Deep Learning Standalone
–
for Chemistry

https://bit.ly/2ZxelbL
Github 저장소 링크

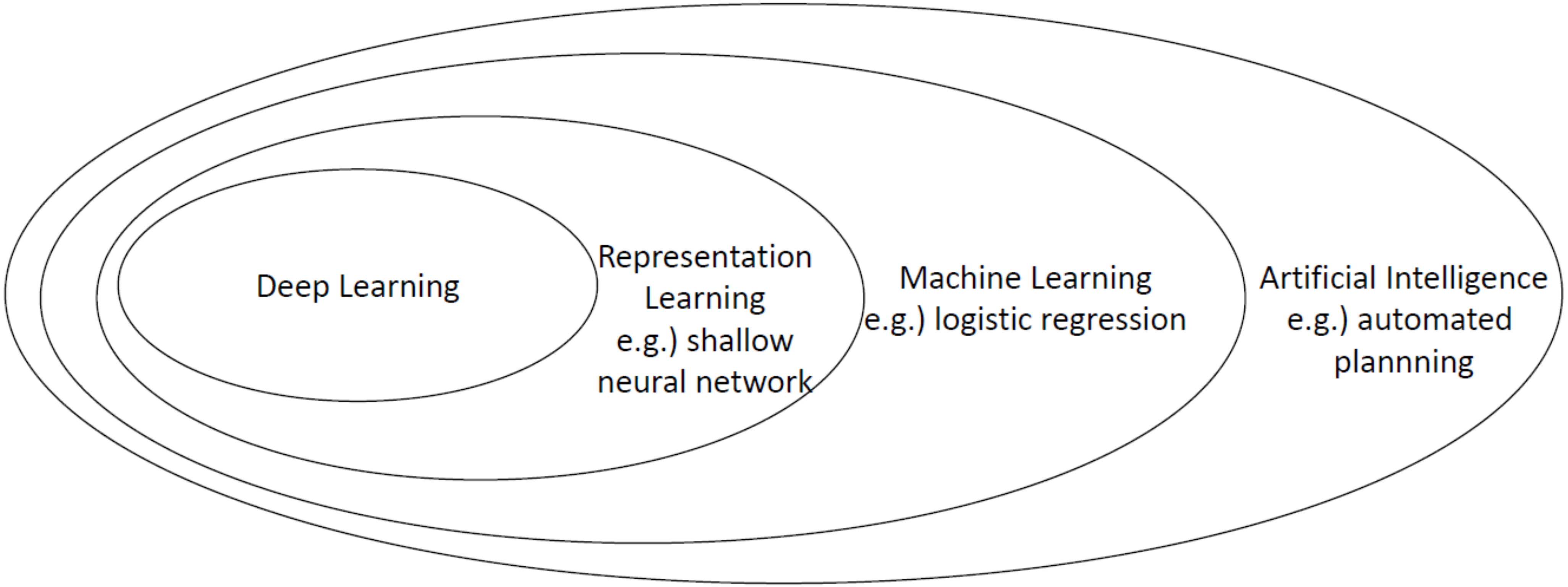
1. ML Basic
2. Pytorch Basic
3. MLP with Fingerprint Representation
4. CNN with SMILES Representation
5. GNN with Graph Representation
6. Experiment Management and Hyperparameter Tuning with Tensorboard
7. Practical Tips

What is Machine Learning?

"A Field of study that gives computer the ability to learn without being explicitly programmed"

– Arthur Samuel, 1959

Deep Learning, Machine Learning, Artificial Intelligence



Categories of ML Problems

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Semi-Supervised Learning

Categories of ML Problems

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Semi-Supervised Learning

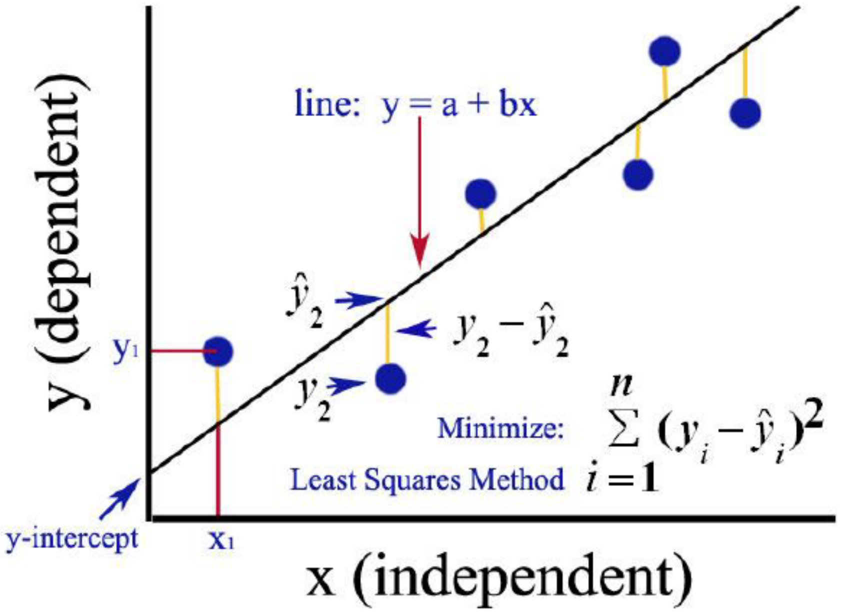
Regression Problem

Price Prediction Based on Gi-Young Style Chart Analysis

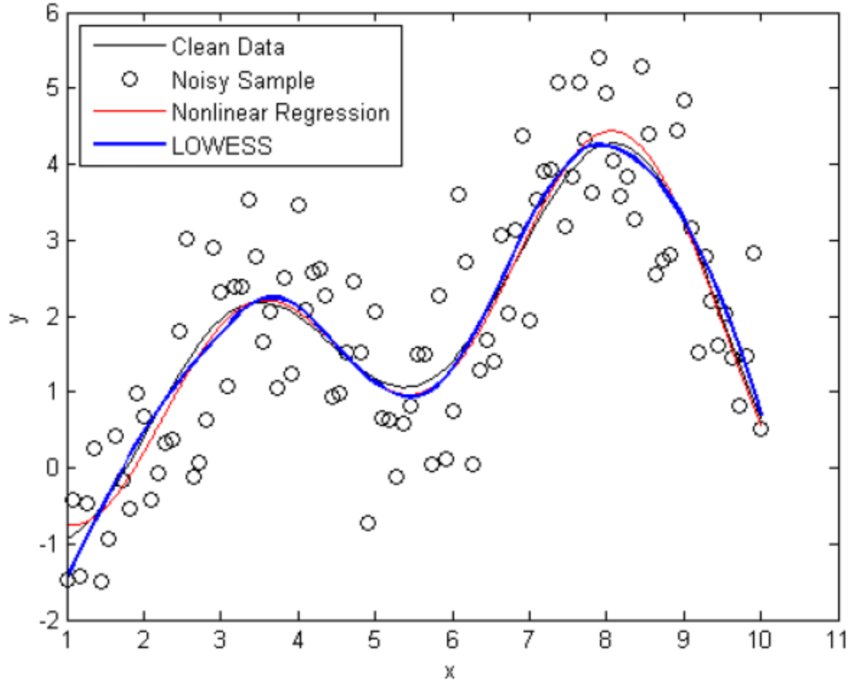
https://coinpan.com/free/22082900

Regression Problem

Fit the prediction function f(x) to the training data,
to predict continuous real value



Linear regression



Nonlinear regression

Categories of ML Problems

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Semi-Supervised Learning

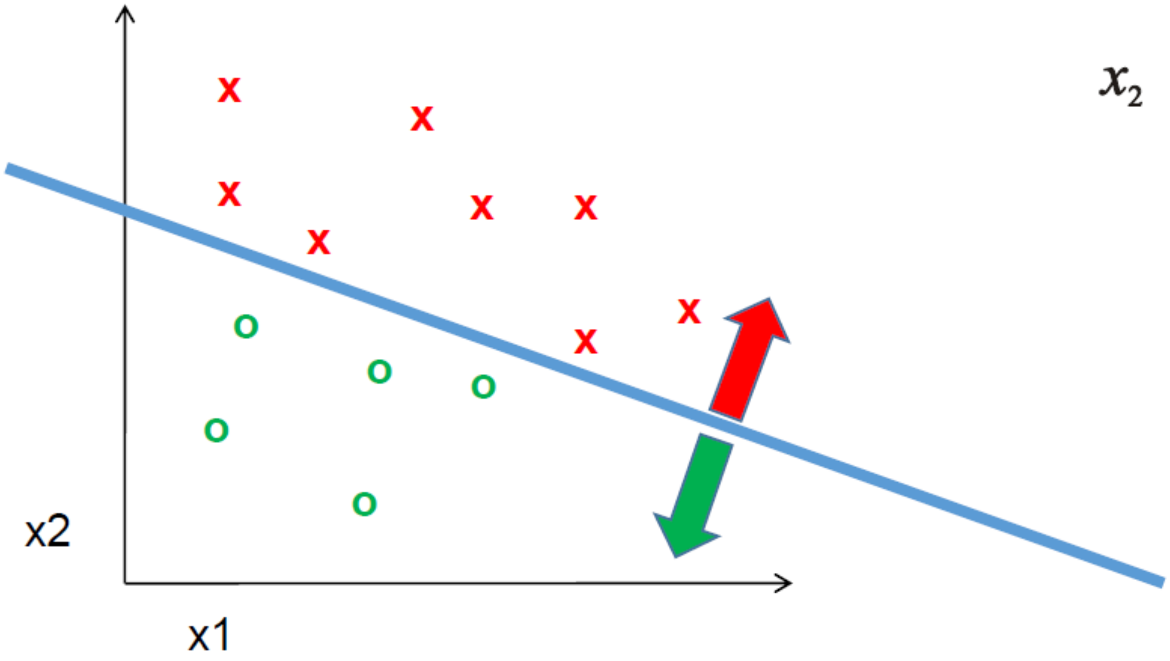
Classification Problem

Chihuahua or Muffin?

