Traitement d'images

Apprentissage profond

Andrea Cherubini

SOURCES: Cours de Nicolas Thome au CNAM

http://cedric.cnam.fr/vertigo/Cours/ml2/

Context: Big data

Superabundance of data: images, videos, audio, text, etc

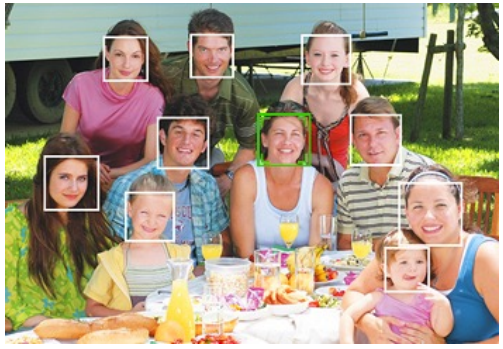
BBC: millions of videos

Facebook: 350M images/day

250M monitoring cameras

Obvious need to access, search, or classify these data

Huge number of applications: visual search, sms, medical imaging, robotics etc

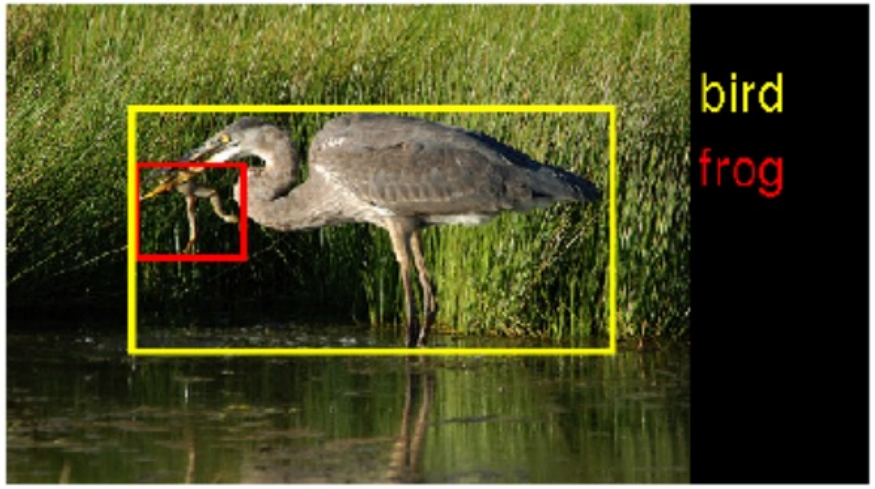


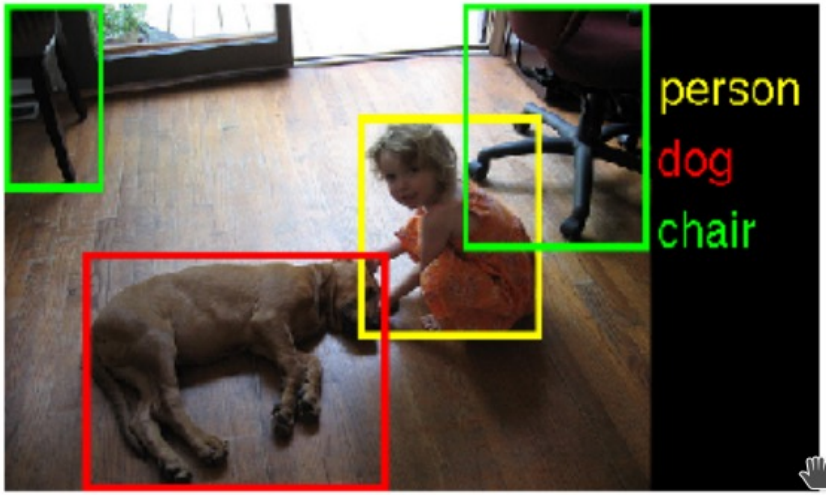
Leading track in Machine Learning / Computer Vision conferences in the last decade

Classification and Recognition

Classification : assign a given data to a given set of pre-defined classes
Recognition (more general than classification). Examples in various fields:

- Object Localization in images
- Ranking for document indexing
- Sequence prediction for text, speech, audio, etc

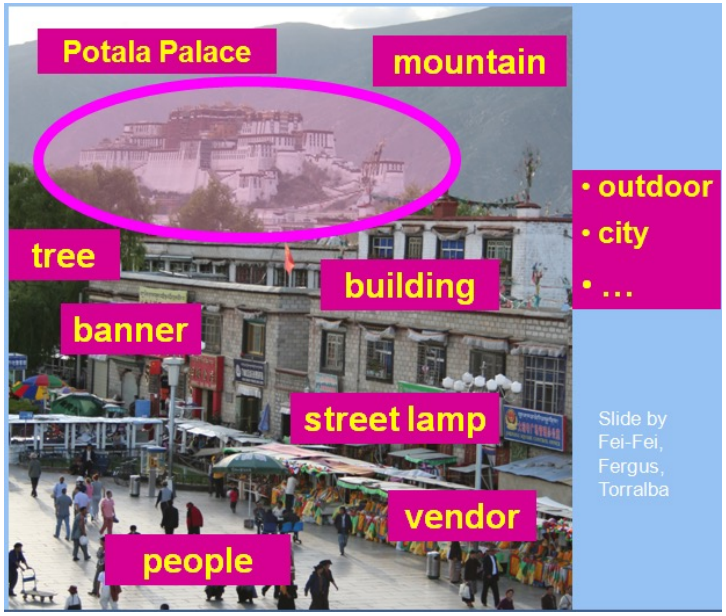




Specifically: Visual recognition

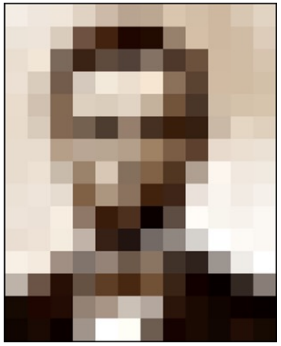
Certainly the most impacted topic by deep learning:

- Scene categorization
- Object localization
- Context & Attribute recognition
- 3D layout, depth estimation
- Rich description of scene, e.g. sentences



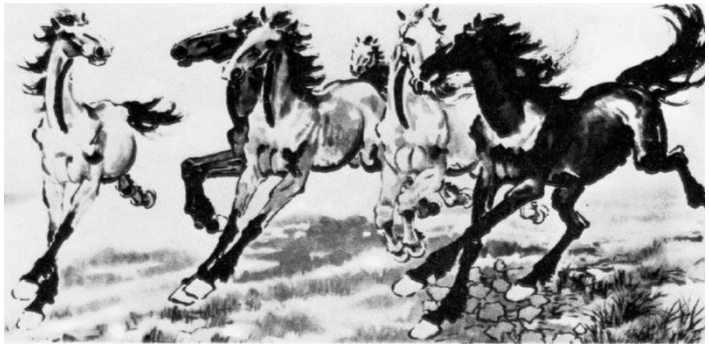
Challenge: “filling” the semantic gap

What we perceives VS what a computer sees



243	239	240	225	206	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	108	208	208	215
235	217	115	212	243	236	247	139	91	209	208	211
233	208	131	222	219	226	196	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
232	232	182	186	184	179	159	123	93	232	235	235
232	236	201	154	216	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	59	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	149	109	138	65	47	156	239	255
190	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

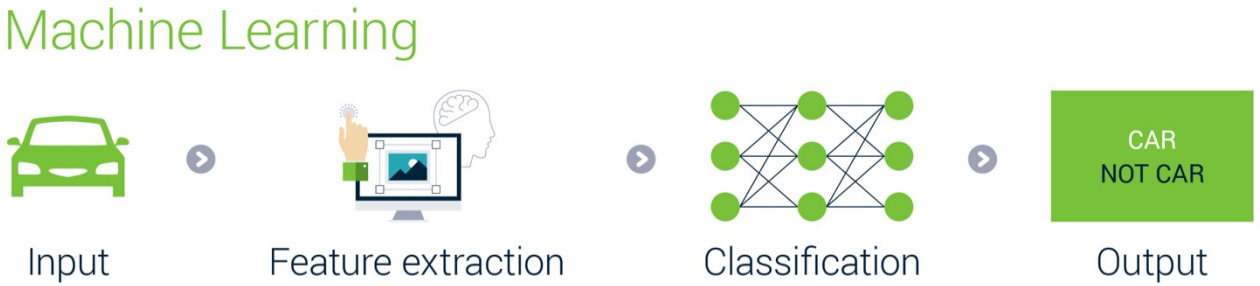
- Illumination variations
- View-point variations
- Deformable objects
- intra-class variance
- etc



