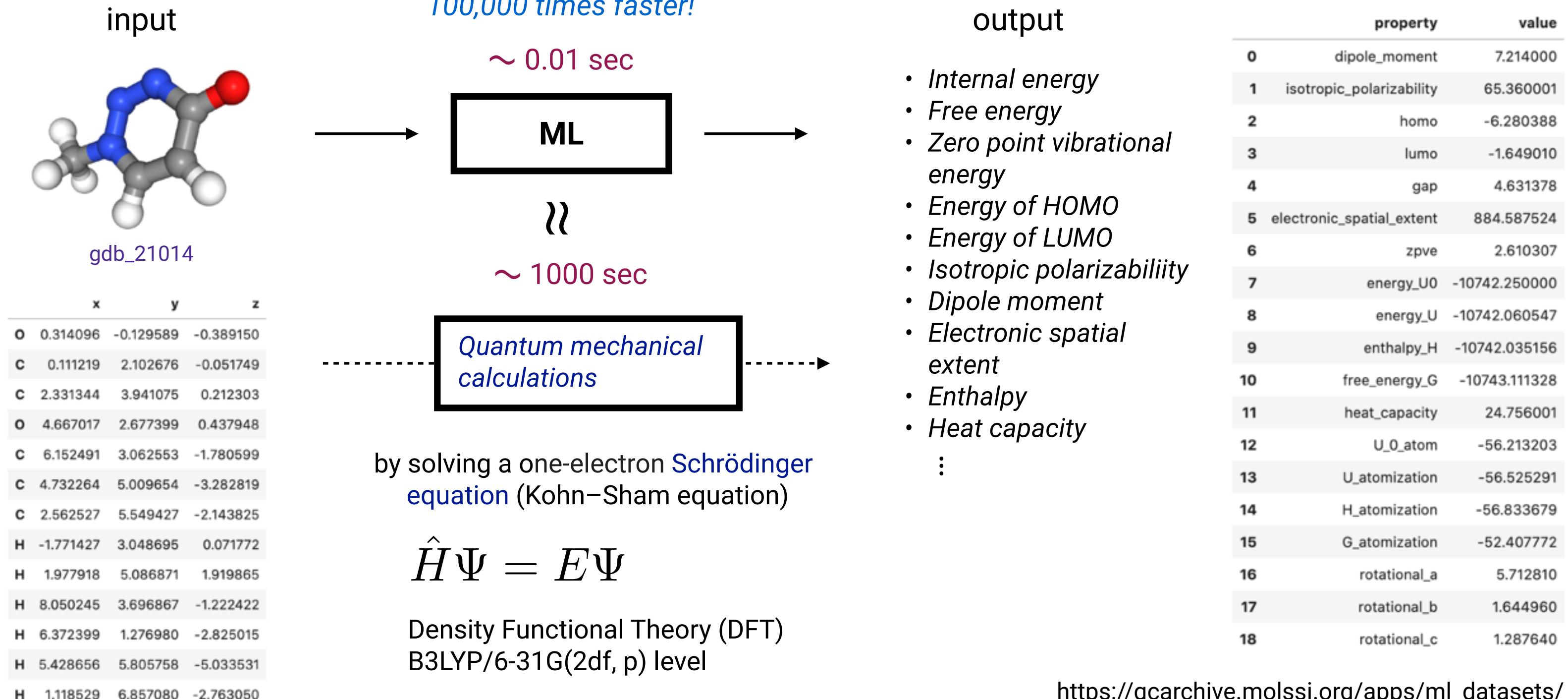
Density Functional Theory (DFT)  
B3LYP/6-31G(2df, p) level

output

- Internal energy
- Free energy
- Zero point vibrational energy
- Energy of HOMO
- Energy of LUMO
- Isotropic polarizability
- Dipole moment
- Electronic spatial extent
- Enthalpy
- Heat capacity
- ⋮

	property	value
0	dipole_moment	7.214000
1	isotropic_polarizability	65.360001
2	homo	-6.280388
3	lumo	-1.649010
4	gap	4.631378
5	electronic_spatial_extent	884.587524
6	zpve	2.610307
7	energy_U0	-10742.250000
8	energy_U	-10742.060547
9	enthalpy_H	-10742.035156
10	free_energy_G	-10743.111328
11	heat_capacity	24.756001
12	U_0_atom	-56.213203
13	U_atomization	-56.525291
14	H_atomization	-56.833679
15	G_atomization	-52.407772
16	rotational_a	5.712810
17	rotational_b	1.644960
18	rotational_c	1.287640

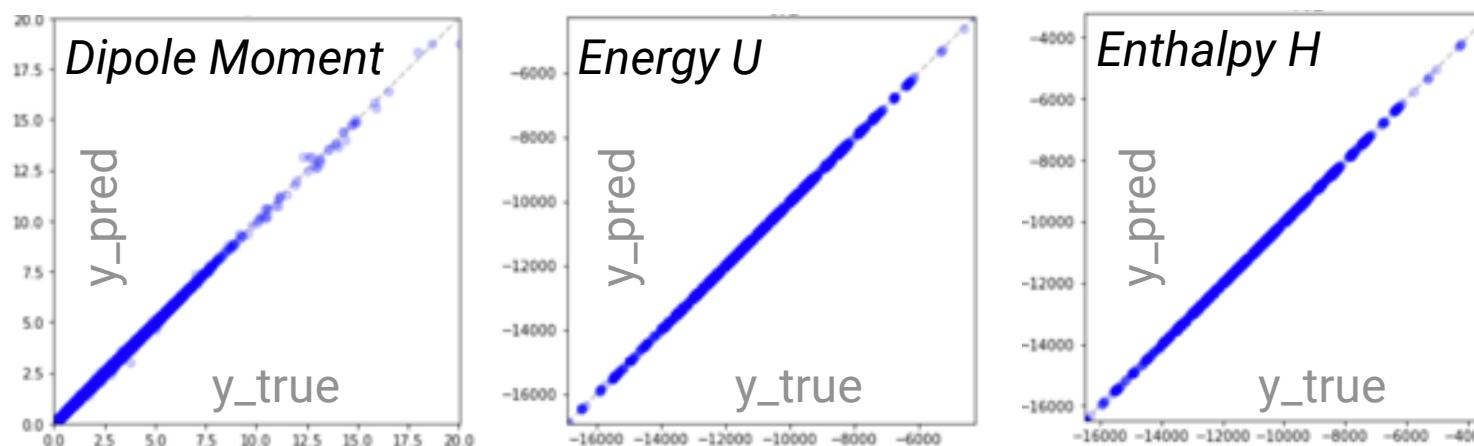
# Use Case 3: Fast Approximation for QM Calculations



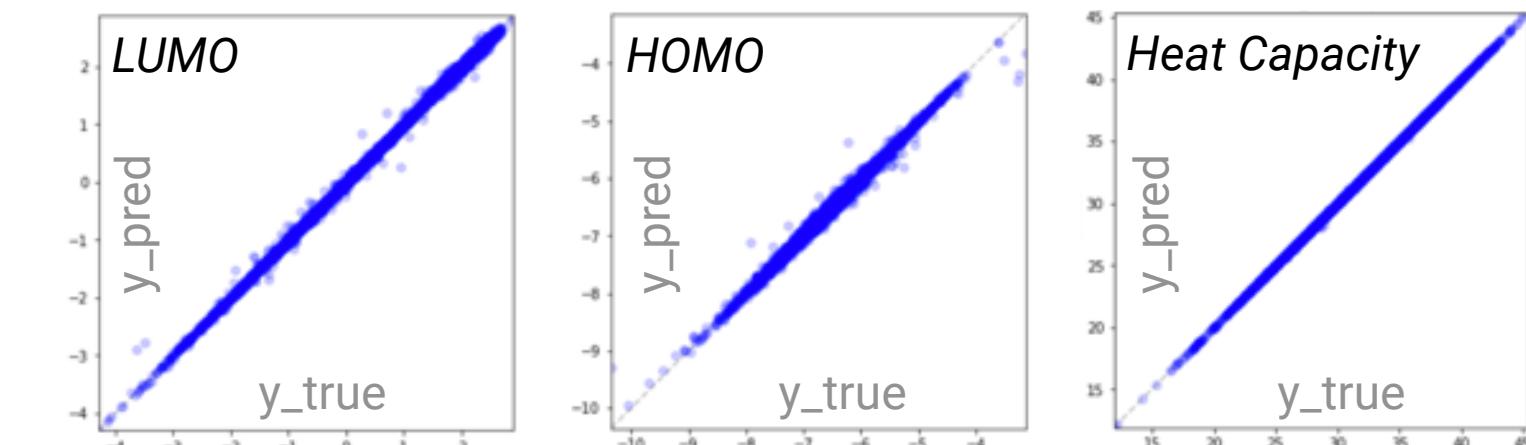
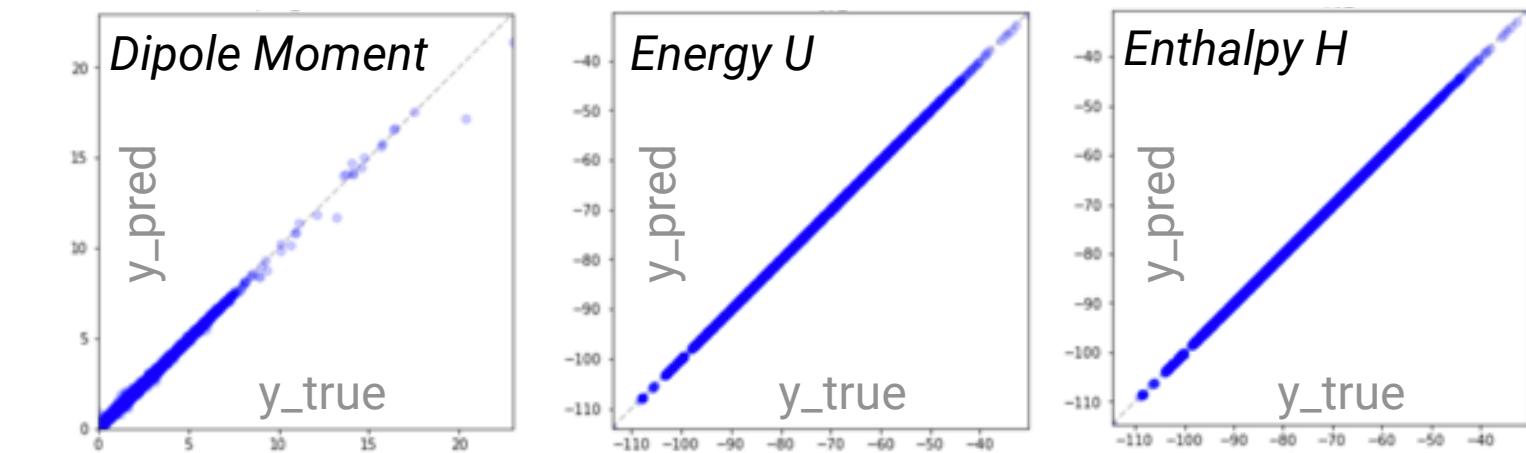
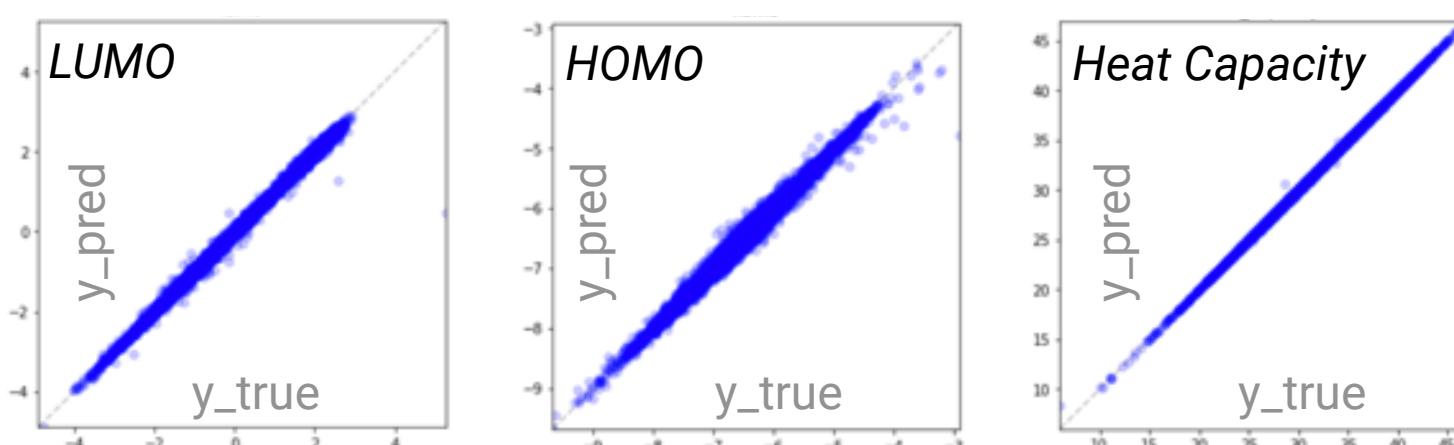
# Use Case 3: Fast Approximation for QM Calculations

GNN predictions are **strikingly accurate**, in particular, for predicting **energies** of a molecule of a conformation or **forces at each atom** to transition towards a more stable conformation!

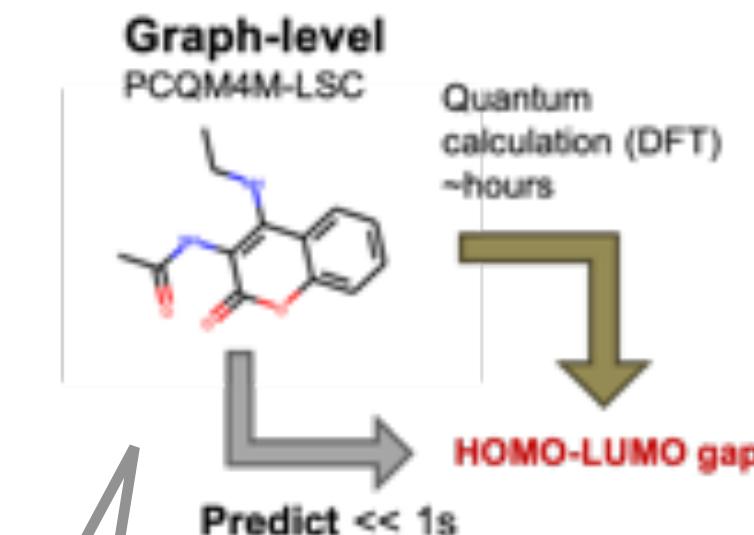
Predictions for Test Data by **SchNet** (Schütt et al, 2017)



Predictions for Test Data by **DimeNet** (Klicpera et al, 2020)



# OGB Large-Scale Challenge (KDDCup 2021)



**Graph Regression:** to predict the DFT-calculated HOMO-LUMO from 2D molecular graphs

Dataset: 3,803,453 graphs from PubChemQC (cf. QM9 = 133,885 graphs)

Rank	Test MAE	Models
1	0.1200 (eV)	<b>10 × GNNs</b> (12-layer <b>Graphomer</b> ) + <b>8 × ExpC*s</b> (5-layer ExpandingConv)
2	0.1204 (eV)	<b>73 × GNNs</b> (11-layer LiteGEMConv + <b>SSL pretraining</b> )
3	0.1205 (eV)	<b>20 × GNNs</b> (32-layer GNN + Noisy Nodes)

# Pretrained models drive applied DL (ImageNet, BERT, ...)

---

Many practical applications lack large data, and *the use of pretrained models* are critical.