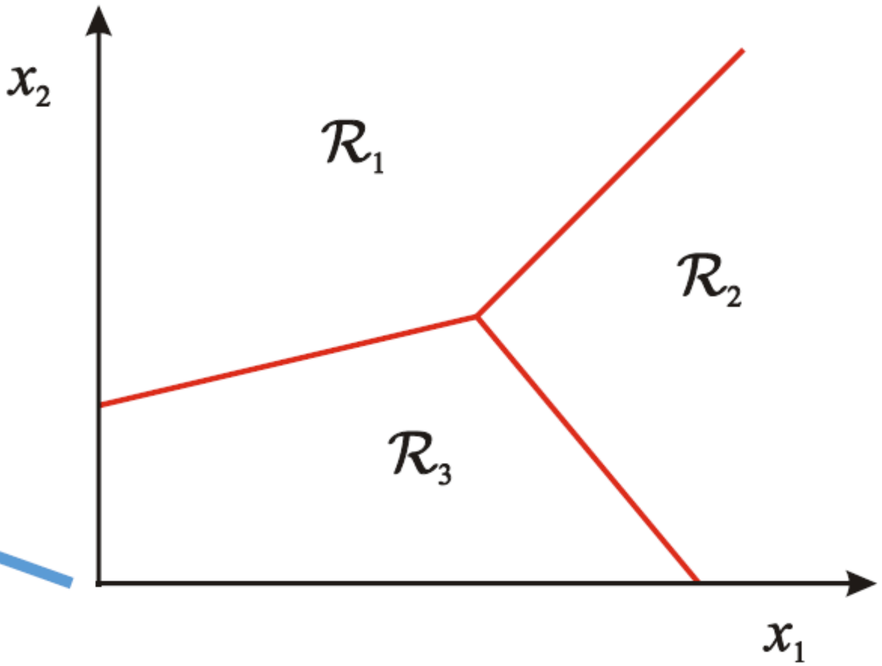
https://medium.freecodecamp.org/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d

Classification Problem

Identifying which of a set of categories a new instance belongs



Binary Classification



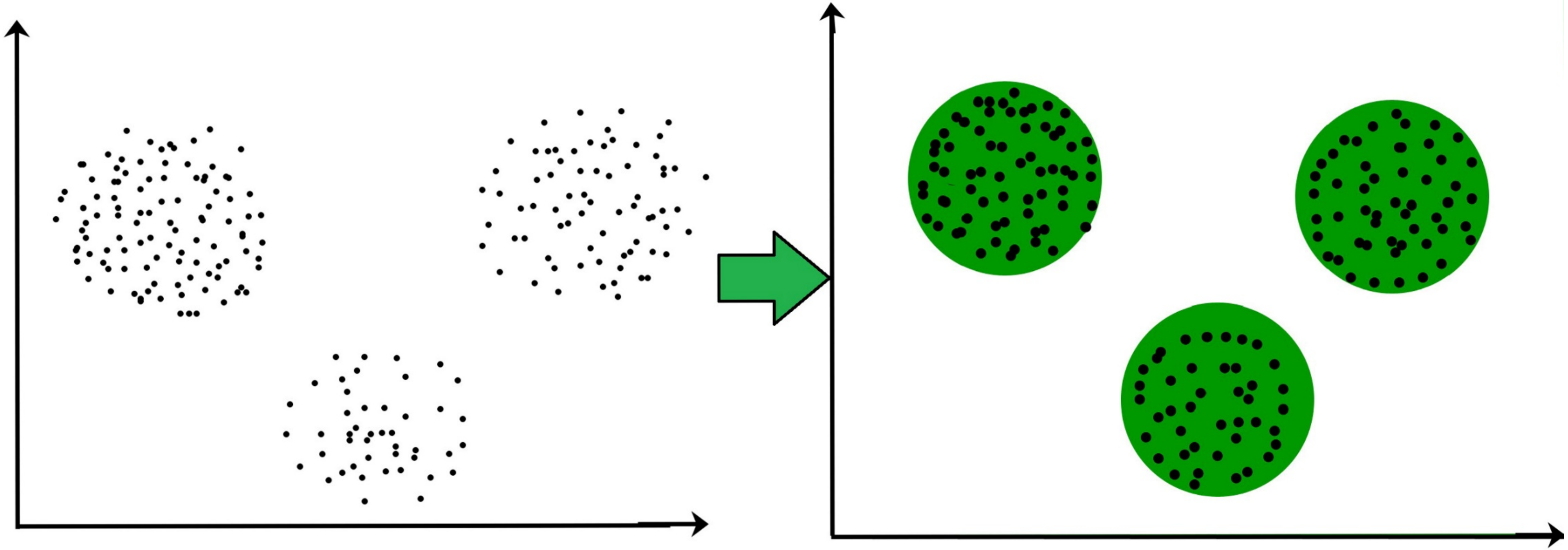
Multi–class Classification

Categories of ML Problems

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Semi-Supervised Learning

Clustering Problem

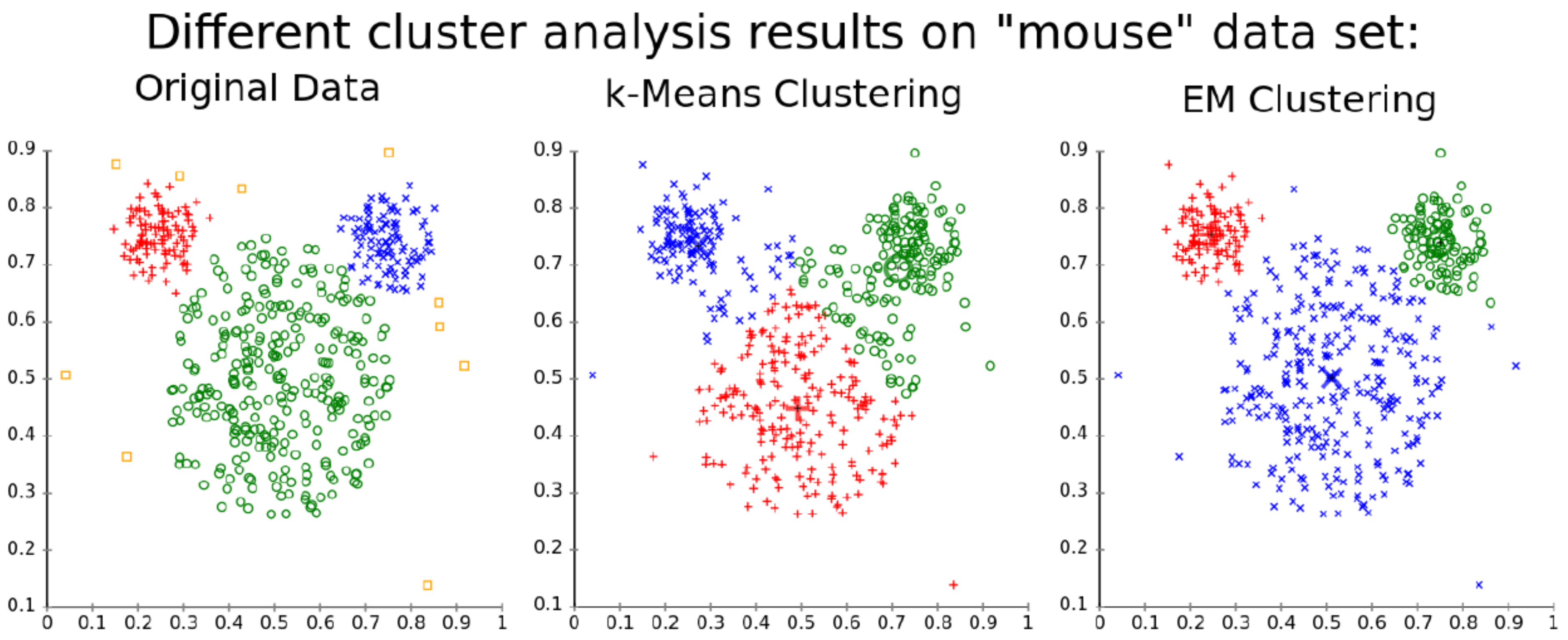


Grouping similar samples into K groups

https://www.geeksforgeeks.org/clustering-in-machine-learning/

Clustering Problem

Automatic grouping of instances, such that the instances that belong to the same clusters are more similar to each other than to those in the other groups