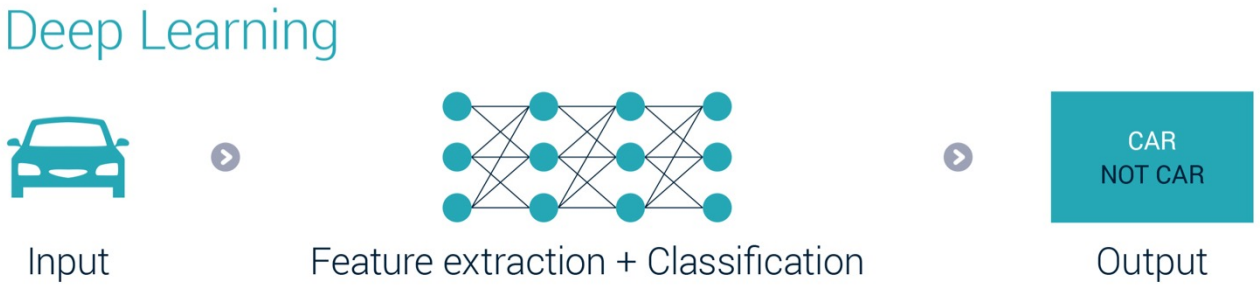
How to design “good” intermediate representations?

Data representation

Before Deep Learning: handcrafted intermediate representations for each task
- Needs expertise (PhD level) in each field
- Weak level of semantics in the representation



Since Deep Learning: automatically learning intermediate representations
+ Experimental performances >> handcrafted features
+ Able to learn high level intermediate representations
+ General learning methodology ➔ field independent, no expertise



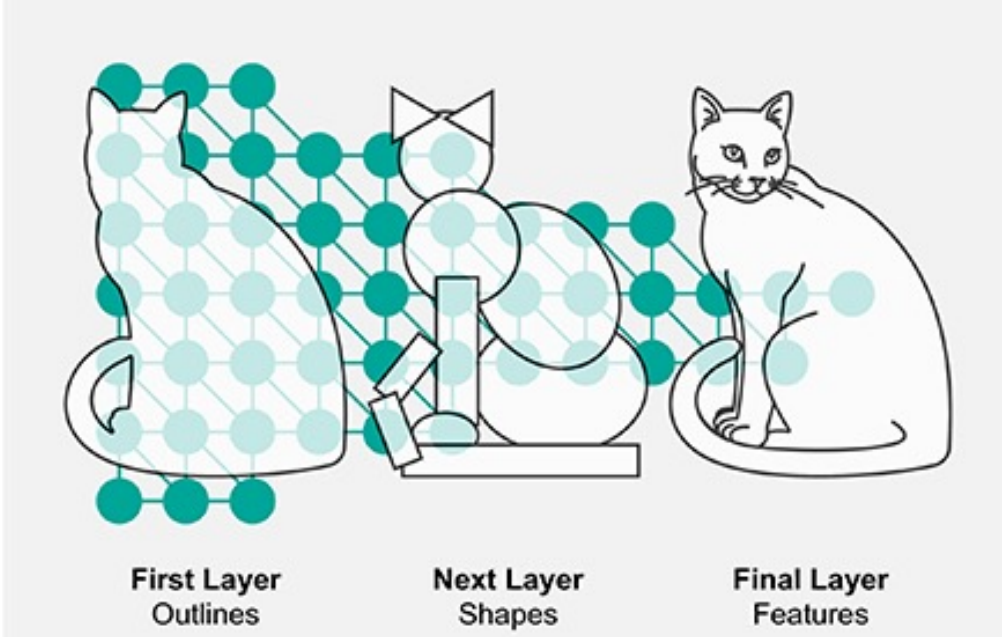
https://verhaert.com/difference-machine-learning-deep-learning/

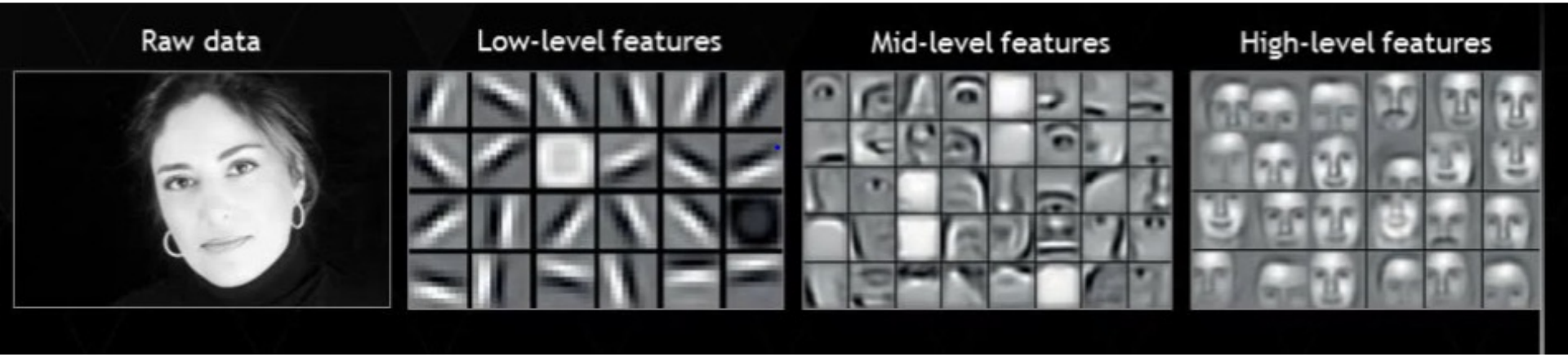
Layers of features

- Learning is done in a hierarchy of layers
- Modeled after the brain's neural networks
- "Deep" describes the number of layers used

Known for

- Speech and image recognition
- Language processing

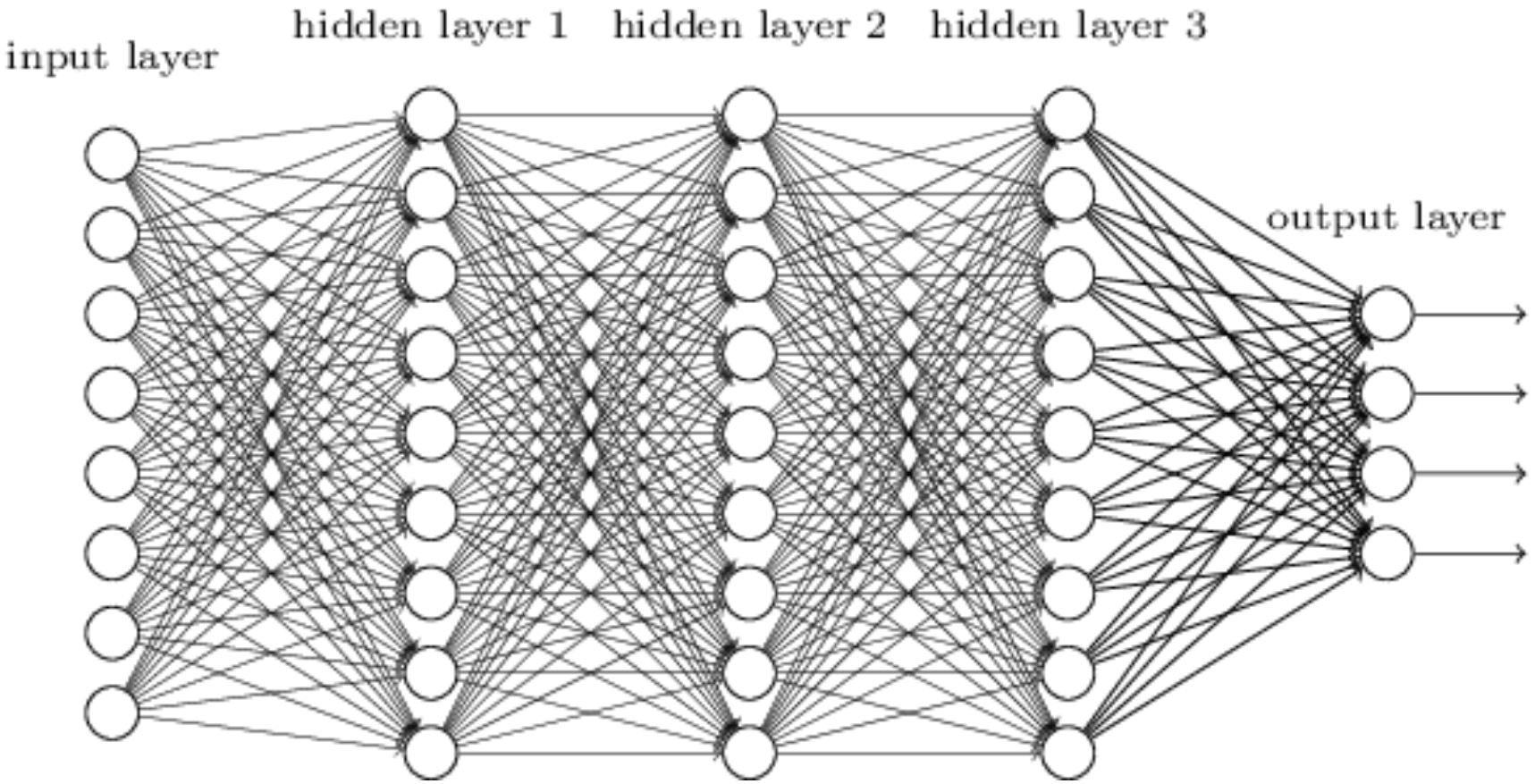




https://mapr.com/blog/artificial-intelligence-and-machine-learning-what-are-they-and-why-are-they-important/

https://www.analyticsvidhya.com/blog/2017/04/comparison-between-deep-learning-machine-learning/

