Introduction to statistical learning

B. Artsybasheff (image source)

Last time

- A brief historical overview of neural networks and deep learning
- Historical origins
- Present successes

Today

- Statistical learning
- Two simple classification models:
  nearest neighbor, linear classifiers
- Beyond classification and supervised learning:
  A brief taxonomy

How can we build an agent to…

Play chess?

Translate between languages?

Recognize object categories?

Fly a drone?

Statistical learning

- Good old-fashioned AI (GOFAI) answer: Program expertise into the agent
  - Never worked (in general)…

Statistical learning

- Good old-fashioned AI (GOFAI) answer:
  Program expertise into the agent
  - Never worked (in general)…
- Modern answer: Program into the agent the *ability to improve performance based on experience*
  - Experience should come from training data or demonstrations
  - Learning is optimizing performance of the agent on the training data, with the hope that it will *generalize* to unseen inputs

Example: Image classification

input **desired output**

apple

pear

tomato

cow

dog

horse

Training data

apple

pear

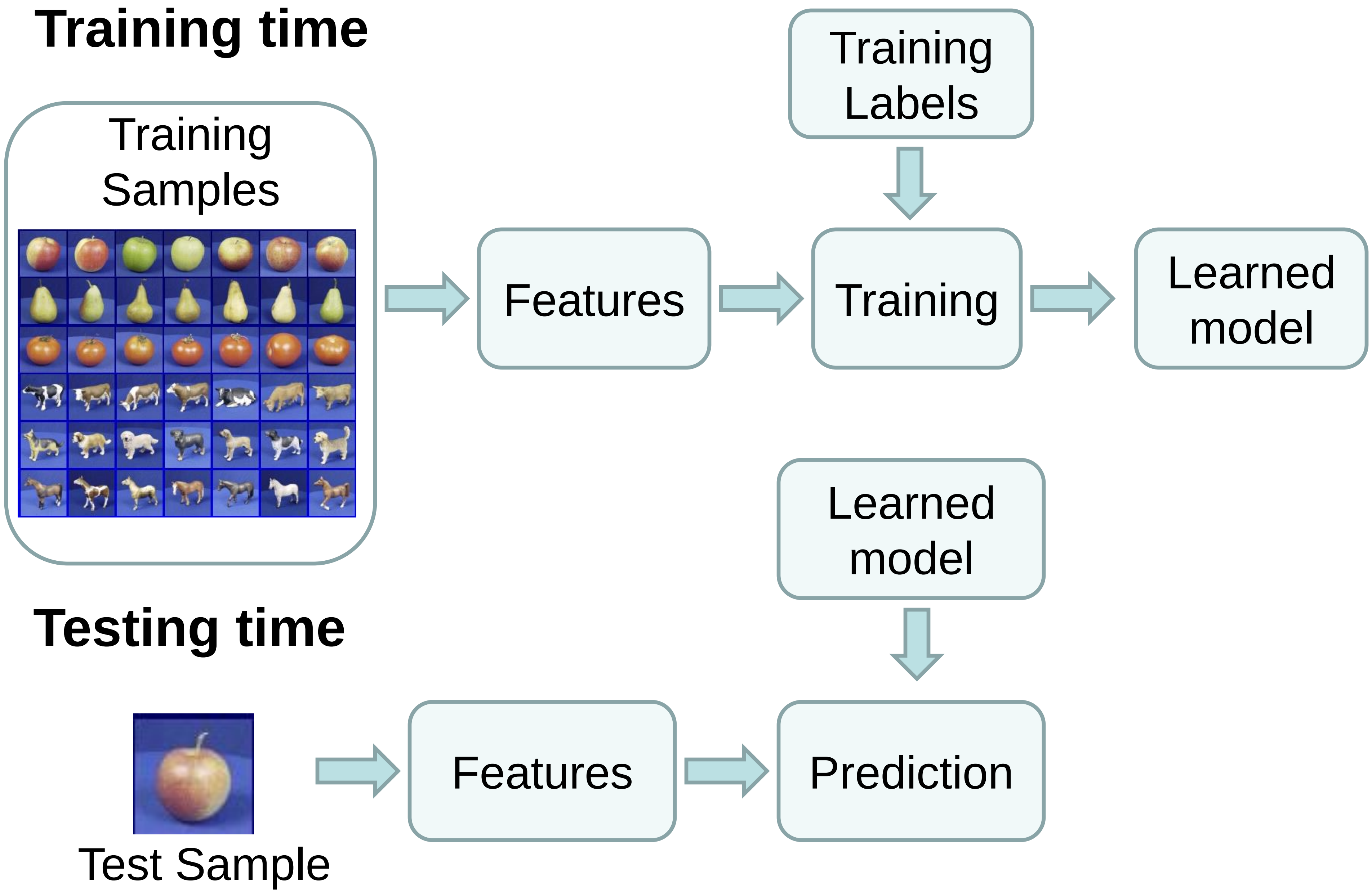
tomato

cow

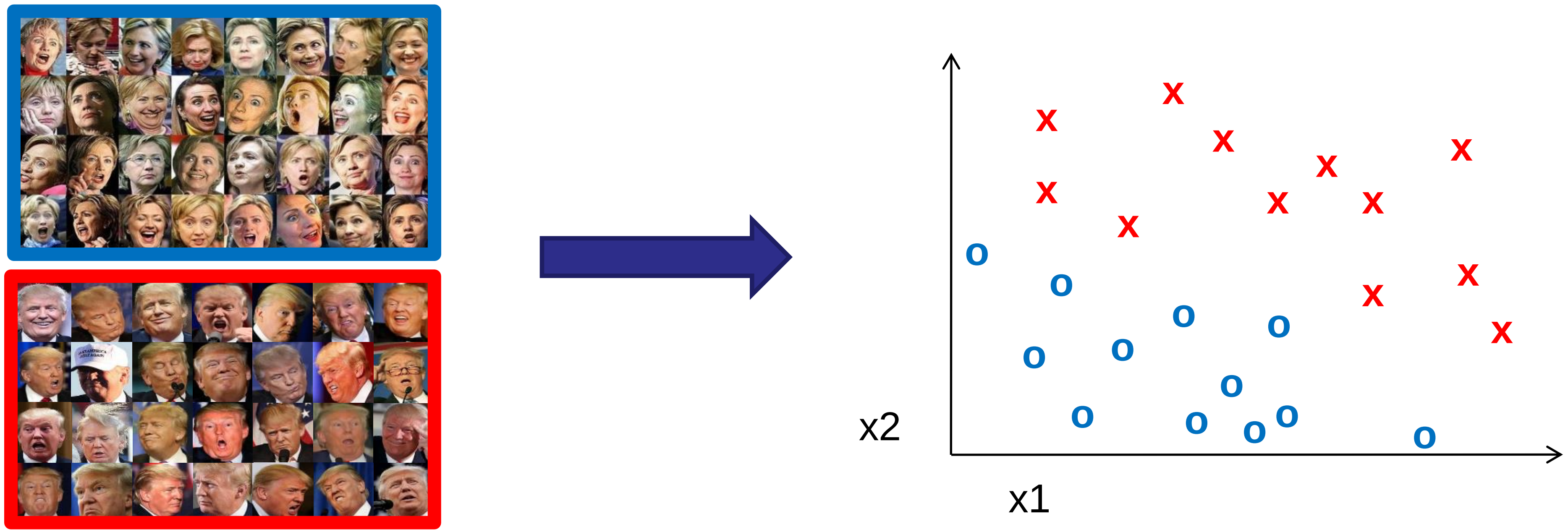
dog

horse

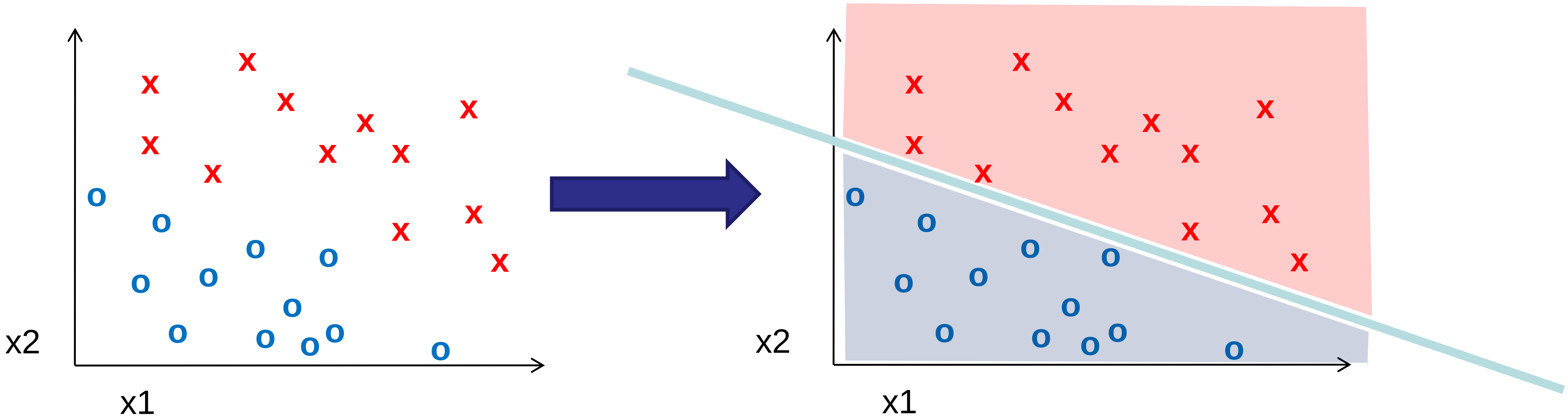