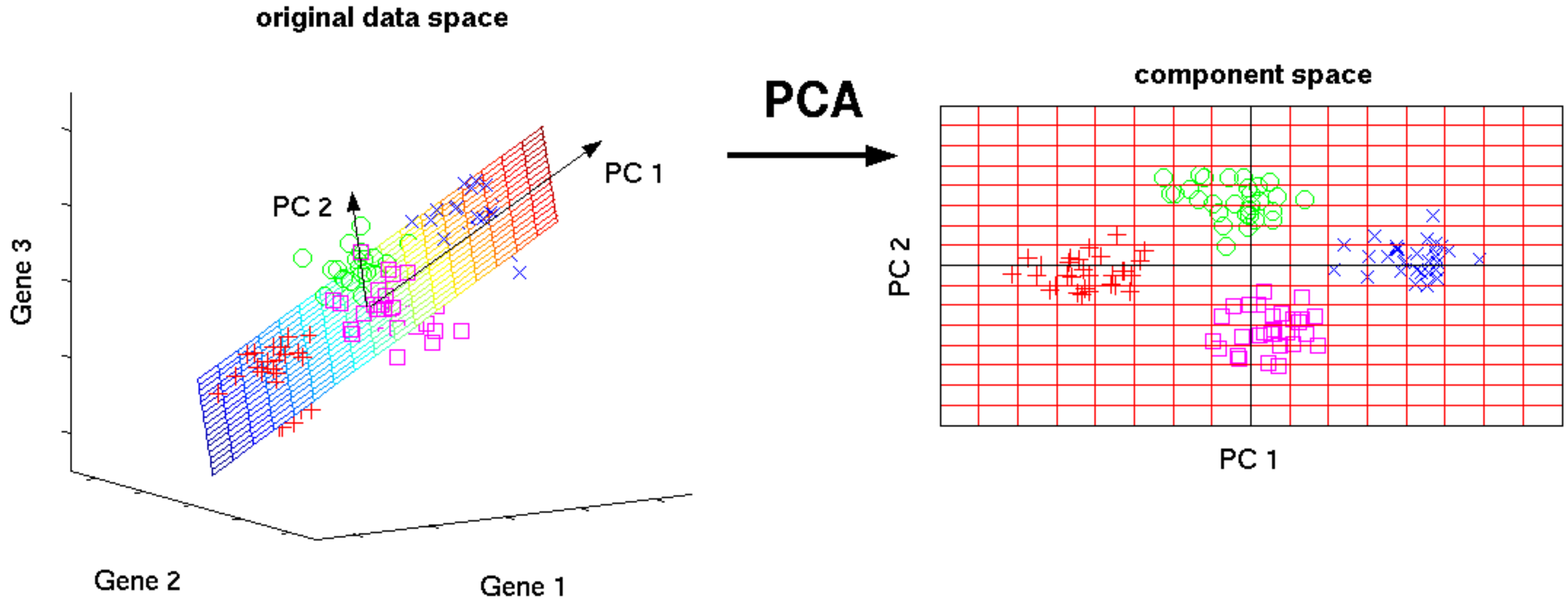
Categories of ML Problems

	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Discrete	Classification	Clustering	Discrete Action Space Agent
Continuous	Regression	Dimensionality Reduction	Continuous Action Space Agent

Semi–Supervised Learning

Dimensionality Reduction Problem

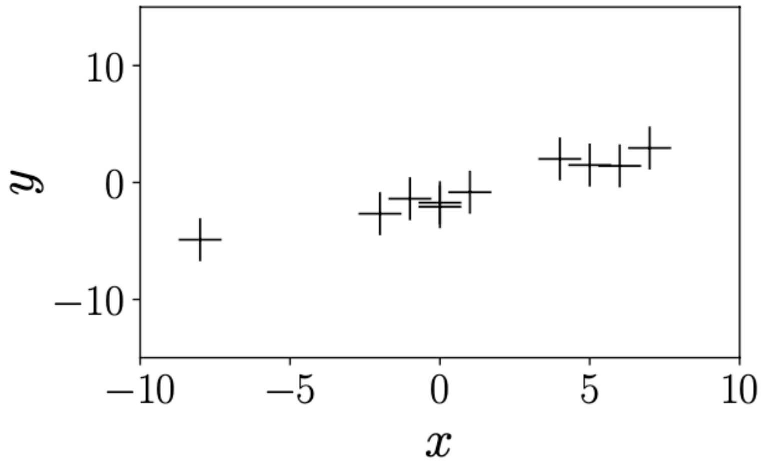
Reduce the dimension of input data, to avoid the effect of the curse of dimensionality



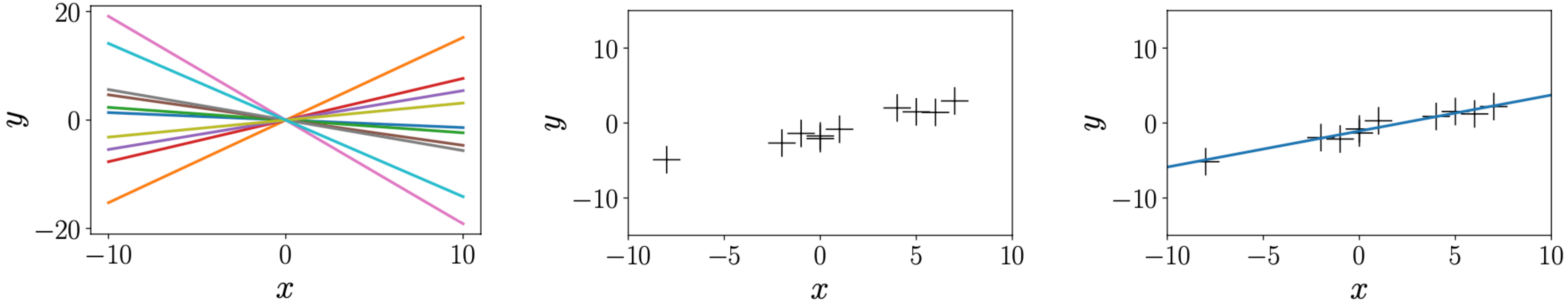
Reduce unnecessary representation axis

https://hackernoon.com/a-laymans-introduction-to-principal-components-2fca55c19fa0