Especially, the models that are **pre-trained on broad data at scale** and are adaptable to a wide range of downstream tasks.

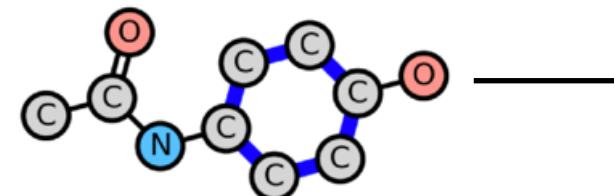
The screenshot shows a red header bar with the arXiv.org logo, a search bar, and navigation links for Help and Advanced. Below the header, the URL 'arXiv.org > cs > arXiv:2108.07258' is displayed. The main content area has a light gray background. It features the title 'Computer Science > Machine Learning' and a subtitle '(Submitted on 16 Aug 2021 (v1), last revised 18 Aug 2021 (this version, v2))'. The main title of the paper is 'On the Opportunities and Risks of Foundation Models'. Below the title is a long list of authors' names, including Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nifforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang et al. (14 additional authors not shown). The abstract begins with a paragraph about the paradigm shift of foundation models and their applications across various fields like law, healthcare, and education.

Call them  
"foundation models"?

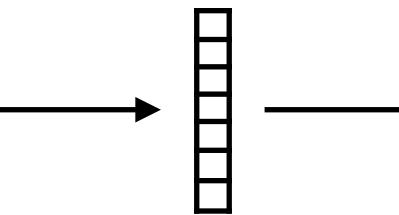
# Large-scale pretraining by self-supervised learning tasks

## Pretraining

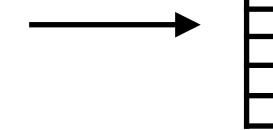
Large-scale Pretext task  
(often self-supervised)



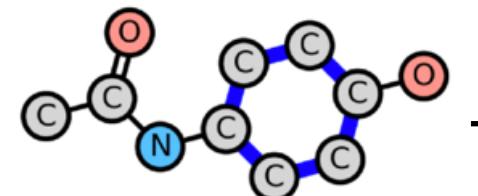
Representation learning module



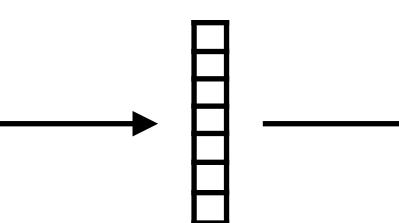
Head for pretraining



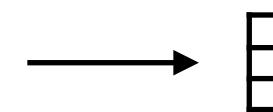
Target downstream task  
(small-data problem)



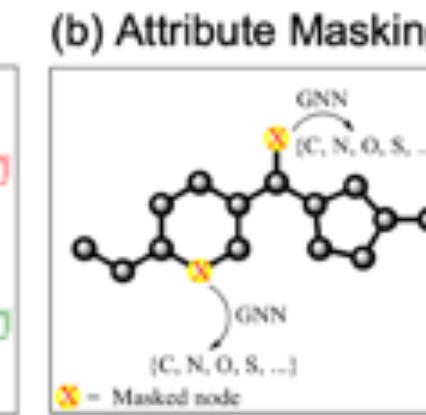
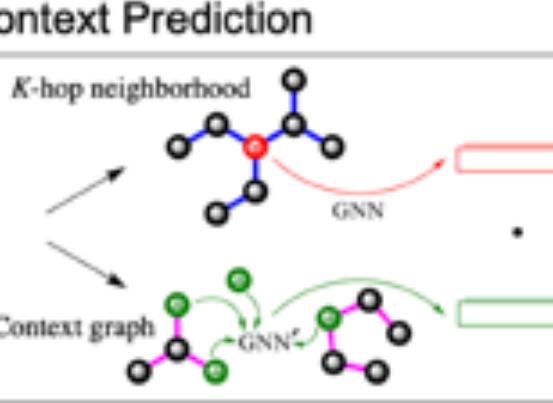
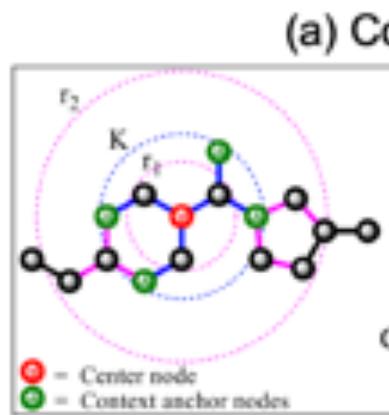
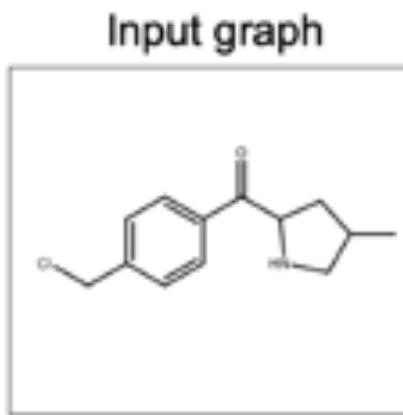
Representation learning module



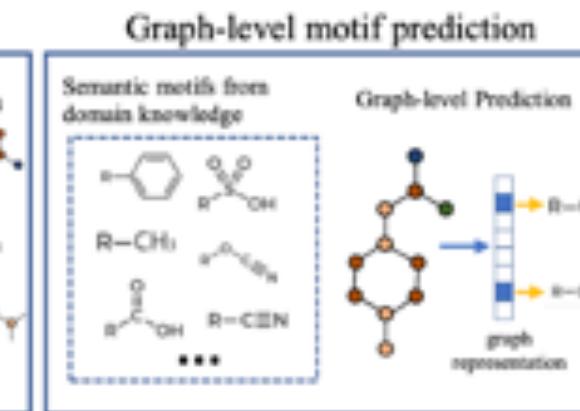
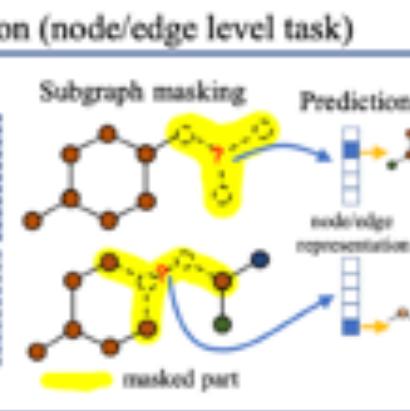
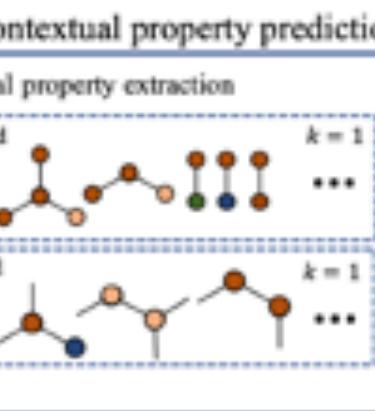
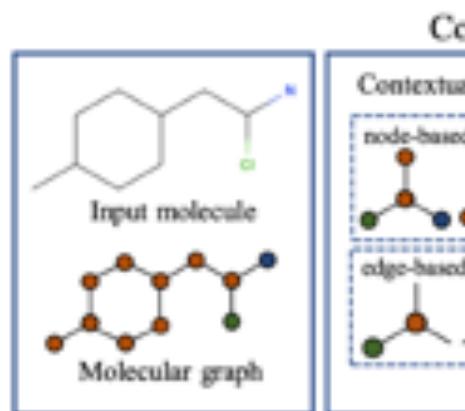
Head for target task



## SSL tasks for molecules



Strategies for Pre-training Graph Neural Networks  
Hu, Liu, Gomes, Zitnik, Liang, Pande, Leskovec (ICLR 2020)  
<https://arxiv.org/abs/1905.12265>



Self-Supervised Graph Transformer on Large-Scale Molecular Data  
Rong, Bian, Xu, Xie, Wei, Huang, Huang (NeurIPS 2020)  
<https://arxiv.org/abs/2007.02835>

# Geometric symmetry: Invariance and equivariance

## Fundamental Requirements for Geometric GNNs:

The xyz coordinates **cannot** be directly used as node features.

We need to consider the **invariance and equivariance under the motion group**

- Euclidean group E(3) : translations, rotations, reflections in 3D
- Special Euclidean group SE(3) : translations, rotations in 3D

E(3)-invariant

- Schütt et al, [SchNet](#). (2017) <https://arxiv.org/abs/1706.08566>
- Unke et al, [PhysNet](#). (2019) <https://arxiv.org/abs/1902.08408>
- Klicpera et al, [DimeNet++](#). (2020) <https://arxiv.org/abs/2011.14115>

**Invariant**

$$f(g \cdot x) = f(x)$$

SE(3)-equivariant

- Anderson et al, [Cormorant](#). (2019) <https://arxiv.org/abs/1906.04015>
- Fuchs et al, [SE\(3\)-Transformers](#). (2021) <https://arxiv.org/abs/2006.10503>

**Equivariant**

$$f(g \cdot x) = g \cdot f(x)$$

E(3)-equivariant

- Thomas et al, [Tensor Field Networks](#). (2018) <https://arxiv.org/abs/1802.08219>
- Köhler et al, [Equivariant Flows \(Radial Field\)](#). (2020) <https://arxiv.org/abs/2006.02425>
- Satorras et al, [E\(n\) Equivariant Graph Neural Networks](#). (2021) <https://arxiv.org/abs/2102.09844>

# Edge-aware Message Passing Neural Networks (MPNNs)

Table 2. Comparison of Previous Approaches (left) with MPNN baselines (middle) and our methods (right)

Target	BAML	BOB	CM	ECFP4	HDAD	GC	GG-NN	DTNN	enn-s2s	enn-s2s-ens5
mu	4.34	4.23	4.49	4.82	3.34	0.70	1.22	-	<b>0.30</b>	0.20
alpha	3.01	2.98	4.33	34.54	1.75	2.27	1.55	-	<b>0.92</b>	0.68
HOMO	2.20	2.20	3.09	2.89	1.54	1.18	1.17	-	<b>0.99</b>	0.74
LUMO	2.76	2.74	4.26	3.10	1.96	1.10	1.08	-	<b>0.87</b>	0.65
gap	3.28	3.41	5.32	3.86	2.49	1.78	1.70	-	<b>1.60</b>	1.23
R2	3.25	0.80	2.83	90.68	1.35	4.73	3.99	-	<b>0.15</b>	0.14
ZPVE	3.31	3.40	4.80	241.58	1.91	9.75	2.52	-	<b>1.27</b>	1.10
U0	1.21	1.43	2.98	85.01	0.58	3.02	0.83	-	<b>0.45</b>	0.33
U	1.22	1.44	2.99	85.59	0.59	3.16	0.86	-	<b>0.45</b>	0.34
H	1.22	1.44	2.99	86.21	0.59	3.19	0.81	-	<b>0.39</b>	0.30
G	1.20	1.42	2.97	78.36	0.59	2.95	0.78	.84 <sup>2</sup>	<b>0.44</b>	0.34
Cv	1.64	1.83	2.36	30.29	0.88	1.45	1.19	-	<b>0.80</b>	0.62
Omega	0.27	0.35	1.32	1.47	0.34	0.32	0.53	-	<b>0.19</b>	0.15
Average	2.17	2.08	3.37	53.97	1.35	2.59	1.36	-	<b>0.68</b>	0.52

## Atom features

Atom type	H, C, N, O, F (one-hot)
Atomic number	# of protonns (int)
Acceptor	0 or 1 (binary)
Donor	0 or 1 (binary)
Aromatic	0 or 1 (binary)
Hybridization	sp, sp2, sp3 (one-hot or null)
# of hydrogens	(integer)

## Bond features

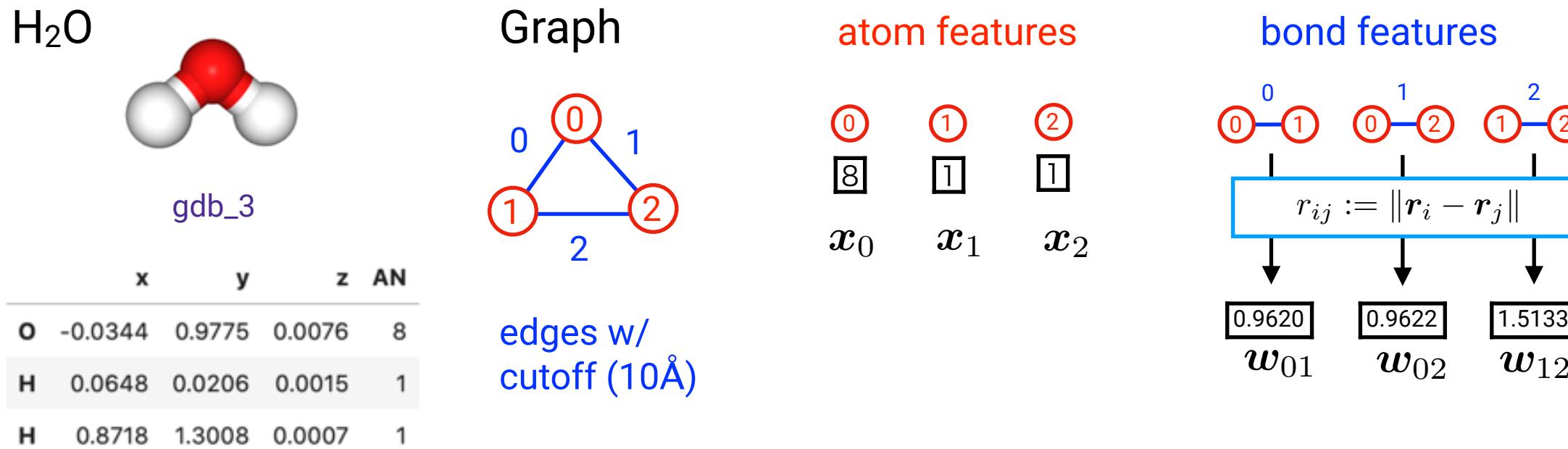
Euclidean distance between atom pair	(real)
Bond types	single, double, triple, aromatic (one-hot)

Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE.  
[Neural Message Passing for Quantum Chemistry.](#)  
[ICML 2017](#)

Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, et al.  
[Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error.](#)  
[J Chem Theory Comput. 2017;13: 5255–5264.](#)