Training and testing



- **Image features**: map images to feature space



- **Classifiers**: map feature space to label space



The basic *supervised learning* framework

$$y = f(x)$$

output

prediction function

input

- **Training** (or **learning**): given a *training set* of labeled examples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, learn a predictor f
- **Testing** (or **inference**): apply f to a new *test example* x and output the predicted value $y = f(x)$

More supervised learning examples: Text classification

Spam classification

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

Sentiment classification

"I love this movie.
I've seen it many times
and it's still awesome."

"This movie is bad.
I don't like it it all.
It's terrible."

Image source

Another example: Grasp classification

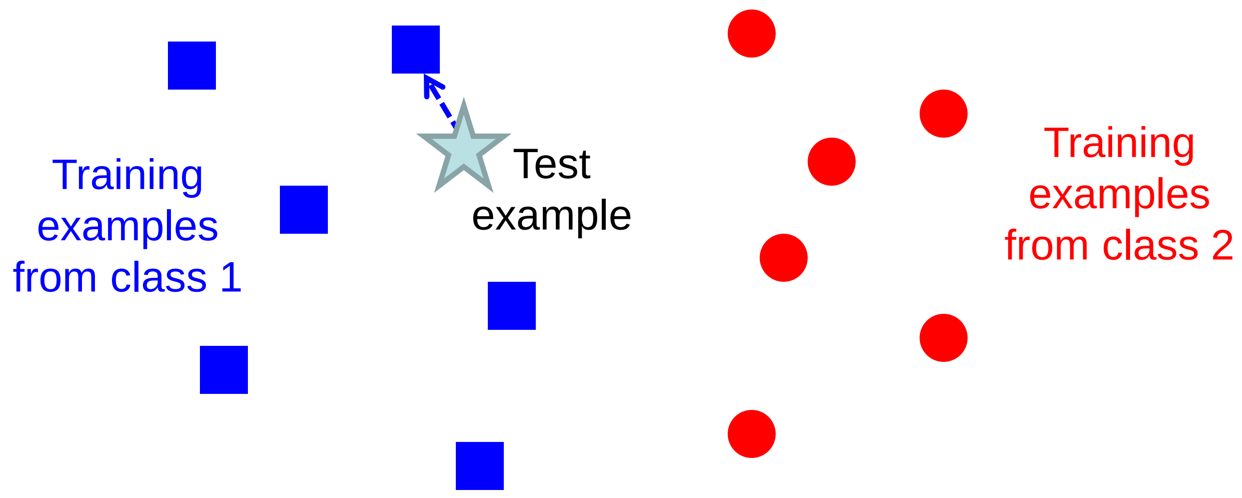
L. Pinto and A. Gupta. [Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours](). ICRA 2016.

[YouTube video]()

Two simple classification models

- Nearest neighbor
- Linear classifiers

Nearest neighbor classifier

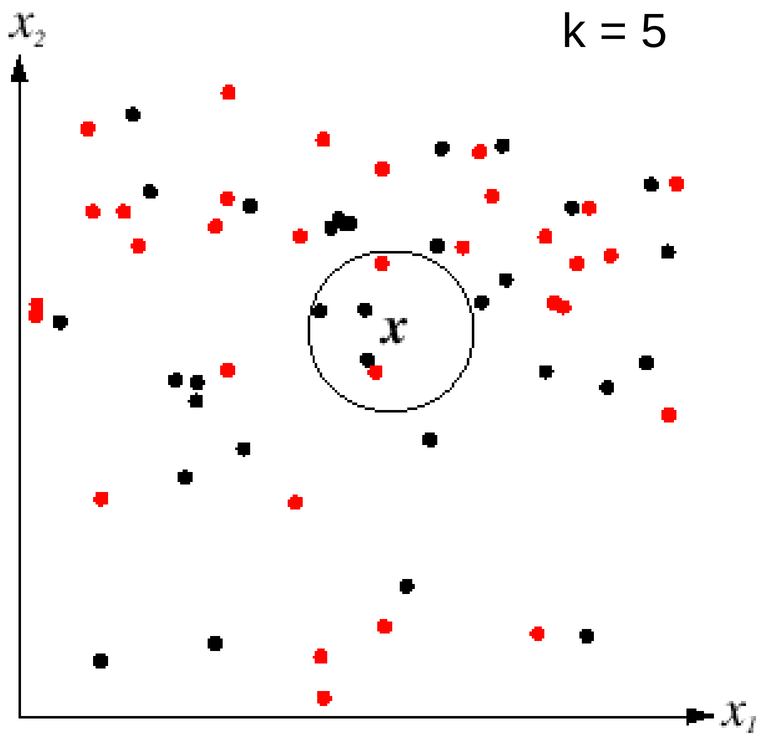


$f(x) =$ label of the training example nearest to x

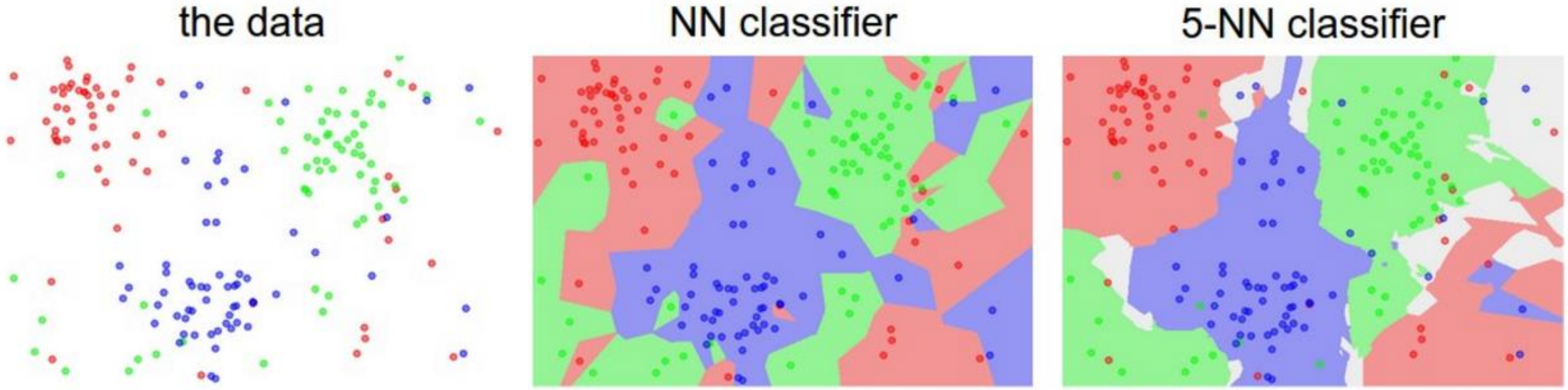
- All we need is a distance function for our inputs
- No training required!

