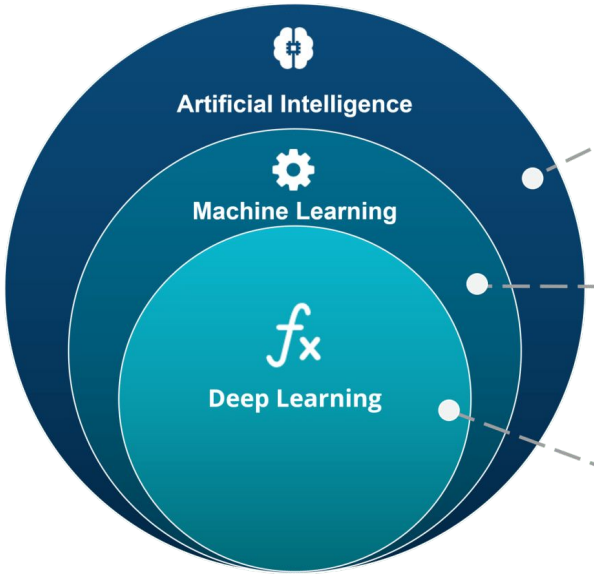
Introduction to ~~Machine Learning~~ Statistical Learning

Bin Liu

6/16/2022

What is AI, ML, DL?



ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour

MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible

What is AI, ML, DL?

A learning algorithm to predict certain desired outputs given the required inputs.

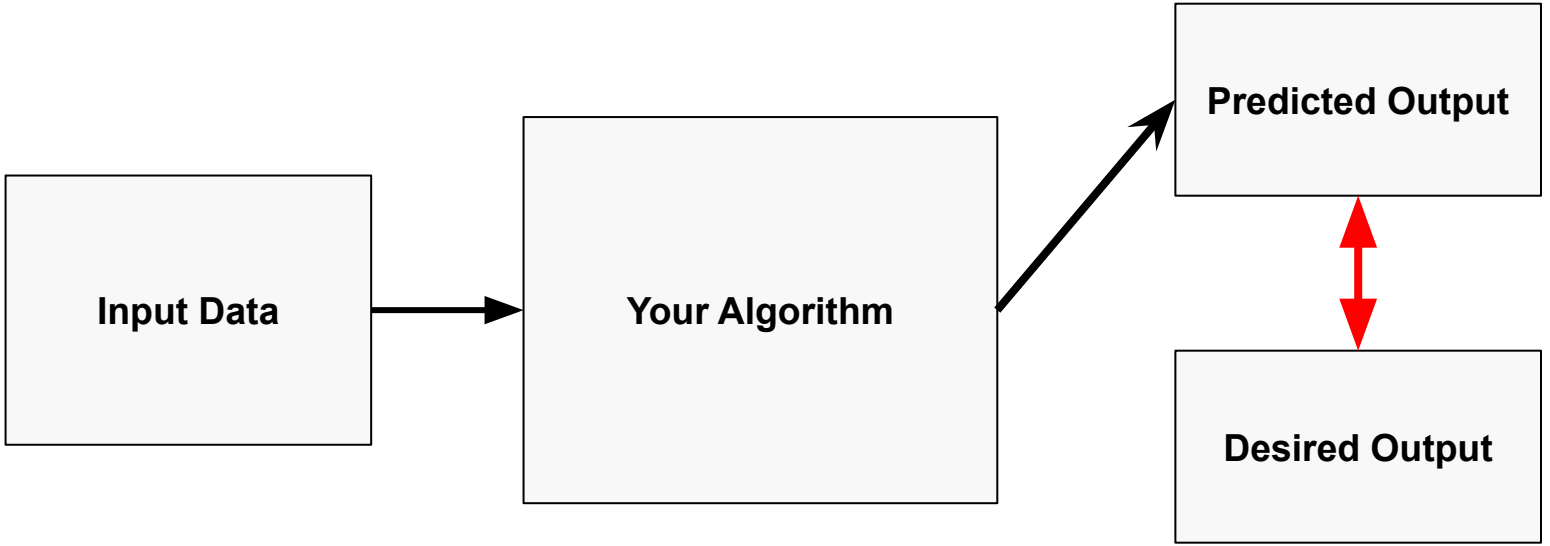
What is AI, ML, DL?

"Garbage in, garbage out"

Your analysis is as good as your data.

What is AI, ML, DL?