→ E(3)-invariant  
xyz are used only as distances

# SchNet (Schütt et al, 2017): Standard Geometric GNN



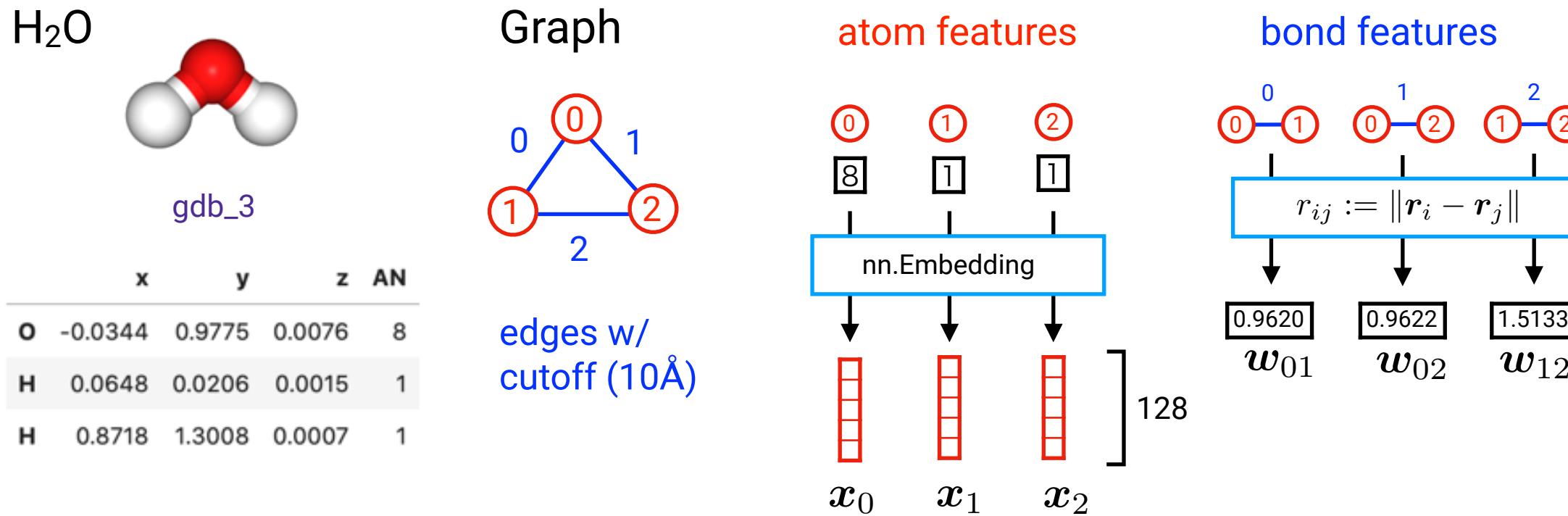
Target	Unit	SchNet	enn-s2s	Proposed
$\epsilon_{\text{HOMO}}$	meV	41	43	<b>36.7</b>
$\epsilon_{\text{LUMO}}$	meV	34	37	<b>30.8</b>
$\Delta\epsilon$	meV	63	69	<b>58.0</b>
ZPVE	meV	1.7	<b>1.5</b>	<b>1.49</b>
$\mu$	Debye	0.033	0.030	<b>0.029</b>
$\alpha$	Bohr <sup>3</sup>	0.235	0.092	<b>0.077</b>
$\langle R^2 \rangle$	Bohr <sup>2</sup>	<b>0.073</b>	0.180	<b>0.072</b>
$U_0$	meV	14	19	<b>10.5</b>
$U$	meV	19	19	<b>10.6</b>
$H$	meV	14	17	<b>11.3</b>
$G$	meV	14	19	<b>12.2</b>
$C_v$	cal/molK	<b>0.033</b>	0.040	<b>0.032</b>



## SchNet-edge-update

Jørgensen PB, Jacobsen KW,  
Schmidt MN. [Neural Message  
Passing with Edge Updates for  
Predicting Properties of  
Molecules and Materials.](#)  
arXiv [stat.ML]. 2018.  
<http://arxiv.org/abs/1806.03146>

# SchNet (Schütt et al, 2017): Standard Geometric GNN

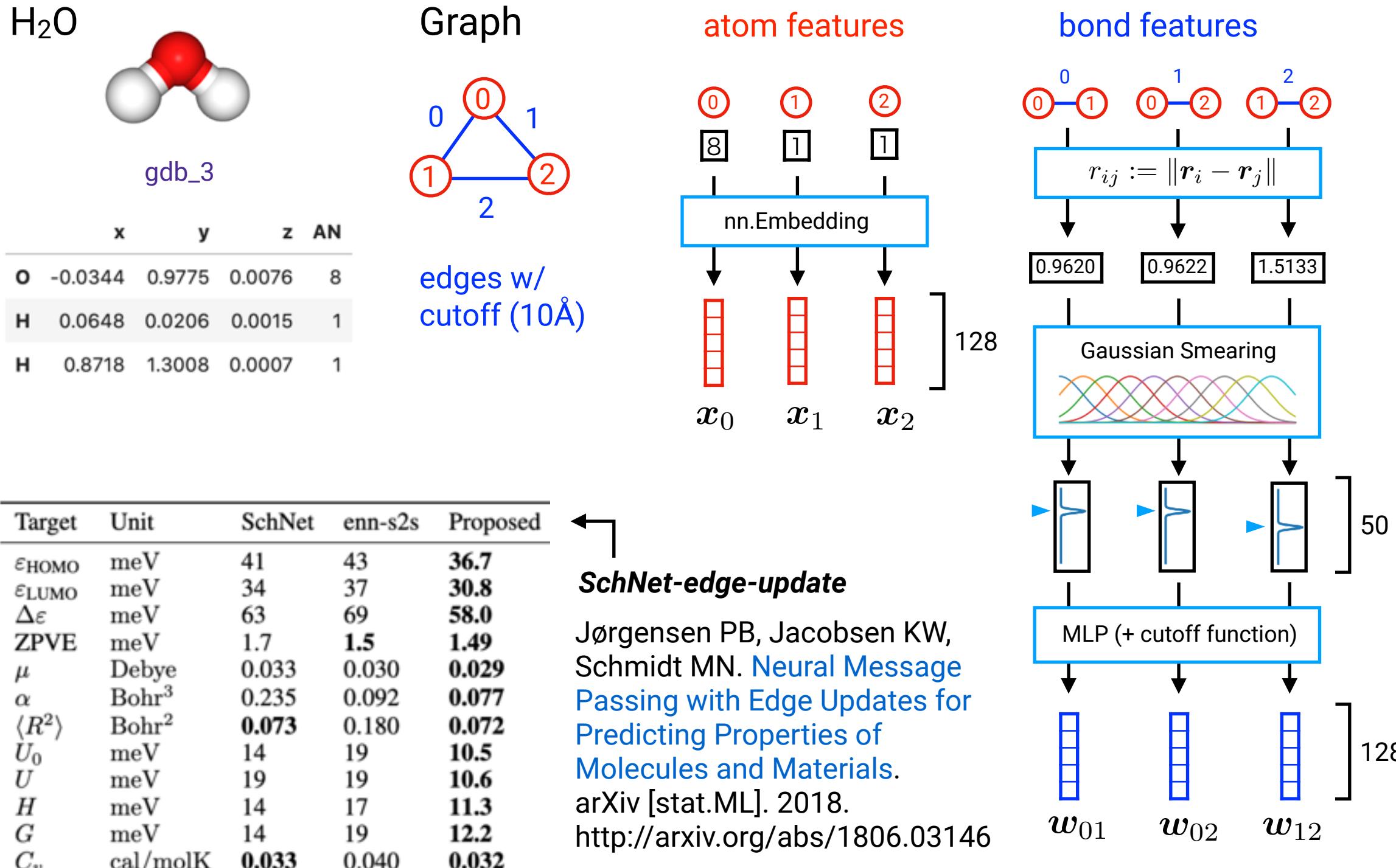


Target	Unit	SchNet	enn-s2s	Proposed
$\varepsilon_{\text{HOMO}}$	meV	41	43	<b>36.7</b>
$\varepsilon_{\text{LUMO}}$	meV	34	37	<b>30.8</b>
$\Delta\varepsilon$	meV	63	69	<b>58.0</b>
ZPVE	meV	1.7	<b>1.5</b>	<b>1.49</b>
$\mu$	Debye	0.033	0.030	<b>0.029</b>
$\alpha$	Bohr <sup>3</sup>	0.235	0.092	<b>0.077</b>
$\langle R^2 \rangle$	Bohr <sup>2</sup>	<b>0.073</b>	0.180	<b>0.072</b>
$U_0$	meV	14	19	<b>10.5</b>
$U$	meV	19	19	<b>10.6</b>
$H$	meV	14	17	<b>11.3</b>
$G$	meV	14	19	<b>12.2</b>
$C_v$	cal/molK	<b>0.033</b>	0.040	<b>0.032</b>

←  
**SchNet-edge-update**

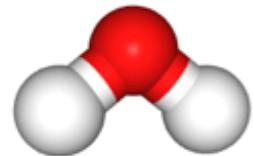
Jørgensen PB, Jacobsen KW,  
Schmidt MN. [Neural Message  
Passing with Edge Updates for  
Predicting Properties of  
Molecules and Materials.](#)  
arXiv [stat.ML]. 2018.  
<http://arxiv.org/abs/1806.03146>

# SchNet (Schütt et al, 2017): Standard Geometric GNN



# SchNet (Schütt et al, 2017): Standard Geometric GNN

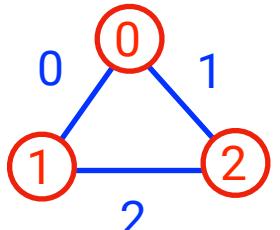
H<sub>2</sub>O



gdb\_3

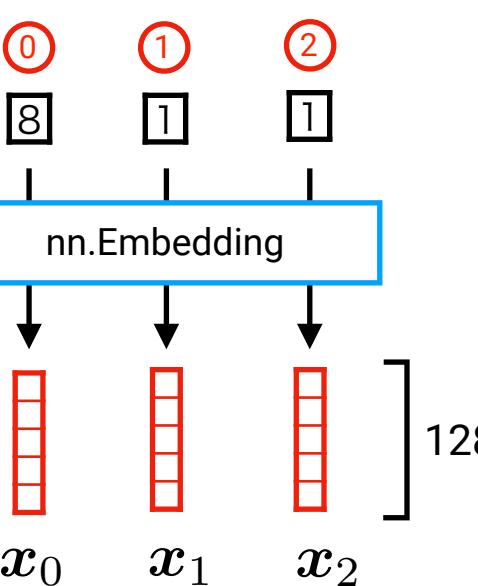
	x	y	z	AN
O	-0.0344	0.9775	0.0076	8
H	0.0648	0.0206	0.0015	1
H	0.8718	1.3008	0.0007	1

Graph

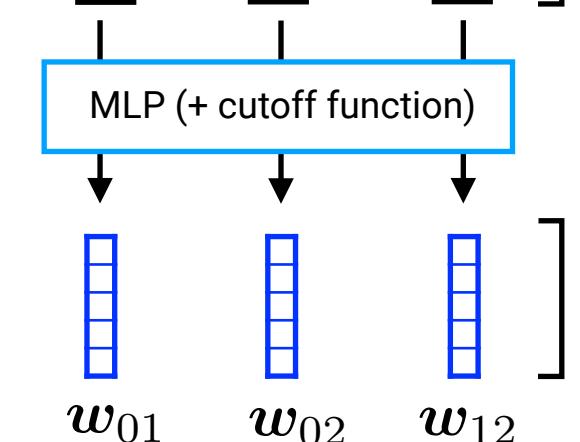
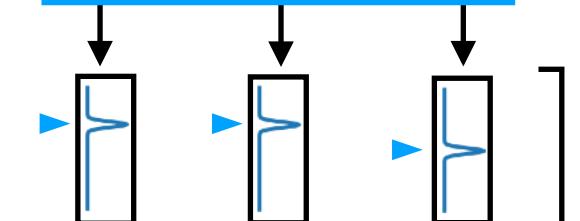
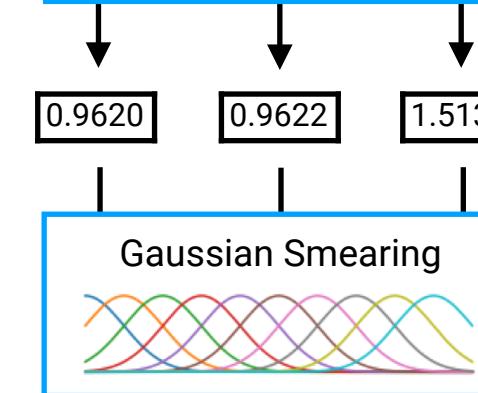
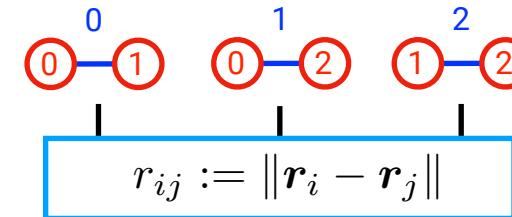


edges w/  
cutoff (10Å)

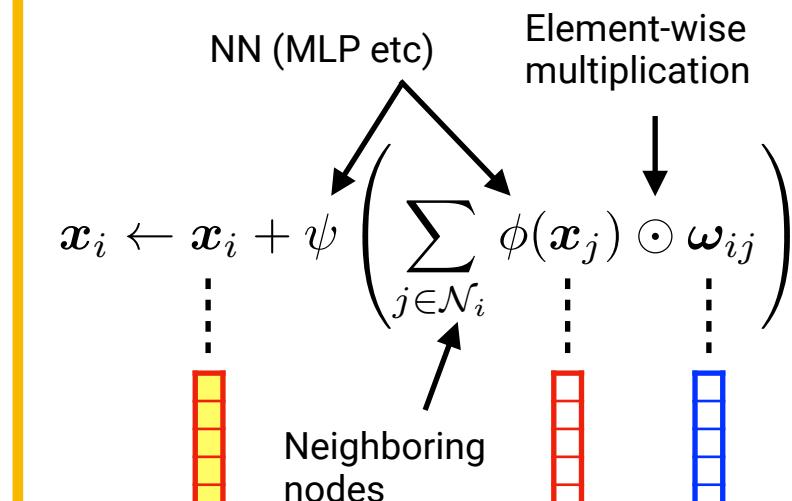
atom features



bond features



Message Passing with  
residual connections



←  
**SchNet-edge-update**

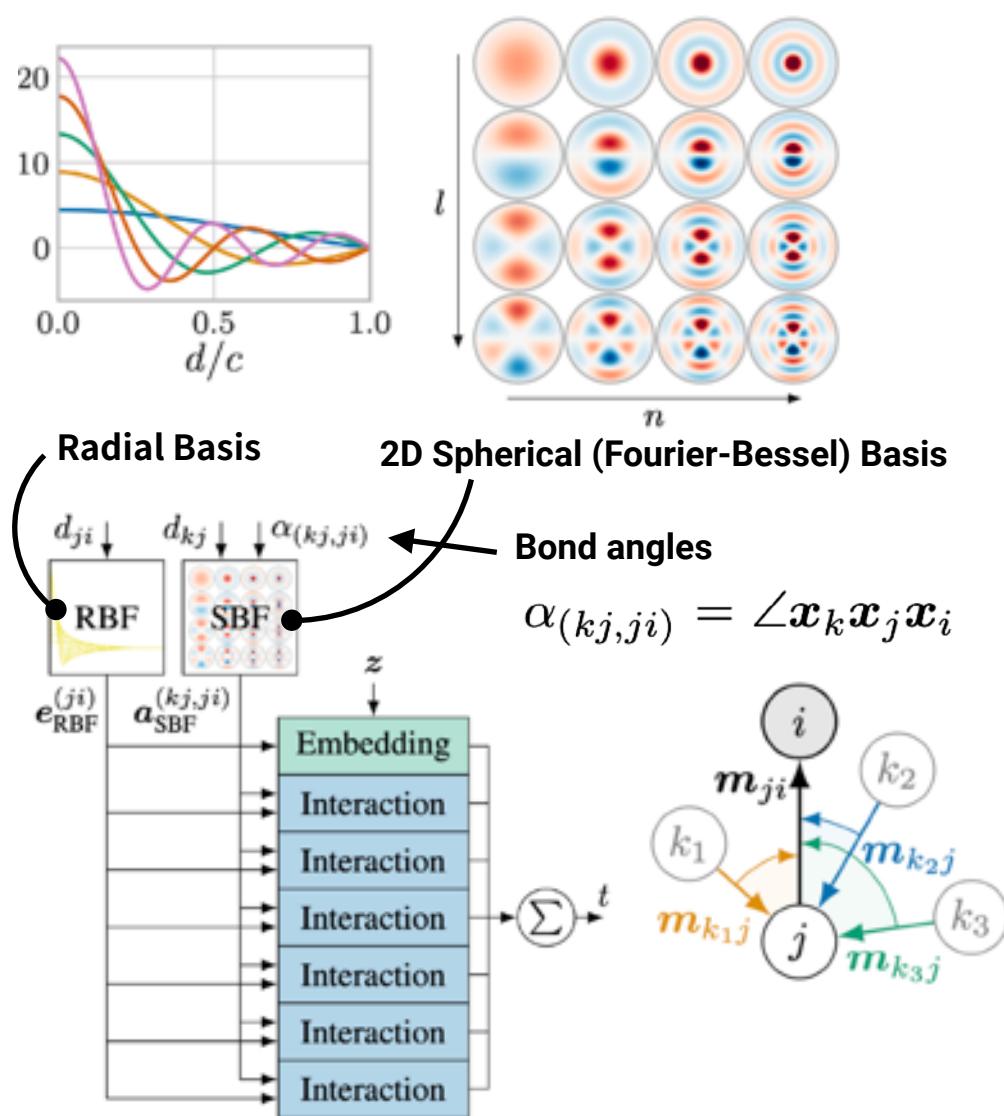
Jørgensen PB, Jacobsen KW,  
Schmidt MN. [Neural Message  
Passing with Edge Updates for  
Predicting Properties of  
Molecules and Materials.](#)  
arXiv [stat.ML]. 2018.  
<http://arxiv.org/abs/1806.03146>