K-nearest neighbor classifier

- For a new point, find the k closest points from training data
- Vote for class label with labels of the k points



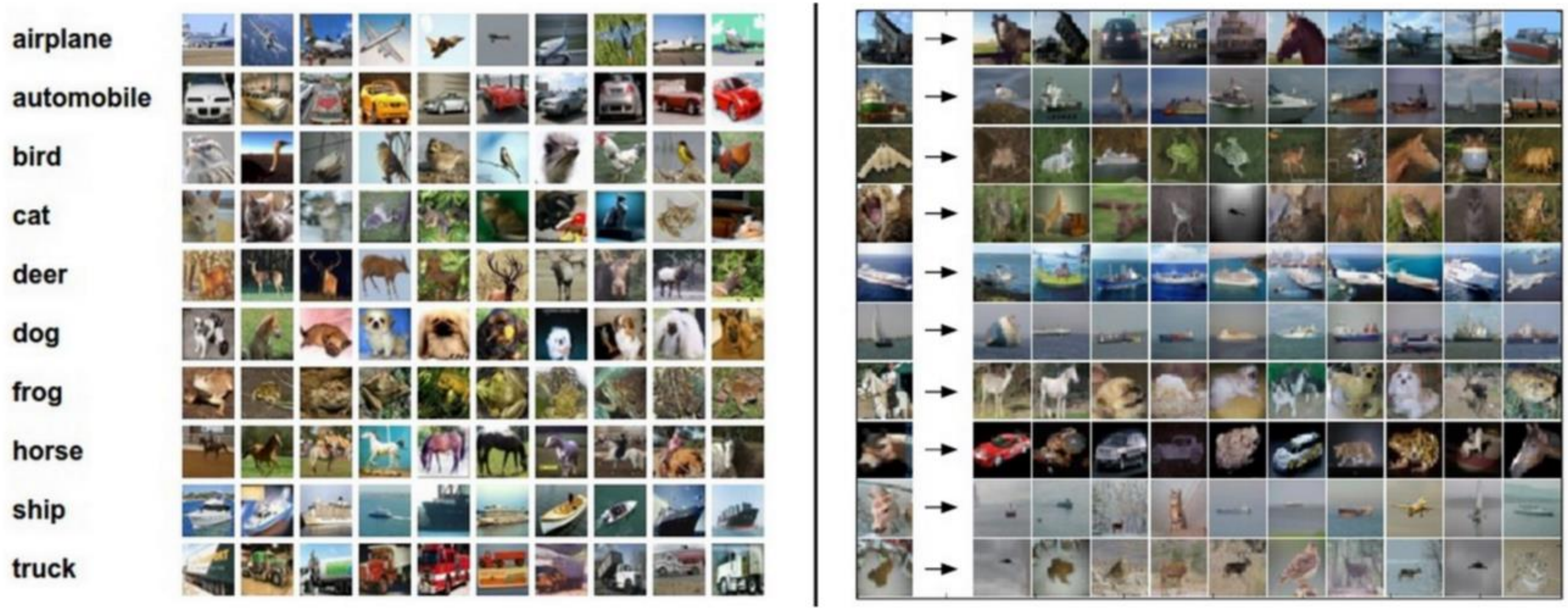
K-nearest neighbor classifier



- K-NN is more robust to *outliers*

Credit: Andrej Karpathy, http://cs231n.github.io/classification/

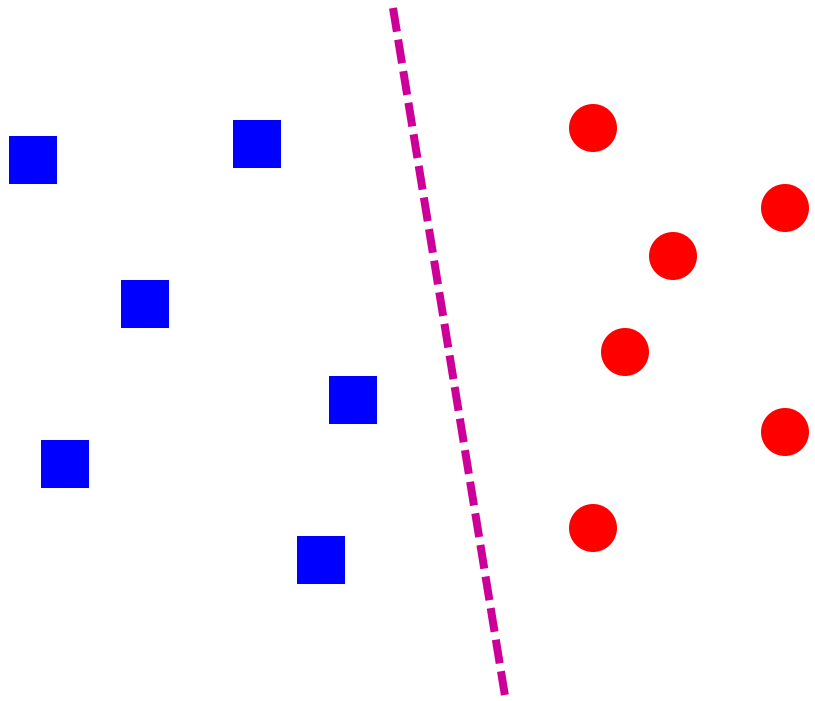
K-nearest neighbor classifier



Left: Example images from the CIFAR-10 dataset. Right: first column shows a few test images and next to each we show the top 10 nearest neighbors in the training set according to pixel-wise difference.

