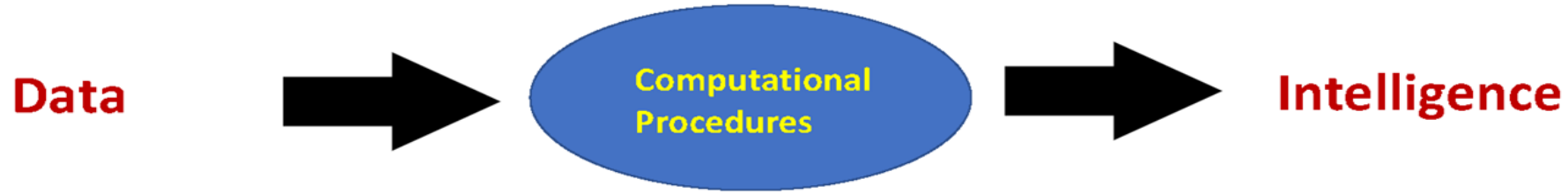
Introduction to the Modern AI/ML/DL

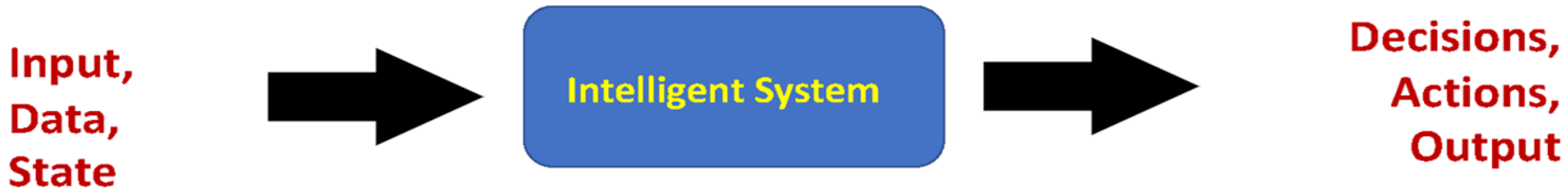
C. V. Jawahar

IIIT Hyderabad

What is Modern “AI” or “ML”?



Data Driven Intelligence



Data Driven Problem Solving

Trivial problem, Simple, well known model. (Y = X/3).
Can find the model with one example.

Complex problem. There could be more parameters (eg. Age, History)

Lot more data needed to solve. Complex biological process

Raju sells samosa. Some one observe the business data.

X = money paid by the customer
Y = no of samosa given in return
(X,Y) = (9,3), (24,8), (15,5)

How many samosa a new customer will get if she gives 30?

Dr. Sheela gives X cups of rice to different patients. They show Y increase in their glucose levels.

(X,Y) = (2,20), (1.5, 27), (1, 8)

For a new patient, if she gives 3 cups of rice, how much glucose will increase?

More Examples

X

Y

In this match , Rajasthan captain Ajinkya Rahane won the toss and decided to bowl first .

इस मैच में राजस्थान के कप्तान अजिंक्य रहाणे ने टॉस जीतकर पहले गेंदबाजी का फैसला किया ।

General Strategy: Given many examples of (X,Y), learn an automated solution to predict Y given a new X.

"Black and White Dog Jumps over Bars"

ILL Posed Problems: Why do they work?

X **Y**

Can Human(experts) do this?

How do they do?

Title: Biscuits

Ingredients:
Flour, butter, sugar, egg, milk, salt.

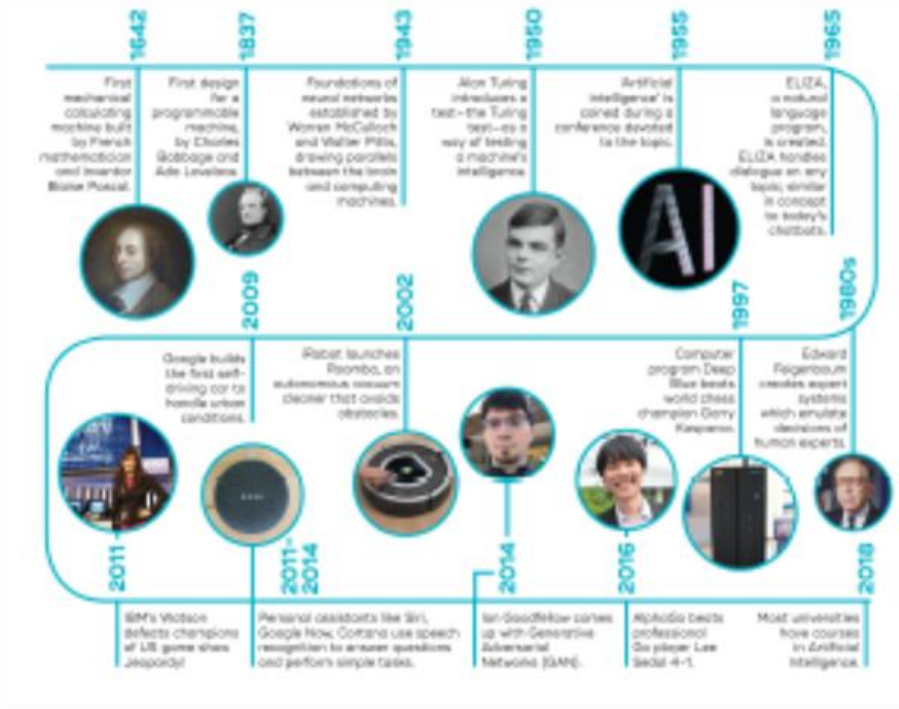
Instructions:
- Preheat oven to 450 degrees.
- Cream butter and sugar.
- Add egg and milk.
- Sift flour and salt together.
- Add to creamed mixture.
- Roll out on floured board to 1/4 inch thickness.
- Cut with biscuit cutter.
- Place on ungreased cookie sheet.
- Bake for 10 minutes.

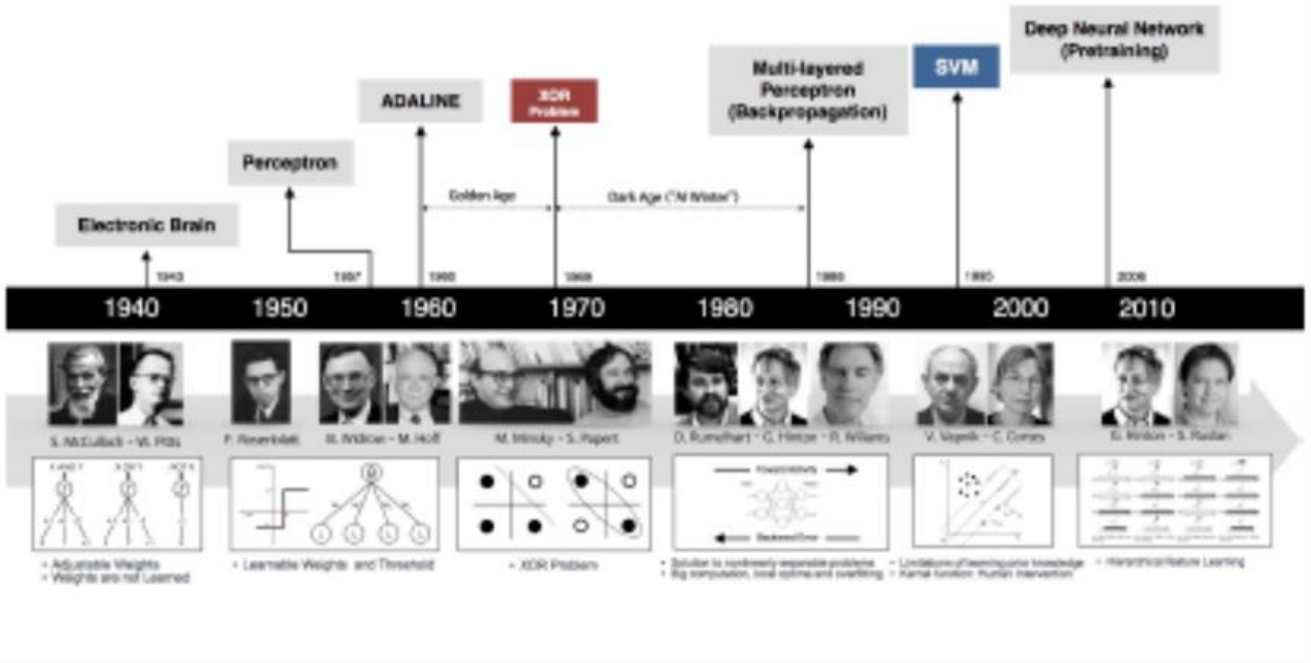
Extensive use of Prior Knowledge.