Basic of ML – Linear Regression



Basic of ML – Linear Regression



Basic of ML – Linear Regression

Hypothesis

Model

Cost

Loss

Optimization

Basic of ML – Linear Regression

Hypothesis
Model

$$H(x) = Wx + b$$

Cost
Loss

Optimization

Basic of ML – Linear Regression

Hypothesis

Model

Cost

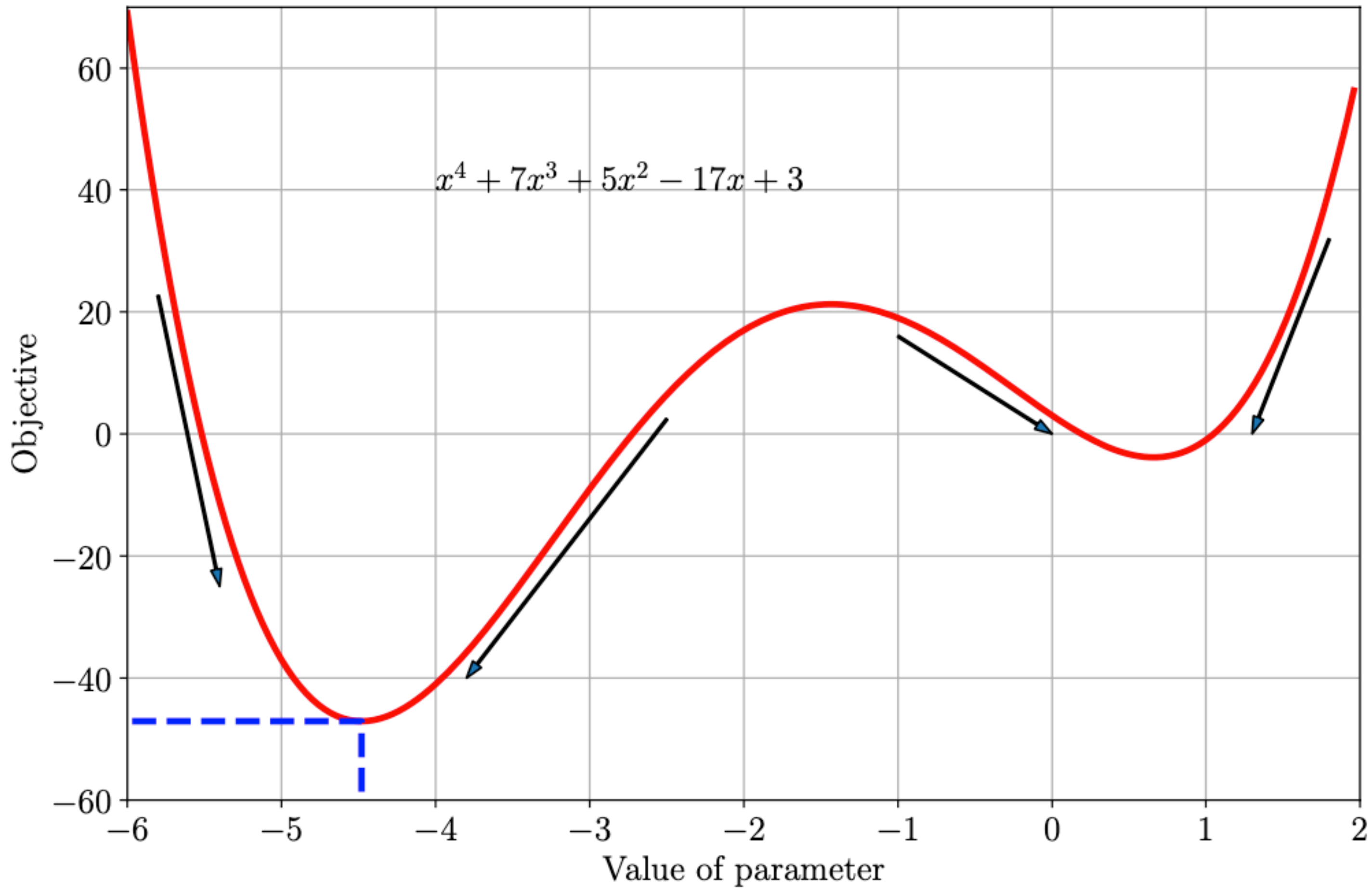
Loss

Optimization

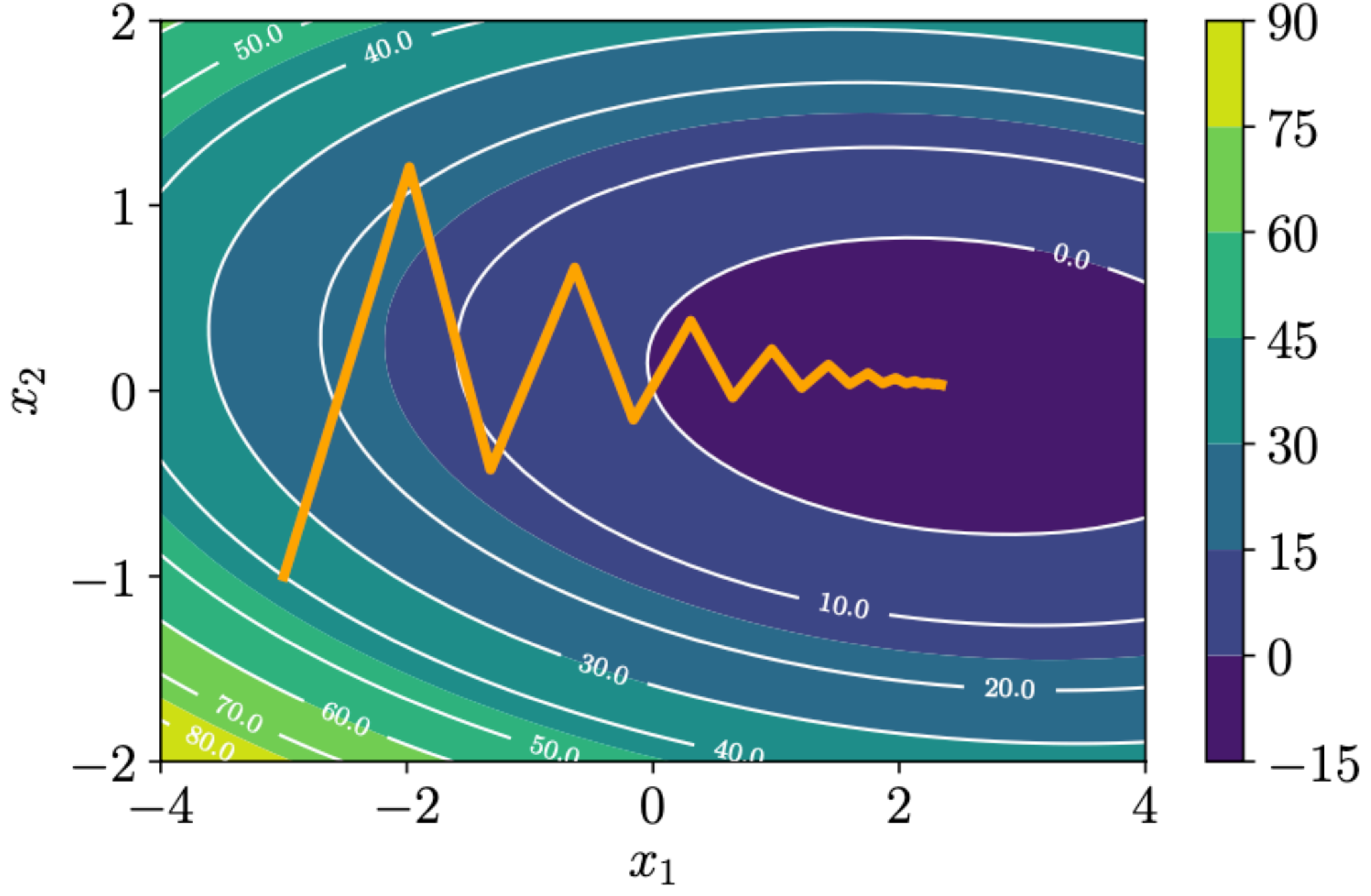
$$H(x) = Wx + b$$

$$cost(W, b) = \frac{1}{m}\sum_{i=1}^{m}(H(x^{(i)}) - y^{(i)})^2$$

Gradient Descent – 1D input



Gradient Descent – 2D input



Basic of ML – Linear Regression

Hypothesis

Model

Cost

Loss

Optimization

$$H(x) = Wx + b$$

$$W := W - \alpha \frac{\partial}{\partial W} cost(W)$$

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x^{(i)}) - y^{(i)})^2$$

Stochastic Gradient Descent

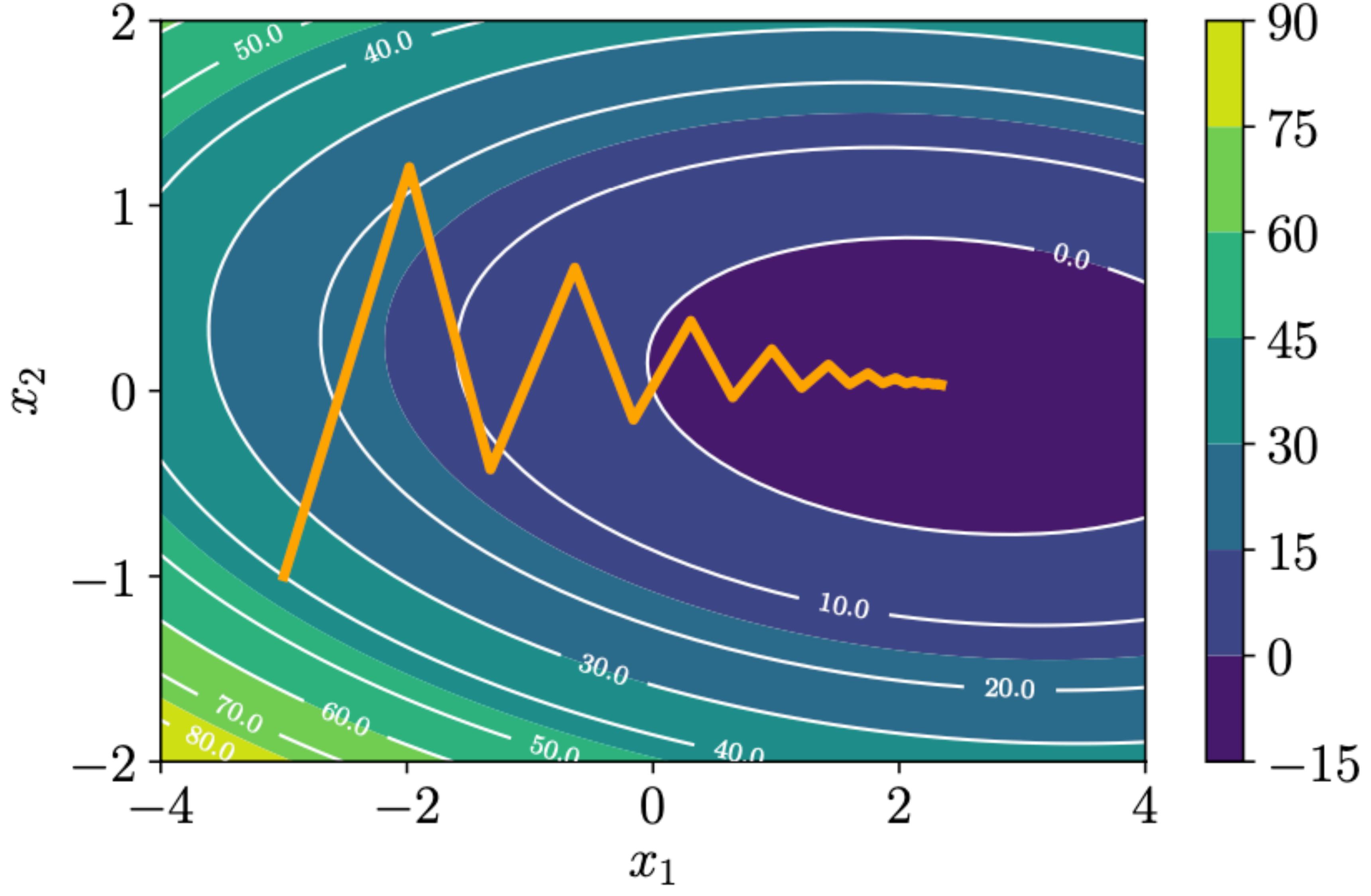
Calculate gradient for small chunk (mini-batch) of whole training dataset, rather than the whole training dataset (batch).

Stochastic since the gradient is not deterministic, but stochastic depending on the mini-batch

Faster than batch gradient descent, while converging similar.

Can avoid local minima by stochasticity.