Credit: Andrej Karpathy, http://cs231n.github.io/classification/

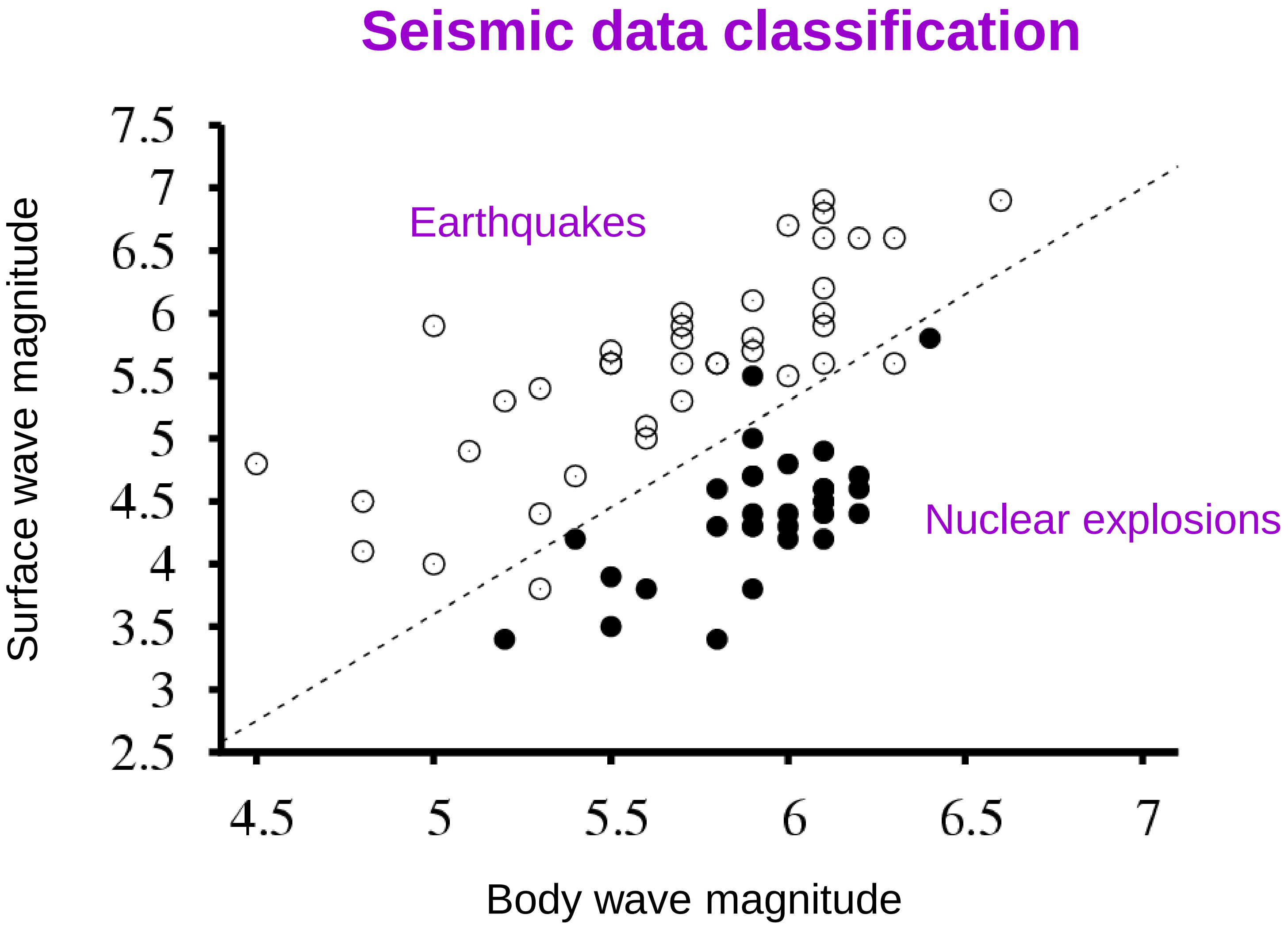
Linear classifier



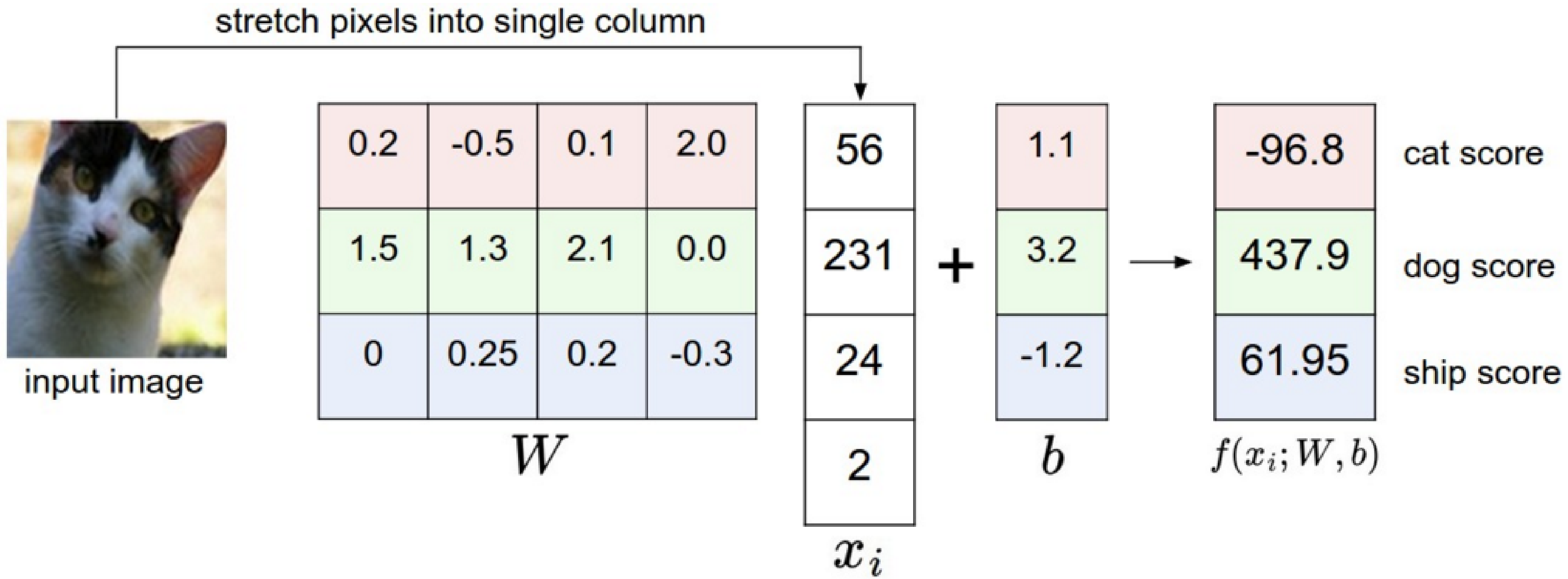
- Find a *linear function* to separate the classes

$$f(x) = \mathrm{sgn}(w_1 x_1 + w_2 x_2 + \ldots + w_D x_D + b) = \mathrm{sgn}(w \cdot x + b)$$