Learning



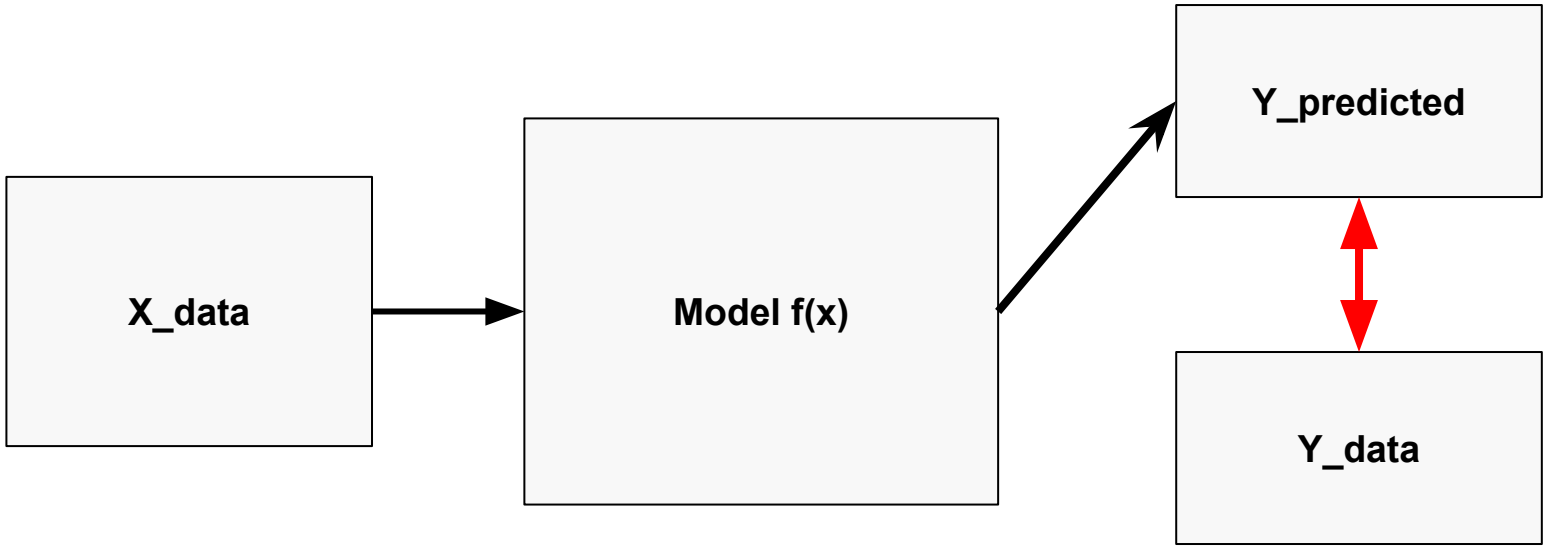
Function approximation

Machine Learning Categories

- **Supervised Learning**
  - Regression
  - Classification
- **Unsupervised Learning**
  - Clustering
  - Dimensionality Reduction
- **Reinforcement Learning**

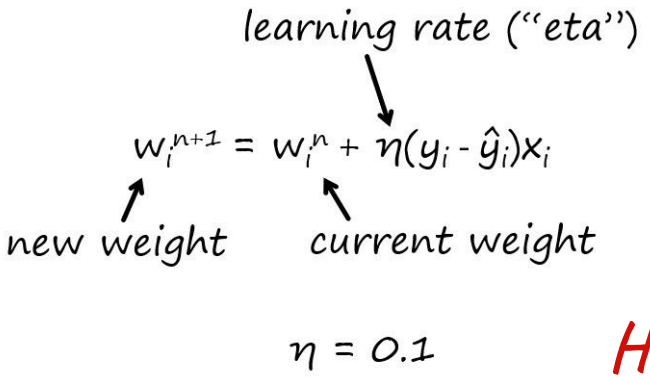
Supervised Learning

Given X_data and Y_data, find f(x) so that f(X) ~= Y_data



Supervised Learning

If f(x) = WX, we can update this function

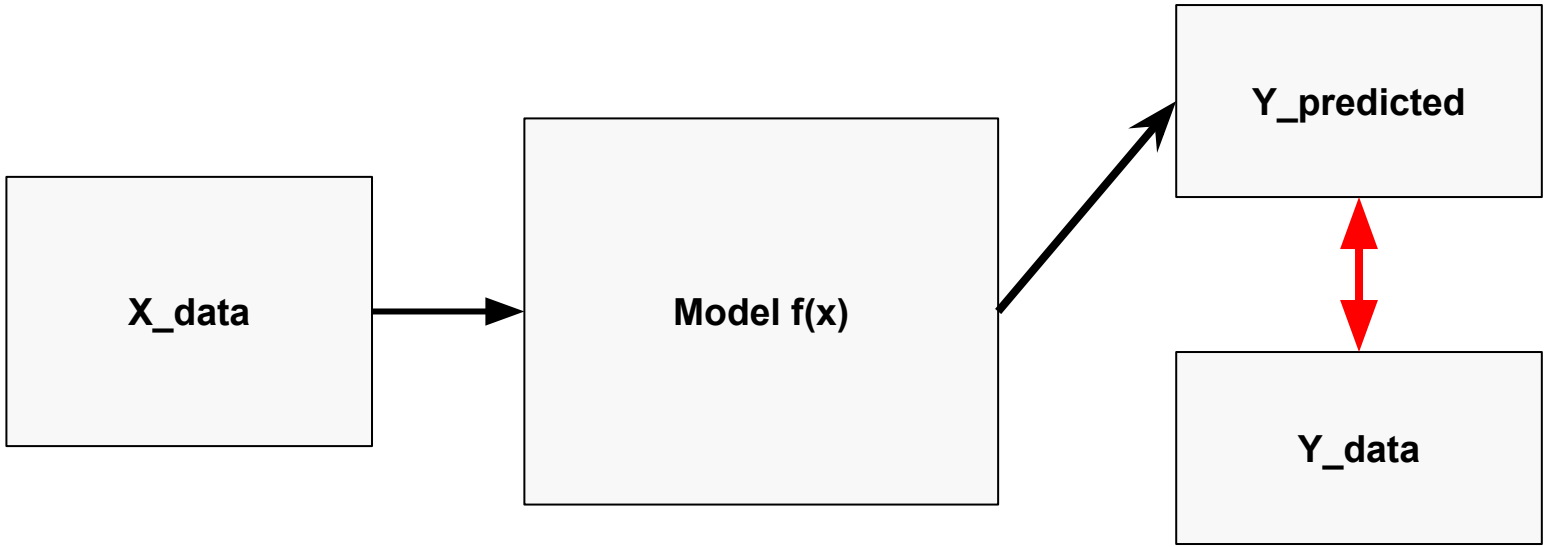


Same for both
- regression,
- classification

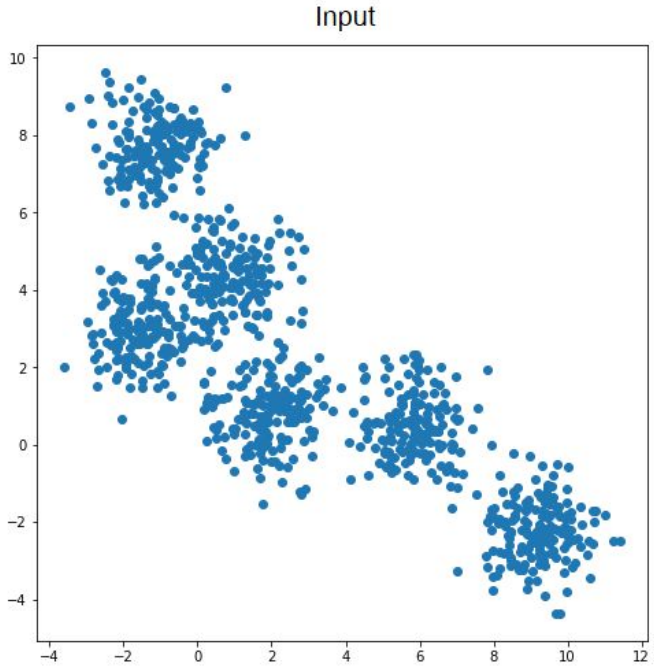
How to generalize this?