Target	Unit	SchNet	enn-s2s	Proposed
$\epsilon_{\text{HOMO}}$	meV	41	43	<b>36.7</b>
$\epsilon_{\text{LUMO}}$	meV	34	37	<b>30.8</b>
$\Delta\epsilon$	meV	63	69	<b>58.0</b>
ZPVE	meV	1.7	<b>1.5</b>	<b>1.49</b>
$\mu$	Debye	0.033	0.030	<b>0.029</b>
$\alpha$	Bohr <sup>3</sup>	0.235	0.092	<b>0.077</b>
$\langle R^2 \rangle$	Bohr <sup>2</sup>	<b>0.073</b>	0.180	<b>0.072</b>
$U_0$	meV	14	19	<b>10.5</b>
$U$	meV	19	19	<b>10.6</b>
$H$	meV	14	17	<b>11.3</b>
$G$	meV	14	19	<b>12.2</b>
$C_v$	cal/molK	<b>0.033</b>	0.040	<b>0.032</b>

# QSAR/QSPR, QM Approximation, Molecule Generations, ...

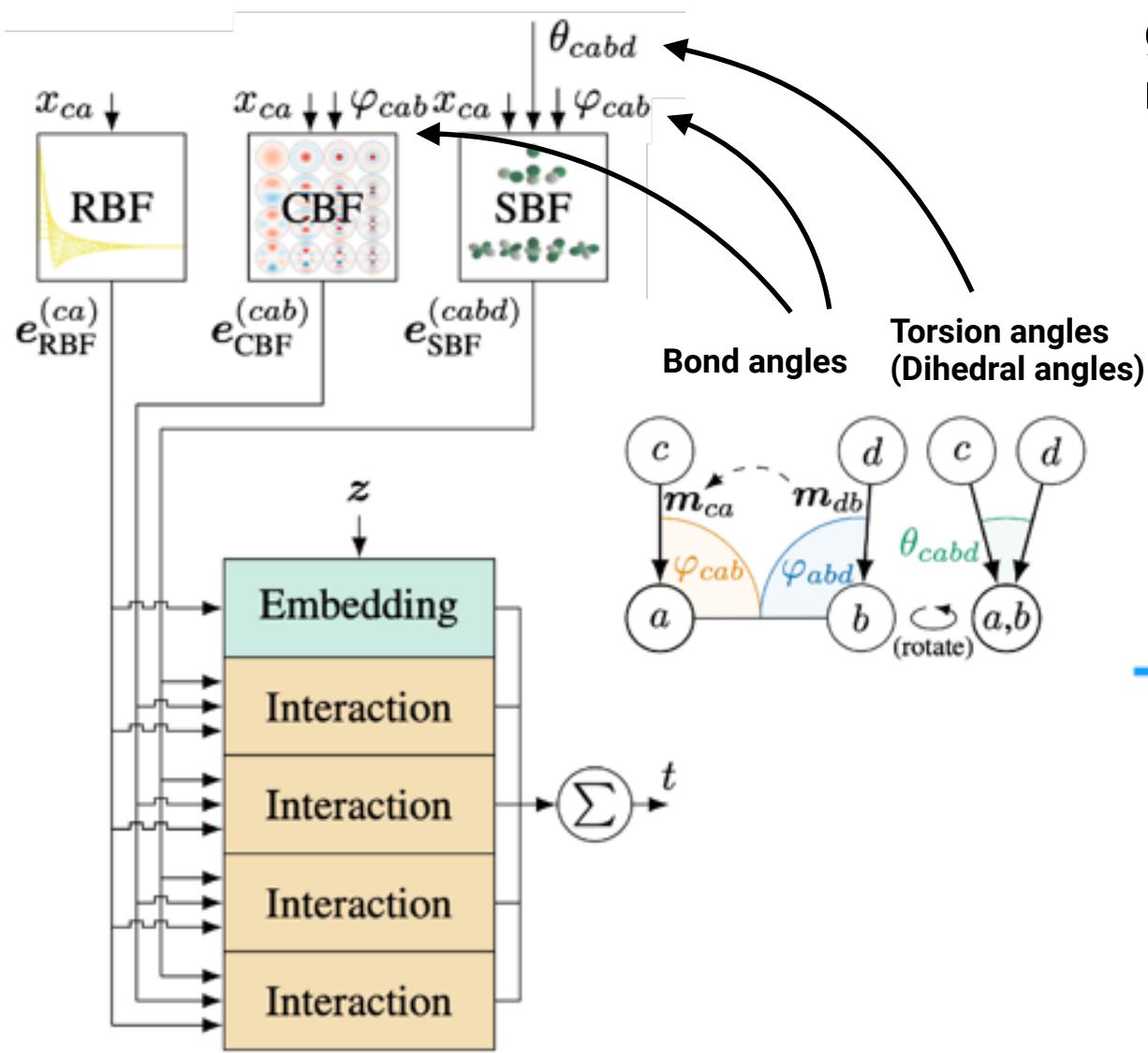
## DimeNet++

Klicpera et al (NeurIPS WS2022)  
<https://arxiv.org/abs/2011.14115>



## GemNet

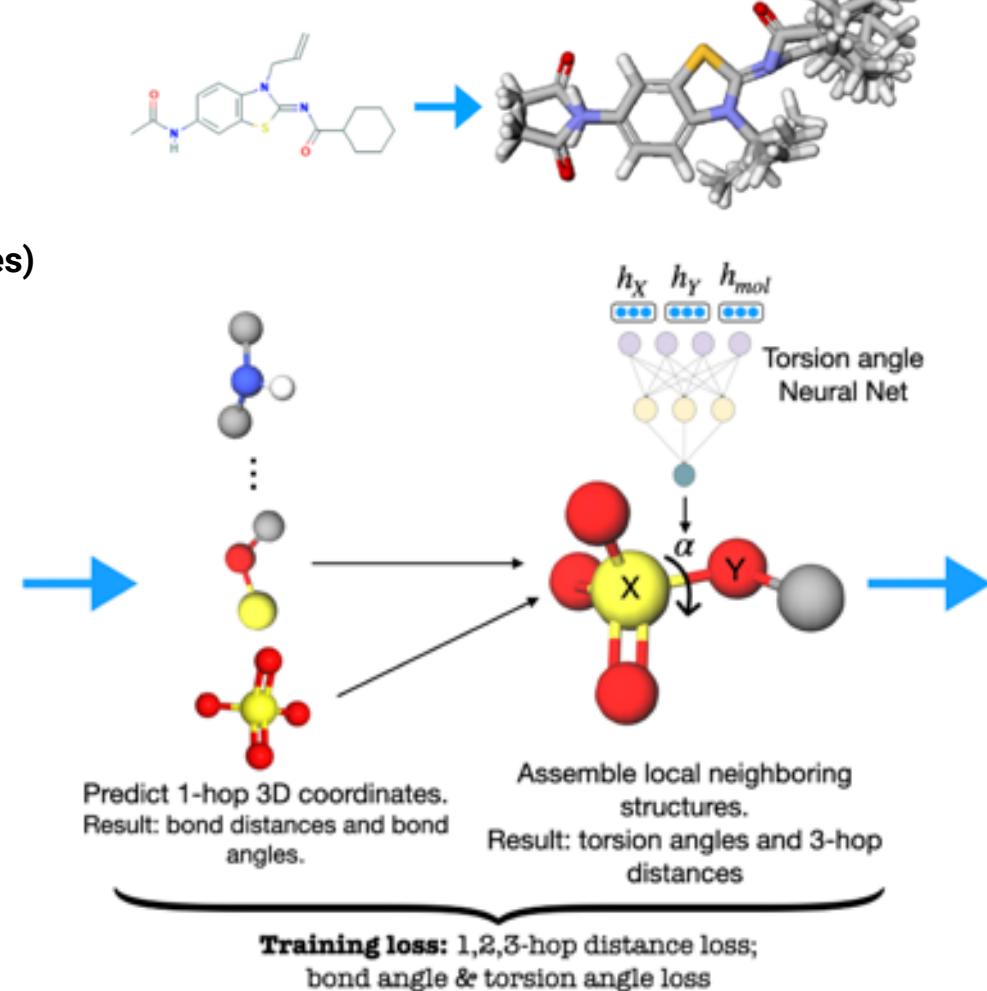
Klicpera et al (NeurIPS2021)  
<https://arxiv.org/abs/2106.08903>



## GeoMol

Ganea et al (NeurIPS2021)  
<https://arxiv.org/abs/2106.07802>

generates distributions of low-energy molecular 3D conformers



# ML × QM: ML Potentials, Force fields, Density functionals

*Machine Learning at the Atomic Scale (Chem. Rev.)*  
<https://pubs.acs.org/toc/chreay/121/16>

## CHEMICAL REVIEWS

[pubs.acs.org/CR](https://pubs.acs.org/CR)

### Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems

John A. Keith,\* Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger, Klaus-Robert Müller,\* and Alexandre Tkatchenko\*

 Cite This: <https://doi.org/10.1021/acs.chemrev.1c00107>

 Read Online



Review

## CHEMICAL REVIEWS

[pubs.acs.org/CR](https://pubs.acs.org/CR)

### Ab Initio Machine Learning in Chemical Compound Space

Bing Huang and O. Anatole von Lilienfeld\*

 Cite This: [Chem. Rev. 2021, 121, 10001–10036](https://doi.org/10.1021/acs.chemrev.1c00107)

 Read Online



Review

*Data Science Meets Chemistry (Acc. Chem. Res.)*  
<https://pubs.acs.org/page/achre4/data-science-meets-chemistry>

## CHEMICAL REVIEWS

[pubs.acs.org/CR](https://pubs.acs.org/CR)

### Physics-Inspired Structural Representations for Molecules and Materials

Felix Musil, Andrea Grisafi, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti\*

 Cite This: [Chem. Rev. 2021, 121, 9759–9815](https://doi.org/10.1021/acs.chemrev.1c00107)

 Read Online



Review

## ACCOUNTS of chemical research

[pubs.acs.org/accounts](https://pubs.acs.org/accounts)

Article

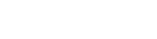
### Learning to Approximate Density Functionals

*Published as part of the Accounts of Chemical Research special issue “Data Science Meets Chemistry”.*

Bhupalee Kalita, Li Li, Ryan J. McCarty, and Kieron Burke\*

 Cite This: [Acc. Chem. Res. 2021, 54, 818–826](https://doi.org/10.1021/acs.accounts.1c00010)