Visualizing linear classifiers

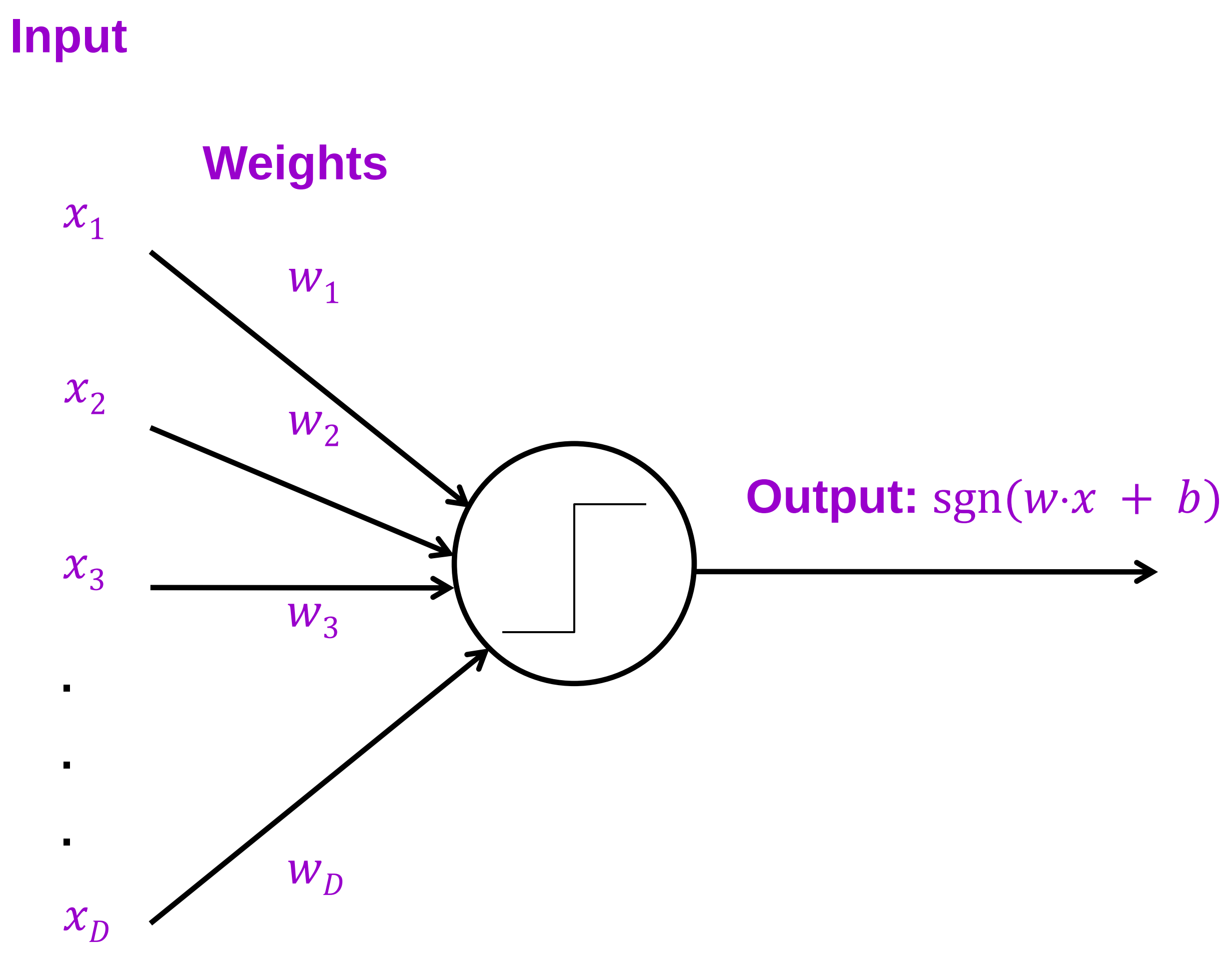


Visualizing linear classifiers

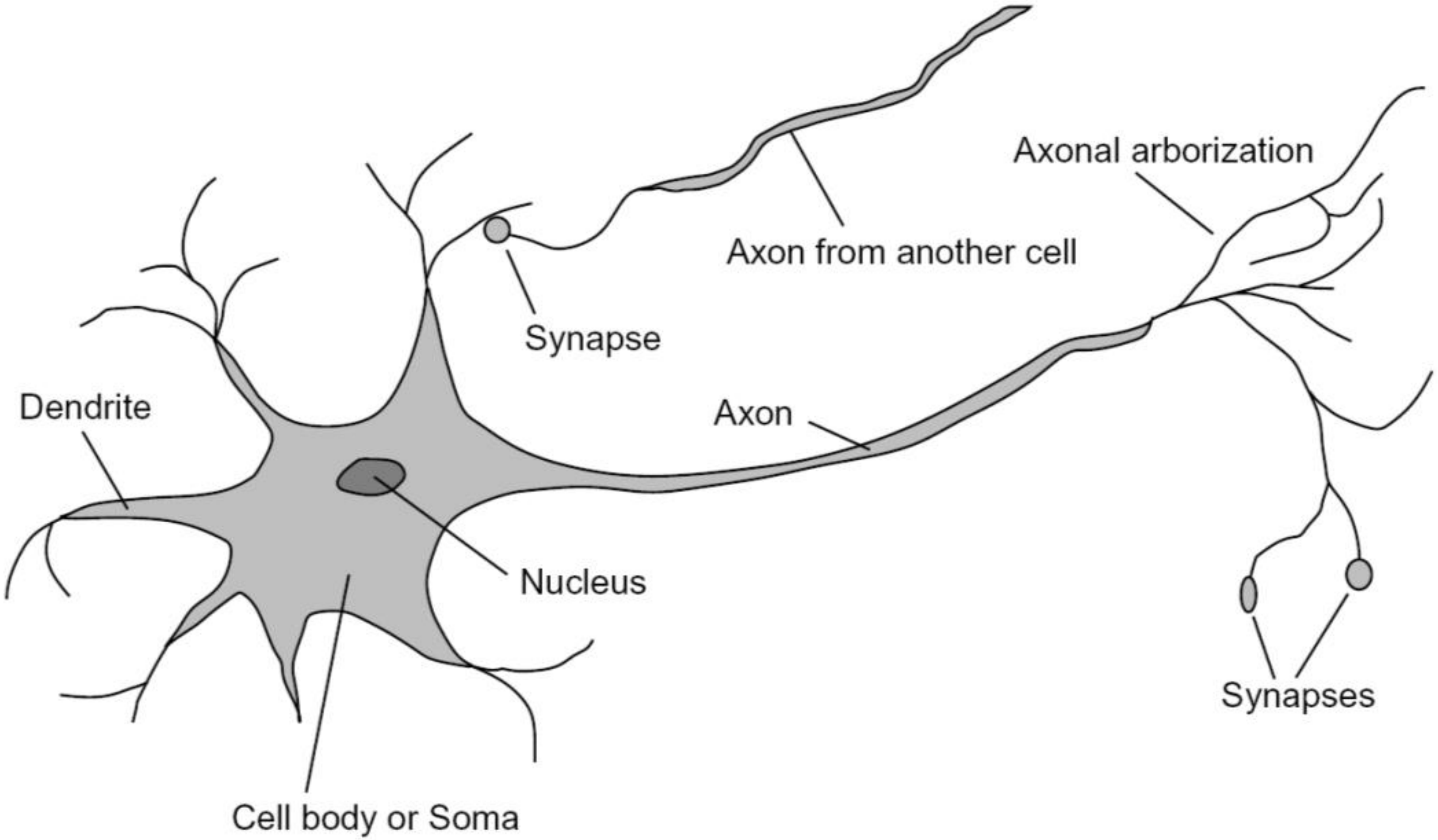


Source: Andrej Karpathy, http://cs231n.github.io/linear-classify/

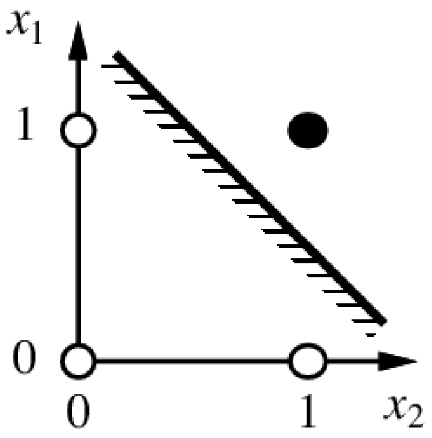
Linear classifier: Perceptron view



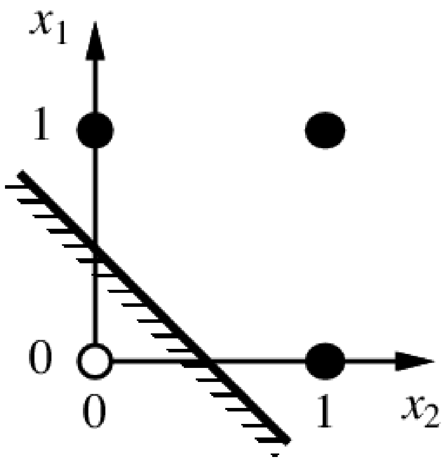
Loose inspiration: Biological neurons



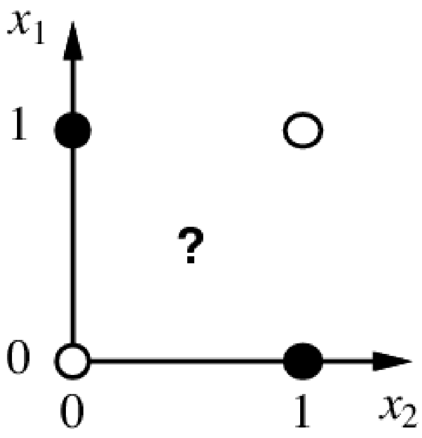
Perceptrons, linear separability and Boolean functions