Unsupervised Learning

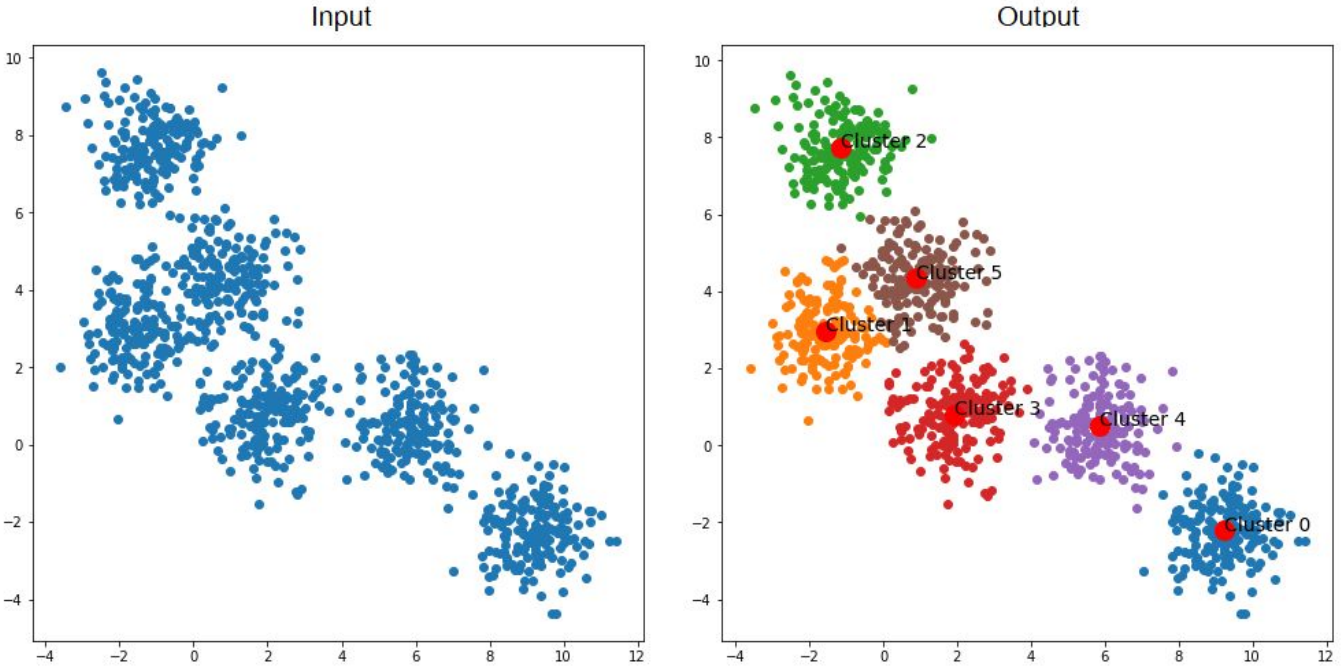
Given X_data only, we don't know anything about Y…