Stochastic Gradient Descent



Basic of ML – Linear Regression

Hypothesis
Model

Cost
Loss

Optimization

$$H(X) = XW$$

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} x_{11}w_1 + x_{12}w_2 + x_{13}w_3 \\ x_{21}w_1 + x_{22}w_2 + x_{23}w_3 \\ x_{31}w_1 + x_{32}w_2 + x_{33}w_3 \\ x_{41}w_1 + x_{42}w_2 + x_{43}w_3 \\ x_{51}w_1 + x_{52}w_2 + x_{53}w_3 \end{pmatrix}$$

Basic of ML – Linear Regression

Hypothesis

Model

Cost

Loss

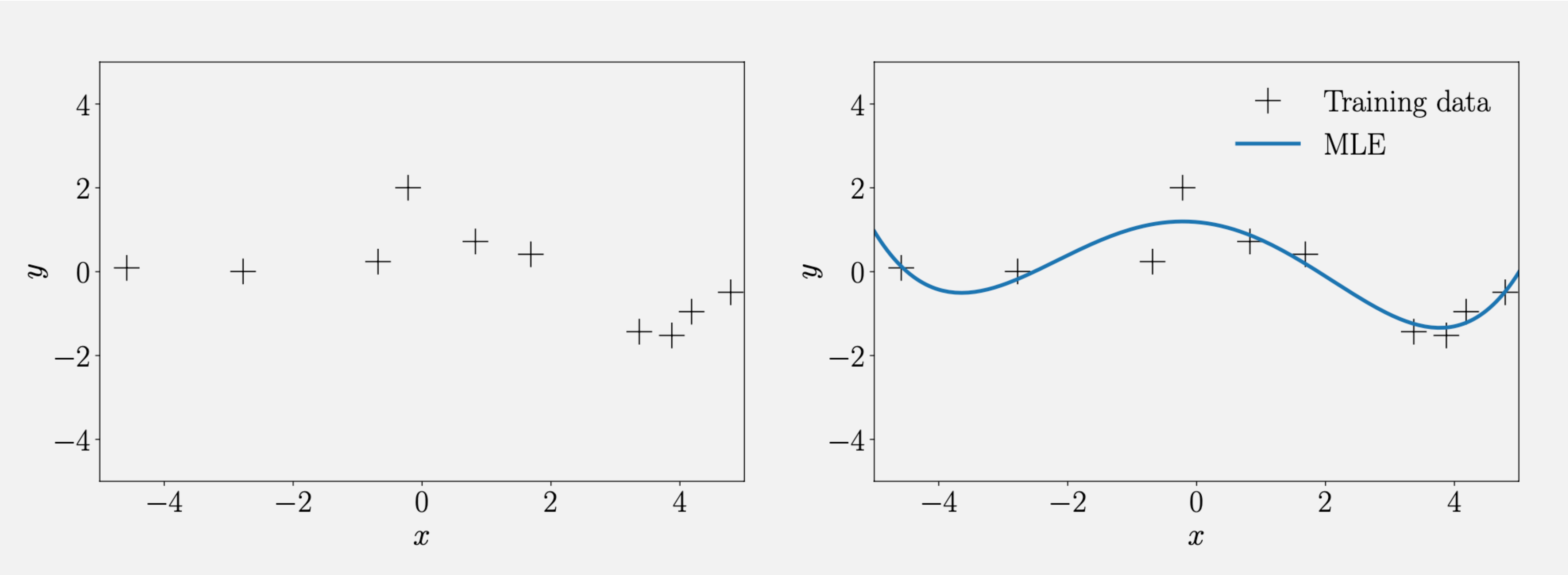
Optimization

$$H(X) = XW$$

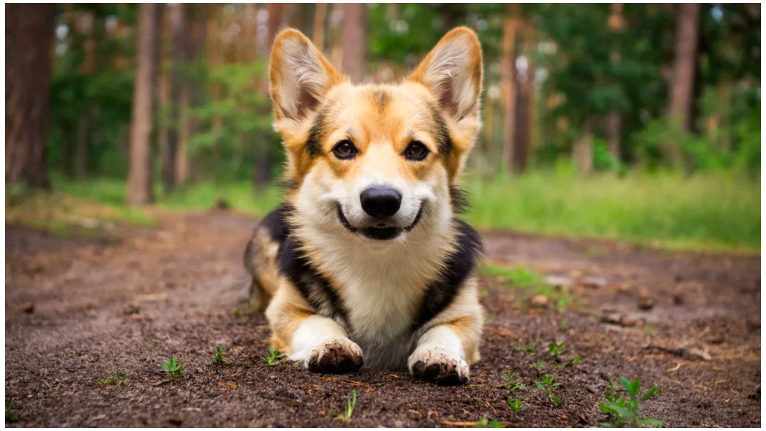
$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} x_{11}w_1 + x_{12}w_2 + x_{13}w_3 \\ x_{21}w_1 + x_{22}w_2 + x_{23}w_3 \\ x_{31}w_1 + x_{32}w_2 + x_{33}w_3 \\ x_{41}w_1 + x_{42}w_2 + x_{43}w_3 \\ x_{51}w_1 + x_{52}w_2 + x_{53}w_3 \end{pmatrix}$$

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x_1^{(i)}, x_2^{(i)}, ..., x_n^{(i)}) - y^{(i)})^2$$

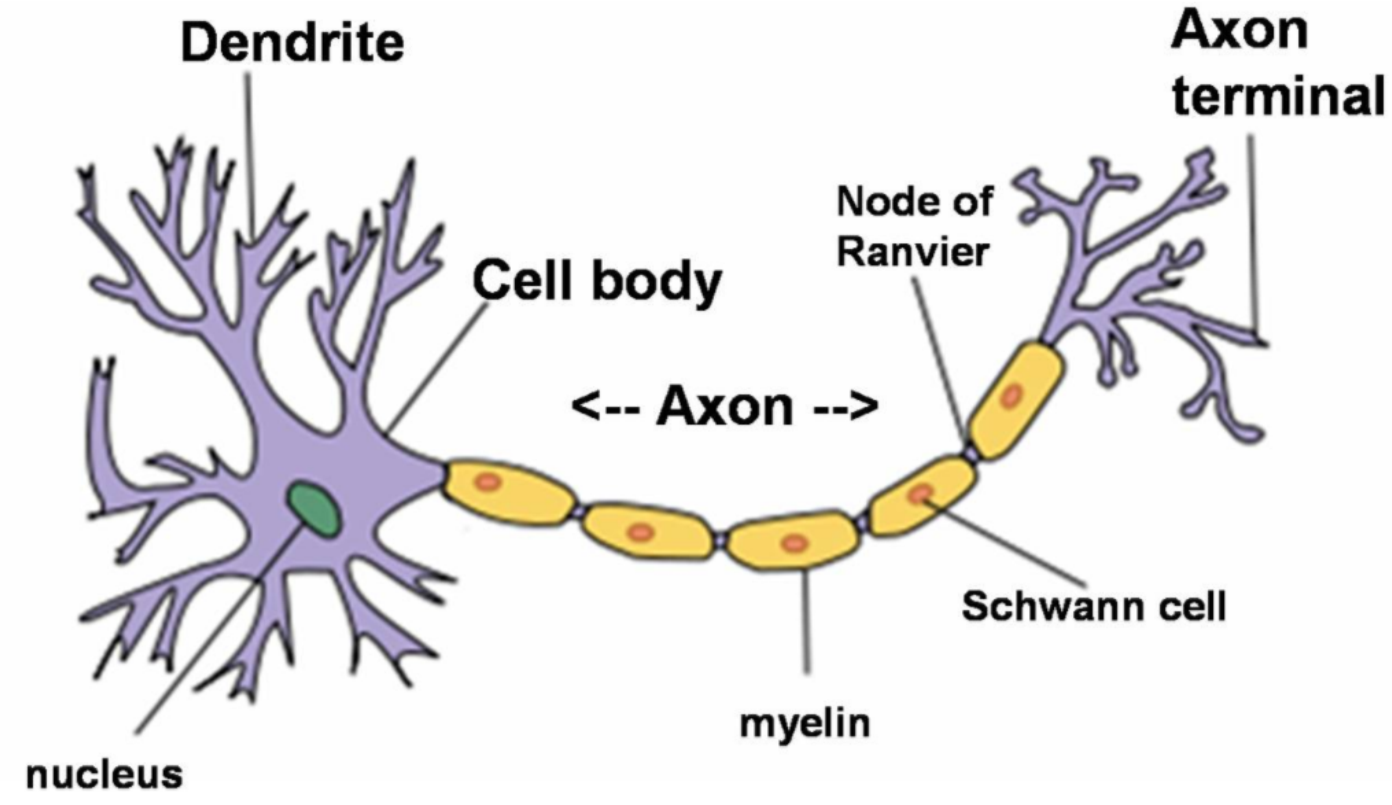
Basic of ML – Linear Regression



Limitation of ML – MNIST/Cat or Dog?