Neural networks



Linear mapping

- Input: vector $\mathbf{x} \in \mathbf{R}^m$
- Output: scalar $s \in \mathbf{R}$
- Linear (affine) mapping

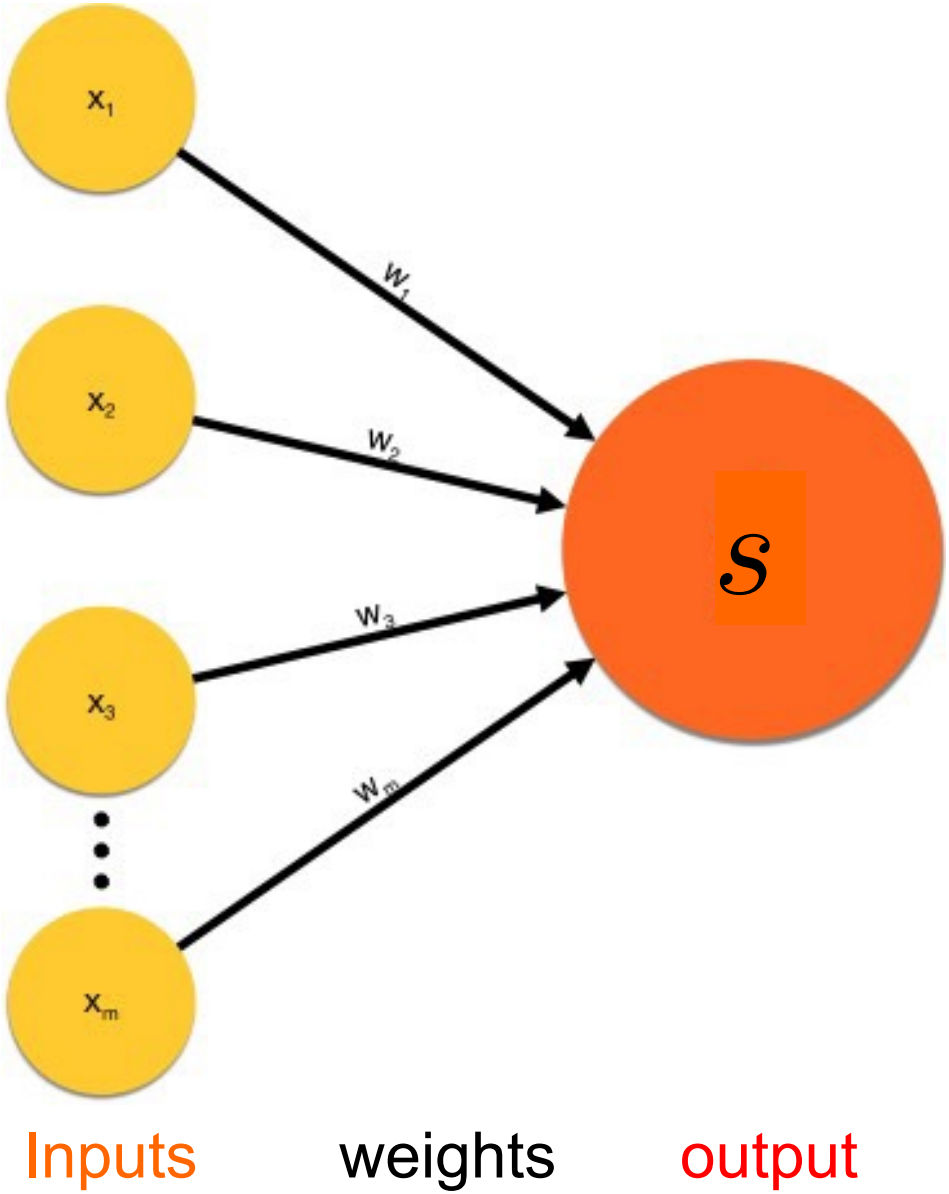
$$s = \mathbf{w}^\top \mathbf{x}$$

with weights:

$$\mathbf{w} \in \mathbf{R}^m$$

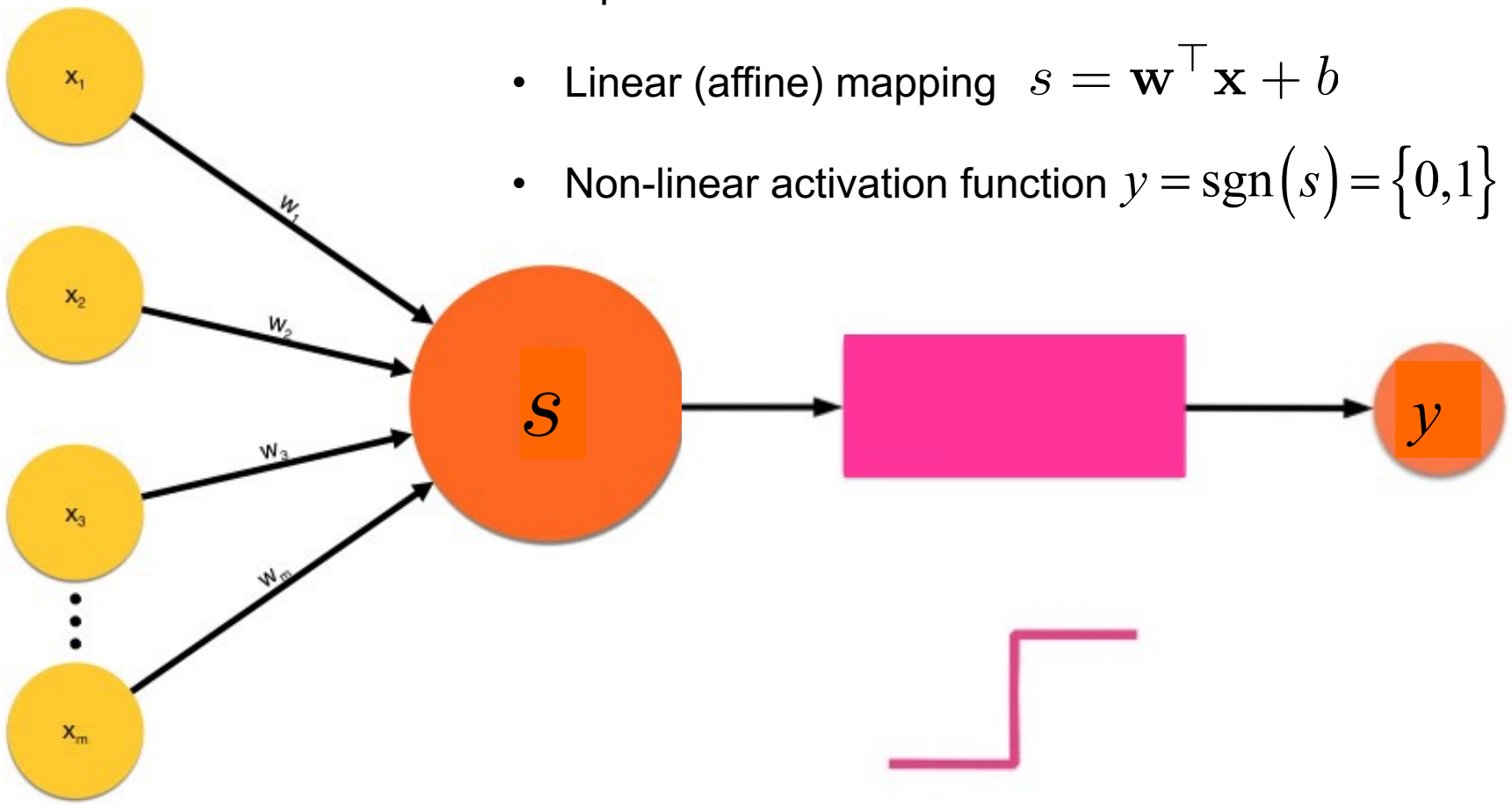
For example:

$$s = 3x_1 - 2x_2 + 5x_3$$

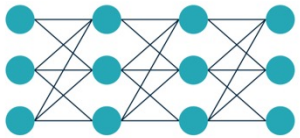


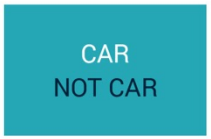
Linear classification

- Input: vector $\mathbf{x} \in \mathbf{R}^m$
- Linear (affine) mapping $s = \mathbf{w}^\top \mathbf{x} + b$
- Non-linear activation function $y = \mathrm{sgn}(s) = \{0,1\}$