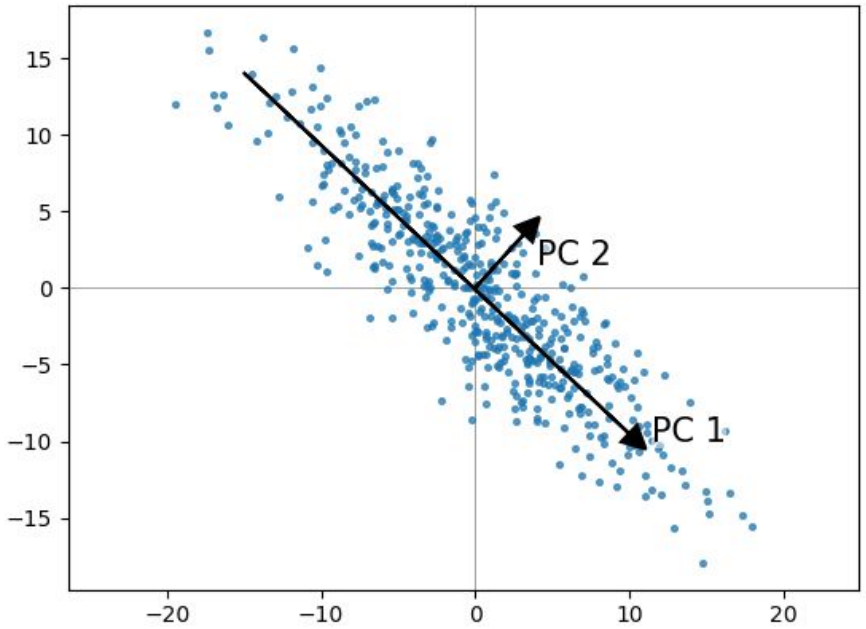
Clustering



Clustering


Input
Output
Cluster 2
Cluster 5
Cluster 1
Cluster 3
Cluster 4
Cluster 0

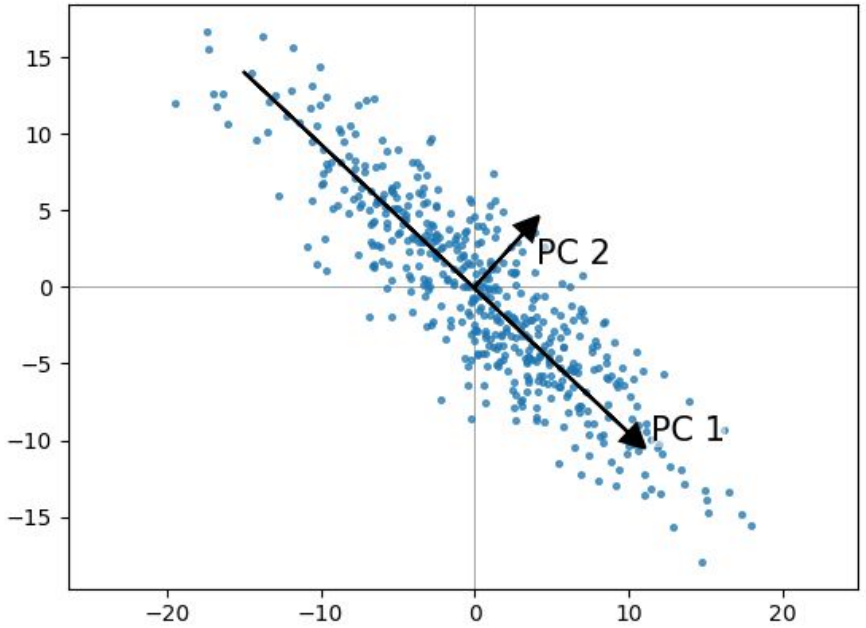
Dimensionality Reduction



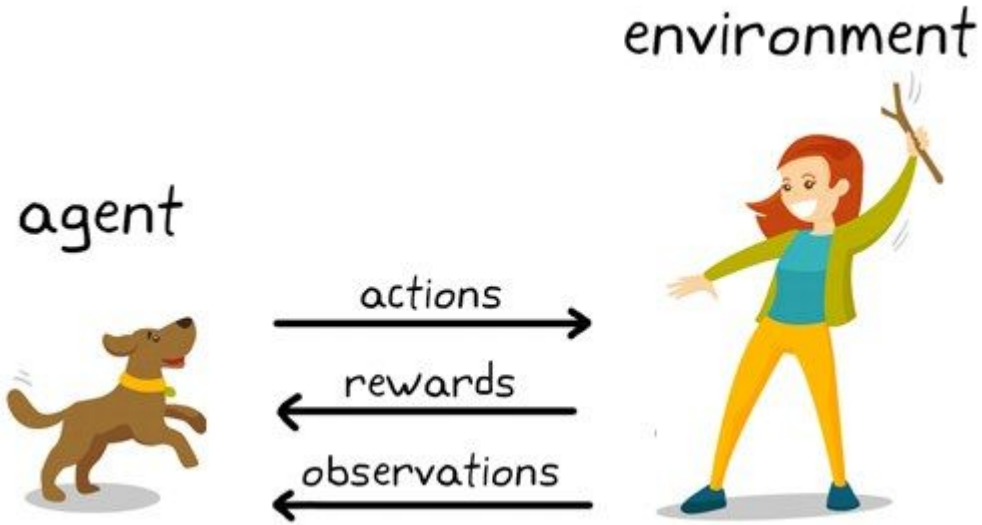
Dimensionality Reduction

Not every data point is useful and meaningful;