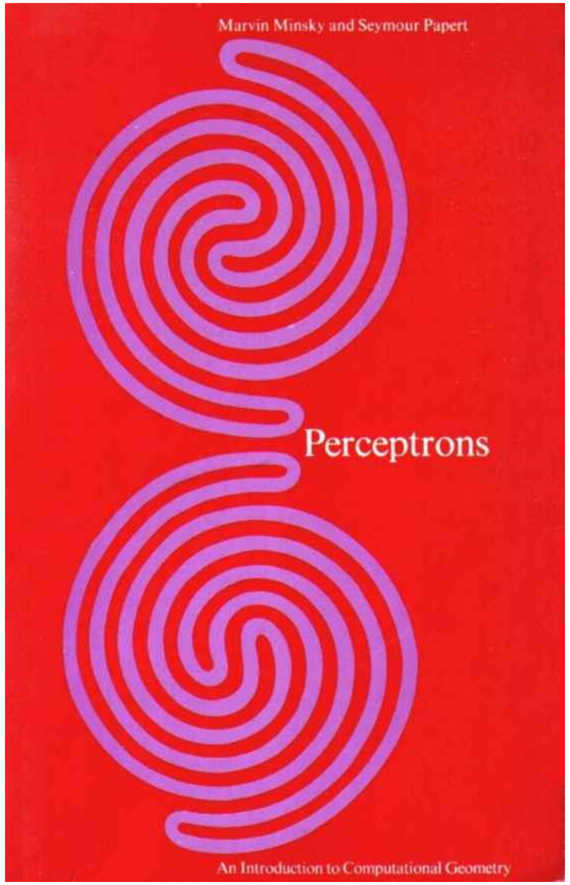
x_1 **and** x_2



x_1 **or** x_2



x_1 **xor** x_2



NN vs. linear classifiers: Pros and cons

- NN pros:
  - + Simple to implement
  - + Decision boundaries not necessarily linear
  - + Works for any number of classes
  - + *Nonparametric* method
- NN cons:
  - – Need good distance function
  - – Slow at test time
- Linear pros:
  - + Low-dimensional *parametric* representation
  - + Very fast at test time
- Linear cons:
  - – Works for two classes
  - – How to train the linear function?
  - – What if data is not linearly separable?

Outline

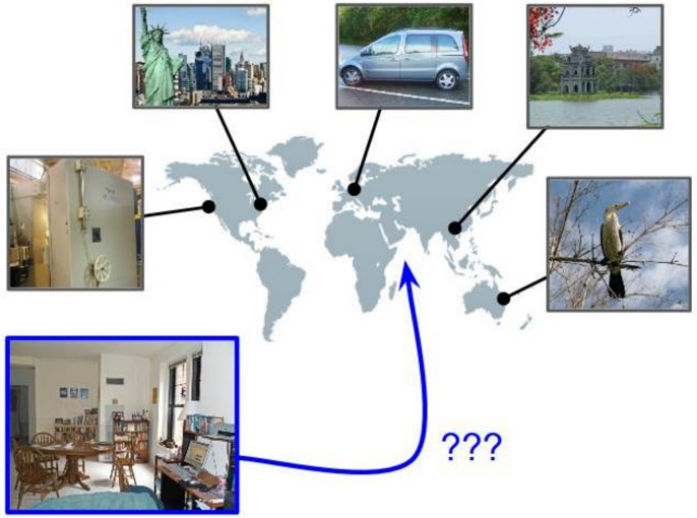
- Statistical learning
- Two simple classification models: nearest neighbor, linear classifiers
- Beyond classification and supervised learning: A brief taxonomy

Beyond classification

- Other prediction scenarios (output types)
  - Regression
  - Structured prediction

Regression

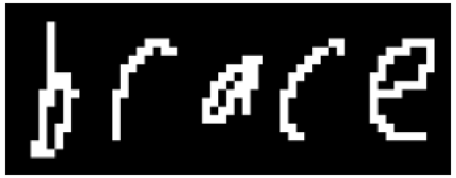
When was that made?