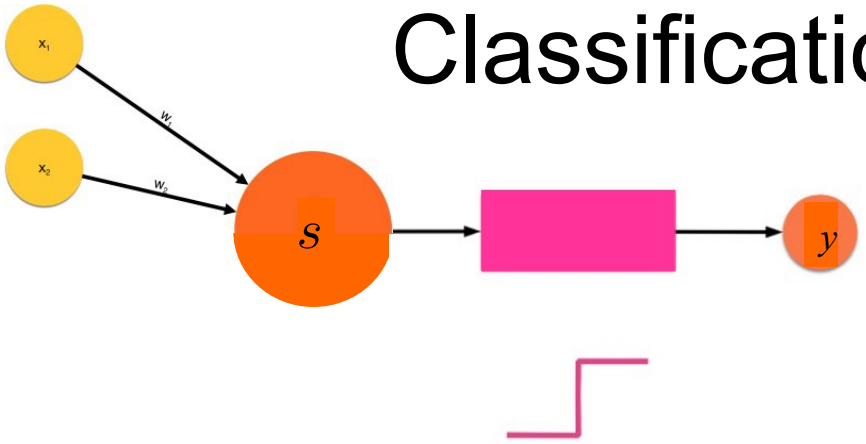


Inputs weights summation+bias activation function output

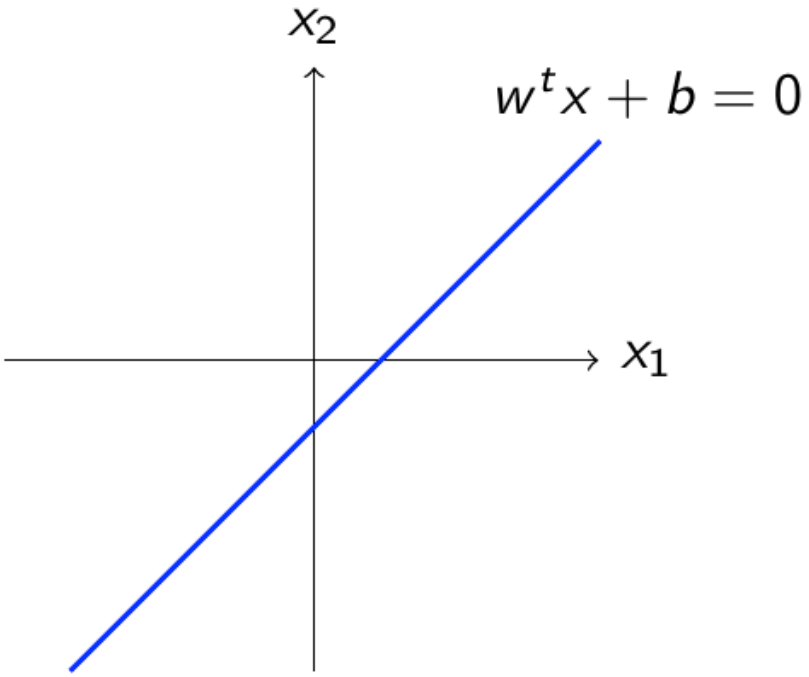




Classification – 2 inputs

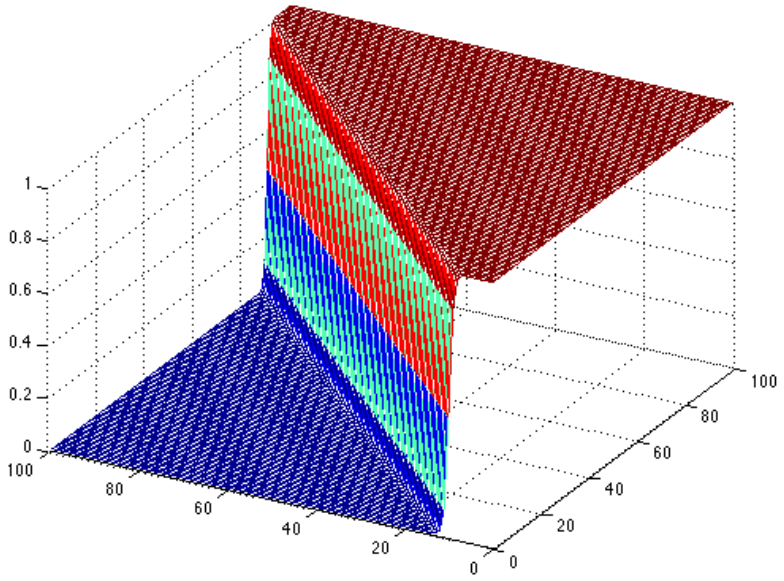


2D hyperplane = line

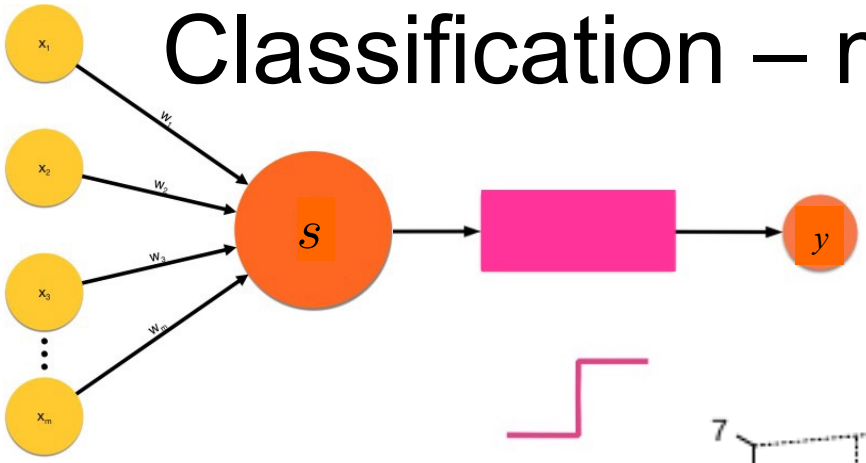


Example for

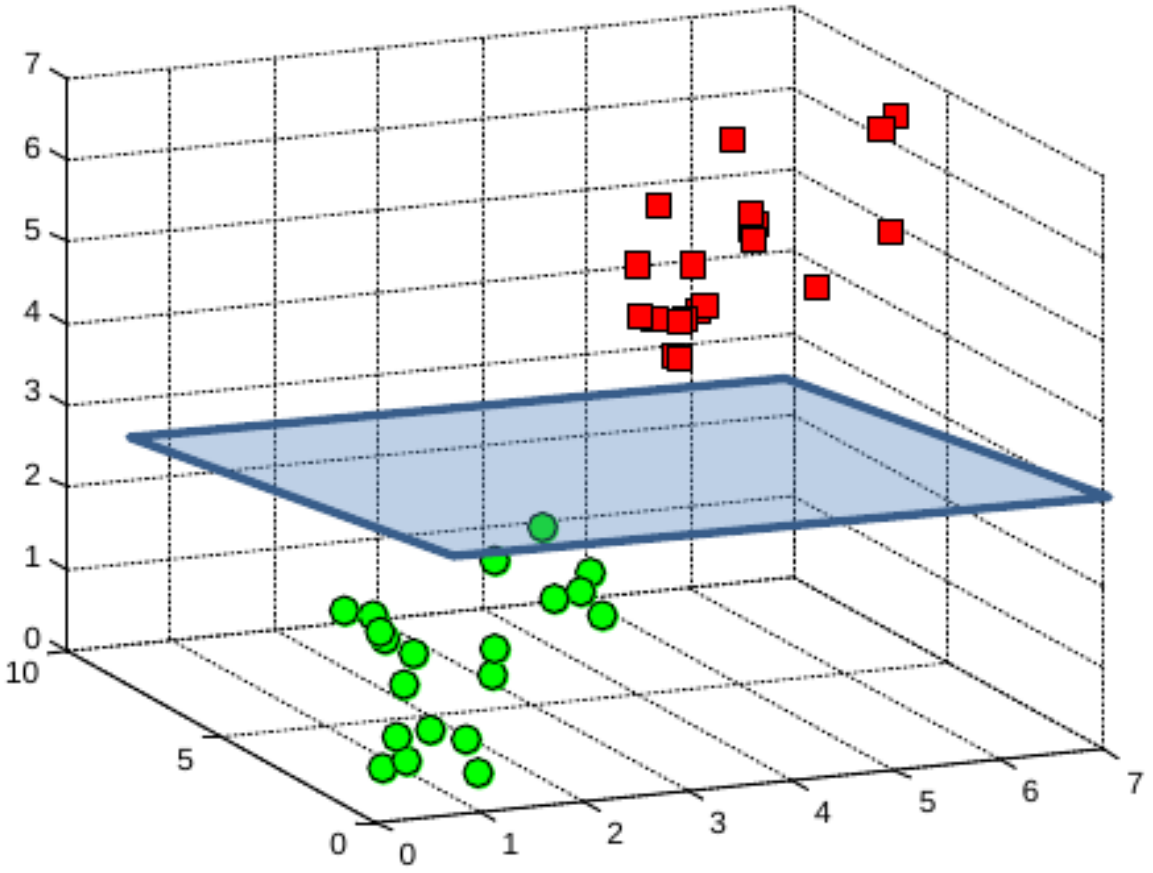
$$\begin{aligned} w_1 &= 1 \\ w2 &= -1 \\ b &= 10 \end{aligned}$$