Need to select data relevant to your outputs

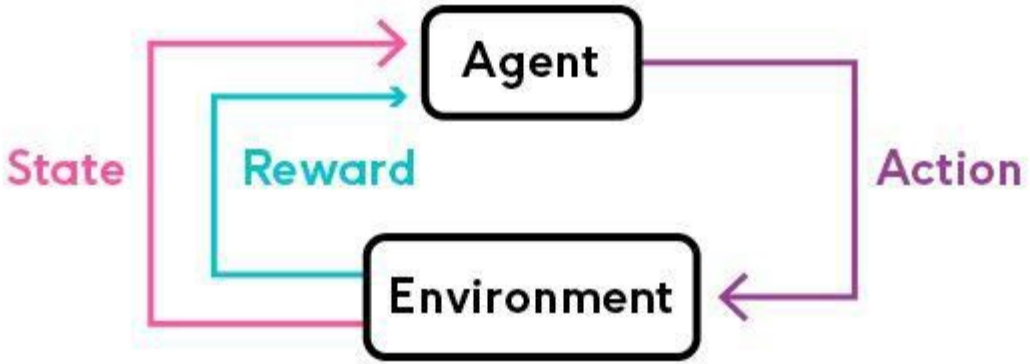


Reinforcement Learning

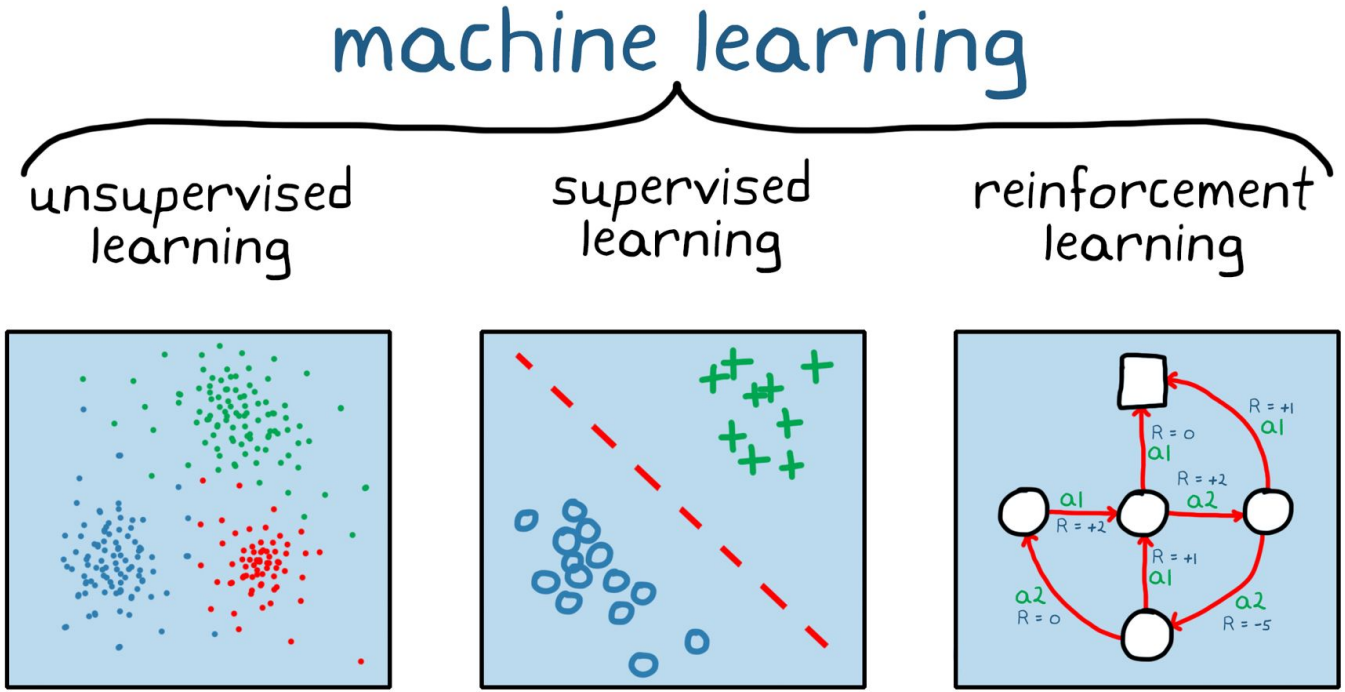


Reinforcement Learning

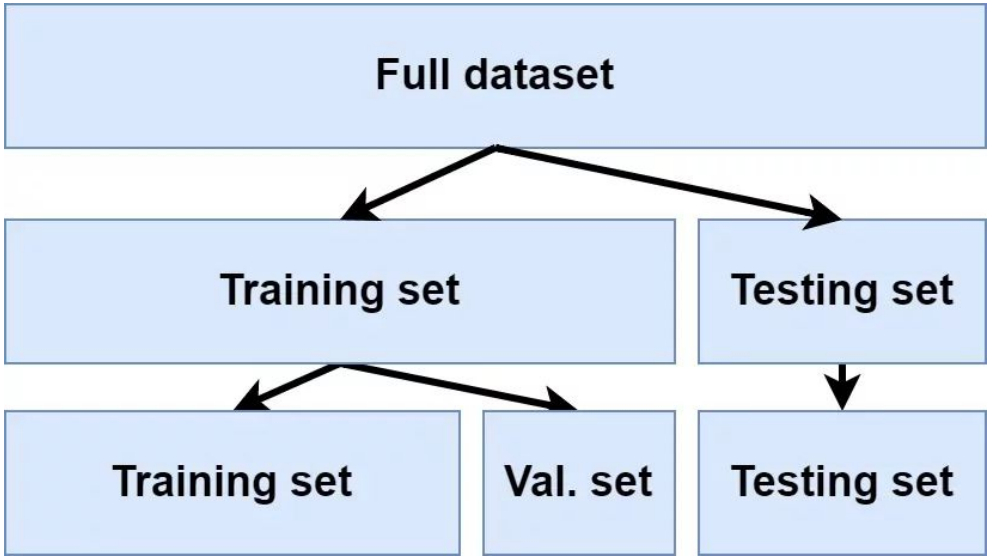
Based on current **state**, choose certain **action** so that the **agent** can get maximum **reward** from the **environment**

