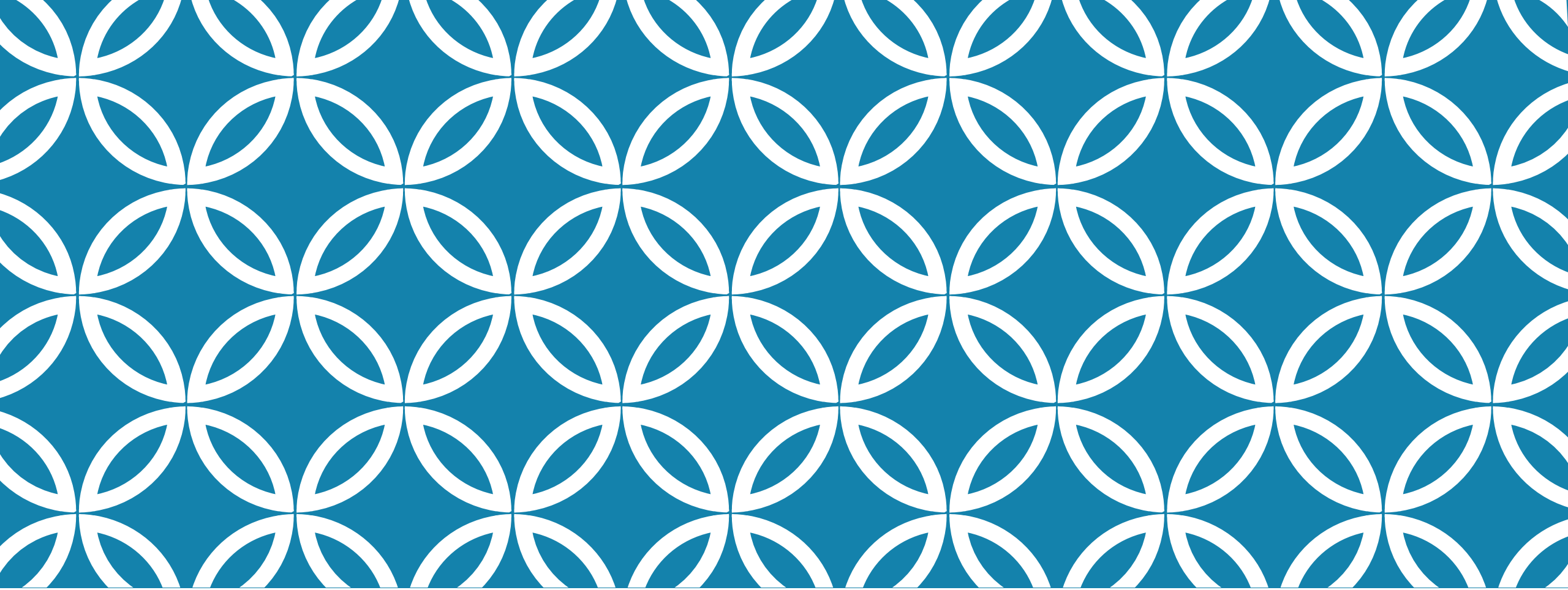


MACHINE LEARNING AND PHYSICAL CHEMISTRY

Theory & Practice

MACHINE LEARNING?

data science

Big Data

machine learning

informatics

What exactly does it do?

General mechanisms to automatically create forecasting and classification models based on data.
I don't think it's very specific to be told that
Let's actually run a machine learning task!