Composition of parts seen in the past.

Inverse Cooking (CVPR 2019)

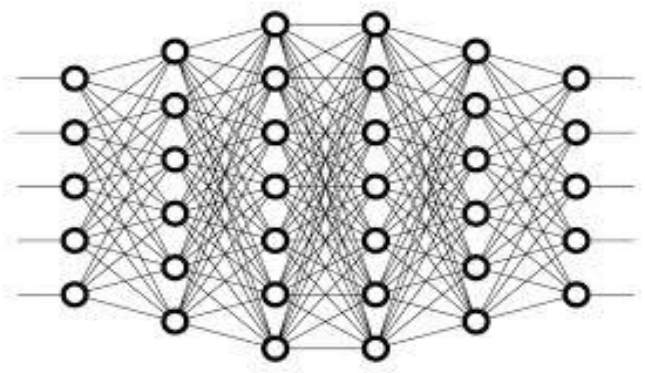
Is AI, ML, DL really New?





Why AI started to work now?

Data; Internet; Connectivity

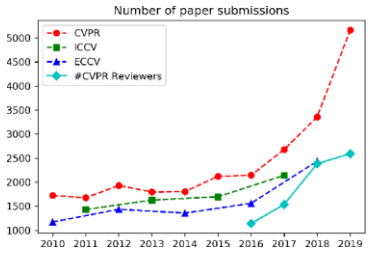


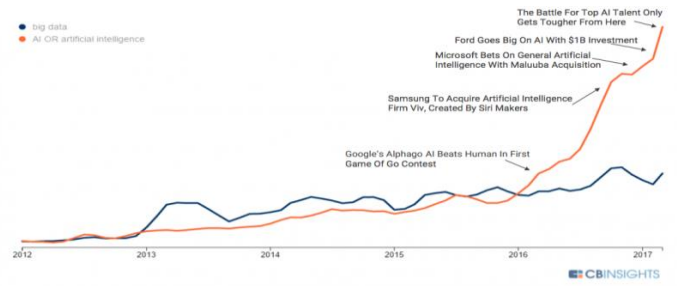
Algorithms, Deep Learning

Compute, Cloud, APIs, Libraries

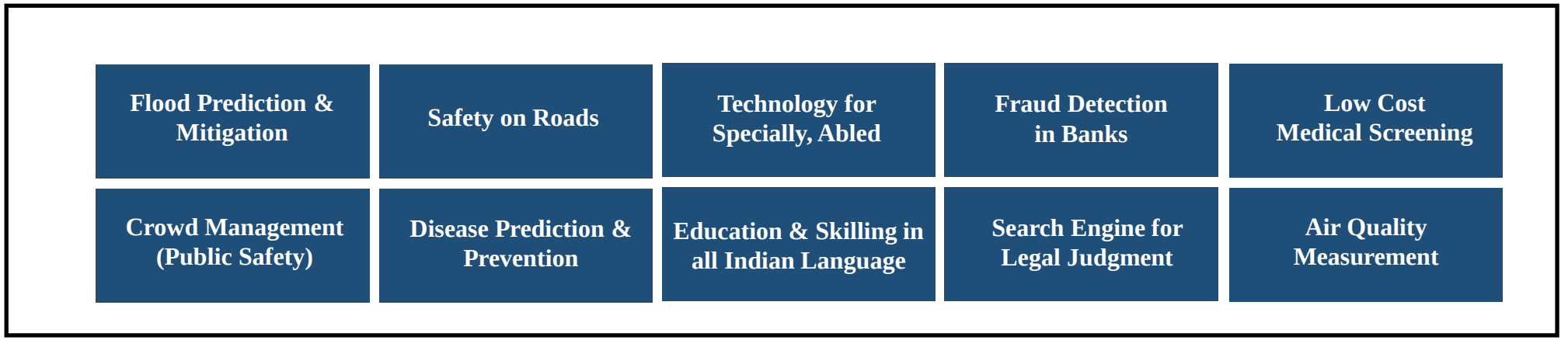
Caffe2 · Microsoft CNTK · TensorFlow · theano