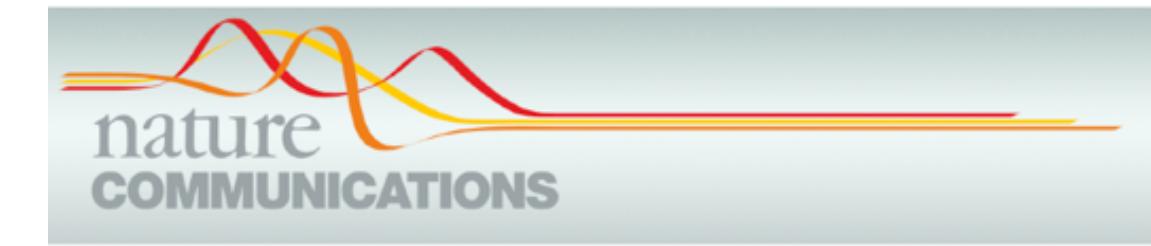
 Read Online



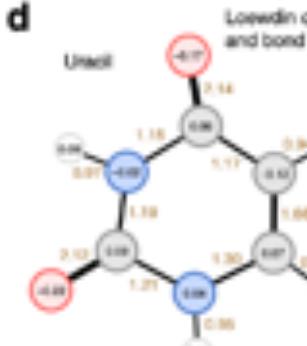
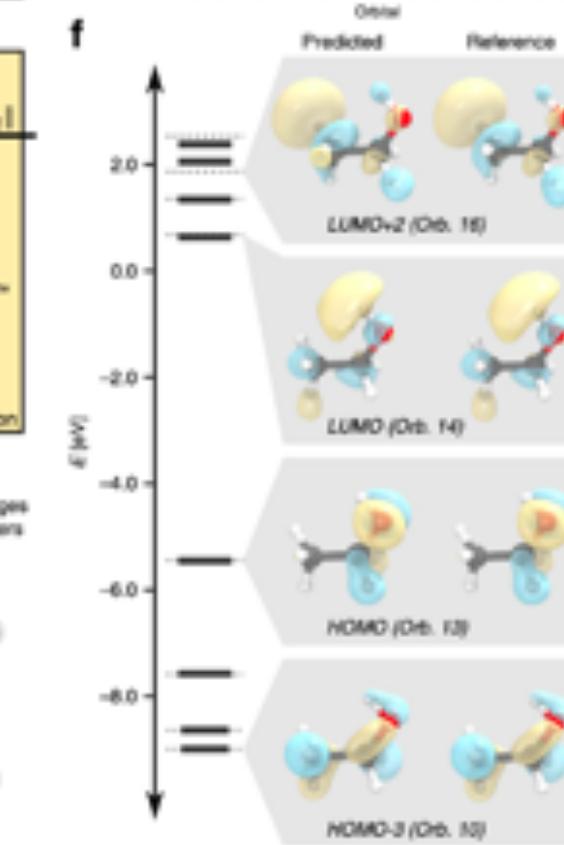
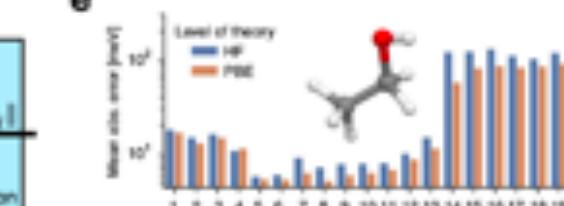
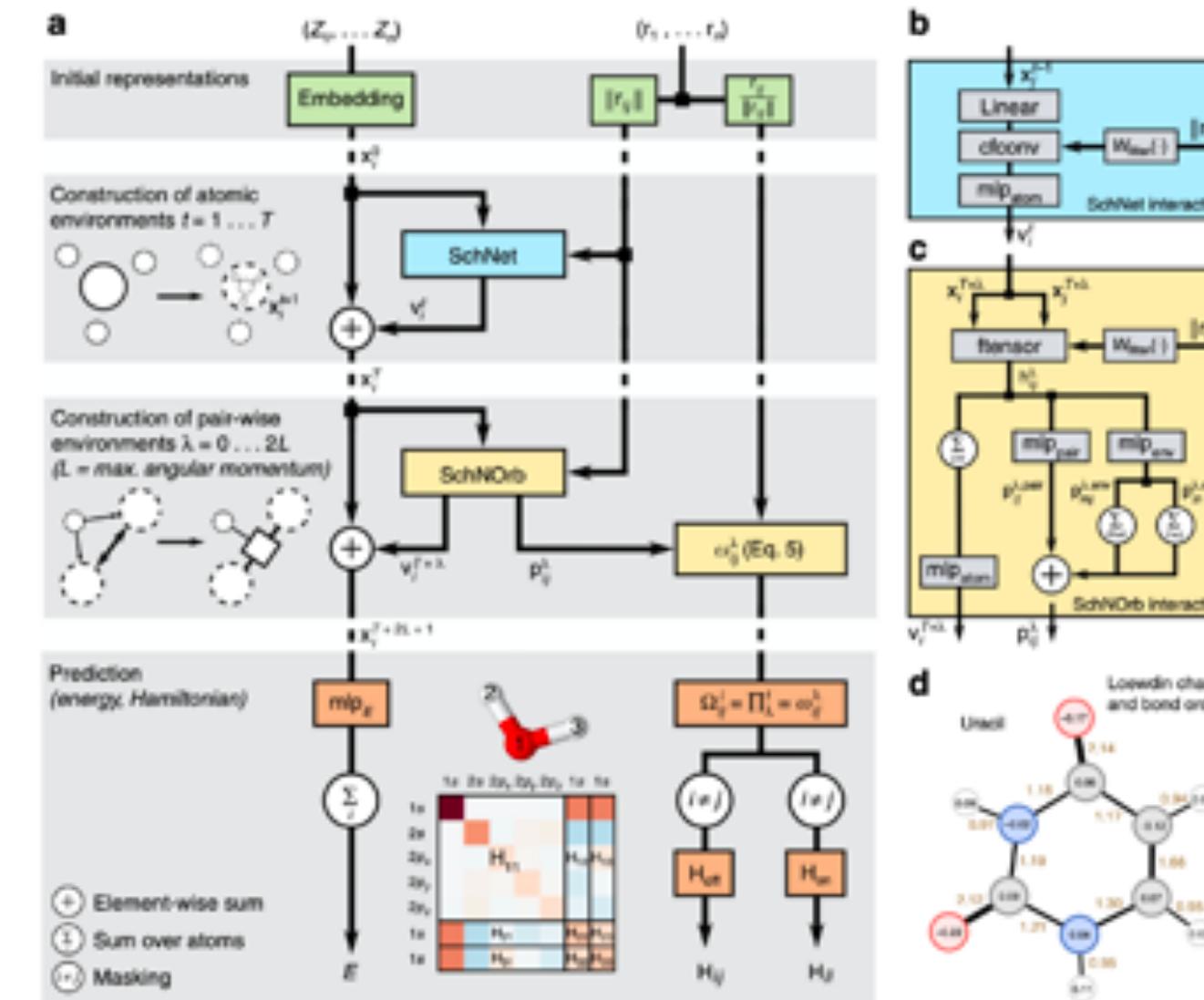
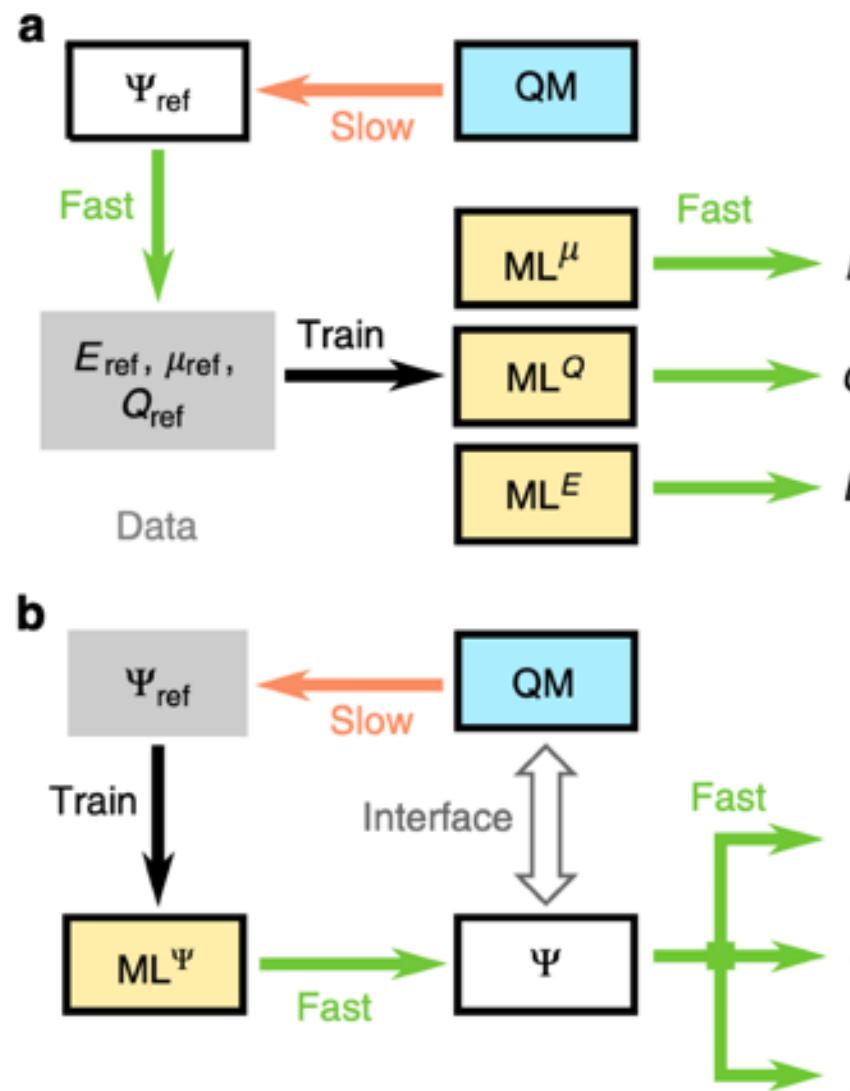
# SchNOrb (Schütt et al, Nat Commn, 2019)

**Unifying machine learning and quantum chemistry  
with a deep neural network for molecular  
wavefunctions**

K. T. Schütt, M. Gastegger, A. Tkatchenko✉, K.-R. Müller✉ & R. J. Maurer✉



Nature Communications 10, Article number: 5024 (2019)



# DM21 Density functional (Kirkpatrick et al, Science, 2021)

Kirkpatrick *et al.*, *Science* **374**, 1385–1389 (2021)

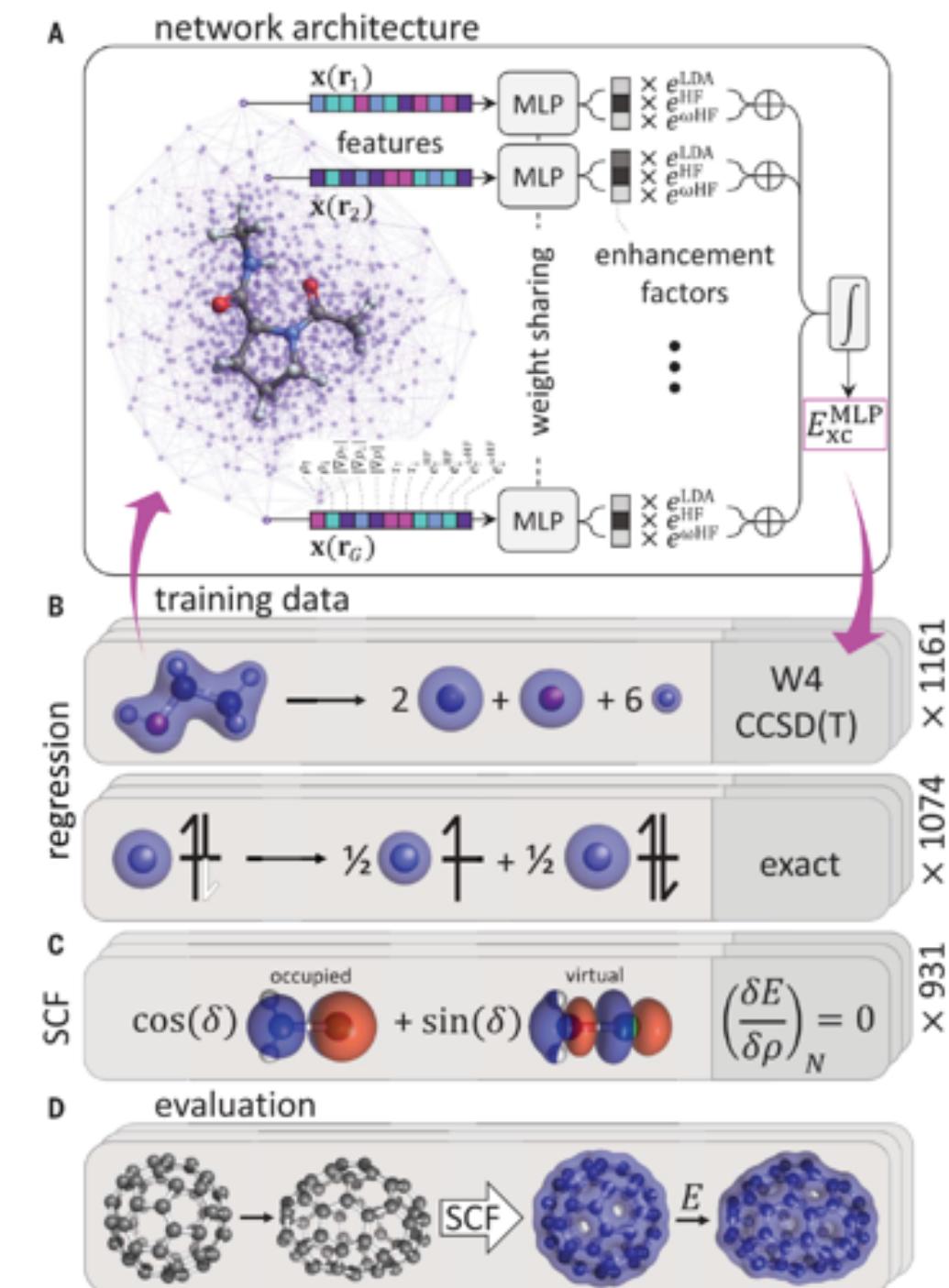
10 December 2021

## QUANTUM CHEMISTRY

# Pushing the frontiers of density functionals by solving the fractional electron problem

James Kirkpatrick<sup>1\*</sup>†, Brendan McMorrow<sup>1†</sup>, David H. P. Turban<sup>1†</sup>, Alexander L. Gaunt<sup>1†</sup>, James S. Spencer<sup>1</sup>, Alexander G. D. G. Matthews<sup>1</sup>, Annette Obika<sup>1</sup>, Louis Thiry<sup>2</sup>, Meire Fortunato<sup>1</sup>, David Pfau<sup>1</sup>, Lara Román Castellanos<sup>1</sup>, Stig Petersen<sup>1</sup>, Alexander W. R. Nelson<sup>1</sup>, Pushmeet Kohli<sup>1</sup>, Paula Mori-Sánchez<sup>3</sup>, Demis Hassabis<sup>1</sup>, Aron J. Cohen<sup>1,4\*</sup>

Density functional theory describes matter at the quantum level, but all popular approximations suffer from systematic errors that arise from the violation of mathematical properties of the exact functional. We overcame this fundamental limitation by training a neural network on molecular data and on fictitious systems with fractional charge and spin. The resulting functional, DM21 (DeepMind 21), correctly describes typical examples of artificial charge delocalization and strong correlation and performs better than traditional functionals on thorough benchmarks for main-group atoms and molecules. DM21 accurately models complex systems such as hydrogen chains, charged DNA base pairs, and diradical transition states. More crucially for the field, because our methodology relies on data and constraints, which are continually improving, it represents a viable pathway toward the exact universal functional.



# "Learn to Simulate"

DeepMind > Research > Learning to Simulate Complex Physics with Graph Networks

 PUBLICATIONS

SHARE  
  

PUBLICATION LINKS  
 DOWNLOAD  
 VIEW PUBLICATION  
 DATASETS & CODE  
 VIDEO SITE  
 → VIEW OPEN SOURCE

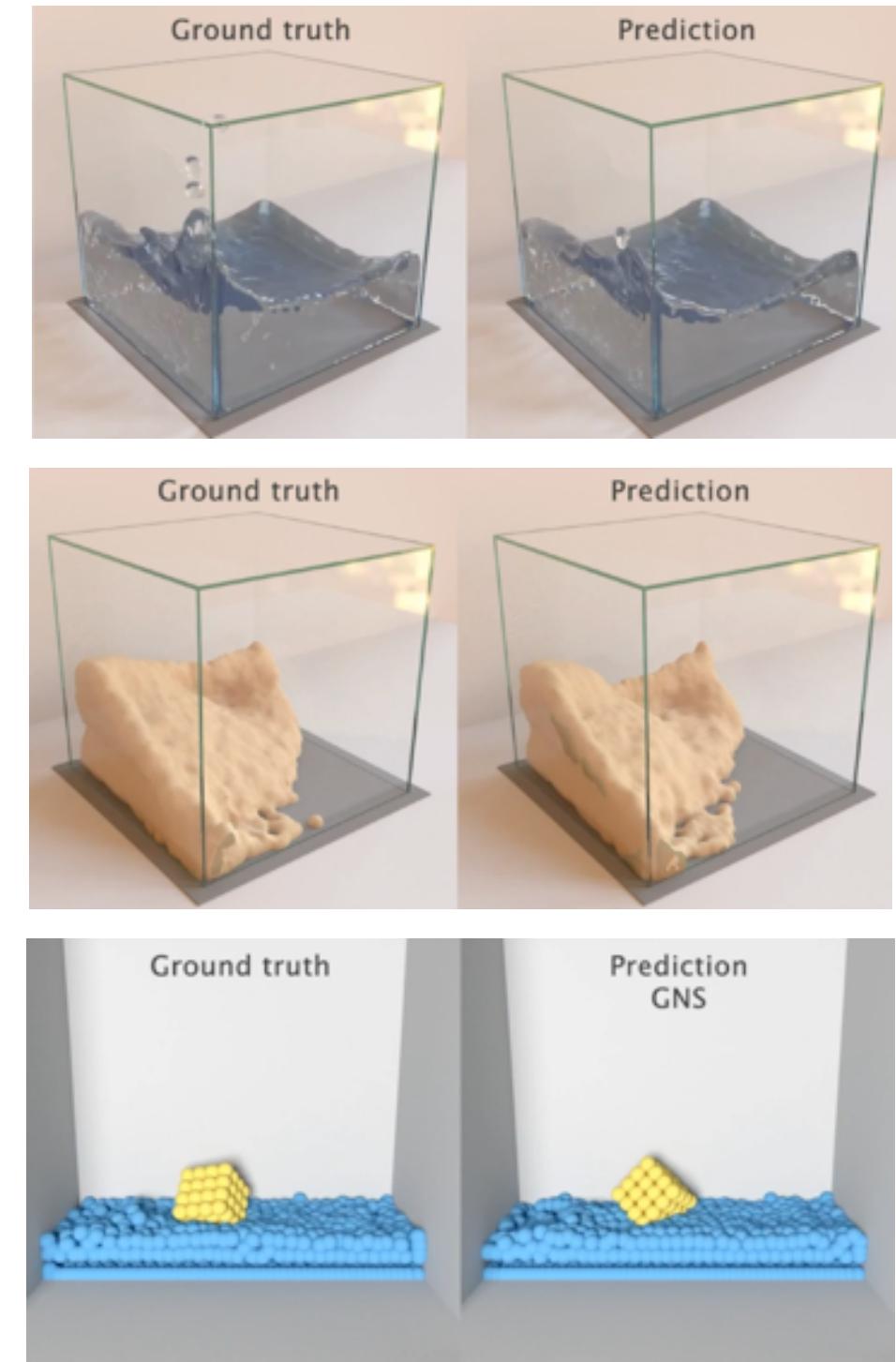
PUBLICATION  
ICML

# Learning to Simulate Complex Physics with Graph Networks

## Abstract

Here we present a machine learning framework and model implementation that can learn to simulate a wide variety of challenging physical domains, involving fluids, rigid solids, and deformable materials interacting with one another. Our framework—which we term “Graph Network-based Simulators” (GNS)—represents the state of a physical system with particles, expressed as nodes in a graph, and computes dynamics via learned message-passing. Our results show that our model can generalize from single-timestep predictions with thousands of particles during training, to different initial conditions, thousands of timesteps, and at least an order of magnitude more particles at test time. Our model was robust to hyperparameter choices across various evaluation metrics: the main determinants of long-term performance were the number of message-passing steps, and mitigating the accumulation of error by corrupting the training data with noise. Our GNS framework advances the state-of-the-art in learned physical simulation, and holds promise for solving a wide range of complex forward and inverse problems.