General Machine Learning Implementation

https://github.com/eminamitani/lecture-Minamitani/blob/master/ML.ipynb

Topics in Chemistry

https://github.com/eminamitani/lecture-Minamitani/blob/master/ML_chemistry.ipynb

APPLICATION IN SOLID STATE PHYSICS: MACHINE LEARNING POTENTIAL

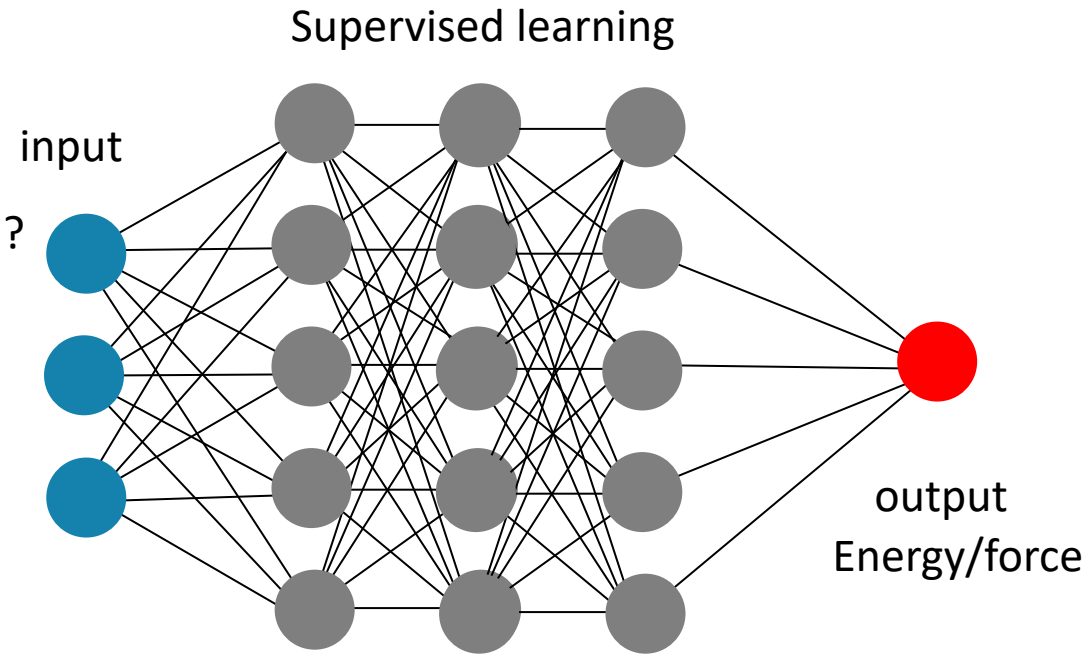
First-principles calculations are accurate but too heavy.

What ab initio calculations do

Use structure as input and return energy or force.

Can we substitute a machine learning model for this function?

Let's make potentials with machine learning.



DIFFERENCE FROM EMPIRICAL POTENTIAL

Experiential Potential

There is a function form determined by physical requirements.

Atom-atom interactions would roughly determine it.

Embedded Atom Method (EAM) Potential

$$E_{tot} = \sum_i E_i(n_i) + \frac{1}{2}\sum_{i \neq j} \phi_{ij}(r_{ij})$$

Energy assigned to each atom

Distance-dependent functions

$$\exp\left[-b^l\left(\frac{r_{ij}}{r_e} - 1\right)\right]$$

Isn't the angle between the bonds also important?

Tersoff Potential

$$V = \frac{1}{2}\sum_i \sum_{i \neq \mathrm{j}} f_c(r_{ij})[F_R(r_{ij}) + b_{ij} F_A(r_{ij})]$$

$$b_{ij} = \left(1 + \beta_i^n \zeta_{ij}^n\right)^{-1/2n} \qquad \zeta_{ij}^n = \sum_{k \neq i,j} f_c(r_{ik}) g(\theta_{ijk})$$

Depends on the angle determined by the coordinates of the three atoms
→Multi-body potential

DIFFERENCE FROM EMPIRICAL POTENTIAL

Empirical potentials: implications are easy to understand & analytical differential formulas can be obtained -> advantage in force calculation, MD

The effort for parameter optimization is tremendous.

Parameters and function forms are fixed.

→It is not clear whether parameters optimized for a specific environment can be used in other systems.

Ex) Carbon atoms with variations in bonding like sp2/sp3

Although diamond structures can be reproduced with Tersoff potentials
With that potential, graphene carbon nanotubes would be a subtle...

It is more flexible and requires less human effort to optimize.
Benefits of Machine Learning Potential

TYPES OF MACHINE LEARNING POTENTIALS

Functional form of potential is not specified.

Gaussian Approximation Potential (GAP)

A. P. Bartok et al. Phys. Rev. Lett. 104, 136403 (2010).

Based on the Kernel Law

(High-dimensional) Neural Network potential (HDNNP)

J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146101 (2007).

neural network

The functional form of the potential is determined to some extent

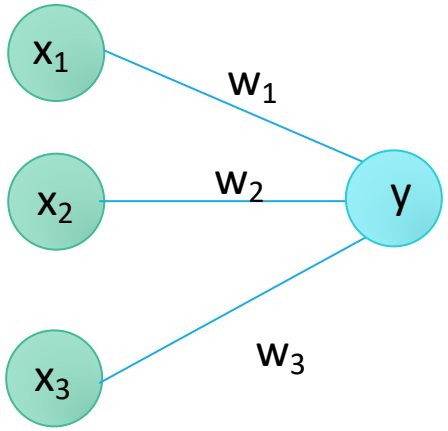
Physically informed artificial neural networks
EAM and many-body potential parameters into a neural network.