Chainer · PYTORCH · dmlc mxnet · K

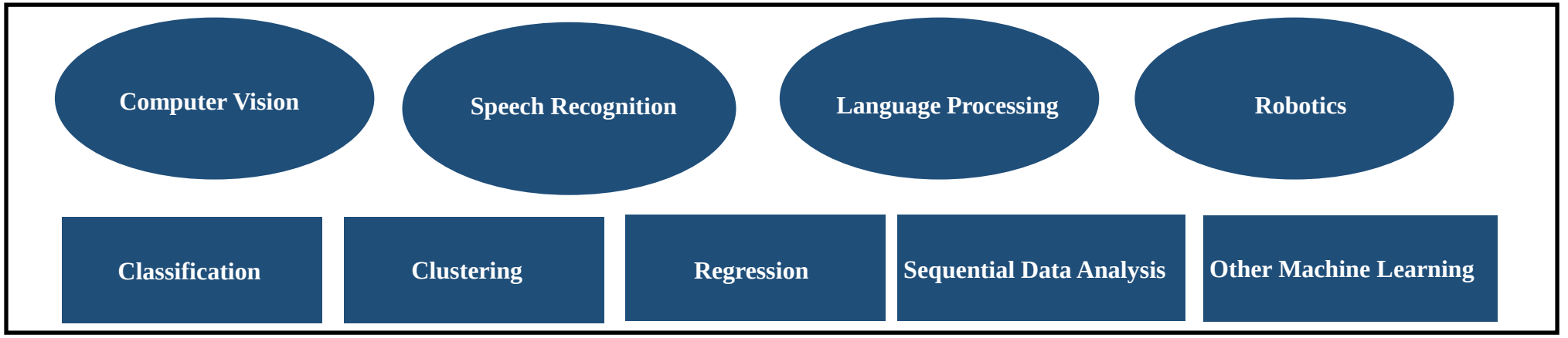




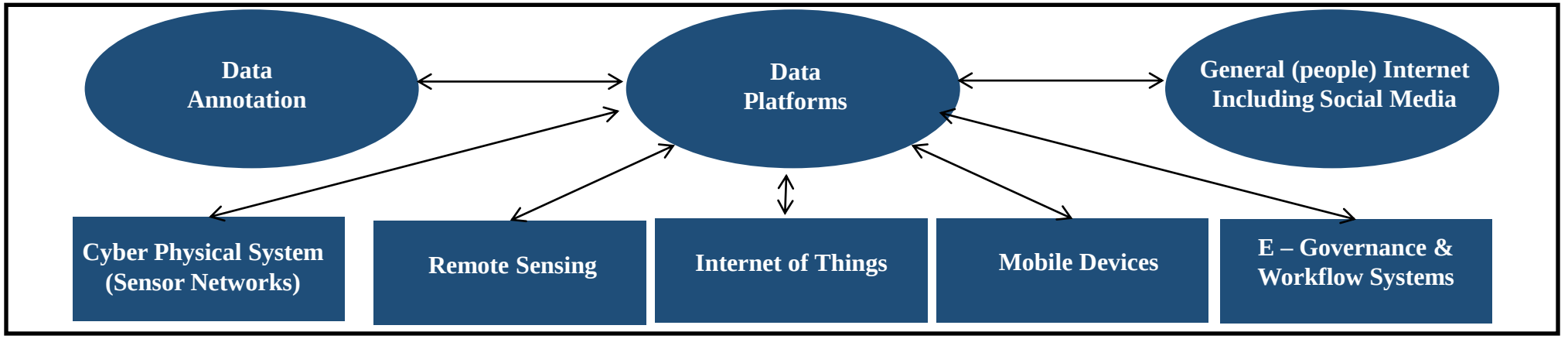
More People, Papers, Results, Funding, People. Positive Feedback.



Domains, Applications and Problems



Algorithms and Areas

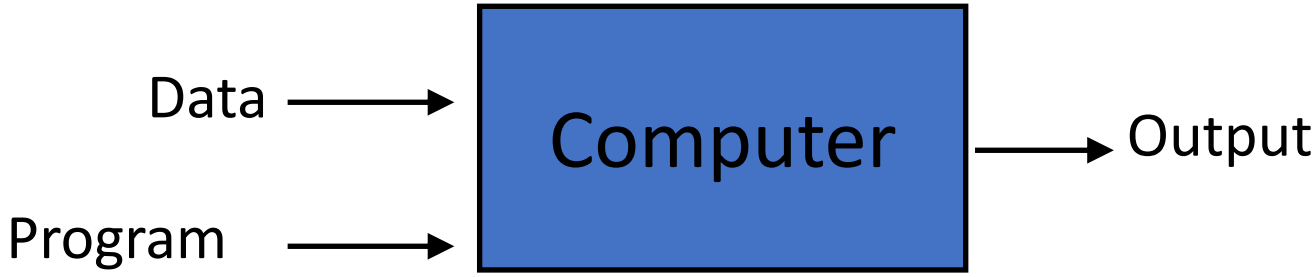


Data and Technology

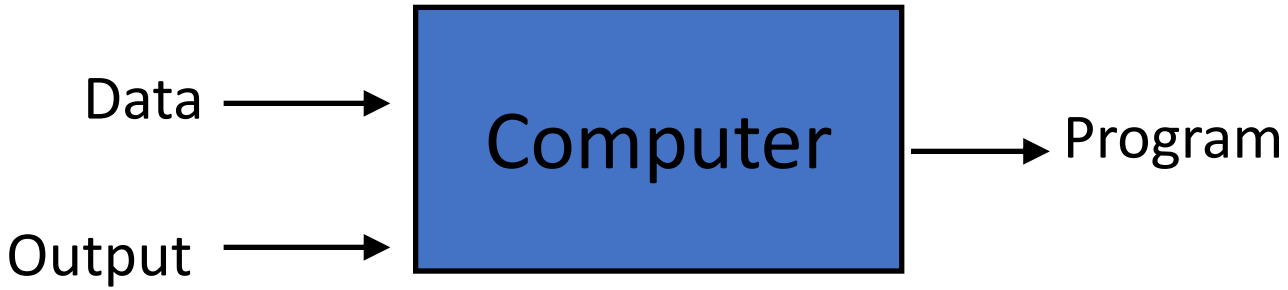
What is Machine Learning?

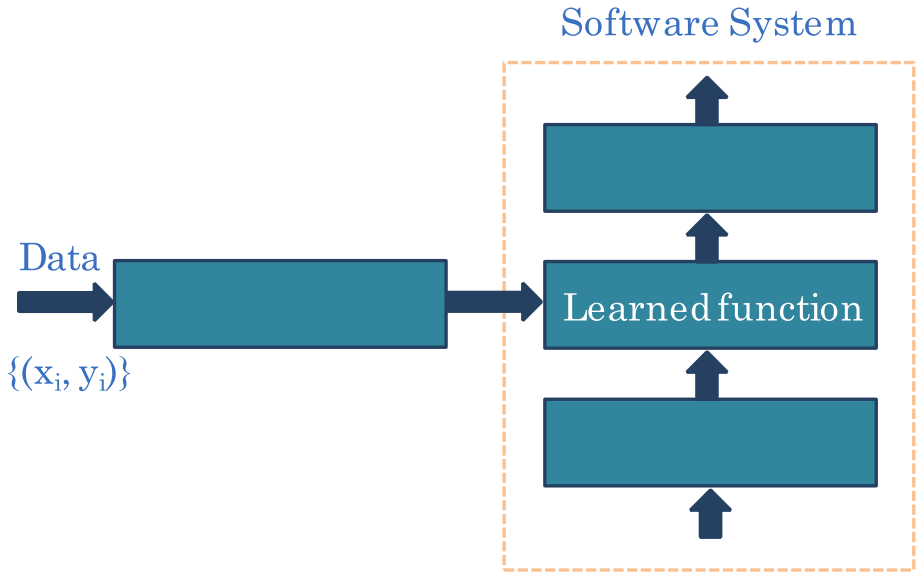
Data Driven Solutions

Traditional Programming



Machine Learning





The machine learning framework

- Apply a prediction function to a feature representation of the “sample” to get the desired output:

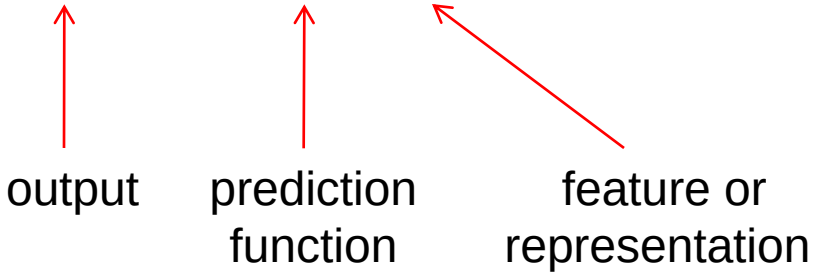
f() = “apple”

f() = “tomato”

f() = “cow”

The machine learning framework

$$y = f(\mathbf{x})$$



- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error.
- **Testing:** apply f to a never-before-seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