IM2GPS

Image colorization

Structured Prediction

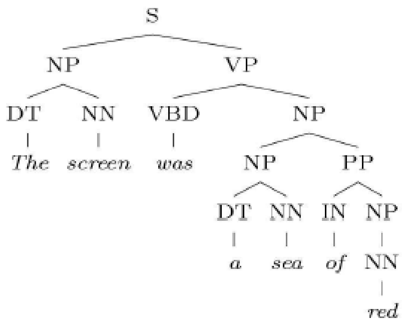


brace

Image

Word

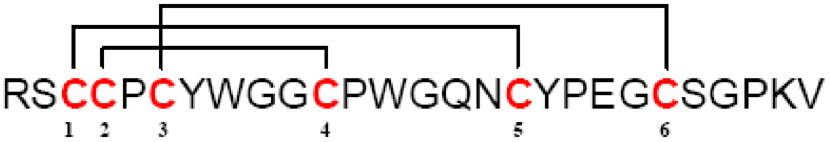
The screen was
a sea of red



Sentence

Parse tree

RSCCPCYWGGCPW
GQNCYPEGCSGPKV



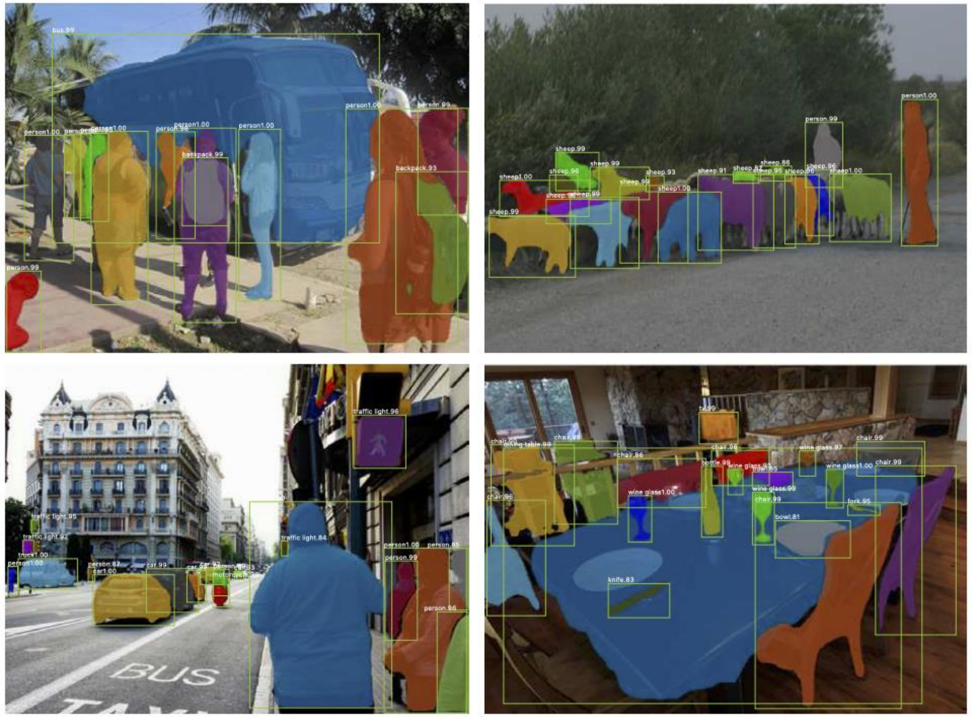
Amino-acid sequence

Bond structure

Source: B. Taskar

Structured and dense prediction for scene understanding

Bounding box prediction,
dense prediction

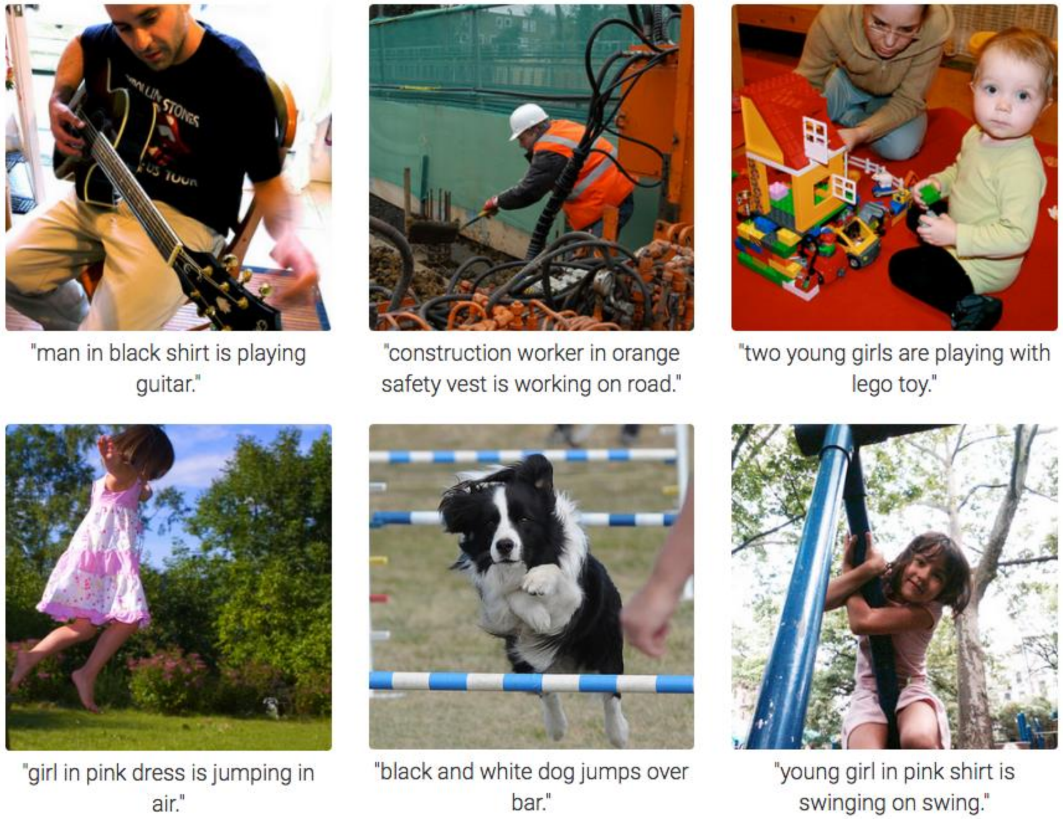


Keypoint prediction

K. He, G. Gkioxari, P. Dollar, and R. Girshick, [Mask R-CNN](), ICCV 2017

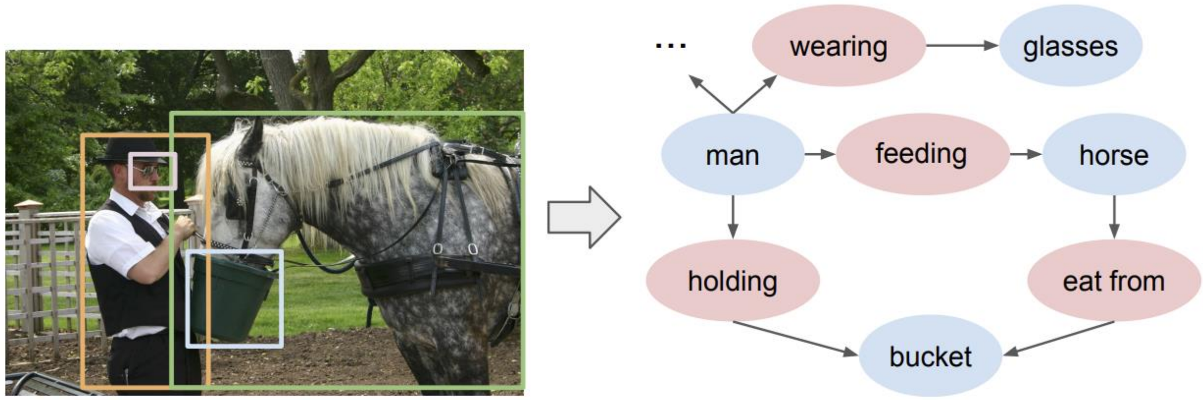
Structured and dense prediction for scene understanding

Image captioning



A. Karpathy, L. Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions. CVPR 2015

Scene graph generation



D. Xu, Y. Zhu, C. Choy, and L. Fei-Fei. Scene Graph Generation by Iterative Message Passing. CVPR 2017

Beyond classification and supervised learning

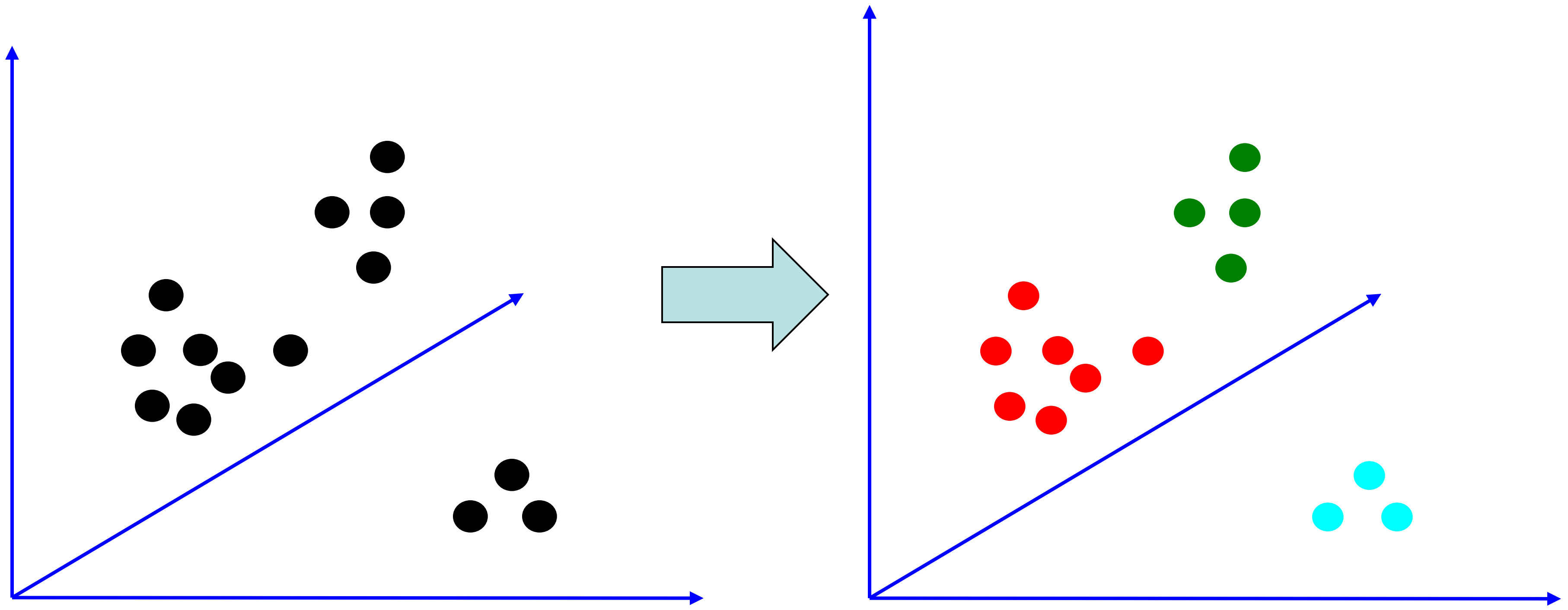
- Other prediction scenarios (output types)
  - Regression
  - Structured prediction
- Other supervision scenarios
  - Unsupervised learning
  - Self-supervised or predictive learning
  - Reinforcement learning
  - Active learning
  - Lifelong learning