G. P. Purja Pun et al. Nat. Commun. 10, 2339 (2019).

Other new architectures:.
Graph neural network-based
SchNet
K. T. Schutt et al, Arxiv: 1706.08566 (2017)
PFP
S. Takemoto et al. Nat. Commun. 21,2991 (2022)

NEURAL NETWORK

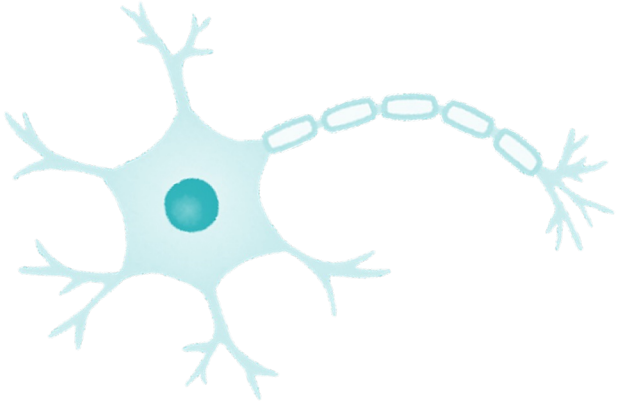
Neural network components: perceptron



$$y = \begin{cases} 0 & (w_1x_1 + w_2x_2 + w_3x_3 \leq \theta) \\ 1 & (w_1x_1 + w_2x_2 + w_3x_3 > \theta) \end{cases}$$

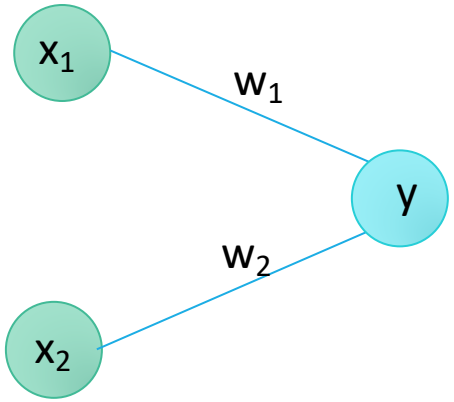
The sum of the weights multiplied by the input is
Return 0 or 1 depending on whether it is greater or less than the threshold

Firing model of neurons



Reference: Deep Learning from Zero by Yasutake Saito (O'Reilly Japan)

WHAT PERCEPTRON CAN DO FOR YOU



AND gate operation

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

w_1 =0.3, w_2 =0.6, θ=0.8

$w_1x_1 + w_2x_2$

0

0.6< θ

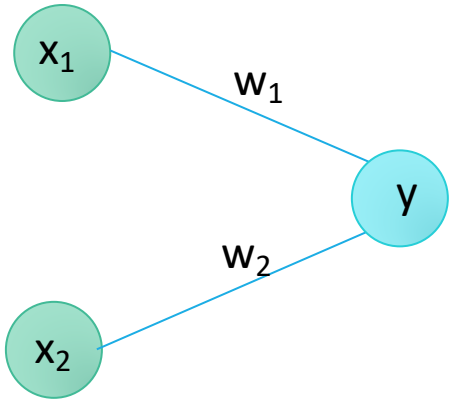
0.3< θ

0.9> θ

Various other combinations can be achieved.

WHAT PERCEPTRON CAN DO FOR YOU

Can also be used for OR and NAND gate operation by simply changing the weights and thresholds.



OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

Ex) w_1=0.3, w_2=0.6, θ=0.2

NAND

x_1	x_2	y
0	0	1
0	1	1
1	0	1
1	1	0

Ex) w_1=-0.3, w_2=-0.2, θ=-0.4

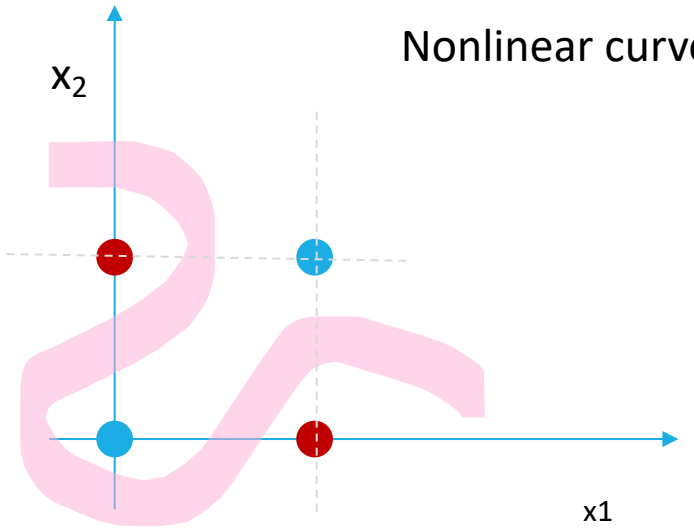
WHAT PERCEPTRON CAN DO FOR YOU

Can you make an XOR gate?

XOR

x_1	x_2	y
0	0	1
0	1	0
1	0	0
1	1	1

XOR -> linear separation not possible

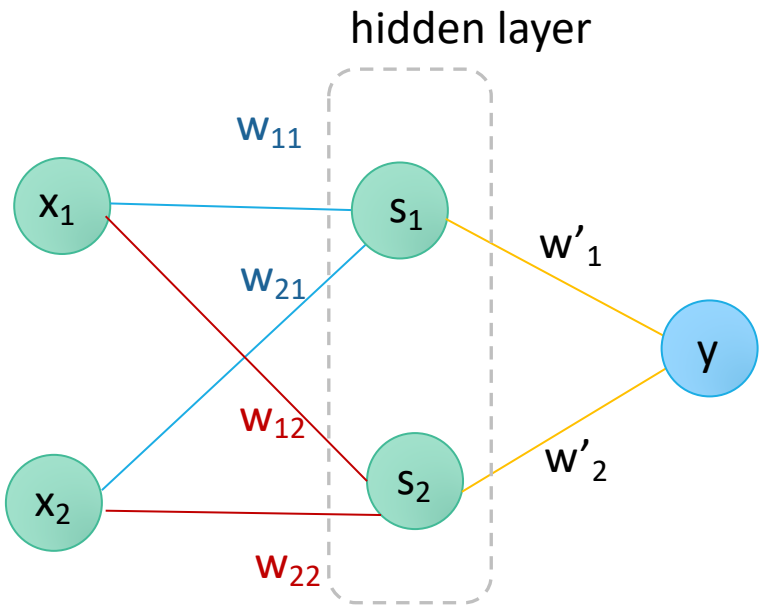


Nonlinear curves

One perceptron is not enough.

multilayer perceptron

MULTILAYER PERCEPTRON



x_1	x_2	s_1	s_2	y
0	0	1	0	1
0	1	0	0	0
1	0	0	0	0
1	1	0	1	1

Possible examples

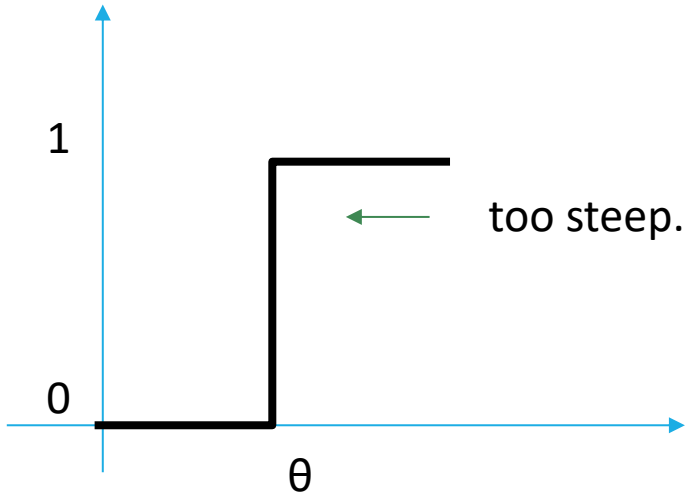
Blue part (w_{11} =-0.5, w_{21} =-0.5, θ=0)
Red area (w12=0.5, w22=0.5, θ=0.8)
Yellow area (w'1=0.4, w'2=0.6, θ=0.3)