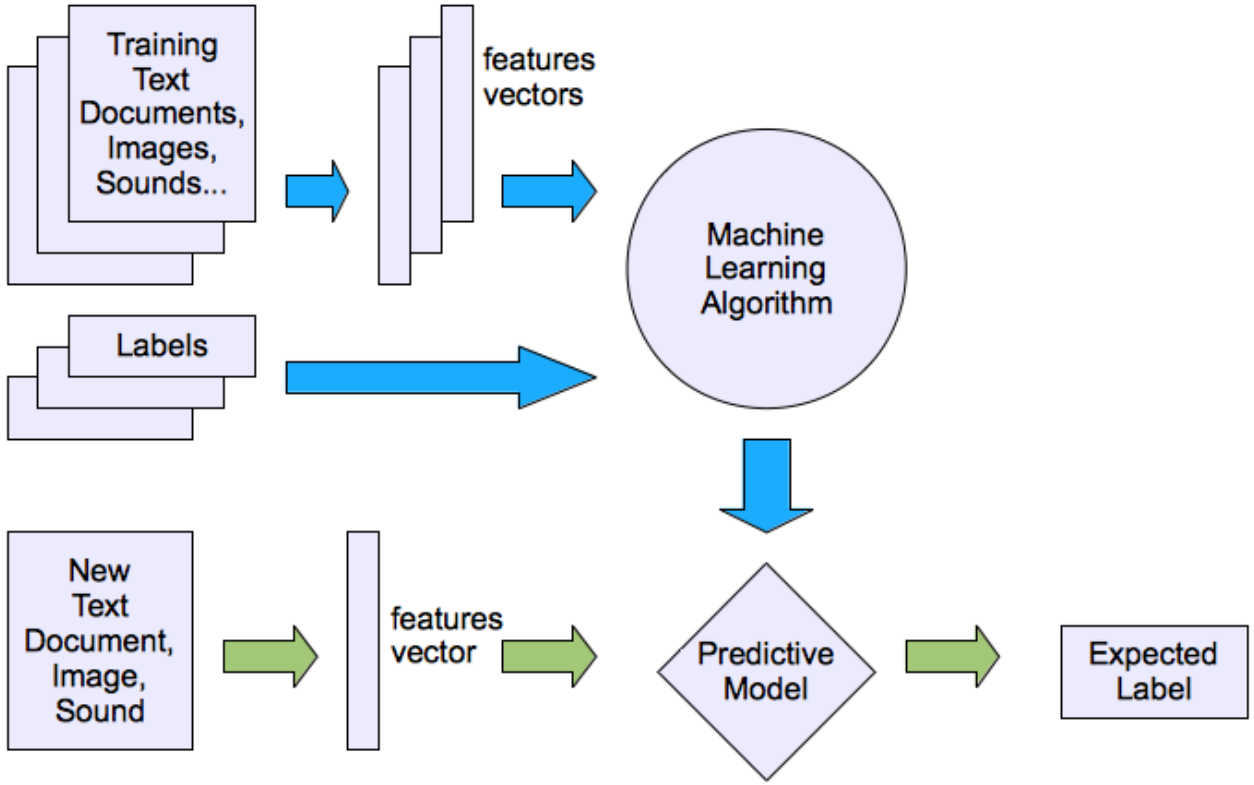
Slide credit: L. Lazebnik

Supervised Learning

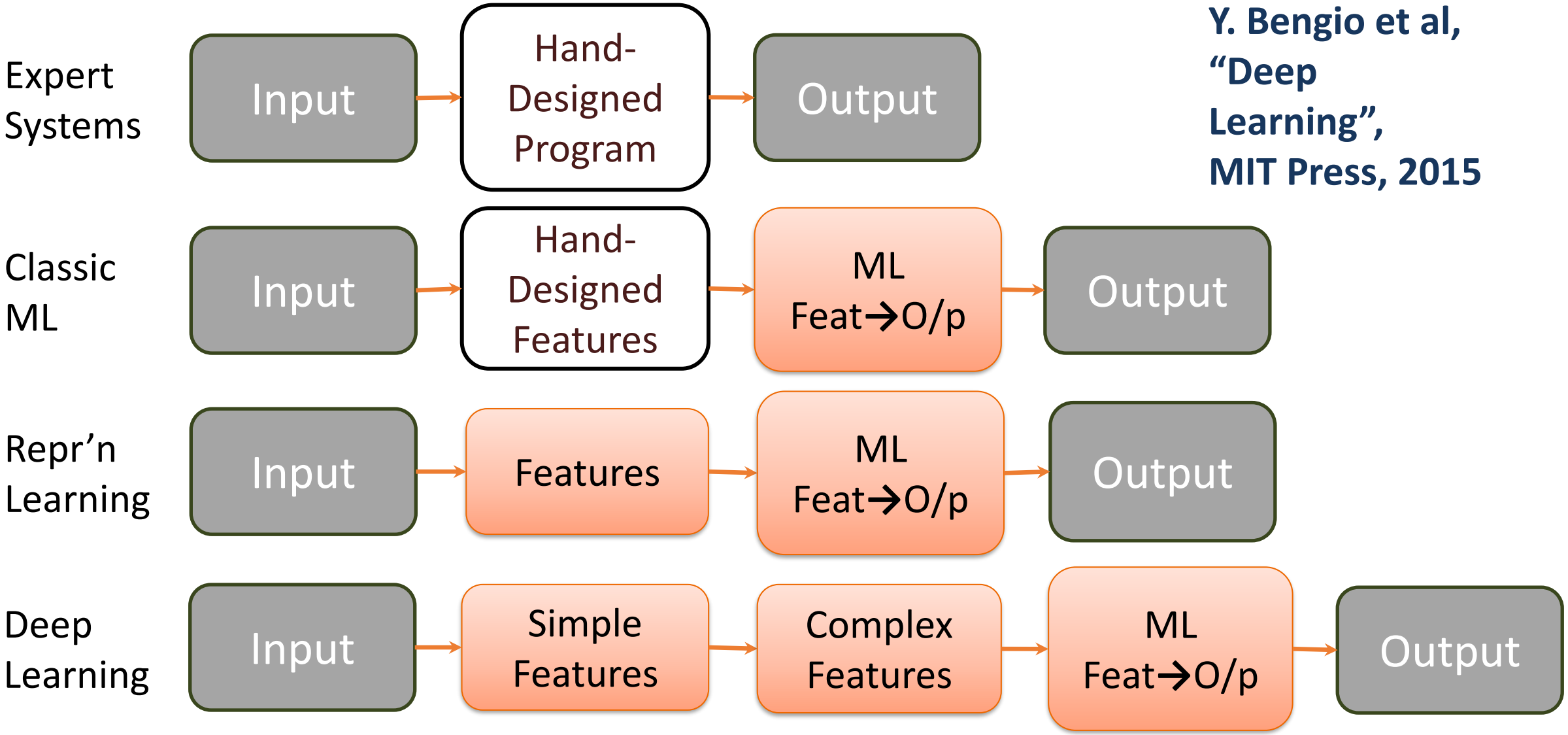
- Input: x (images, text, emails…)
- Output: y (spam or non-spam…)
- (Unknown) Target Function
  - f: X → Y (the "true" mapping / reality)
- Data
  - (x_1,y_1), (x_2,y_2), …, (x_N,y_N)
- Model / Hypothesis Class
  - g: X → Y
  - $y = g(x) = \text{sign}(w^T x)$
- Learning = Search in hypothesis space
  - Find best g in model class.