Datasets and example model and training code available.

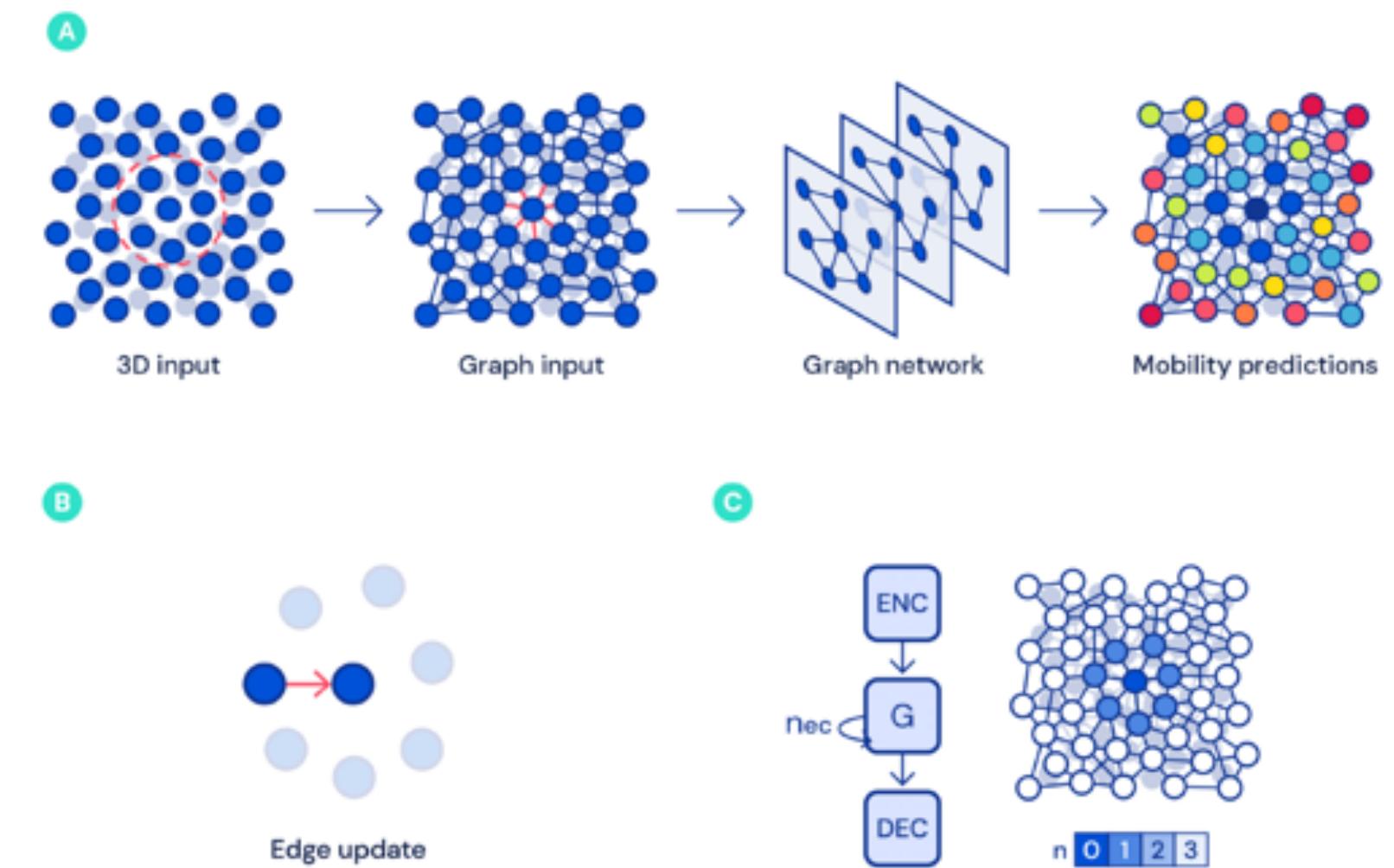
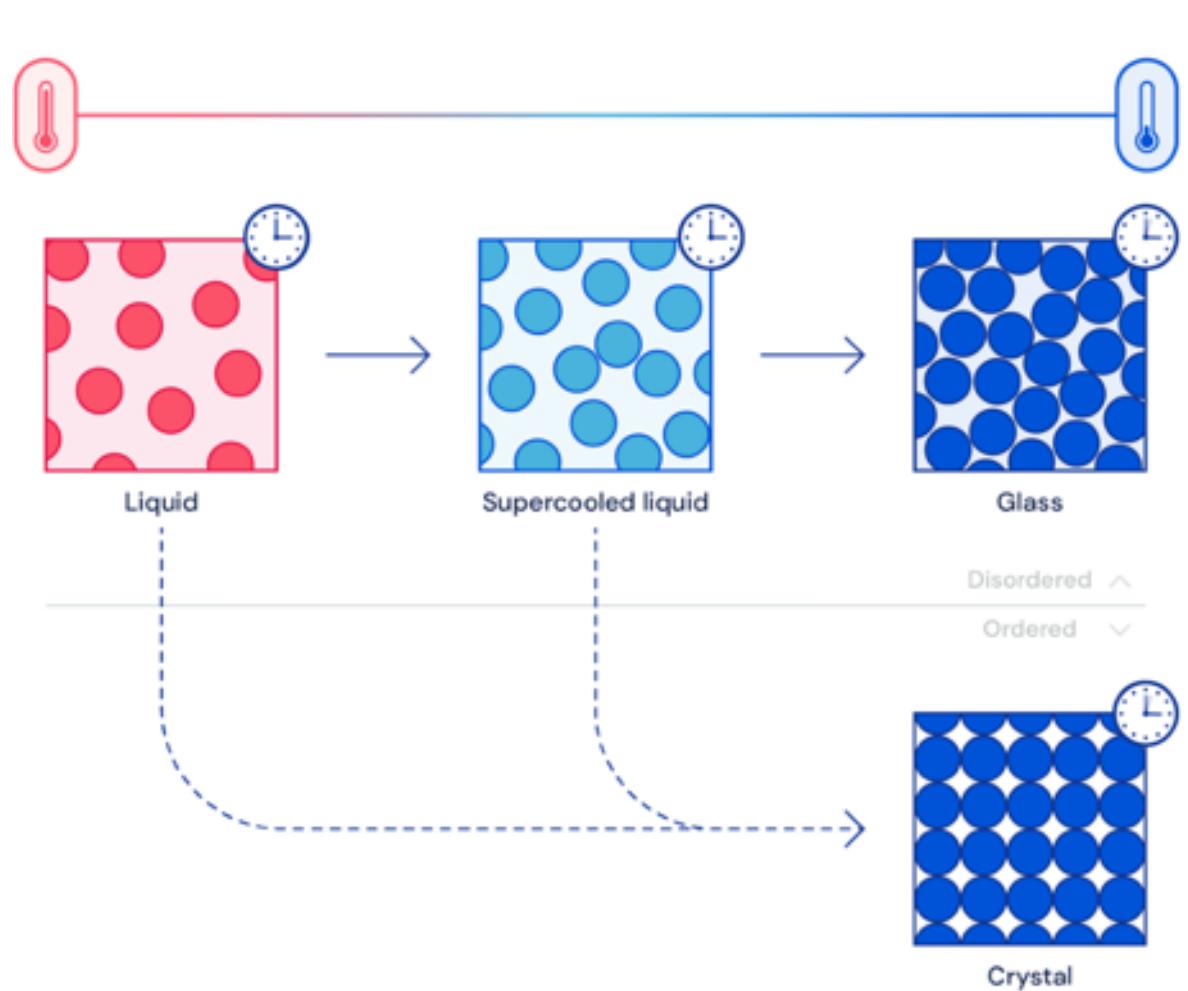


# Learn to simulate long-time evolution of a glassy system

Bapst V, Keck T, Grabska-Barwińska A, Donner C, Cubuk ED, Schoenholz SS, et al.

Unveiling the predictive power of static structure in glassy systems.

Nature Physics. 2020;16: 448–454.



# DL/GNNs × Simulations

How to fuse ML (empiricism) and simulations (rationalism) is one of the recent hot topics.

Annu. Rev. Phys. Chem. 71:361–90 (2020)



*Annual Review of Physical Chemistry*

Machine Learning for  
Molecular Simulation

Frank Noé,<sup>1,2,3</sup> Alexandre Tkatchenko,<sup>4</sup>  
Klaus-Robert Müller,<sup>5,6,7</sup> and Cecilia Clementi<sup>1,3,8</sup>

Acc. Chem. Res. 54(7):1575–1585 (2021)

ACCOUNTS  
of chemical research

[pubs.acs.org/accounts](https://pubs.acs.org/accounts)

Article

**Development of Multimodal Machine Learning Potentials: Toward a Physics-Aware Artificial Intelligence**

Published as part of the Accounts of Chemical Research special issue “Data Science Meets Chemistry”.

Tetiana Zubatiuk and Olexandr Isayev\*

PNAS (2020)

**The frontier of simulation-based inference**

Kyle Cranmer<sup>a,b,1</sup>, Johann Brehmer<sup>a,b</sup>, and Gilles Louppe<sup>c</sup>

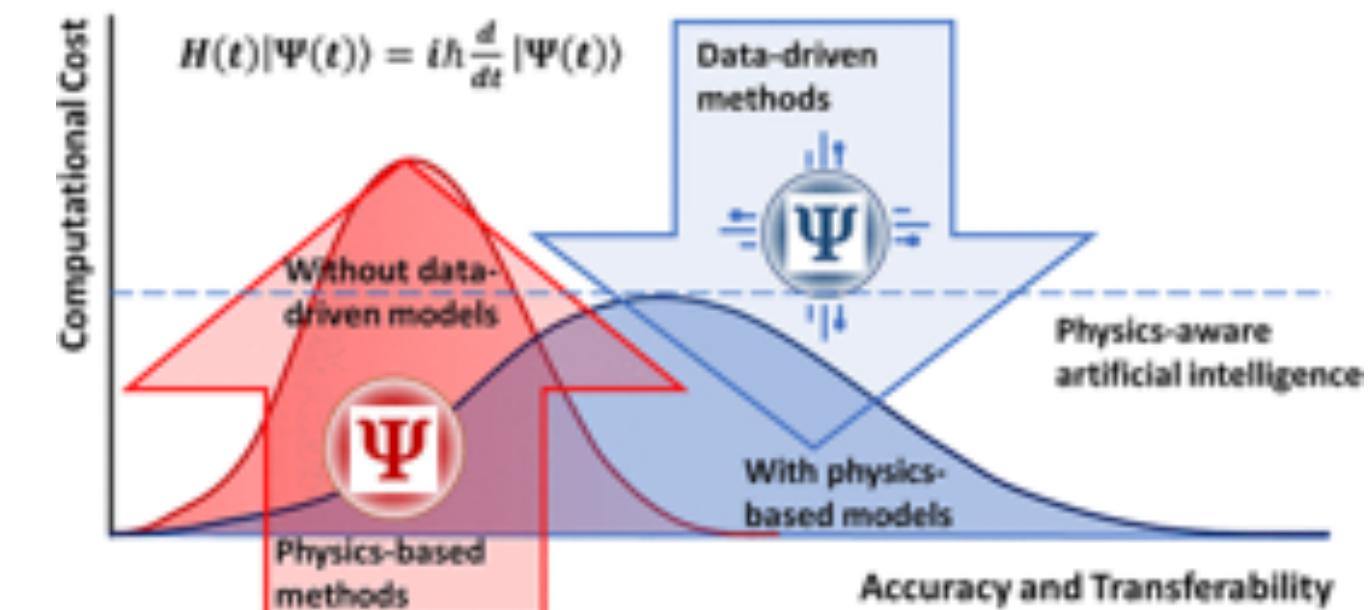
<sup>a</sup>Center for Cosmology and Particle Physics, New York University, New York, NY 10003; <sup>b</sup>Center for Data Science, New York University, New York, NY 10011;  
<sup>c</sup>Montefiore Institute, University of Liège, B-4000 Liège, Belgium

Edited by Jitendra Malik, University of California, Berkeley, CA, and approved April 10, 2020 (received for review November 4, 2019)

Many domains of science have developed complex simulations to describe phenomena of interest. While these simulations provide high-fidelity models, they are poorly suited for inference and lead to challenging inverse problems. We review the rapidly developing field of simulation-based inference and identify the forces giving additional momentum to the field. Finally, we describe how the frontier is expanding so that a broad audience can appreciate the profound influence these developments may have on science.

the simulator—is being recognized as a key idea to improve the sample efficiency of various inference methods. A third direction of research has stopped treating the simulator as a black box and focused on integrations that allow the inference engine to tap into the internal details of the simulator directly.

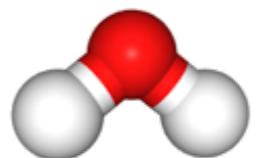
Amidst this ongoing revolution, the landscape of simulation-based inference is changing rapidly. In this review we aim to provide the reader with a high-level overview of the basic ideas



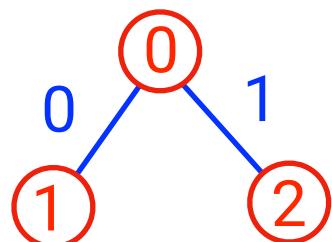
# Toward more effective inductive bias?