The perceptron can be multi-layered to represent non-linear functions
→ Universal Approximation Theorem

(If there is one hidden layer, MLP can approximate any continuous function)

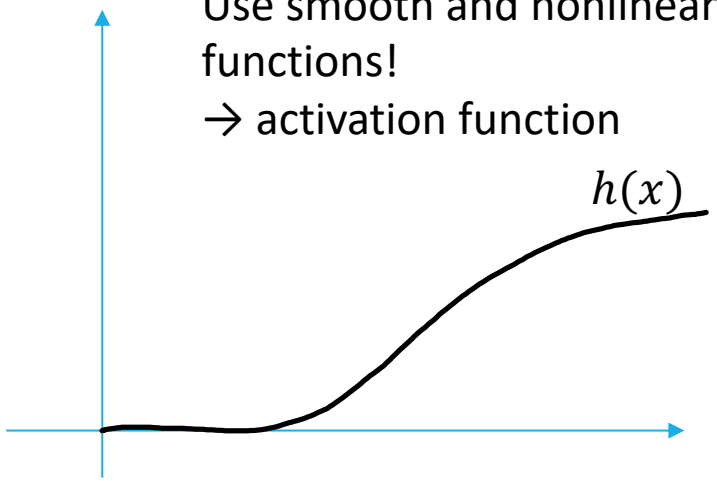
ACTIVATION FUNCTION

The perceptron was switching between 0 and 1 depending on whether it was above or below a certain threshold.



$$y = \begin{cases} 0 & (w_1x_1 + w_2x_2 \leq \theta) \\ 1 & (w_1x_1 + w_2x_2 > \theta) \end{cases}$$

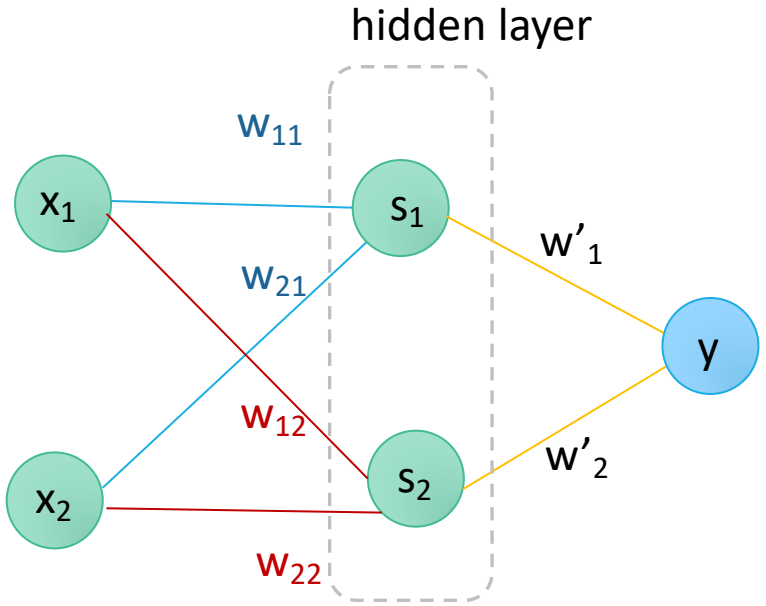
Use smooth and nonlinear functions!
→ activation function



$$y = h(w_1x_1 + w_2x_2)$$

NEURAL NETWORKS AND MACHINE LEARNING

There are an infinite number of weights/thresholds (biases) that realize the relationship between inputs and outputs.



By smoothing the activation function, the back-propagation method using derivatives can optimize weight and biases

Output obtained from a network for a given parameter and the error determined by the true value: the result of the loss function calculation.

$Loss(y, y_{true})$ e.g. Least squares error

Update parameters using their partial derivatives

$$w_{11} = w_{11} - \eta \frac{\delta(Loss(y, y_{true}))}{\delta w_{11}}$$ gradient method

Repeat this to obtain a good approximation network -> machine learning

HOW TO MAKE MACHINE LEARNING "POTENTIAL"?

Model predicts energy from structure

It is not so good that the inputs of the structure are Cartesian coordinates as they are.

physical requirement

The energy of the system is invariant to spatial translation, rotation, and replacement of homologous atoms.