Machine learning structure

- Supervised learning

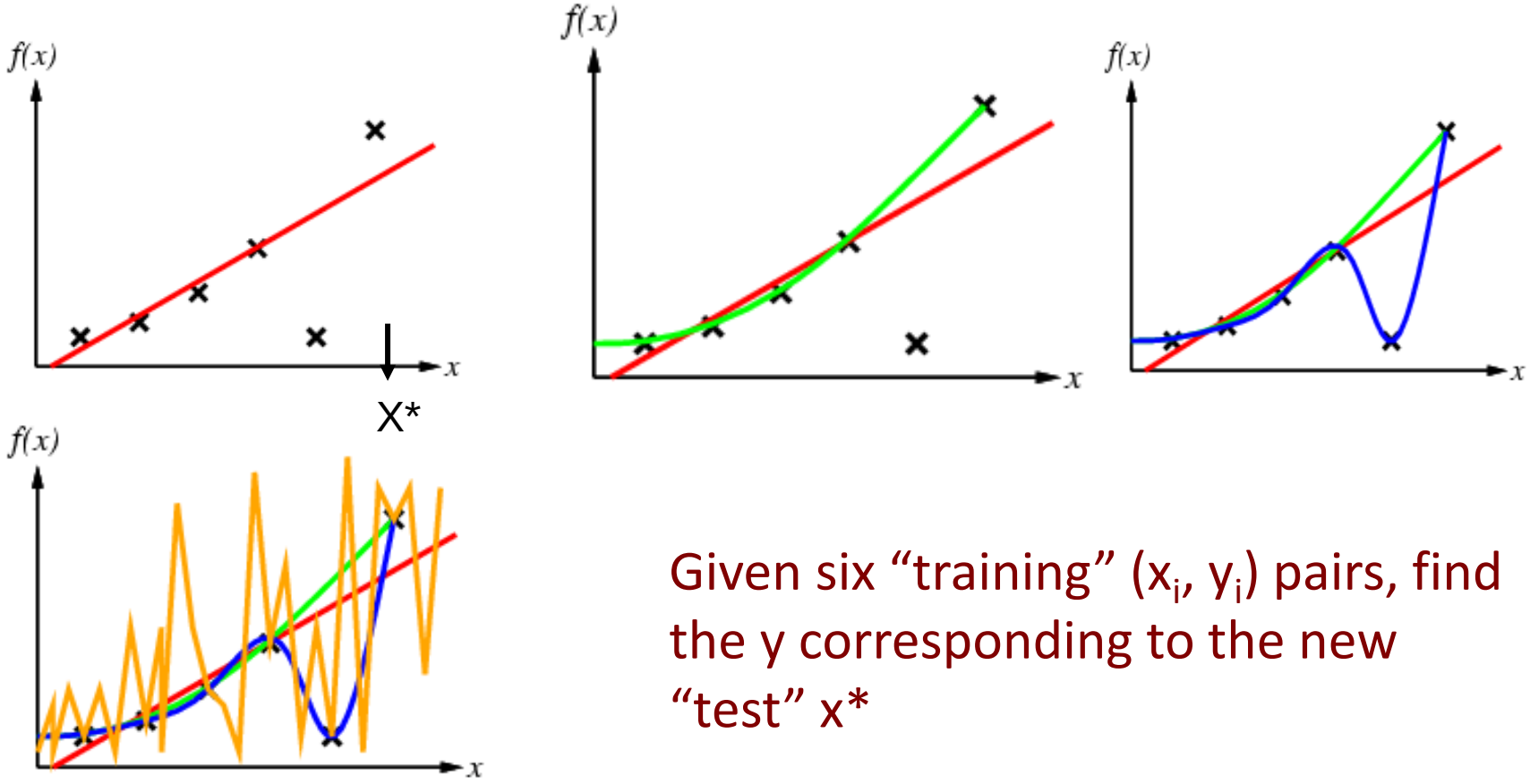


Evolution of Learning



Y. Bengio et al, "Deep Learning", MIT Press, 2015

A Simple Fitting/Predicting Problem



Given six "training" (x_i, y_i) pairs, find the y corresponding to the new "test" x*

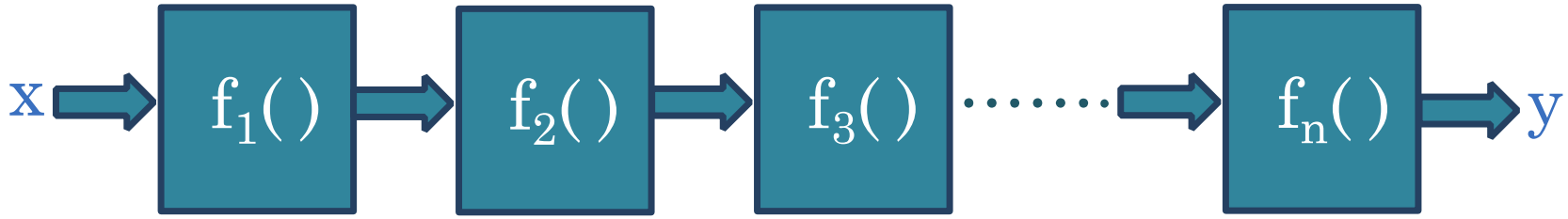
Which curve is the best?

Occam’s Razor

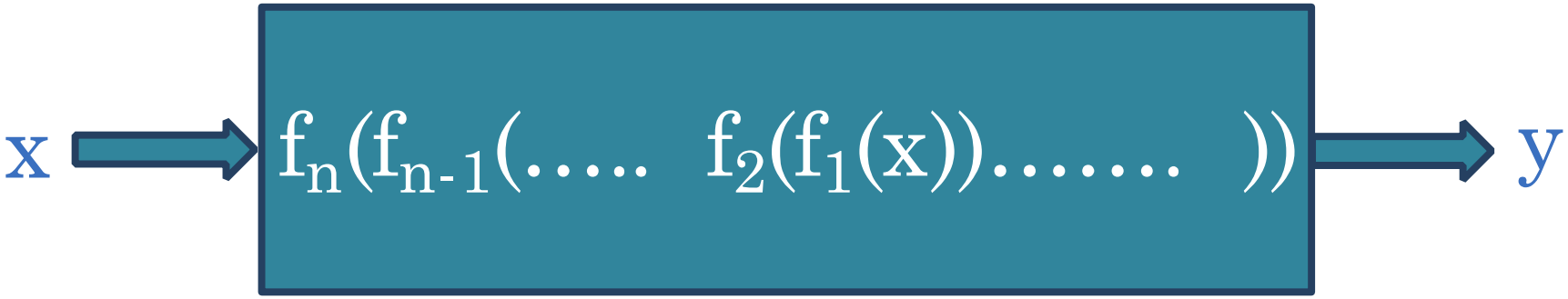
Select the simplest hypothesis (solution) that suits the data.

Eg.Minimize Sum of “fit error” and “degree of the polynomial”

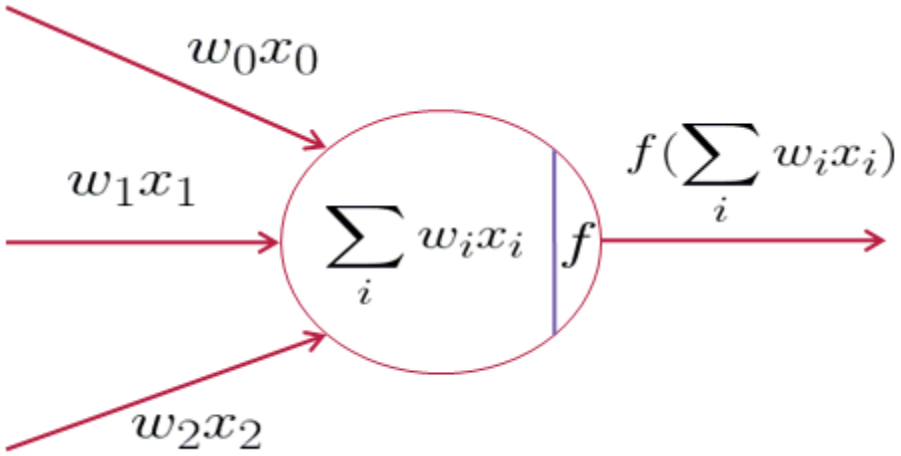
What is Deep Learning?