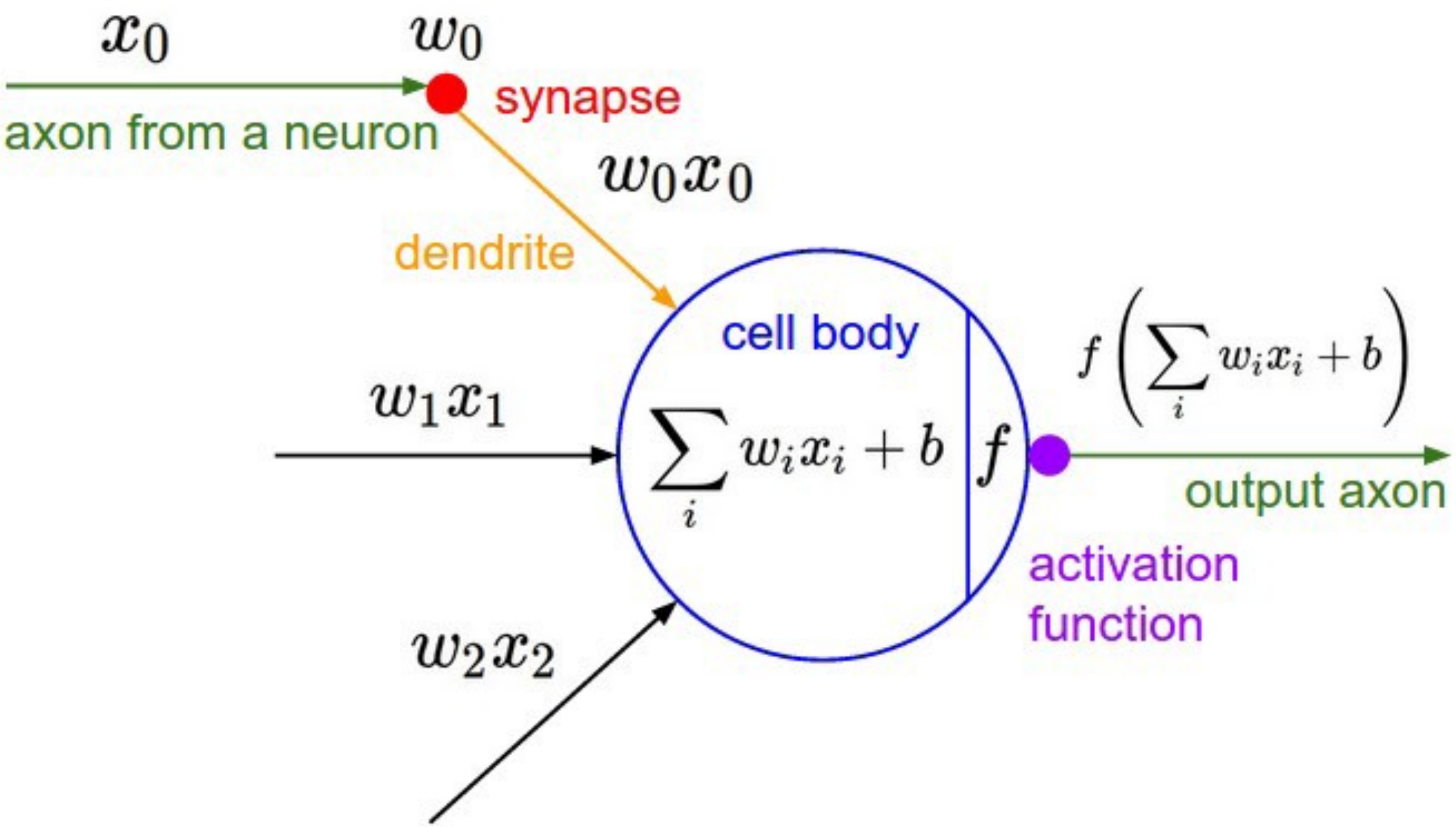


Structure of Neuron



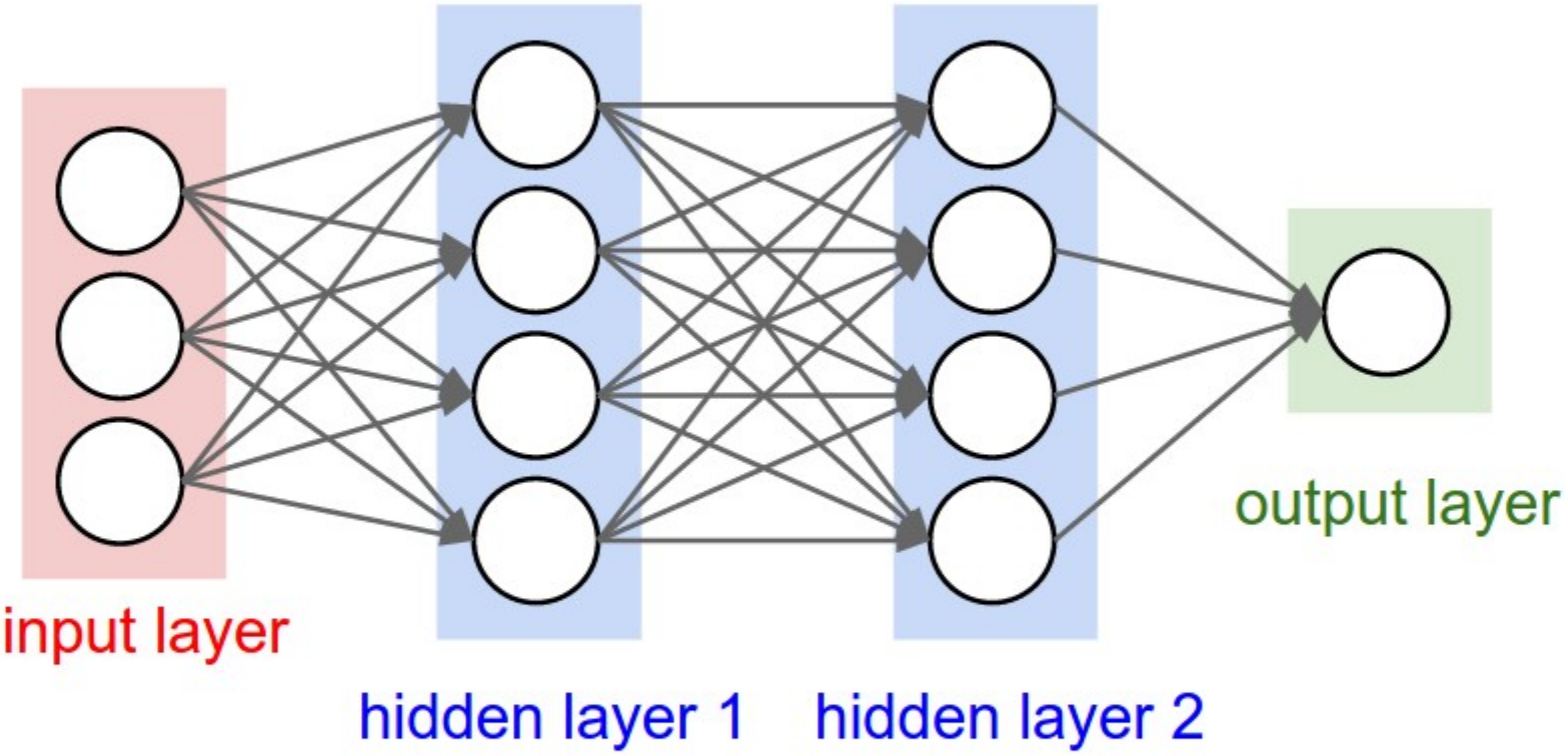
http://www.machineintellegence.com/deep-learning-basics/

Modeling Neuron (1957)