---

Water molecule H<sub>2</sub>O



A molecular graph (RDKit)



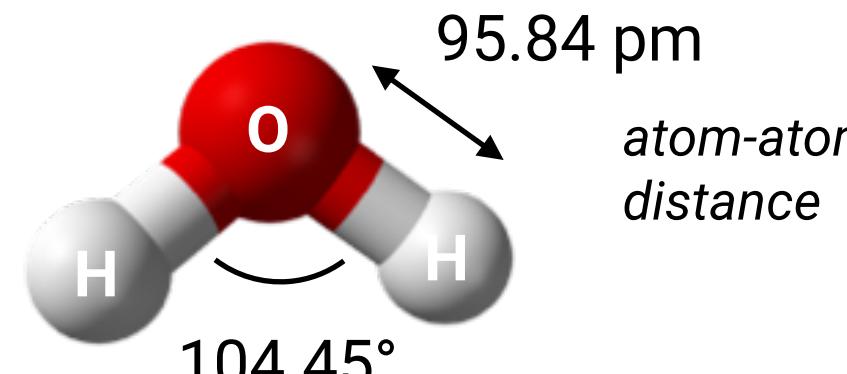
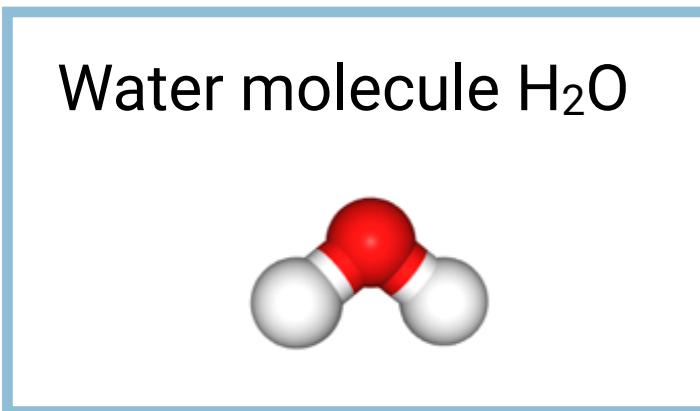
	0	1	2
AtomicNum	8	1	1
TotalDegree	2	1	1
TotalNumHs	0	0	0
FormalCharge	0	0	0
deltaMass	0	0	0
IsInRing	0	0	0

*Atom  
Invariants*

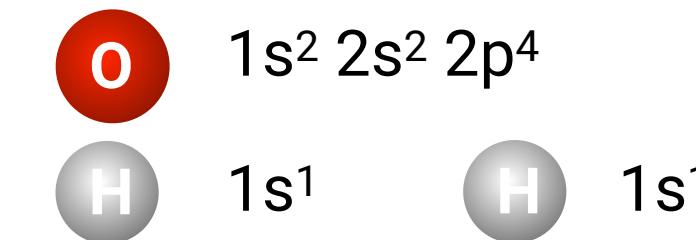
	0	1
BondType	0	0
Stereo	0	0

*Bond  
Invariants*

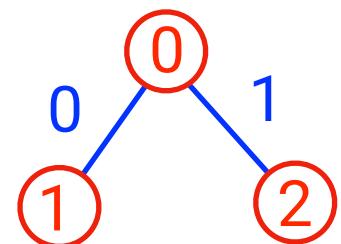
# Toward more effective inductive bias?



8 + 1 + 1 = 10 electrons



A molecular graph (RDKit)

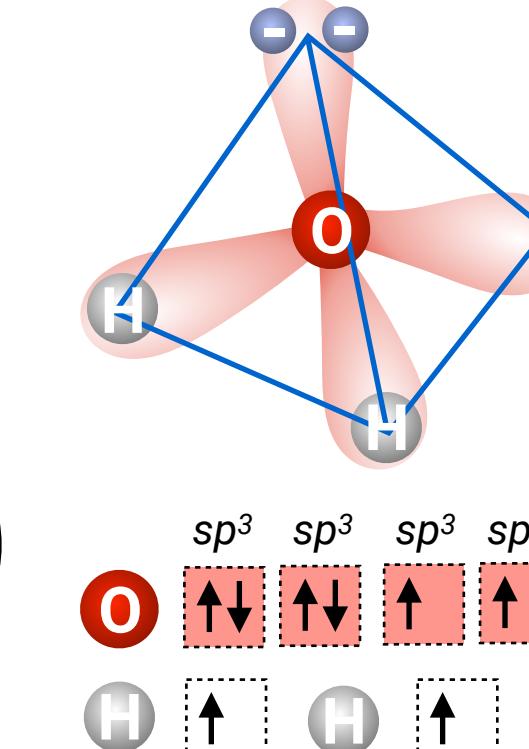
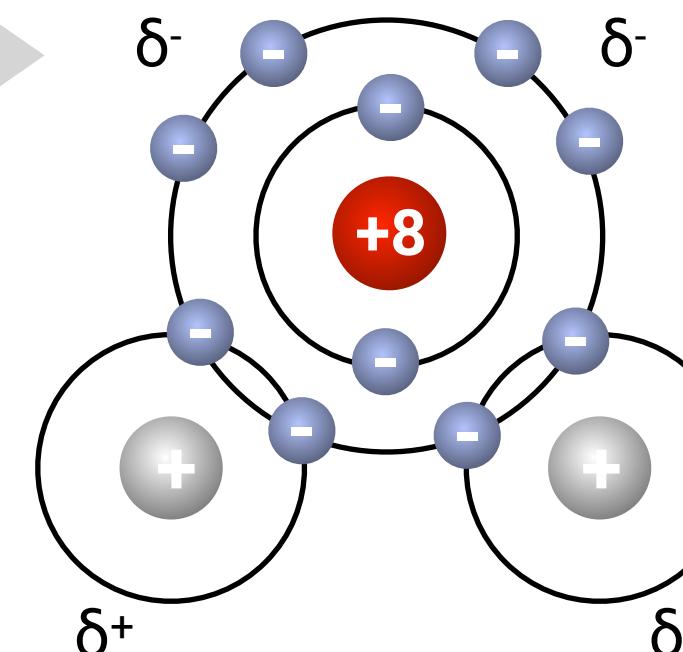


Atom Invariants

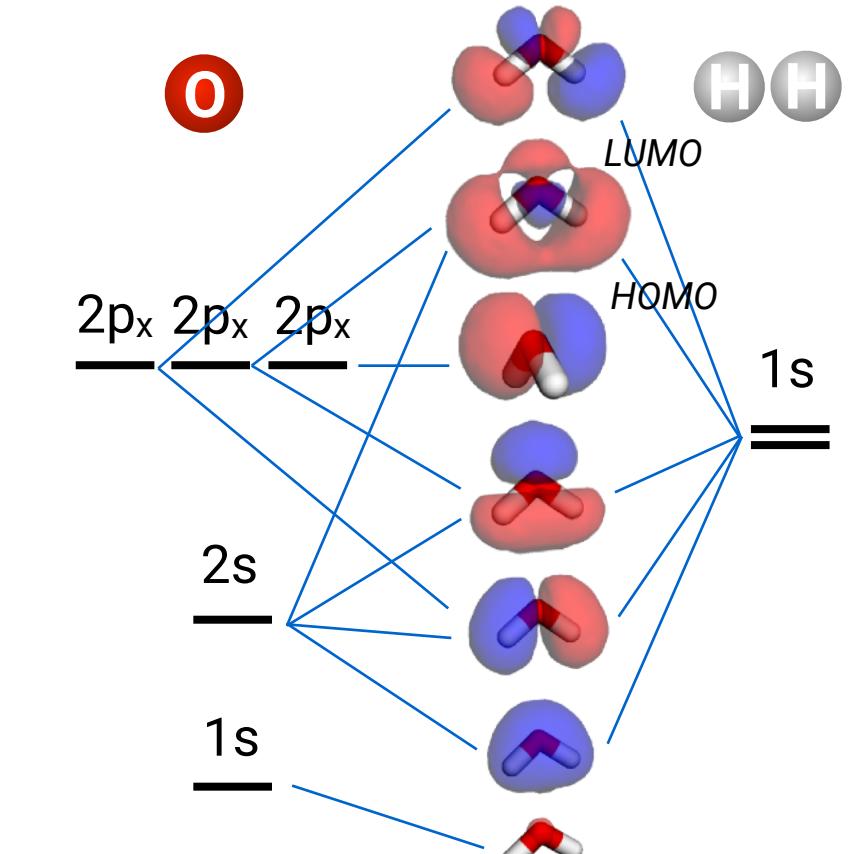
AtomicNum	8	1	1
TotalDegree	2	1	1
TotalNumHs	0	0	0
FormalCharge	0	0	0
deltaMass	0	0	0
IsInRing	0	0	0

Bond Invariants

BondType	0	0
Stereo	0	0



$sp^3$  hybridization



molecular orbitals

# DL/GNNs × Symbolic Tasks, Logical Inference, Planning

Now we can use machine learning also for symbolic, logical, algorithmic tasks!

## Neural Abstract Machines & Program Induction

<https://uclnlp.github.io/nampi/>

- **Differentiable Neural Computers / Neural Turing Machines** (Graves+ 2014)
- **Memory Networks** (Weston+ 2014)
- **Pointer Networks** (Vinyals+ 2015)
- **Neural Stacks** (Grefenstette+ 2015, Joulin+ 2015)
- **Hierarchical Attentive Memory** (Andrychowicz+ 2016)
- **Neural Program Interpreters** (Reed+ 2016)
- **Neural Programmer** (Neelakantan+ 2016)
- **DeepCoder** (Balog+ 2016)
- 



Can we also use explicit chemical knowledges?



Computer-Aided Synthetic Planning



International Edition: DOI: 10.1002/anie.201506101  
German Edition: DOI: 10.1002/ange.201506101

## Computer-Assisted Synthetic Planning: The End of the Beginning

Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik, Karol Molga, Piotr Dittwald, Michał Startek, Michał Bajczyk, and Bartosz A. Grzybowski\*

Angew. Chem. Int. Ed. 2016, 55, 5904–5937



International Edition: DOI: 10.1002/anie.201912083  
German Edition: DOI: 10.1002/ange.201912083

## Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning

Tomasz Badowski, Ewa P. Gajewska, Karol Molga, and Bartosz A. Grzybowski\*

Angew. Chem. Int. Ed. 2019, 58, 1–7

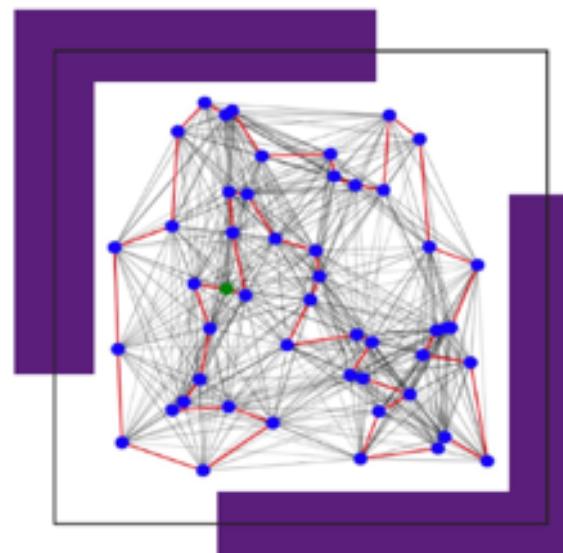
# DL/GNNs × Combinatorial Optimization and Reasoning

<https://www.ipam.ucla.edu/dlc2021>

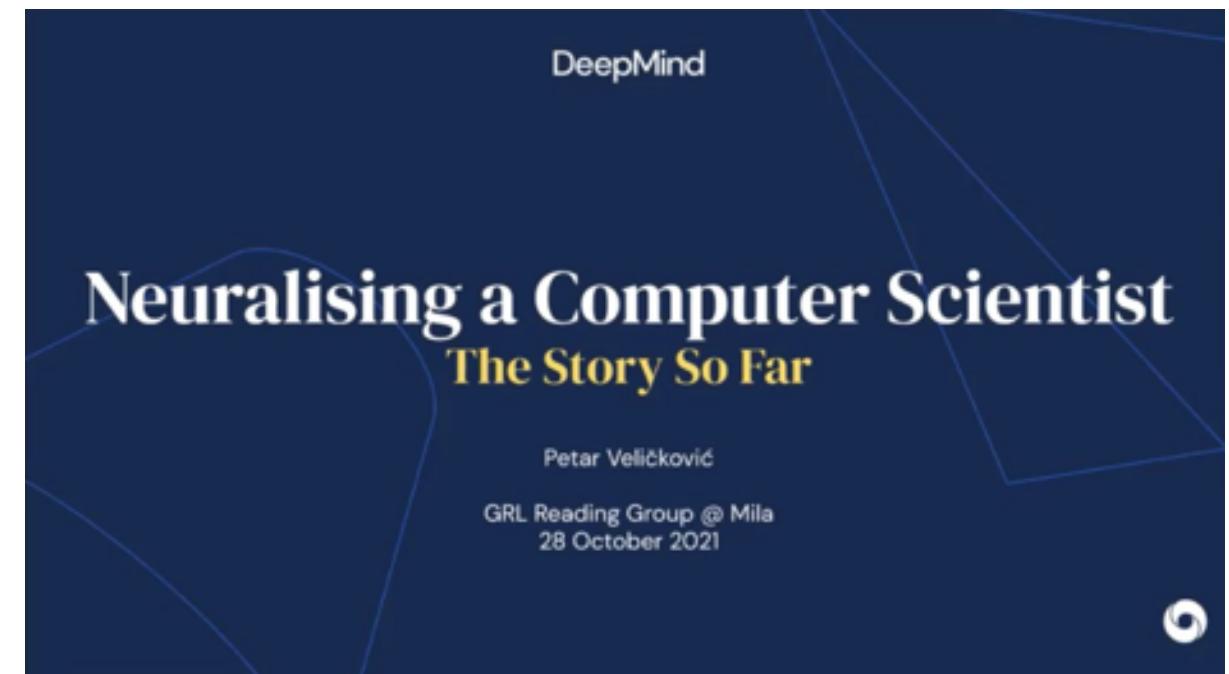


## Deep Learning and Combinatorial Optimization

February 22 - 25, 2021



<https://youtu.be/QOBoZaDZYUI>



<https://arxiv.org/abs/2102.09544>

arXiv.org > cs > arXiv:2102.09544

Computer Science > Machine Learning

[Submitted on 18 Feb 2021 (v1), last revised 27 Apr 2021 (this version, v2)]

### Combinatorial optimization and reasoning with graph neural networks

Quentin Cappart, Didier Chételat, Elias Khalil, Andrea Lodi, Christopher Morris, Petar Veličković

Combinatorial optimization is a well-established area in operations research and computer science. Until recently, its methods have focused on solving problem instances in isolation, ignoring the fact that they often stem from related data distributions in practice. However, recent years have seen a surge of interest in using machine learning, especially graph neural networks (GNNs), as a key building block for combinatorial tasks, either directly as solvers or by enhancing exact solvers. The inductive bias of GNNs effectively encodes combinatorial and relational input due to their invariance to permutations and awareness of input sparsity. This paper presents a conceptual review of recent key advancements in this emerging field, aiming at researchers in both optimization and machine learning.