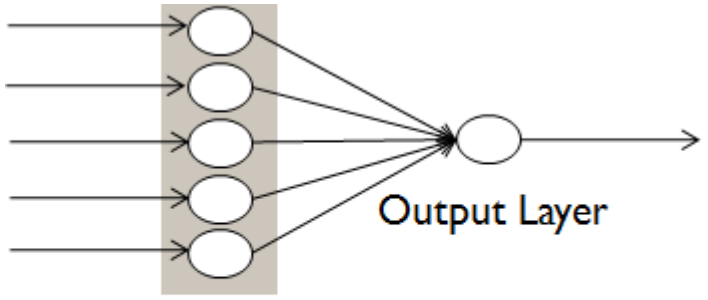


Each function could be simple.

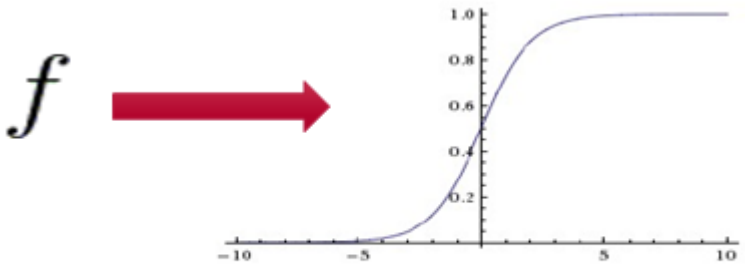


Neuron, Perceptron and MLP

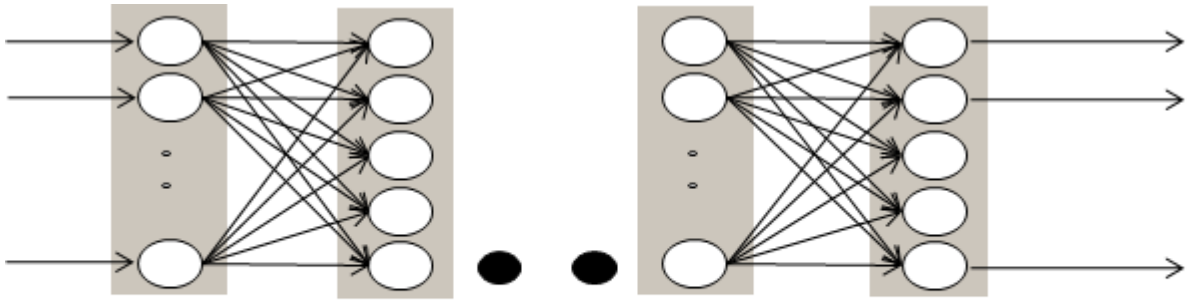




Perceptron

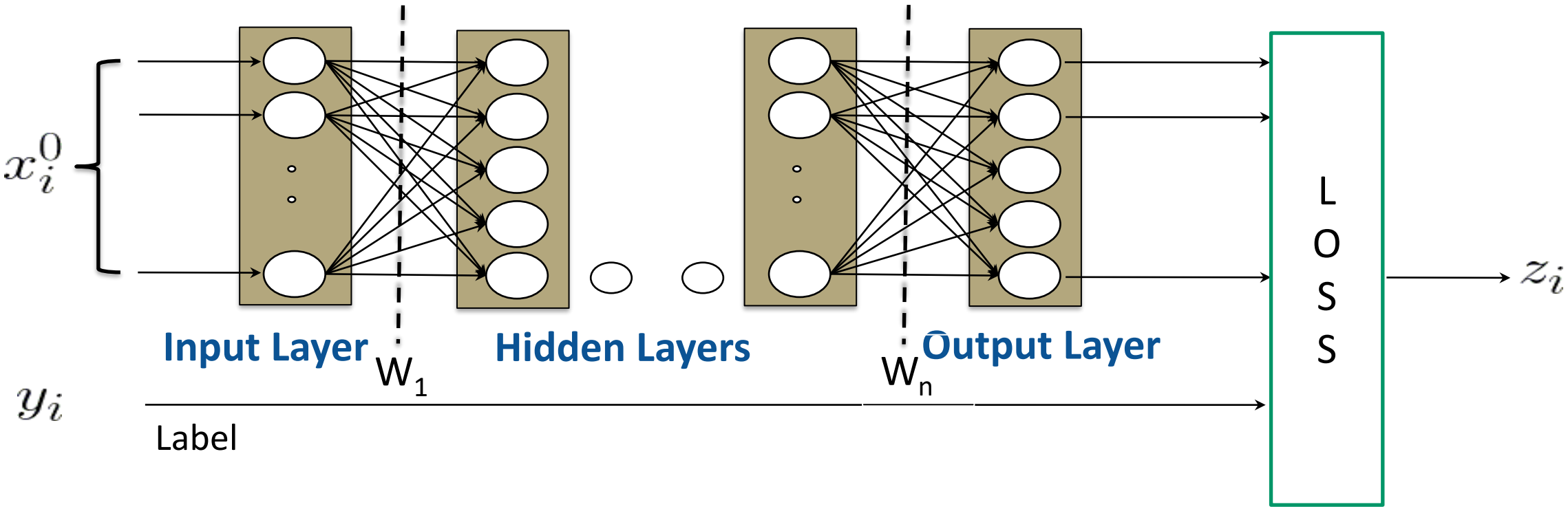


E.g. Sigmoid Activation Function



Multi Layer Perceptron

Loss or Objective



Objective: Find out the best parameters which will minimize the loss.

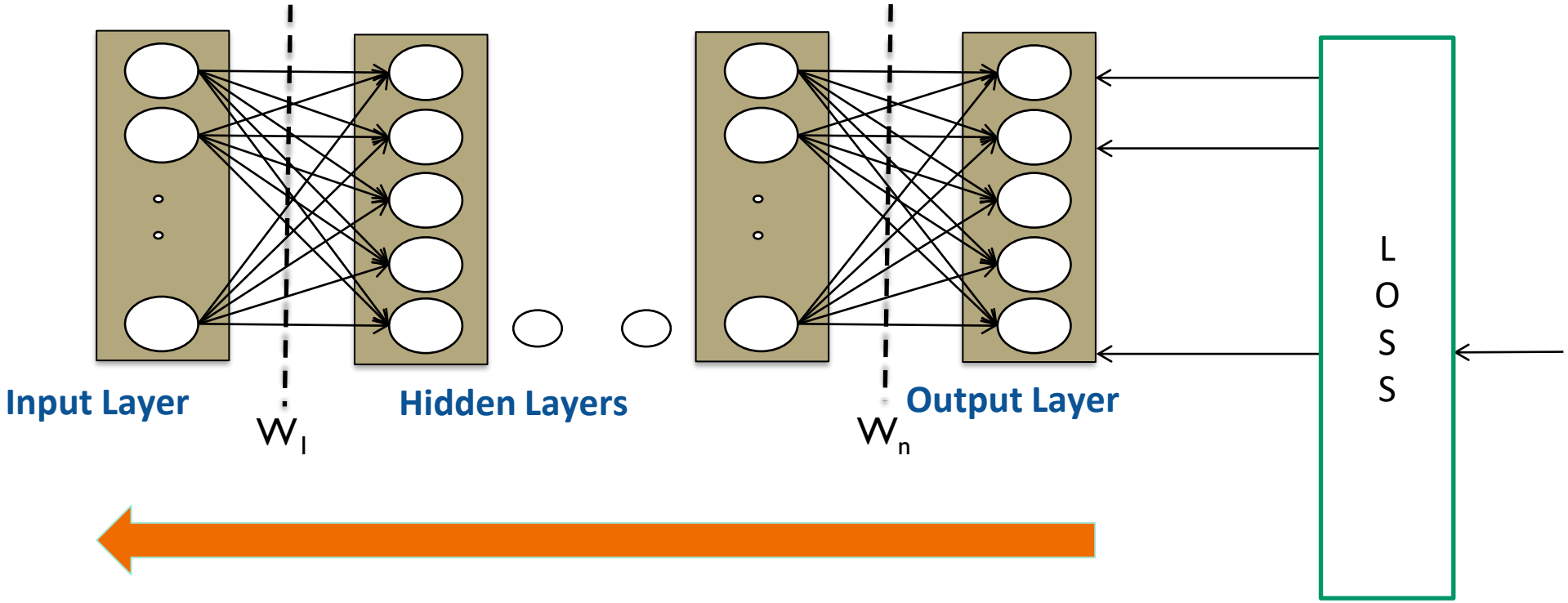
$$W^* = arg\min_W \sum_{i=1}^{N} L(x_i^n, y_i; W)$$