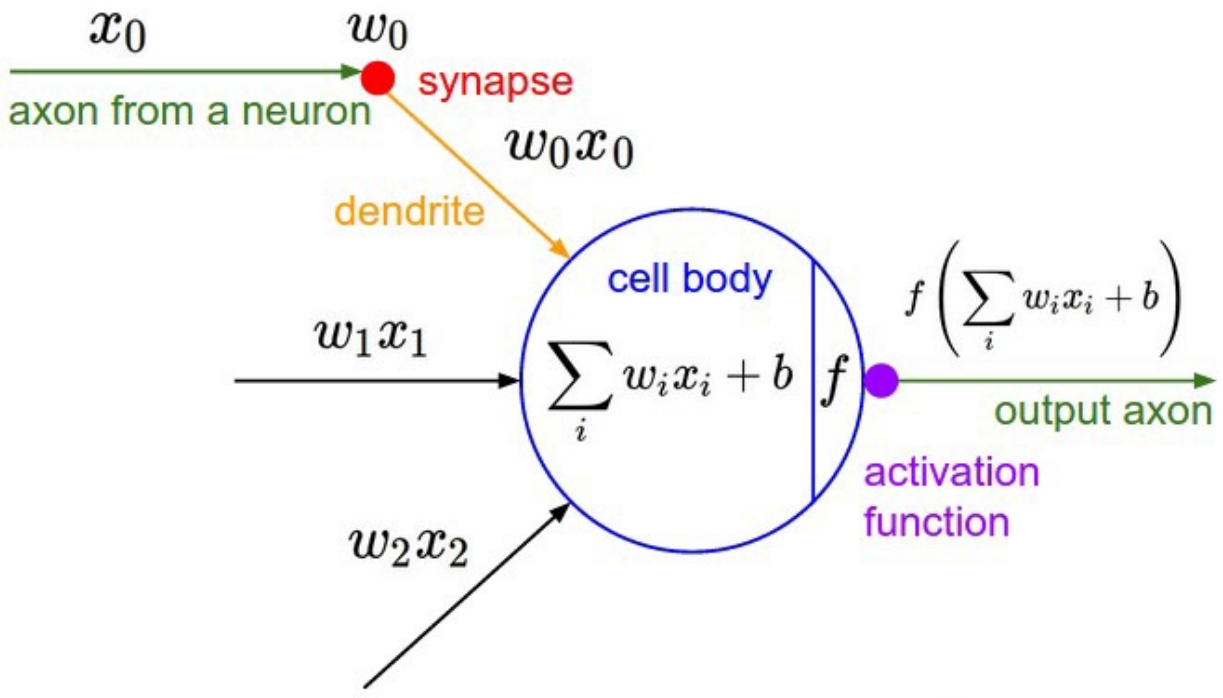
Classification – more than 2 inputs

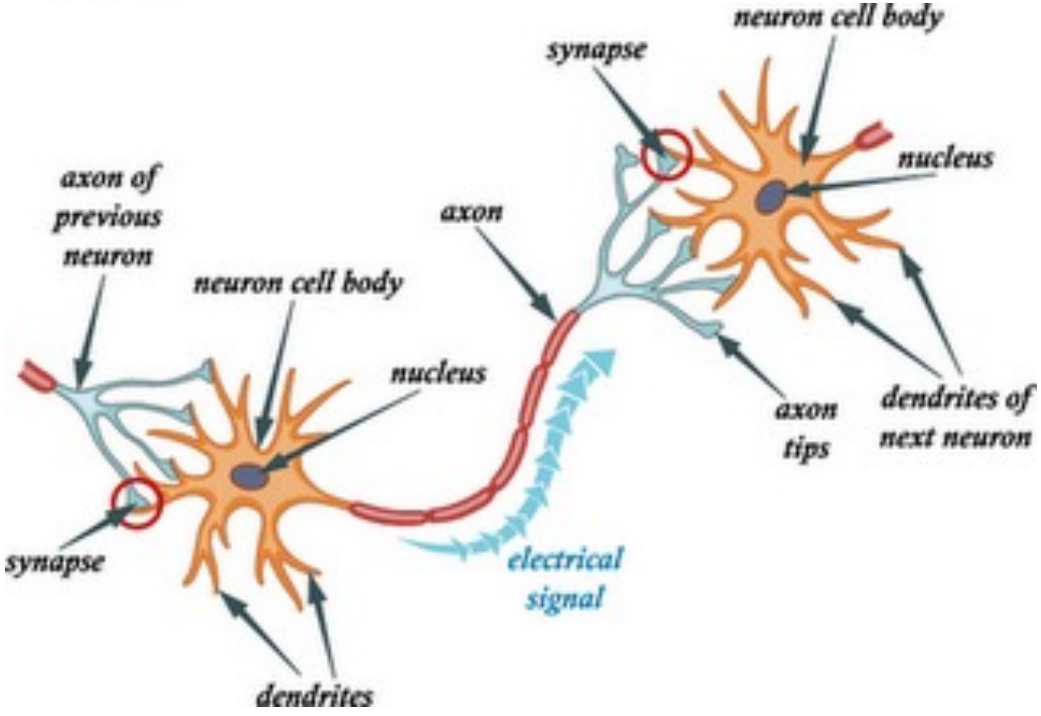


3D hyperplane=plane

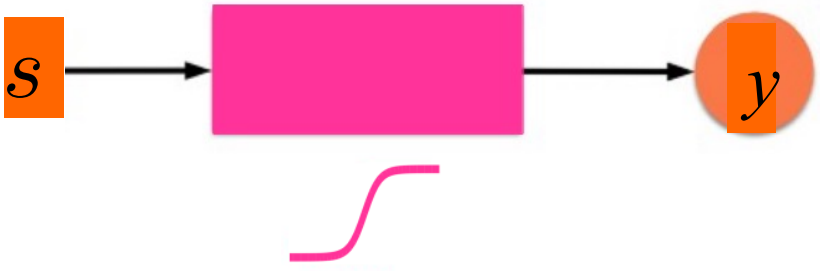


Biological analogy

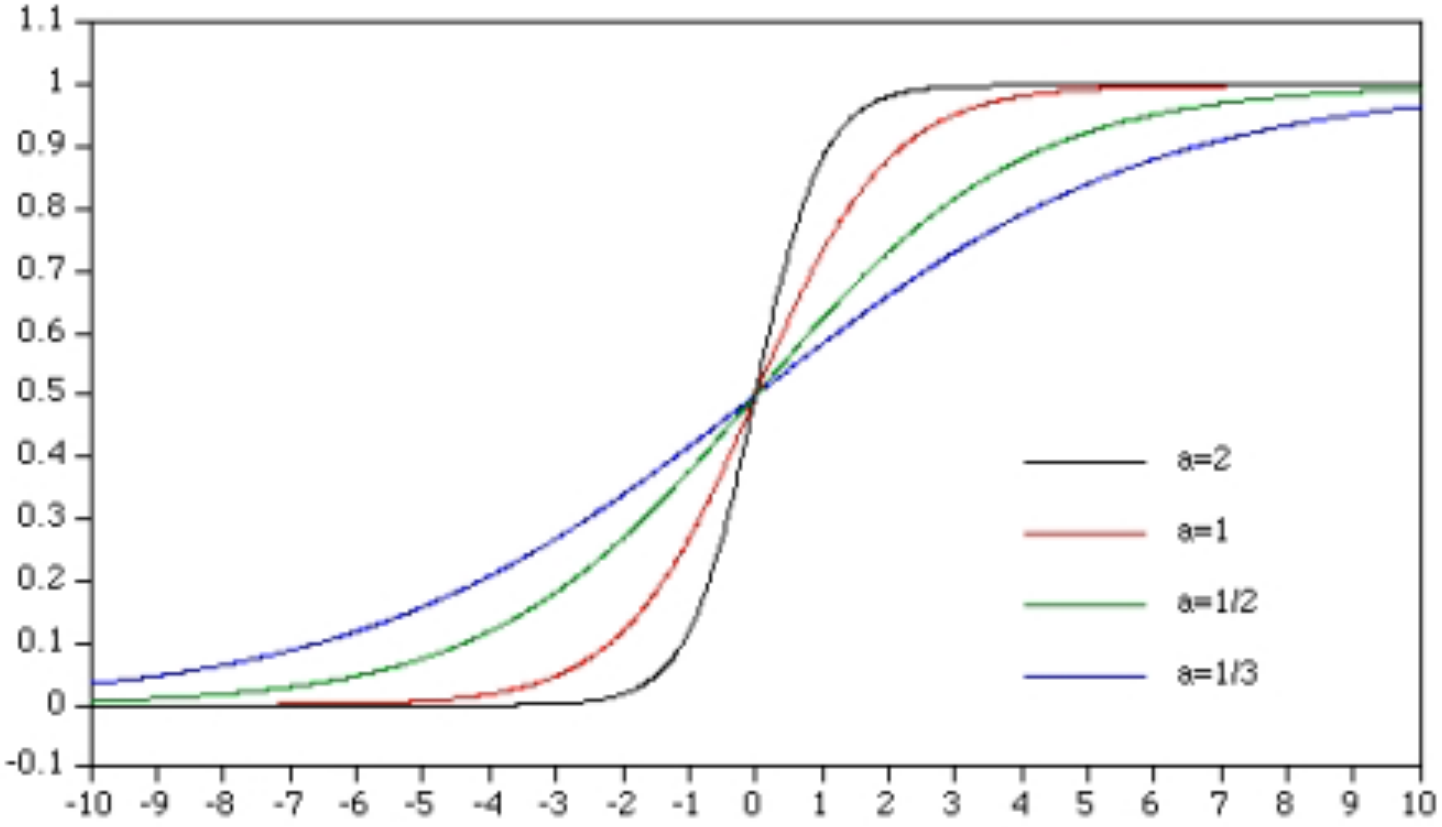




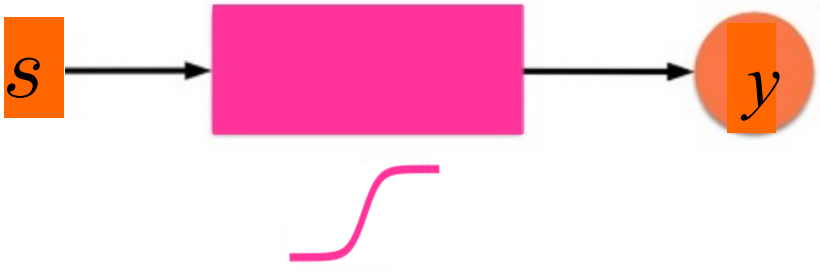
Sigmoid activation function