<https://doi.org/10.1016/j.patter.2021.100273>

## Patterns

### Opinion

## Neural algorithmic reasoning

Petar Veličković<sup>1,\*</sup> and Charles Blundell<sup>1</sup>

<sup>1</sup>DeepMind, London, Greater London, UK

\*Correspondence: [petarv@google.com](mailto:petarv@google.com)

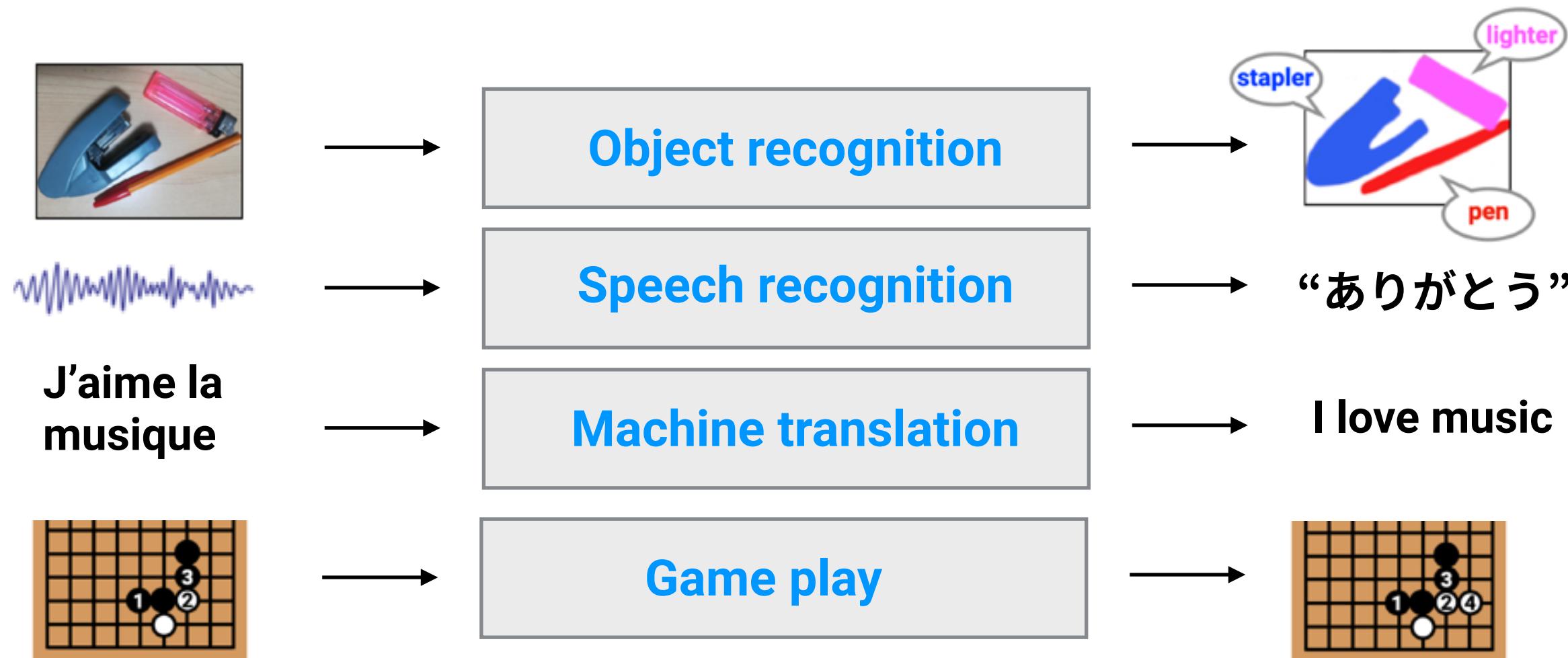
<https://doi.org/10.1016/j.patter.2021.100273>

**One final remark on "the true dark side" ...**

# The huge gap between prediction and understanding

**Prediction does not directly bring us Understanding nor Discovery.**

Current ML already provides practical applications below, but how we recognize objects and speech sounds, and how we acquire and use languages is still not well understood...

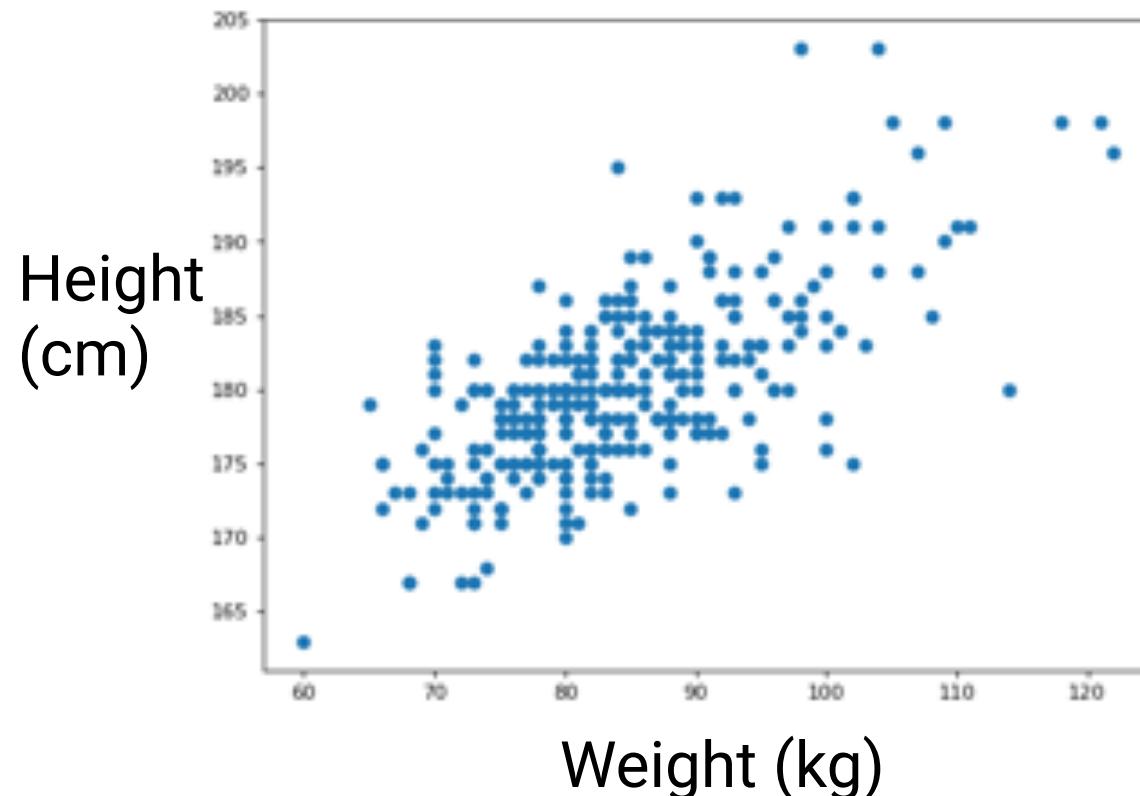


# Observational vs experimental (interventional) studies

To find out what happens to a system when you interfere with it,  
you have to interfere with it (not just passively observe it).

George Box

**Height and Weight of Pro baseball players  
(2016 NPB Website)**



We see correlation, but these data **never** tell us  
"increase weight" does not imply "grow in height"

Applied Stats 101  
**Correlation does not imply causation**  
An Inconvenient Truth  
**All we can directly access is correlation**

**We need experiments!** Passively observing  
data and applying ML to it is always insufficient.

# Toward causal representation learning

---

PROCEEDINGS OF THE IEEE | Vol. 109, No. 5, May 2021 <https://arxiv.org/abs/2102.11107>



# Toward Causal Representation Learning

*This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.*

By BERNHARD SCHÖLKOPF<sup>✉</sup>, FRANCESCO LOCATELLO<sup>✉</sup>, STEFAN BAUER<sup>✉</sup>, NAN ROSEMARY KE,  
NAL KALCHBRENNER, ANIRUDH GOYAL, AND YOSHUA BENGIO<sup>✉</sup>

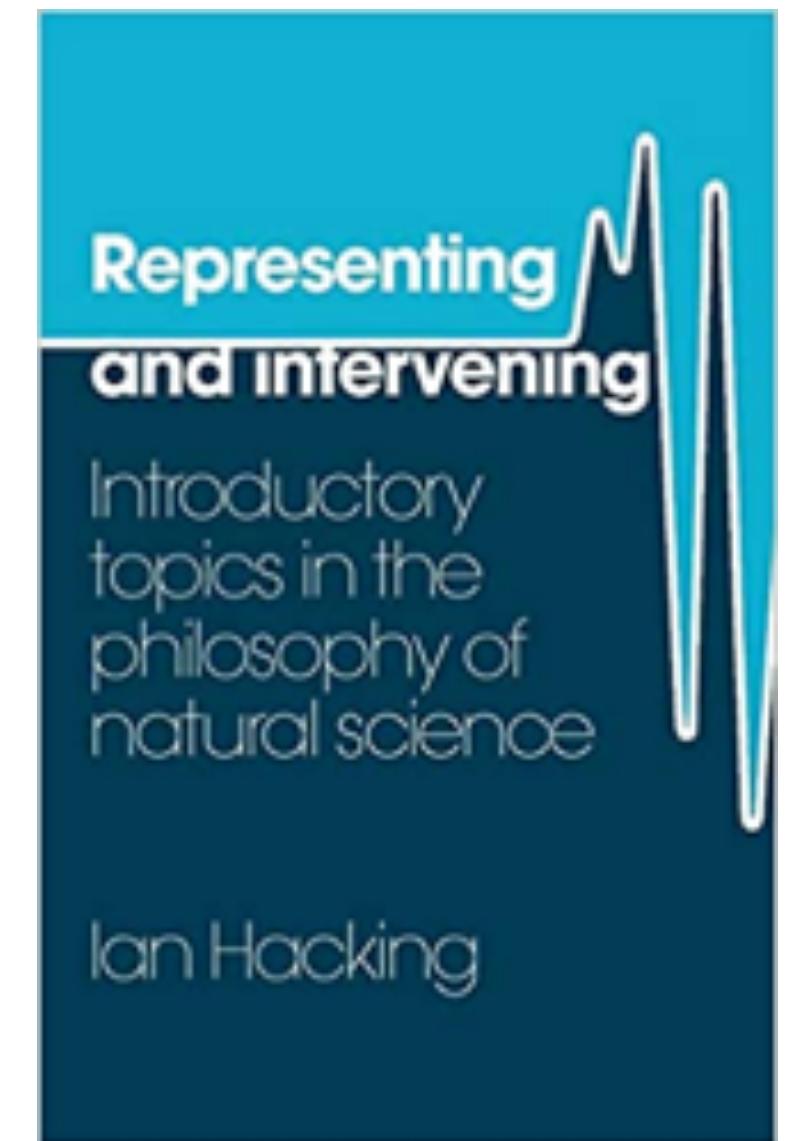
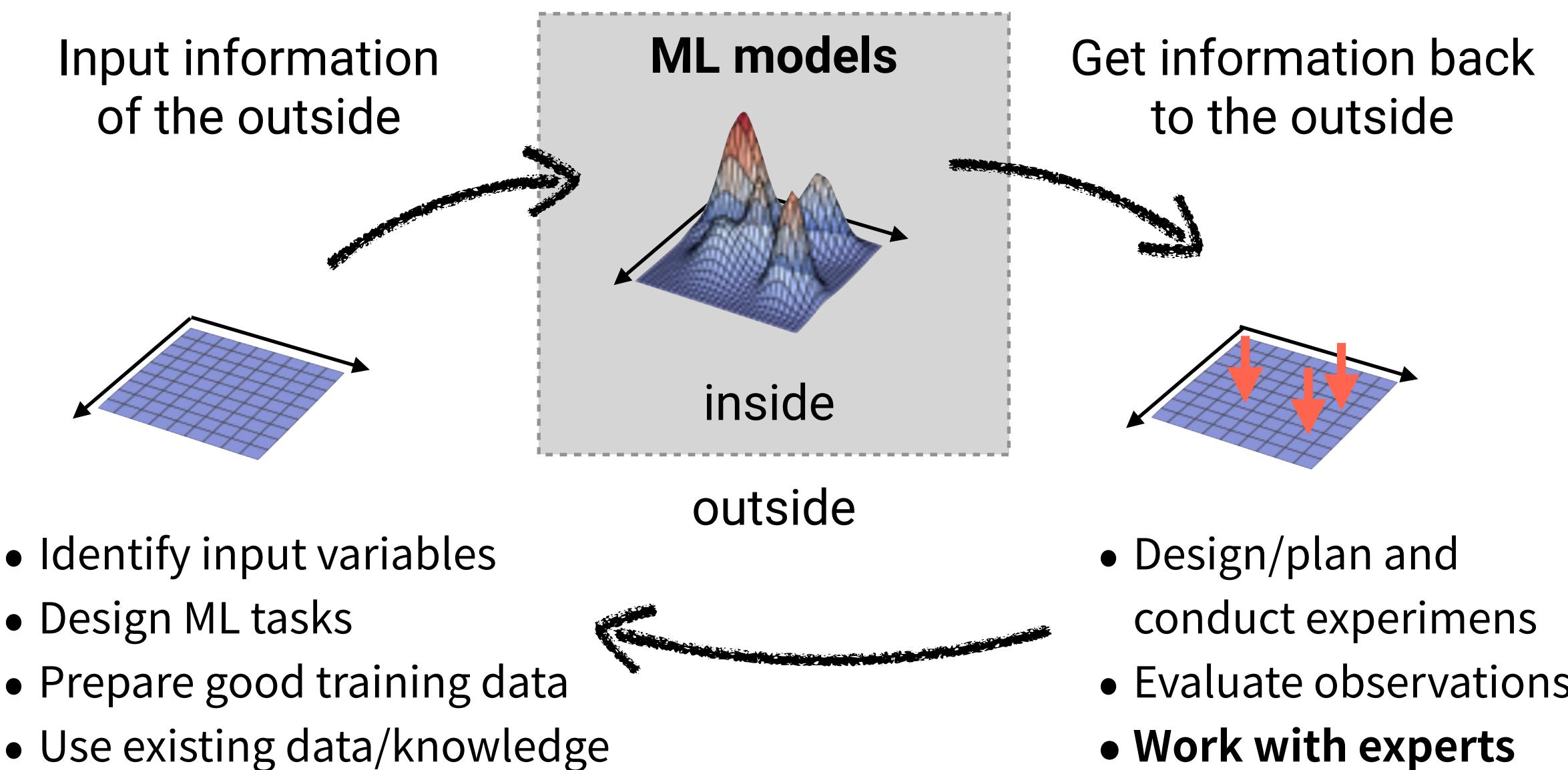
**ABSTRACT** | The two fields of machine learning and graphical causality arose and are developed separately. However, there is, now, cross-pollination and increasing interest in both fields to benefit from the advances of the other. In this article, we review fundamental concepts of causal inference and relate them to crucial open problems of machine learning, including transfer and generalization, thereby assaying how causality can contribute to modern machine learning research. This also applies in the opposite direction: we note that most work in causality starts from the premise that the causal variables are given. A central problem for AI and causality is, thus,

## I. INTRODUCTION

If we compare what machine learning can do to what animals accomplish, we observe that the former is rather limited at some crucial feats where natural intelligence excels. These include transfer to new problems and any form of generalization that is not from one data point to the next (sampled from the same distribution), but rather from one problem to the next—both have been termed *generalization*, but the latter is a much harder form thereof, sometimes referred to as *horizontal*, *strong*, or *out-of-distribution* generalization. This shortcoming is not too

# From machine learning to machine discovery

The hard problems are happening **at the interface** between inside and outside of ML.



# The lesson: Science is a human activity, after all

---

We need to seriously take **all "human factors"** into consideration. All difficulty is our fault. We want "understanding" and "discovery", but neither machines nor nature do.

- **Interpretability**
  - All information needs to be simple enough so that we feel like we understand it within our (poor) cognitive capability.
- **Partiality of information**
  - We cannot observe everything, nor model everything. Any data is some finite, partial, and inevitably biased snapshot.
- **Finitude of time**
  - Discovery needs to be done within the time limit of one's life (or the extinction of mankind)

# Summary

**This slide is available at**  
<https://itakigawa.github.io/news.html>

A quick review on the **dark side** and **light side** of ML  
from both viewpoints as an ML algorithm researcher and an ML practitioner/user

1. What actually ML is?
2. The **dark side**: Modern aspects of ML
3. The **light side**: Deep learning for molecules

**May the ML Force be with you...**

Science is built up of facts, as a house is built of stones;  
but an accumulation of facts is no more a science than  
a heap of stones is a house.

*Henri Poincaré "Science and hypothesis"*