It is difficult to learn this symmetry if the coordinates that the atoms have are input as they are

Request as a learning model

The model must be able to handle systems with different numbers of atoms

HOW TO EMBED ATOMIC COORDINATES INTO VECTOR DATA

smooth overlap atomic positions (SOAP)

1. set up a Gaussian function centered on the atomic coordinates

2. rewritten as a linear combination of the homogeneous distribution function + spherical harmonic function

The overlap of atoms is

Coefficient of linear combination

$$p(r_i)_{nn'l}^{Z_1Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m c_{nlm}^{Z_1}(r_i)^* c_{n'lm}^{Z_2}(r_i)$$

moments ( ) that can be represented by

S. De et al. Phys. Chem. Chem. Phys. 18, 13745-13769 (2016).

A. P. Bartók, R. Kondor and G. Csányi, Phys. Rev. B, 87, 184115 (2013).

HOW TO EMBED ATOMIC COORDINATES INTO VECTOR DATA

Atom centered symmetry function (ACSF)

J. Behler, J. Chem. Phys. 134, 074106 (2011).

Distribution of the distance between two atomic pairs

$$G_i^2 = \sum_{j \neq i} e^{-\eta\left(R_{ij}-R_s\right)^2} f_c(R_{ij})$$

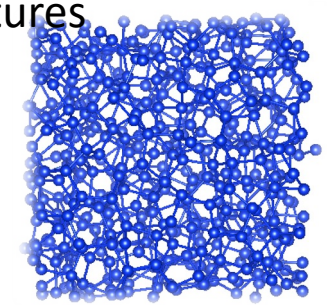
Angular distribution made by three atoms

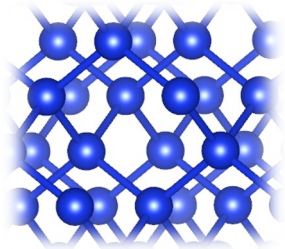
$$G_i^4 = 2^{1-\zeta} \sum_{j \neq i, k \neq i, j} (1 + \lambda \cos\theta_{ijk}) e^{-\eta\left(R_{ij}^2+R_{ik}^2\right)} \; f_c\left(R_{ij}\right) \cdot f_c(R_{ik})$$

Typical computation of these descriptors
→ Can be done with open source packages
https://singroup.github.io/dscribe/latest/

HDNNP OVERVIEW

1. collect first-principles calculations of various structures





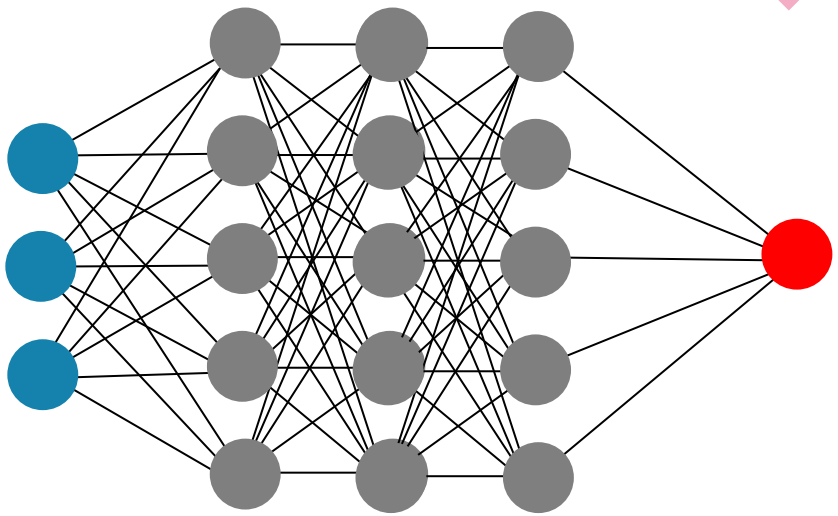
Structure & Energy/Force Pairs

2. process vector data with symmetry functions, etc.