$$y = \frac{1}{1+\exp(-as)}$$



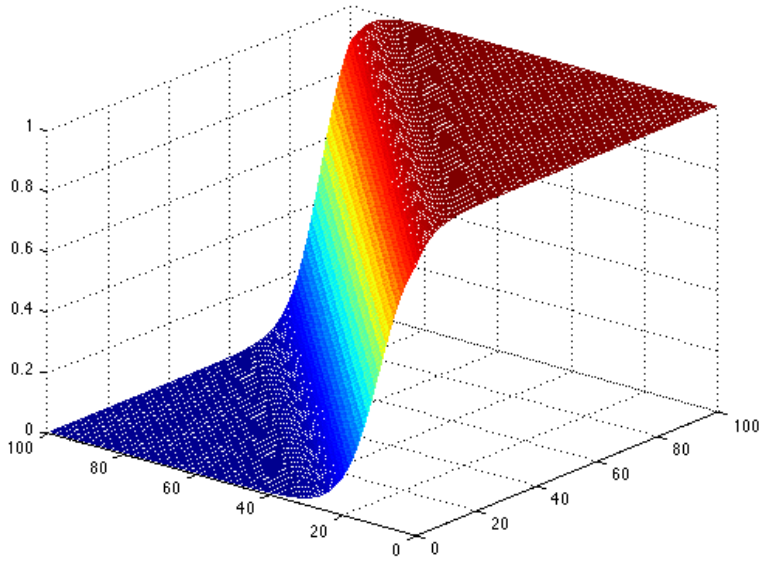
Sigmoid activation function

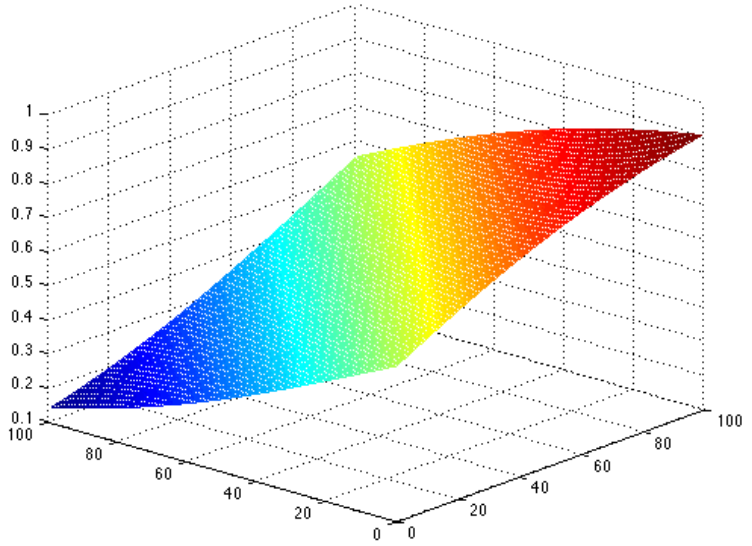


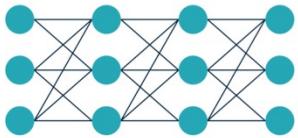
Example for

$$\begin{aligned} w_1 &= 1 \\ w2 &= -1 \\ b &= 10 \end{aligned}$$

The output is now a real number (probability %)