Weight vector

$$z_i = \frac{1}{2} \parallel x_i^n - y_i \parallel_2^2$$

E.g. Squared Loss

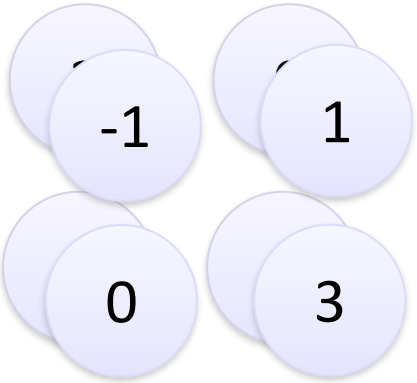
Back Propagation



Solution: Iteratively update W along the direction where loss decreases.

Each layer's weights are updated based on the derivative of its output w.r.t. input and weights

CNN



A new image

Convolution

Max Pooling

Convolution

Max Pooling

Can repeat many times

Smaller than the original image

The number of channels is the number of filters

The whole CNN

cat dog ……

Convolution

Max Pooling

A new image

Convolution

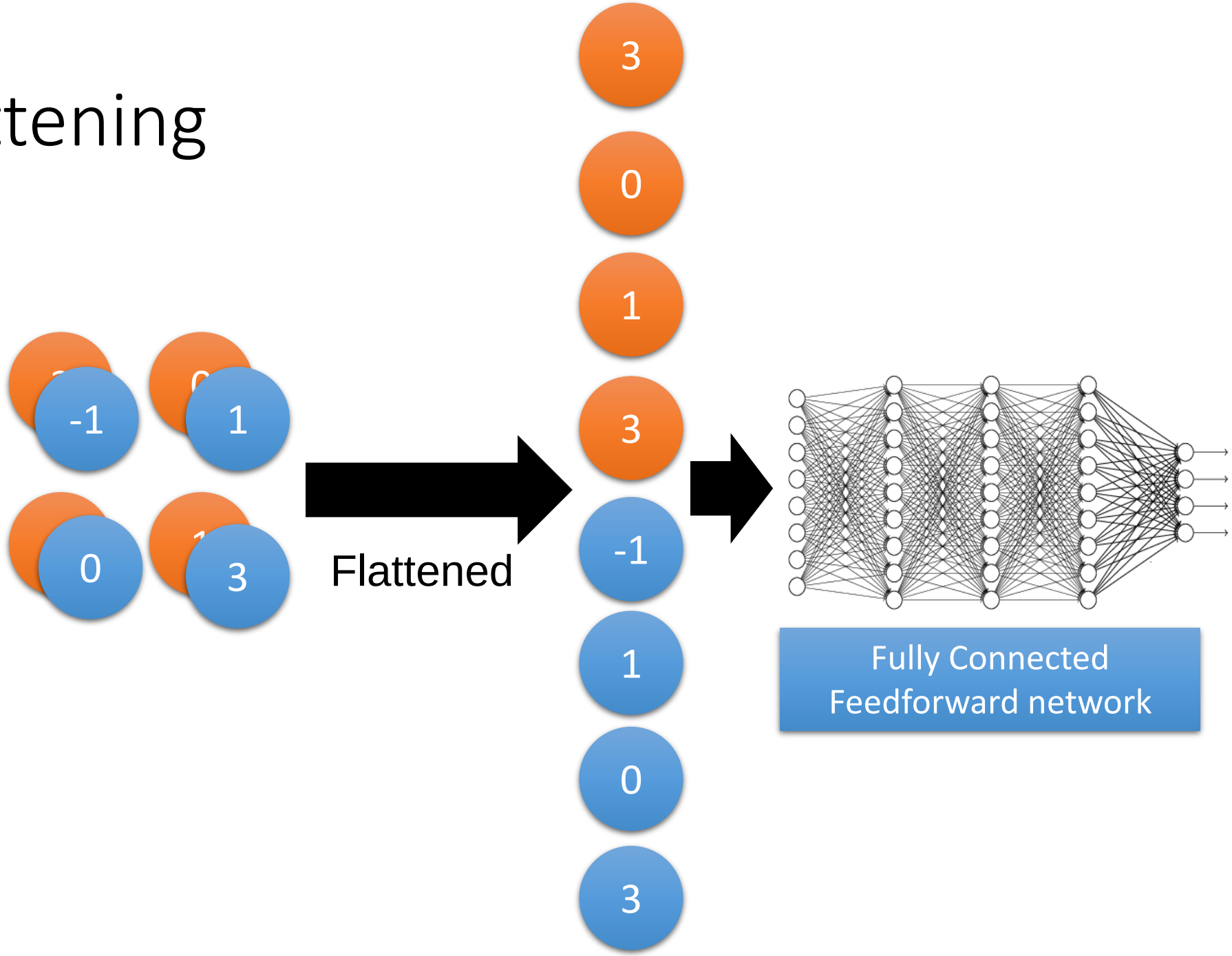
Max Pooling

A new image

Flattened

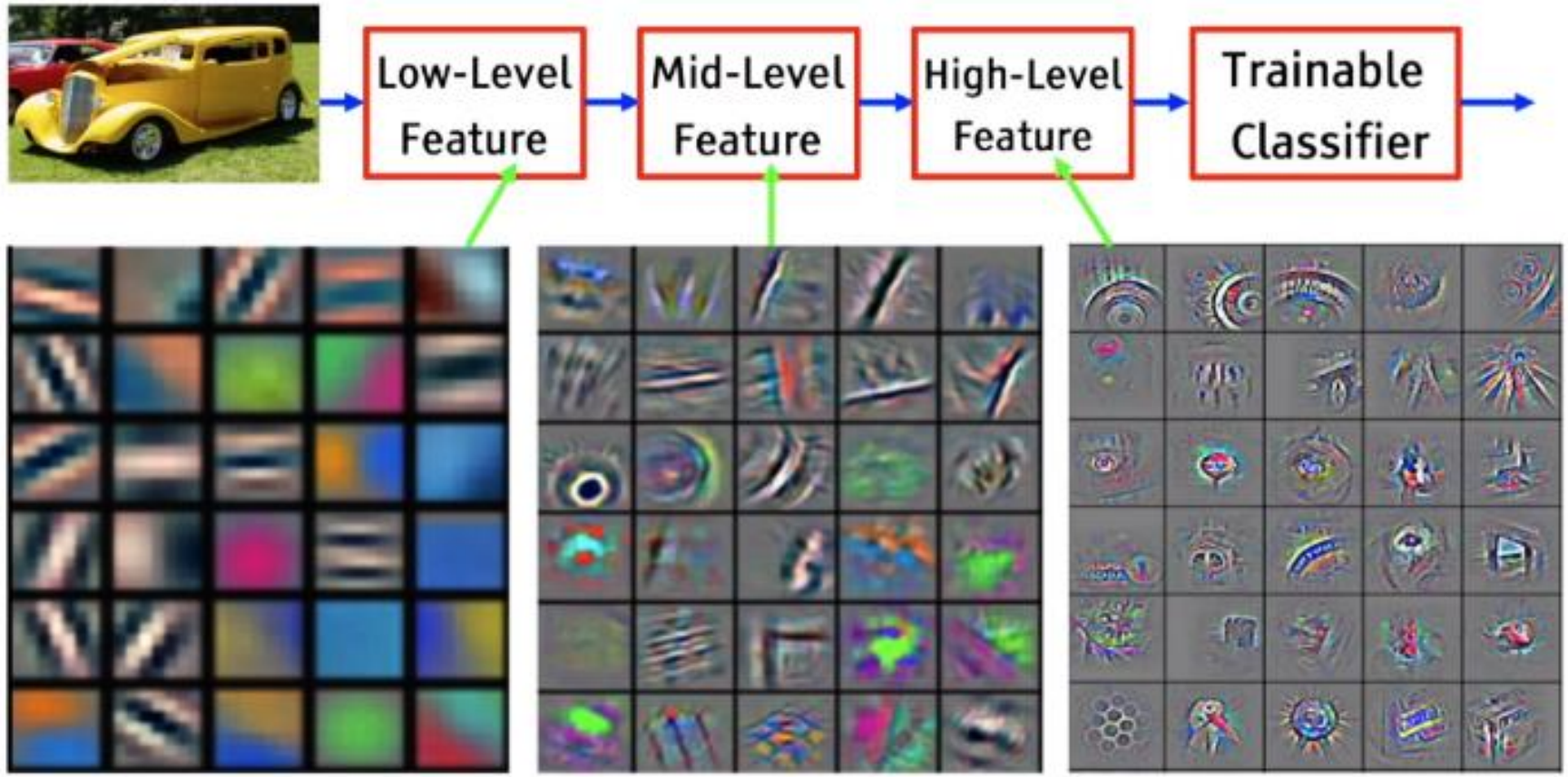
Fully Connected Feedforward network

Flattening