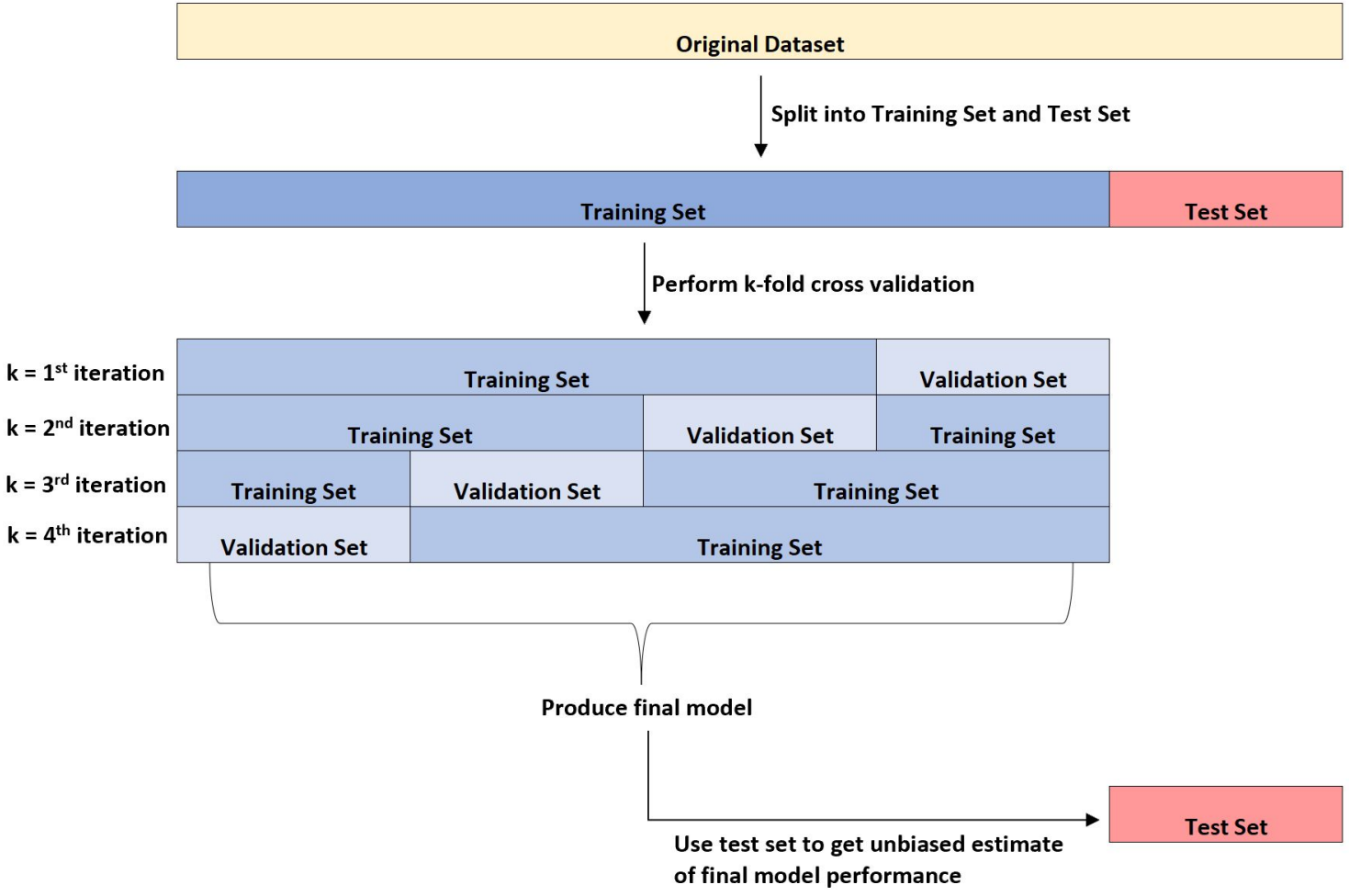


What is AI, ML, DL?

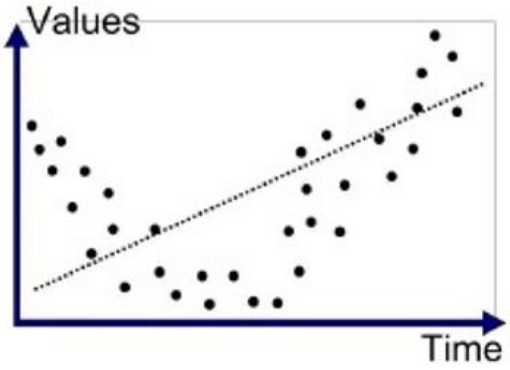


About DATA

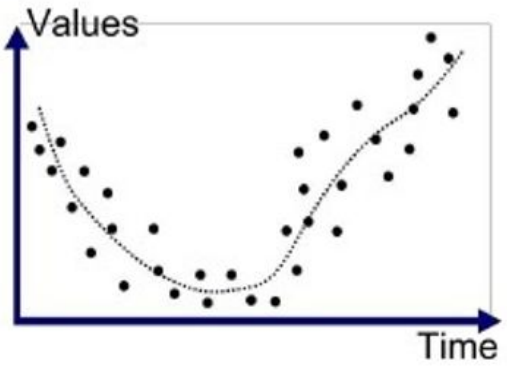

Full dataset
Training set
Testing set
Training set
Val. set
Testing set


Original Dataset
Split into Training Set and Test Set
Training Set
Test Set
Perform k-fold cross validation
k = 1st iteration
k = 2nd iteration
k = 3rd iteration
k = 4th iteration
Training Set
Validation Set
Training Set
Validation Set
Training Set
Training Set
Validation Set
Training Set
Validation Set
Training Set
Produce final model
Test Set
Use test set to get unbiased estimate of final model performance

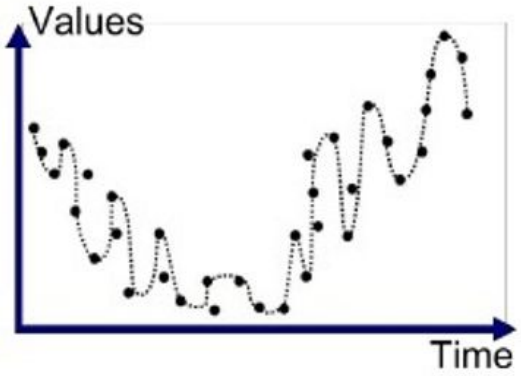
Remember Underfitting/Overfitting?



Underfitted

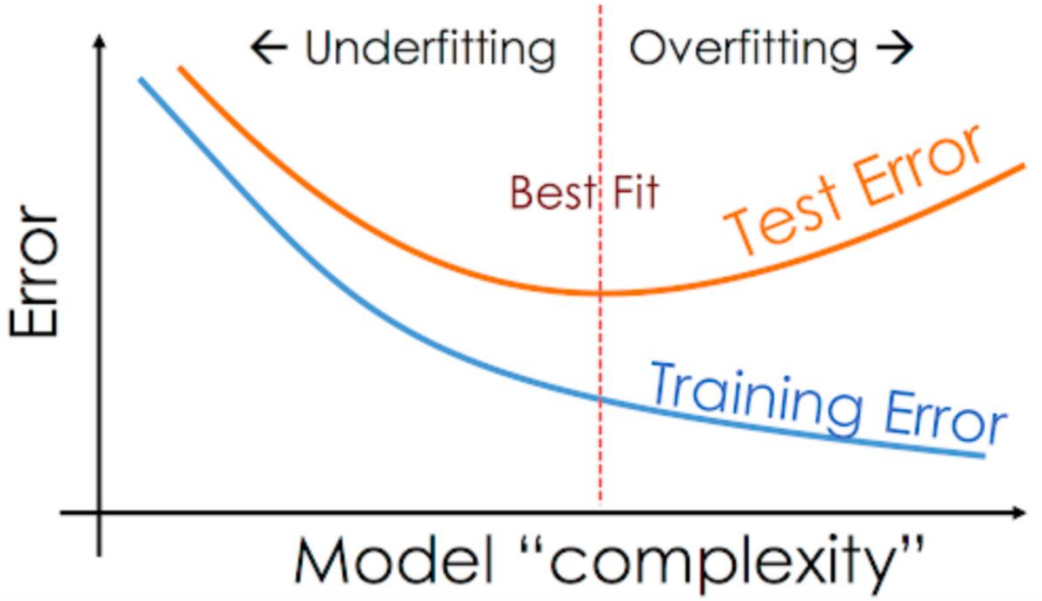


Good Fit/Robust



Overfitted

Why do we keep a test set?