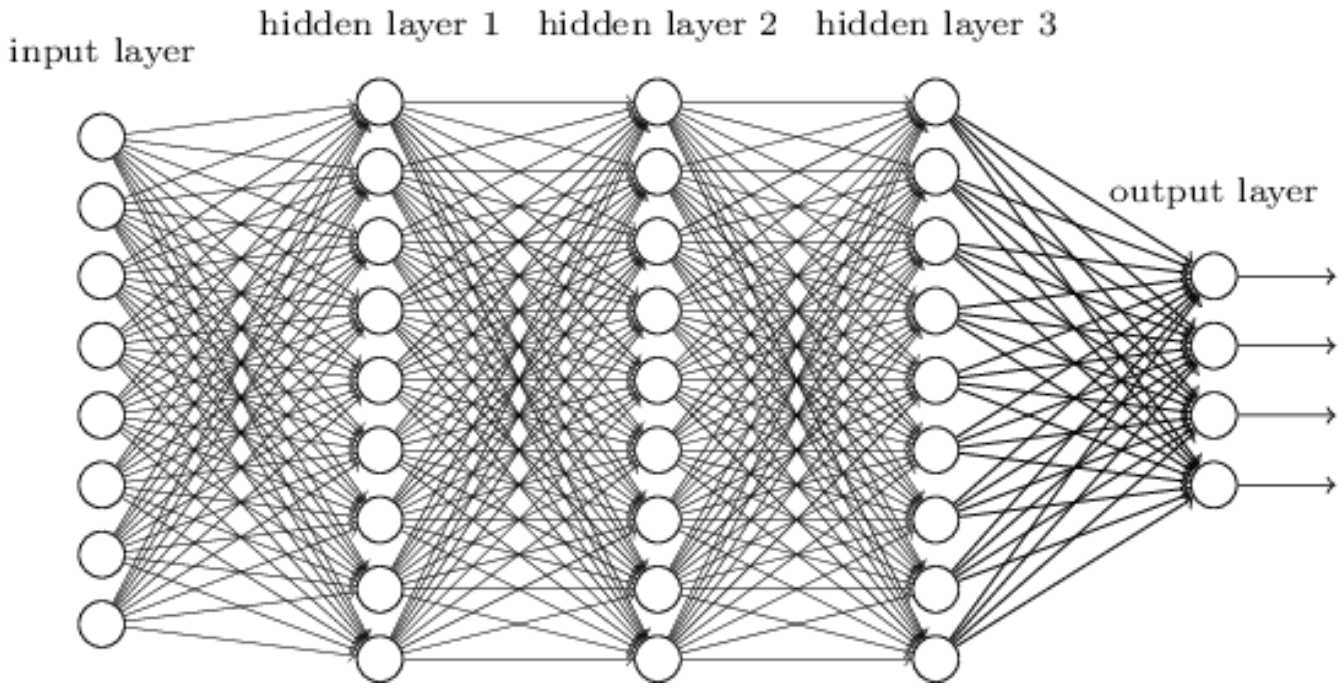




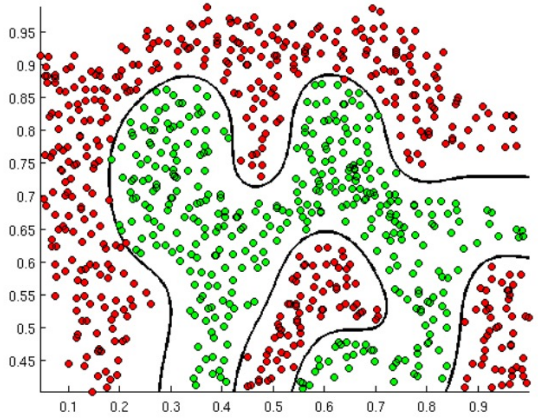


81% CAR

Neural network



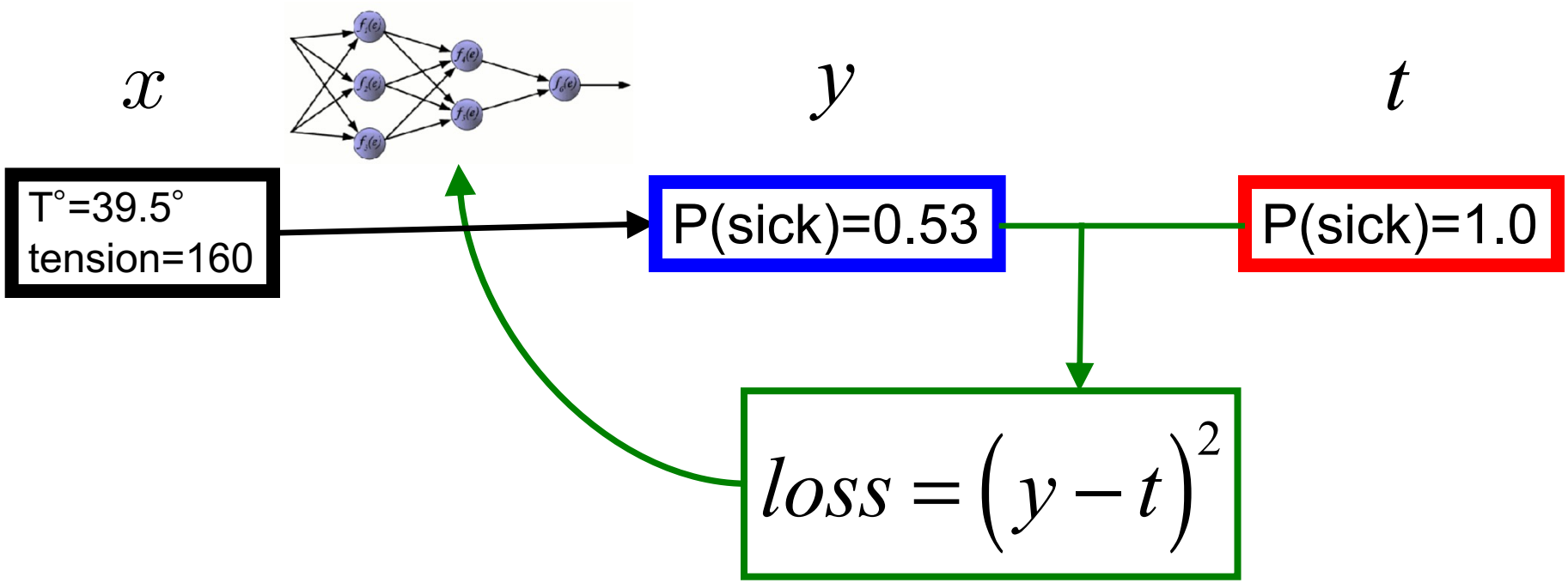
- Multiple outputs: multi-class classification (plane, car, dog…)
- Multiple (hidden) layers: can address non linear classification



Training the neural network

Finding weights and biases

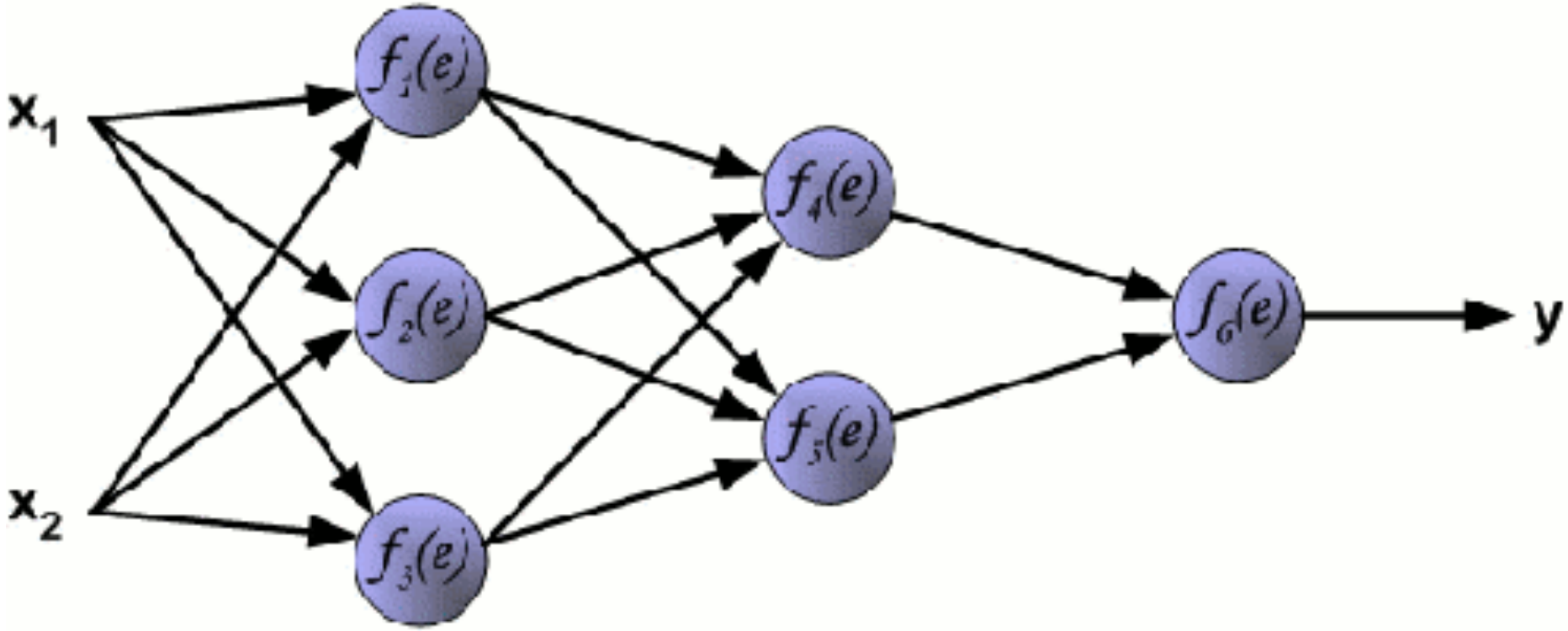
Supervised learning: provide inputs *x* with known (desired) targets/labels *t* (ground truth)