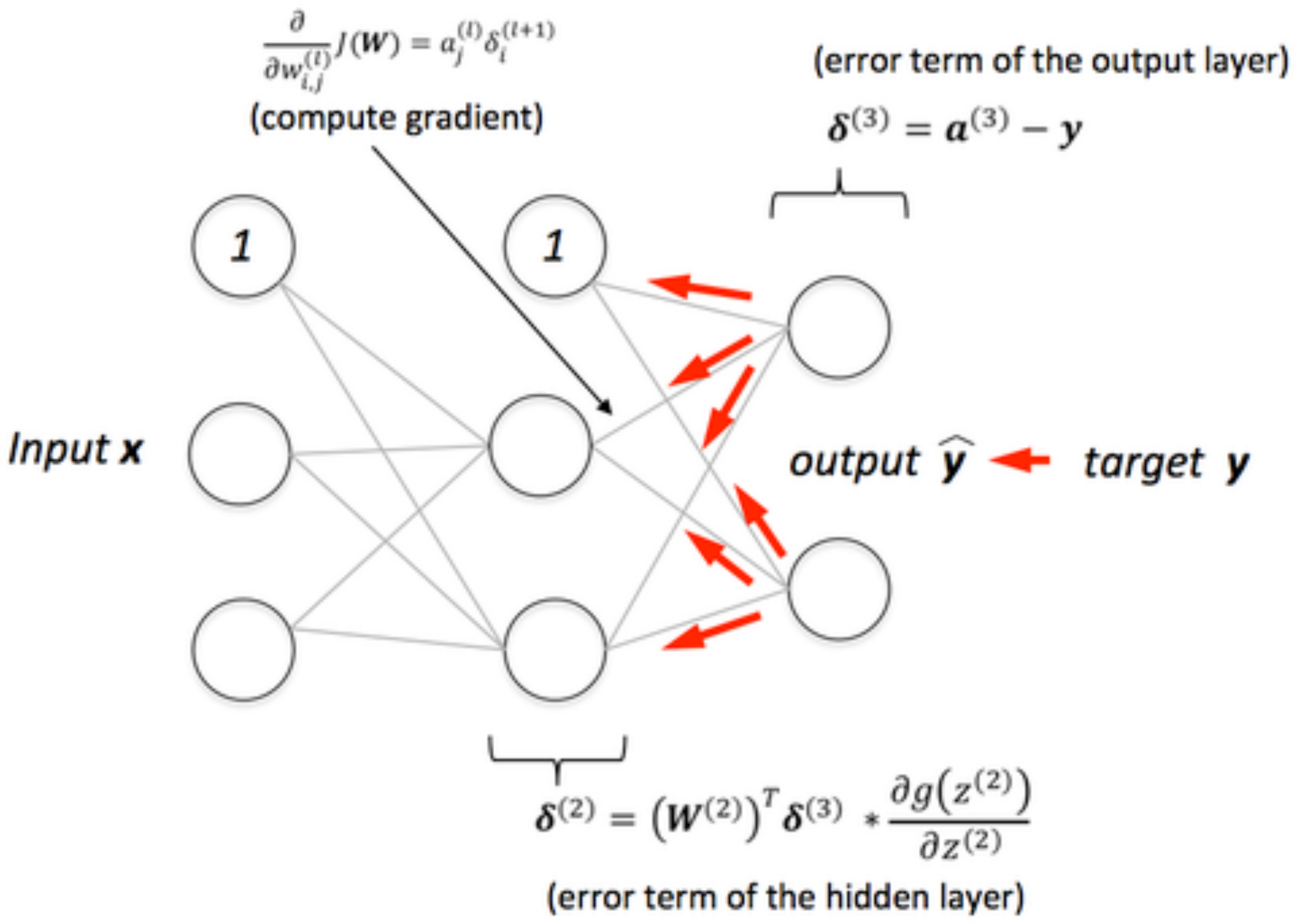
Stanford, cs231n 2017

Multilayer Perceptron (1969)



https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part2.html

Backpropagation (1986)



https://sebastianraschka.com/faq/docs/visual-backpropagation.html

Backpropagation

Backpropagation

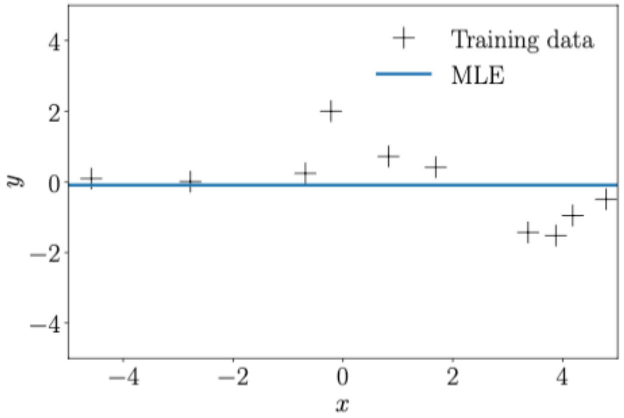
Universal Approximation Theorem

A feed-forward network with single hidden layer is sufficient to represent any function, but the required hidden unit might be infinitely large and may fail to learn.

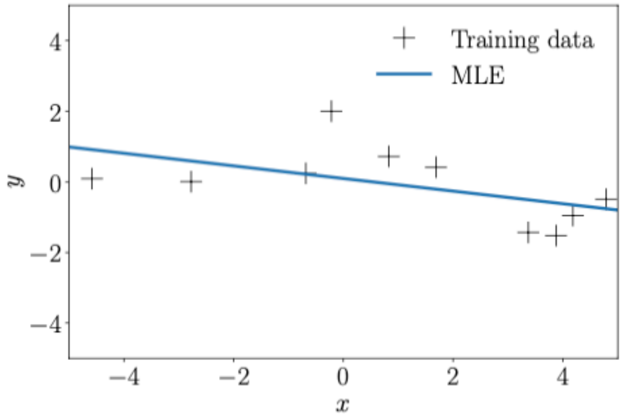
Using deeper model can reduce the number of required units for representing desired function.

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSoyZlgfUiZItscMLUpP7ZiZTXc_KKZqxPOSYFviXwFz2KoztBO