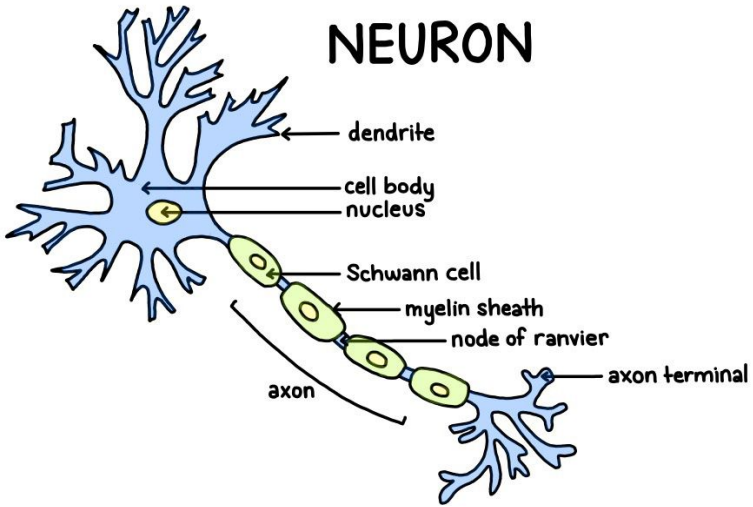
Congratulations! You've known everything about how ML works…

What about Deep Learning?