Compare output ***y*** and target ***t*** to update weights and biases

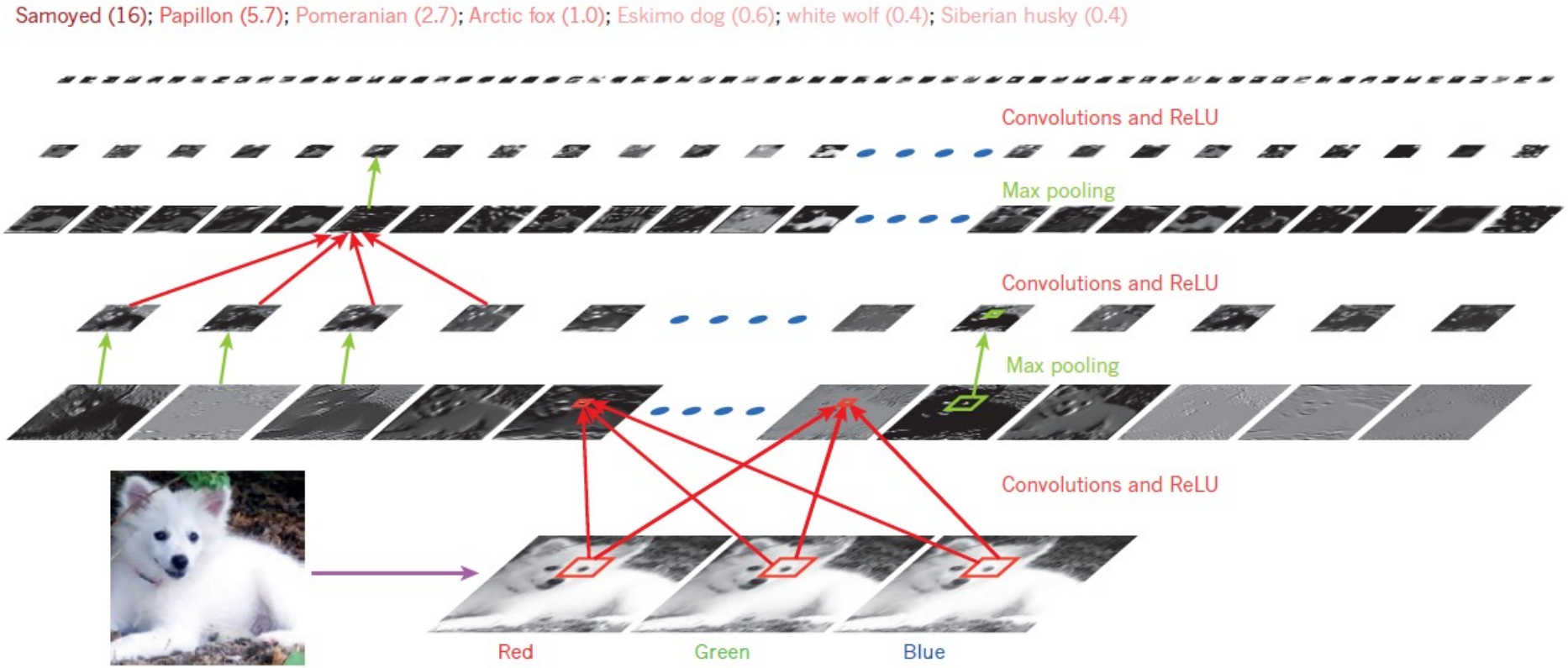
Training the neural network

Finding weights and biases



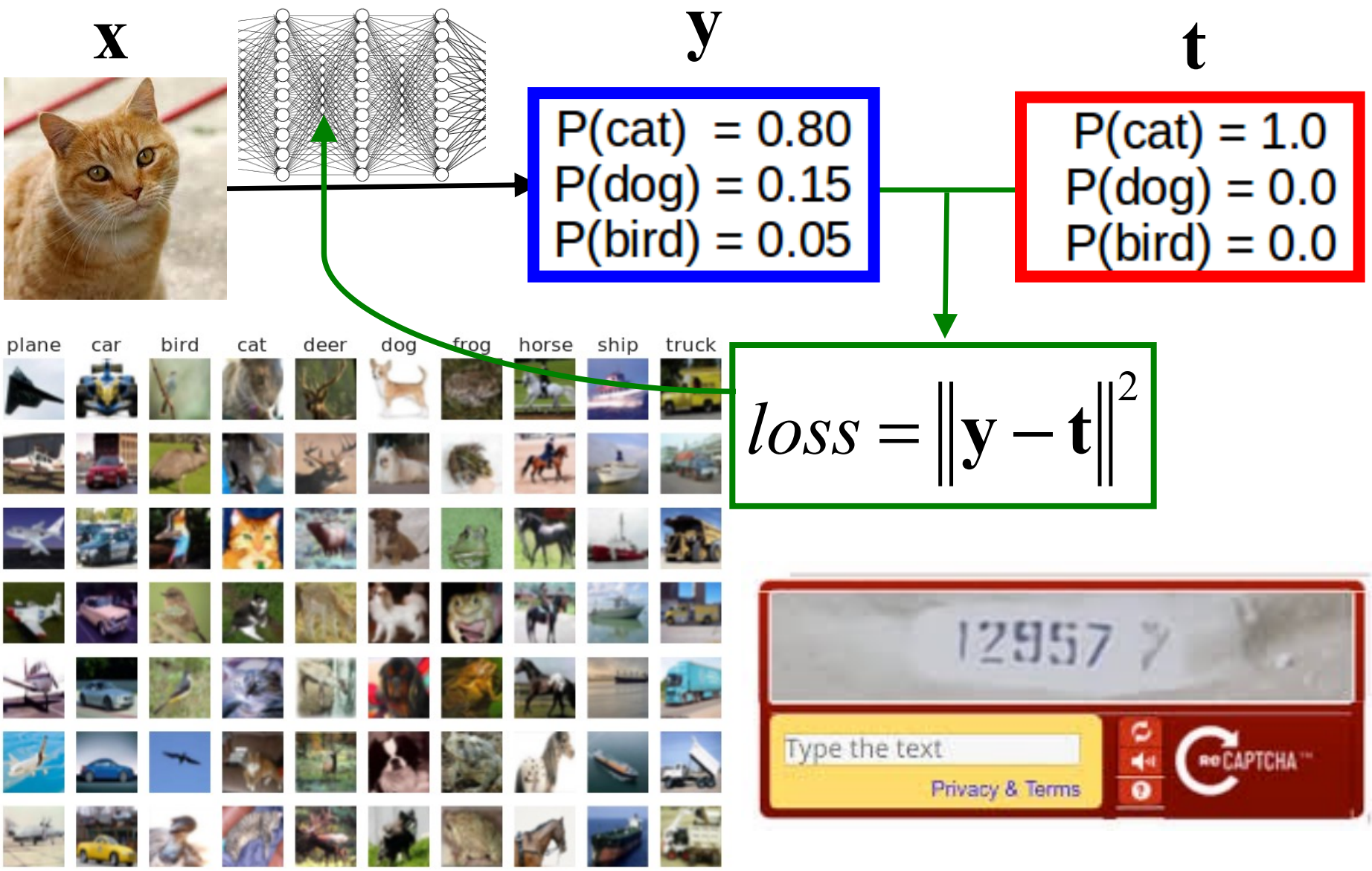
https://giphy.com/gifs/neural-networks-4LiMmbAcvgTQs

Convolutional neural network



Y. LeCun, Y. Bengio and G. Hinton; Deep Learning, Nature, vol 521, 28 may 2015

Training the convolutional neural network



Try at home!

Libraries

- C++: Tensorflow
- Python: Keras (interface for tensorflow), pytorch
- Matlab: toolbox

Pre-trained CNN

- YOLO (You Only Look Once)
- Inception V3
- ResNet50
- VGG19

Thank you for your attention!

Historique des réseaux de neurones

- 1957 : proposition du perceptron par Frank Rosenblatt
- 1967 : demonstration par Marvin Minsky que le perceptron est incapable de traiter des donnees non lineairement separables, perte d'interêt pour les approches neuronales
- 1986 : Rumelhart, Hinton et Williams demontrent l'utilisation de la retropropagation des gradients pour l'entrainement du perceptron multicouche
- 1995-2005 : developpement des SVM, perte d'interêt pour les reseaux de neurones
- 2006 : premieres architectures profondes de reseaux de neurones
- 2012 : resultats en reconnaissance d'objets (Toronto, ImageNet) et de la parole (Microsoft) demontre le potentiel de technologie disruptive de l'apprentissage profond
- 2014 : explosion d'investissements prives en apprentissage automatique, en particulier en apprentissage profond