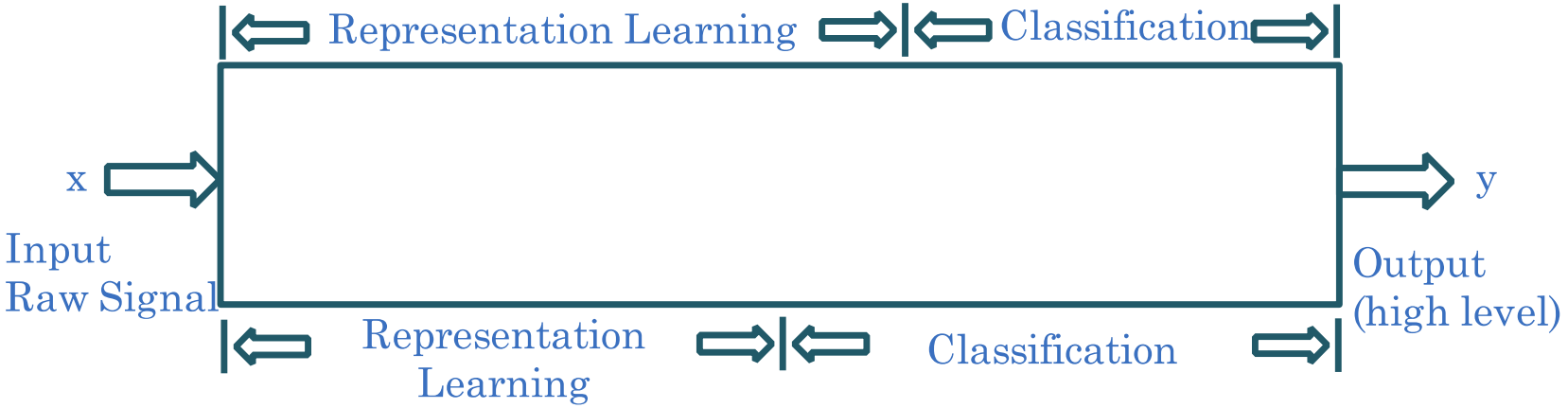

-1
1
0
3
Flattened
3
0
1
3
-1
1
0
3
Fully Connected
Feedforward network

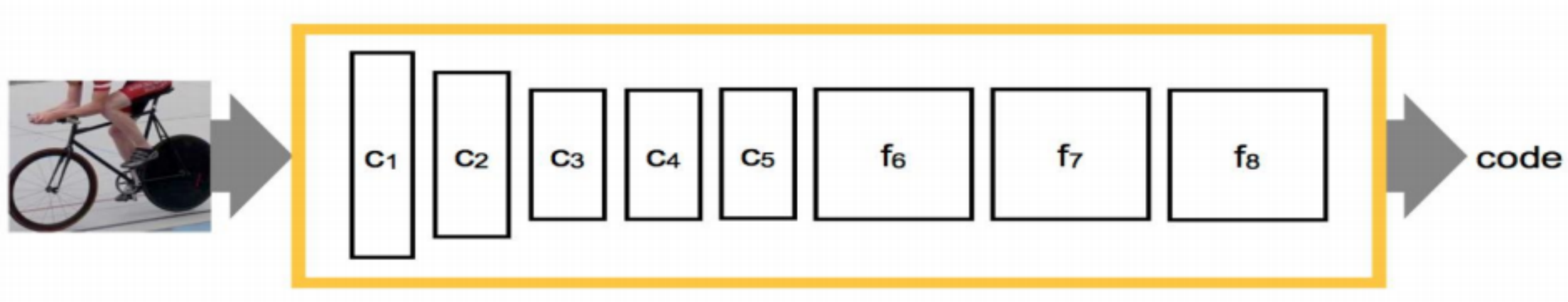
Deep Learnt Features

- It's deep if it has more than one stage of non-linear feature trans



Re-Usability





Quick Tour: Terminologies

- Training, Testing/Inference, Validation, Overfitting, Generalization
- Optimization, Regularization, Backpropagation, Initialization
- MLP, CNN, RNN, LSTM, GRU, Transformer
- Fully Connected, Convolutional, Attention, Normalization,
- Weight, Parameters, Sigmoid, Relu, Activation,
- Trained Model, Fine Tune, Deploy, Edge Implementation
- PyTorch, TensorFlow, Python, GPU, Docker..
- Supervised, Self Supervised, Reinforcement, .. Learning
- SVM, Random Forest, MLP,..

Thank you!

Questions?