[( ),…]

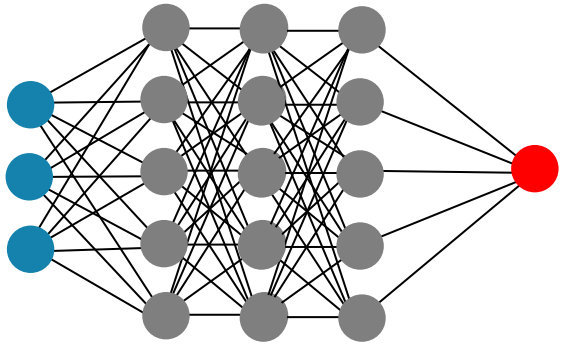
By shaking the hyperparameters
Create high-dimensional data

3. machine learning to optimize weights and biases



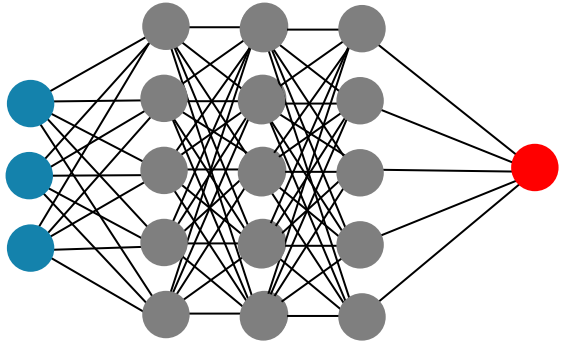
WHAT IS THE HIGHER DIMENSION?

Atomic 1 data



Atomic 1 energy

Atom 2 data



Atomic 2 energy

⋮ ⋮ ⋮

Networks are the same for the same type of elements
The entire network is represented by adding up the network

→Higher dimension.

Sum total →
Total energy of the system

(Calculate the loss function with this quantity)

Independence on the number of atoms
& guarantee constancy in the replacement of the same atom.

HDNNP PACKAGE

The *Atomic Energy Network* (**ænet**)

http://ann.atomistic.net/

n2p2

https://compphysvienna.github.io/n2p2/

Simple-NN

https://github.com/MDIL-SNU/SIMPLE-NN_v2

There are many others, but this is a sampling of those for which some documentation is available.

Including GAP and SchNet bases will give you more options.

For GAP, a package called QUIP is maintained.
https://libatoms.github.io/GAP/index.html
VASP6 also has the ability to create GAPs

Advanced packages are hard to read, I want a quick overview
→Toy models are available.

https://github.com/eminamitani/sample_NNP

RUNNING THE HDNNP TOY MODEL

I'll download the file to my hand.

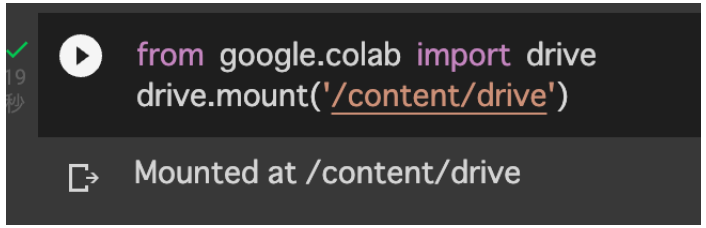
git clone git@github.com:eminamitani/sample_NNP.git (also includes pre-computed data for the symmetry function)

Have a python3 environment at hand → virtual environment + pytorch installation

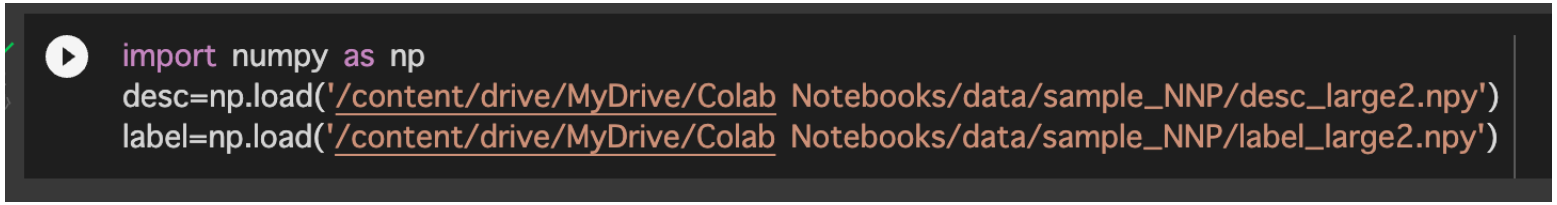
No python3 environment at hand → Use Google Collaboratory

Upload desc.npy, label.npy to your Google Drive
Upload sample_NNP.ipynb or sample_NNP2.ipynb (GPU version) at Google Collaboratory

Mount Google Drive at the beginning



Change data read path



Please adapt to your own environment.

It should work with these two changes