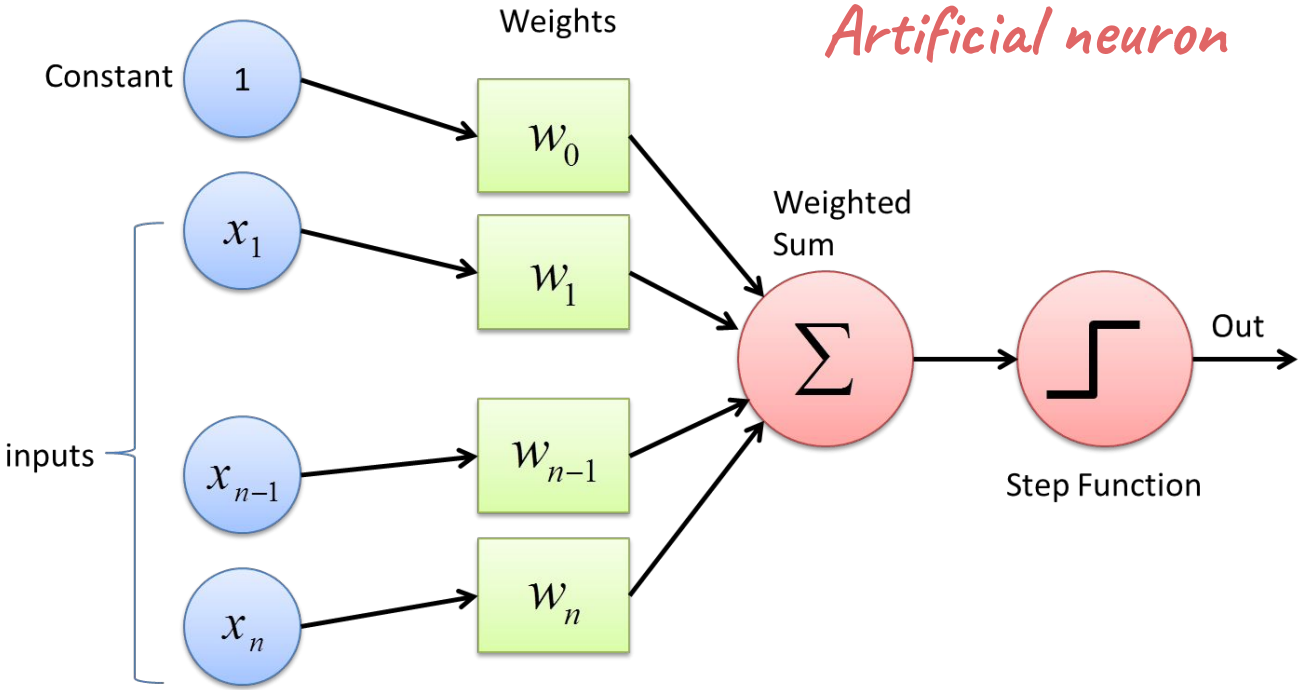
Deep Learning

It all starts with one neuron,

but not the real one…

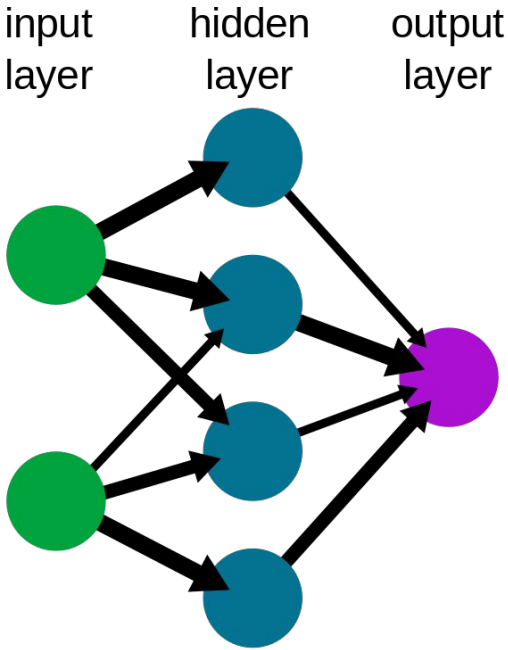


Perceptron Algorithm

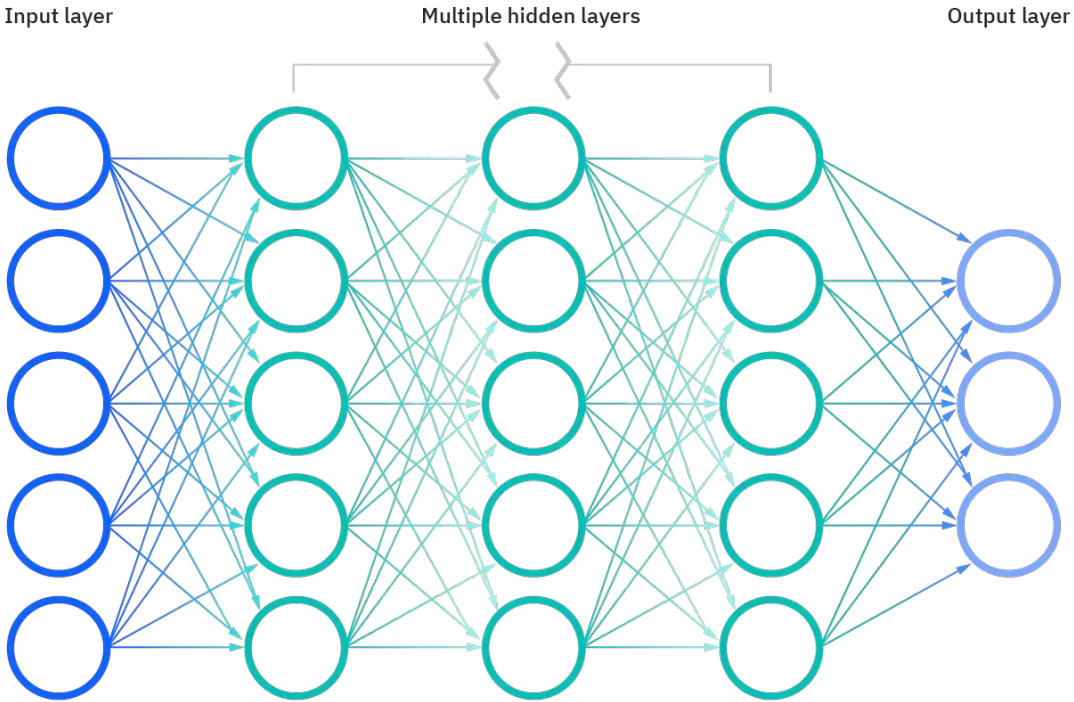


Neural Network

Assemble lots of perceptrons all together….



Deep Neural Network



Universal Function Approximation

Neural Networks has a kind of universality i.e. no matter what f(x) is, there is a network that can approximately approach the result and do the job! This result holds for any number of inputs and outputs.

How to start?

Advanced calculus, probability theory, random processes, pattern recognition…

A strong background makes your optimization process easier.. Do not treat any of these algorithms as black box!

How to start?

Coding framework:

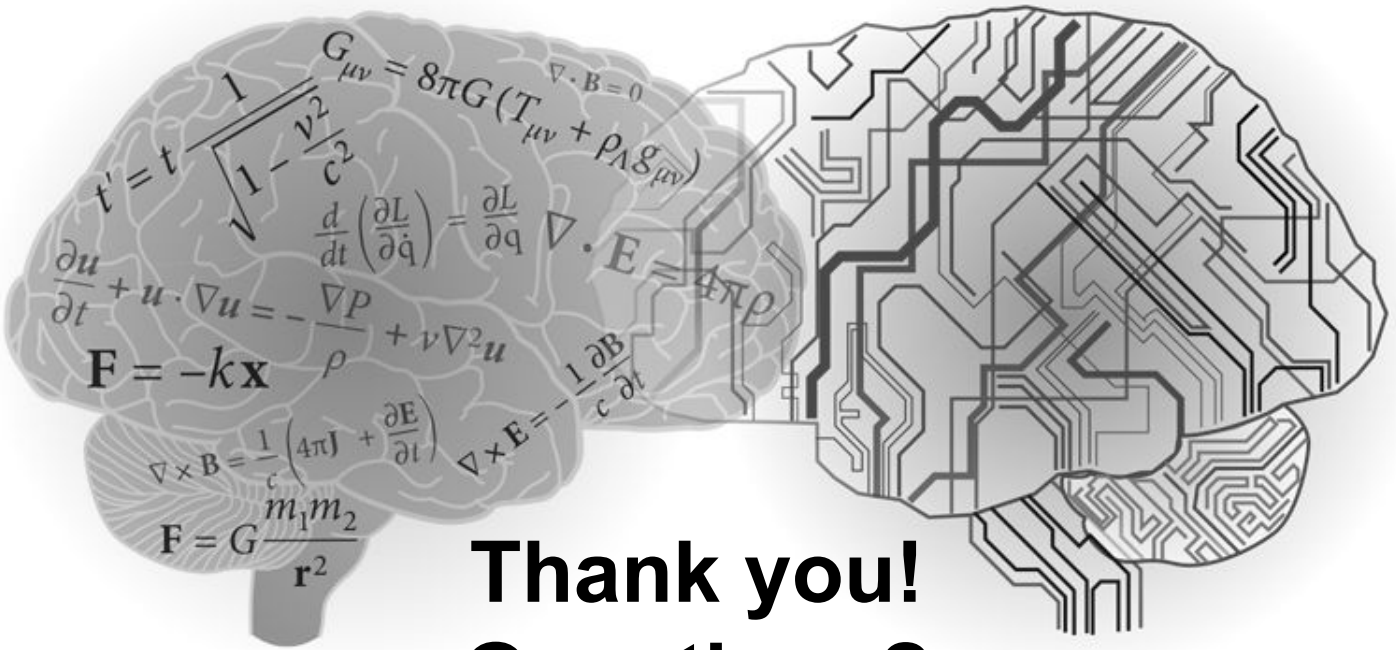
Best for ML only

From Google

From Meta

How to start?


scikit
learn
ONNX
PyTorch
TensorFlow
PaddlePaddle
DL4J
mxnet



Thank you!
Questions?