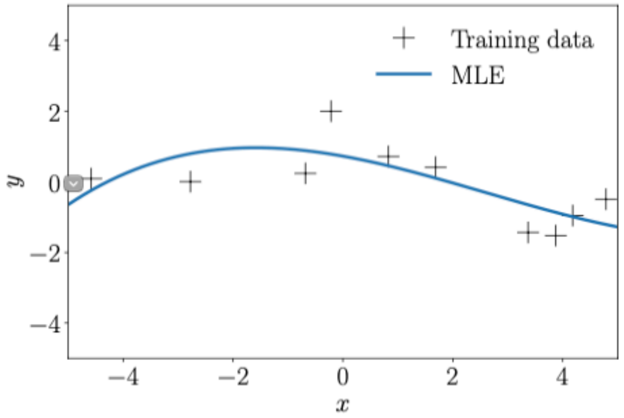
Overfitting – Linear Regression



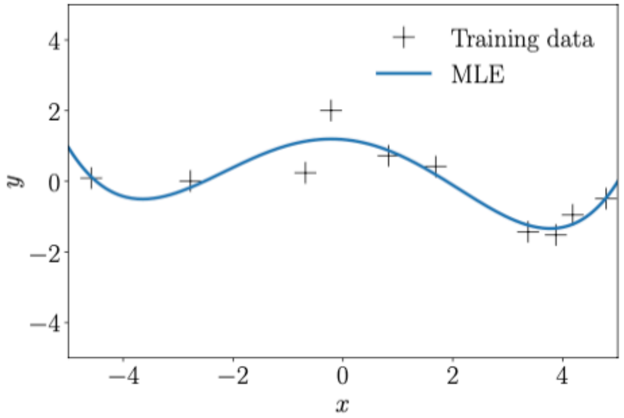
(a) $M = 0$



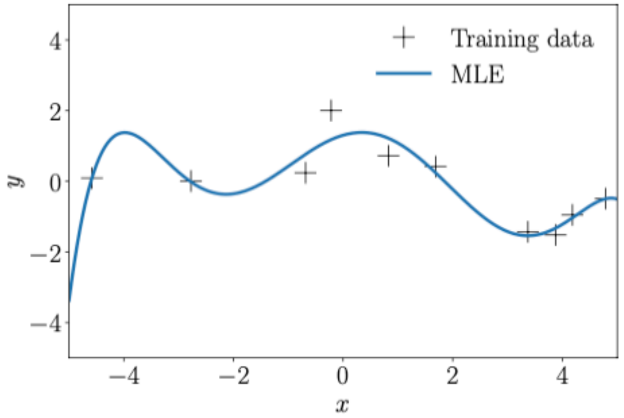
(b) $M = 1$



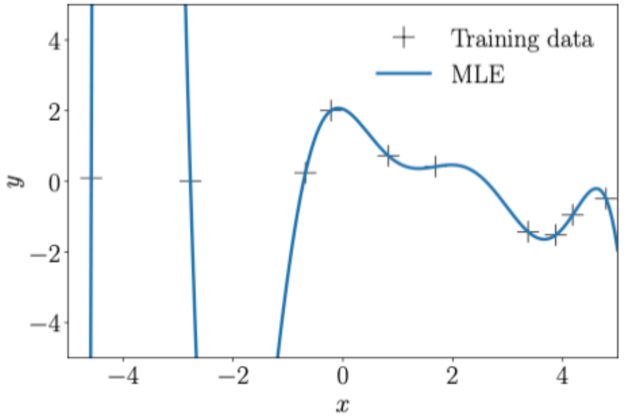
(c) $M = 3$



(d) $M = 4$

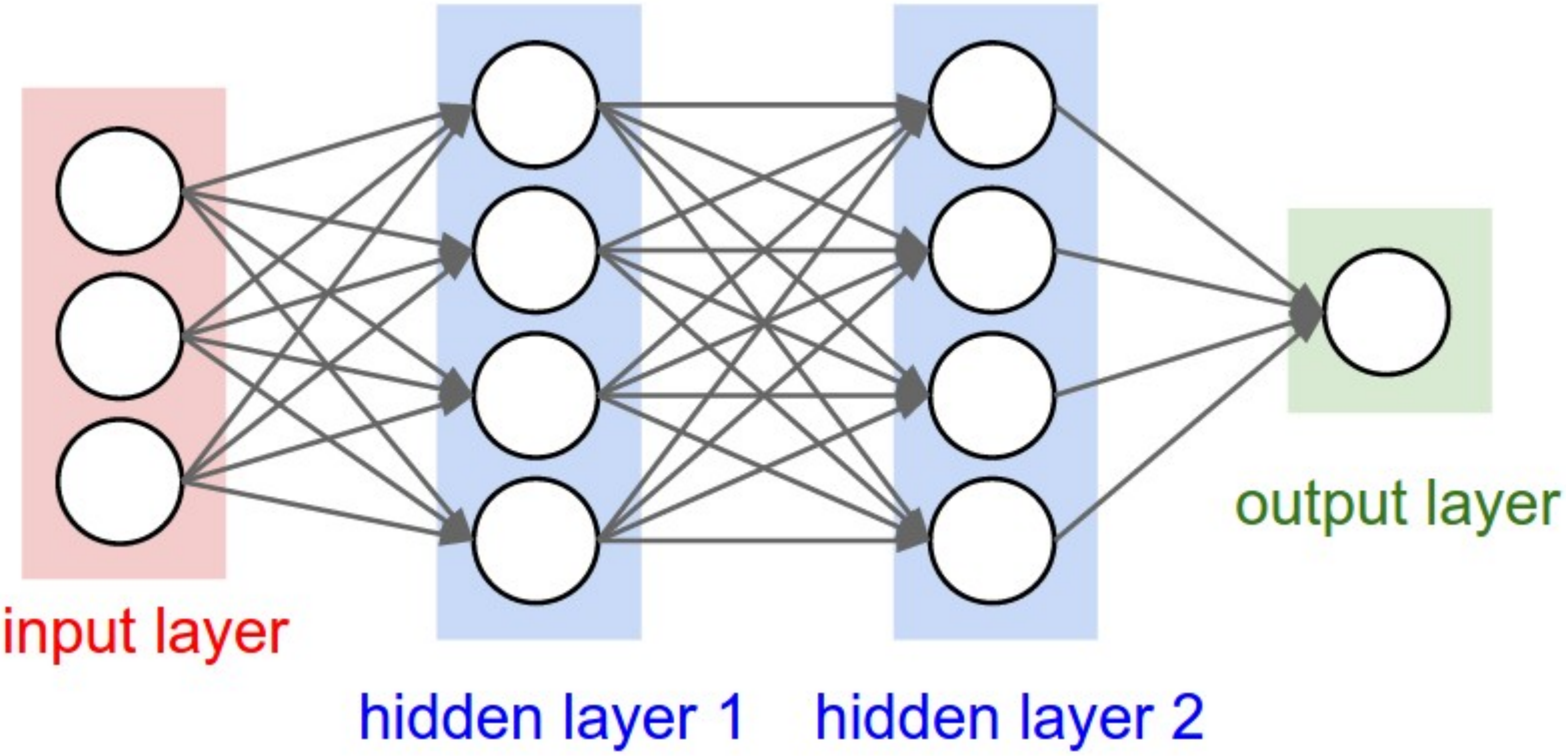


(e) $M = 6$



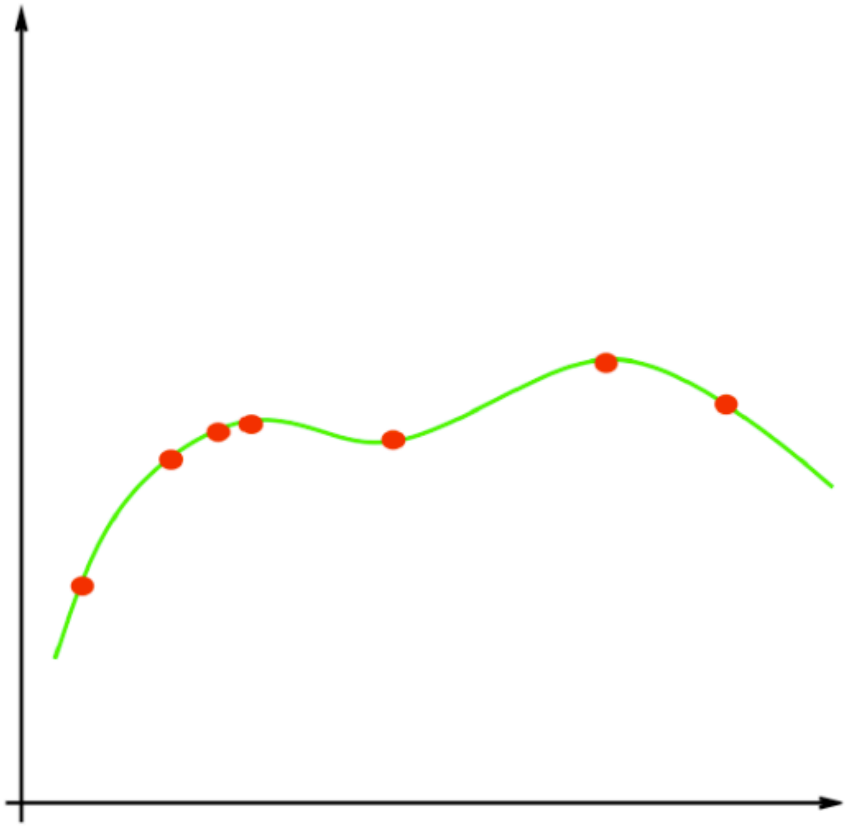
(f) $M = 9$

Overfitting – MLP

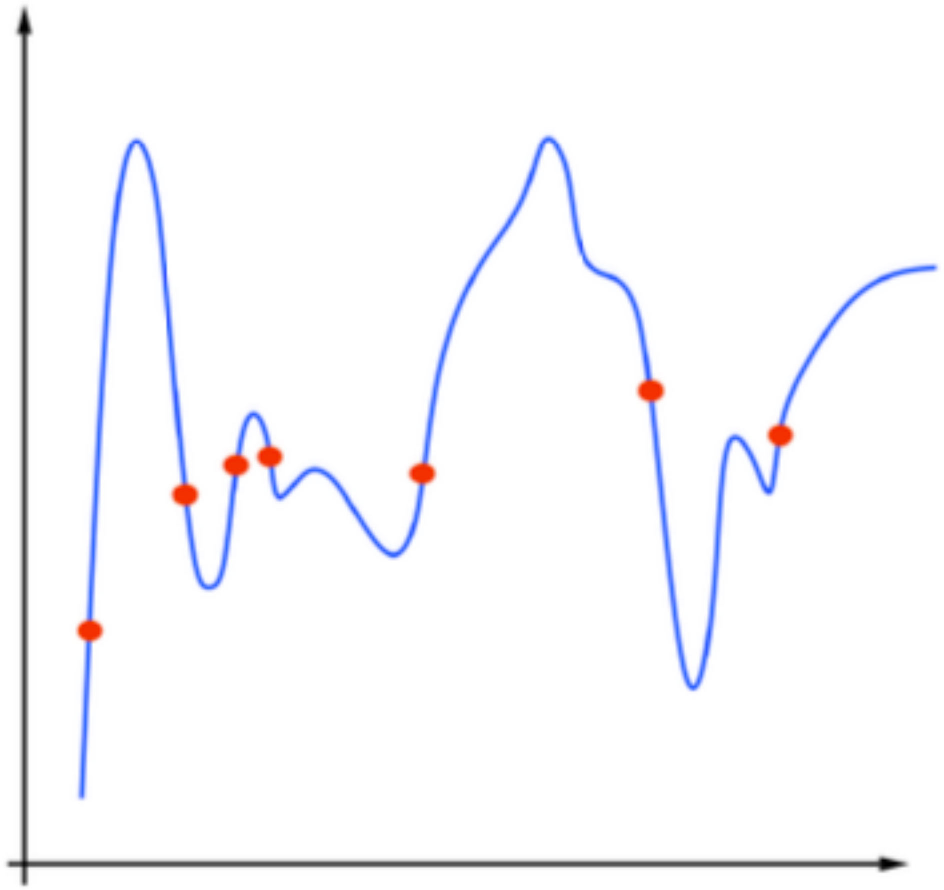


https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part2.html

Overfitting – True Distribution

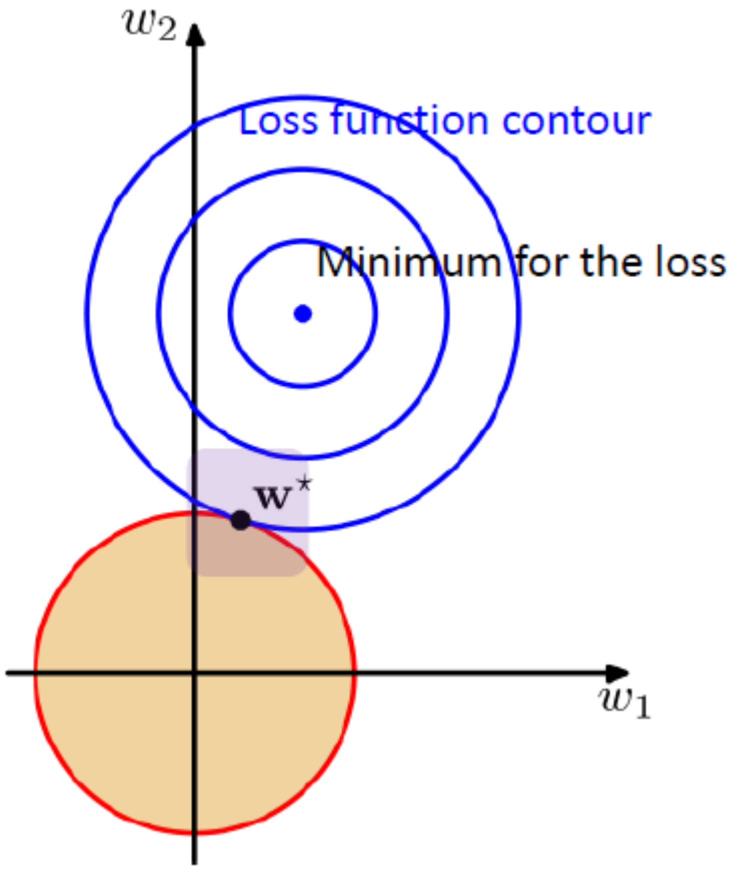


Overfitting – True Distribution



Training, Validation, and Test Set

L2 Regularization



Hyperparameter Tuning