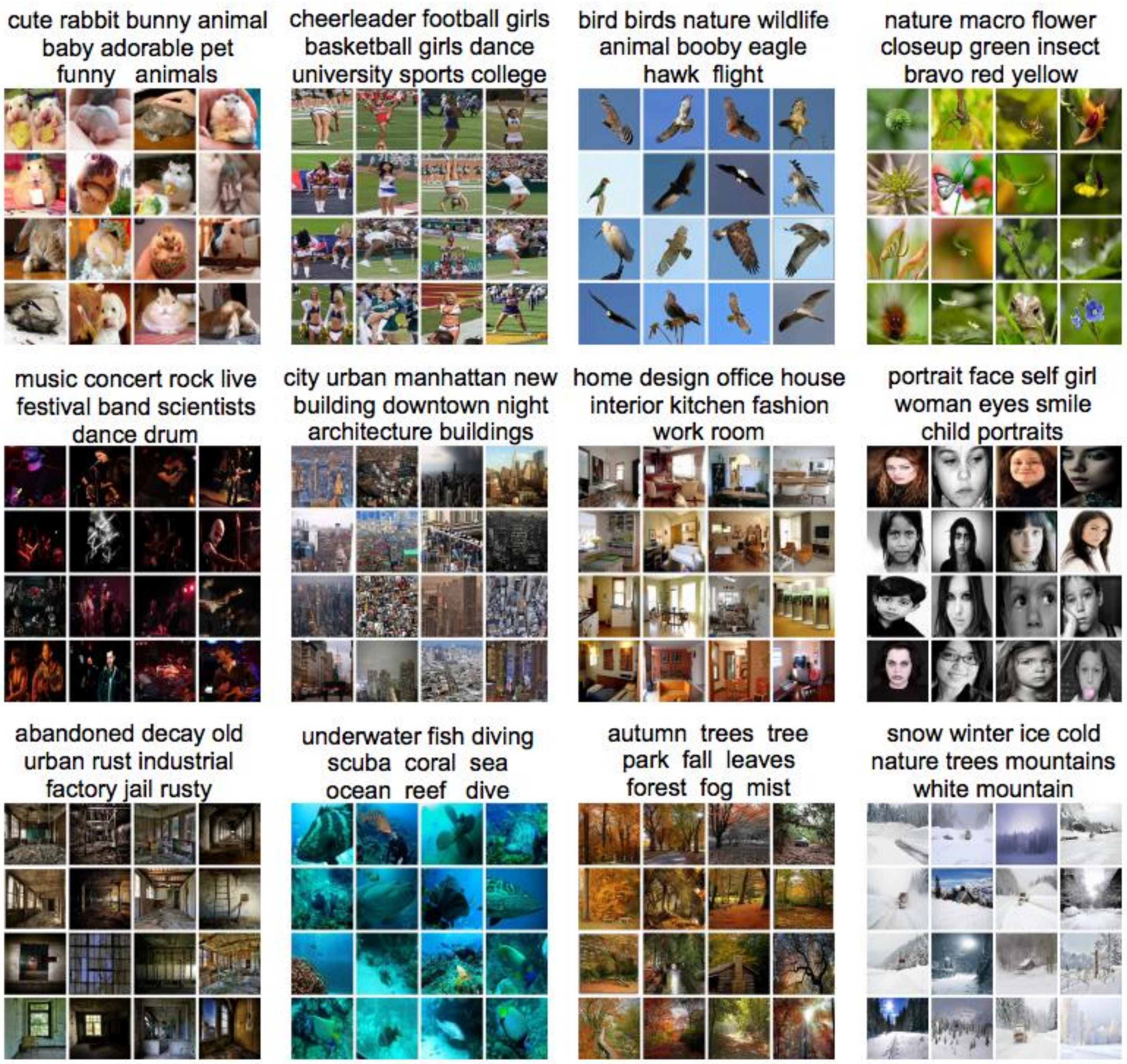
Unsupervised Learning

- **Idea:** Given only *unlabeled* data as input, learn some sort of structure
  - The goal is less clearly defined than in supervised learning
  - Also known as exploratory/descriptive data analysis

Unsupervised Learning

- **Clustering**
  - Discover groups of “similar” data points

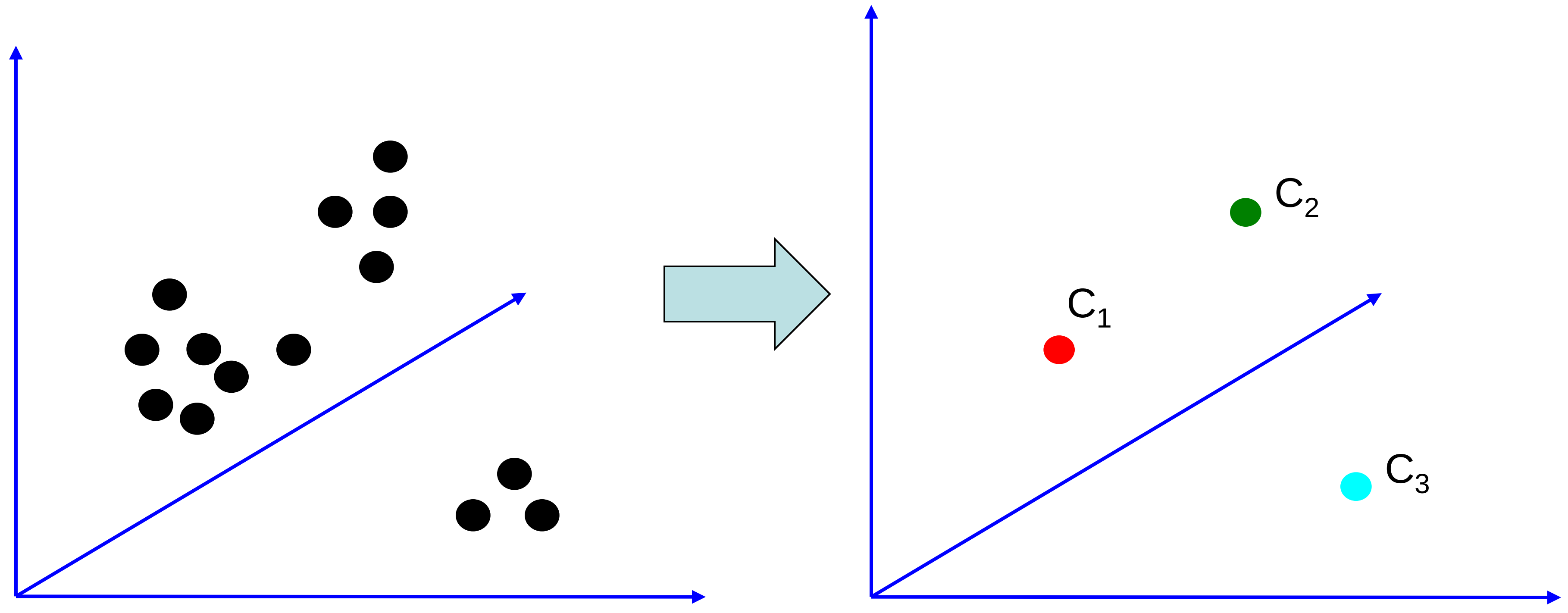




Y. Gong, Q. Ke, M. Isard, and S. Lazebnik. A Multi-View Embedding Space for Modeling Internet Images, Tags, and Their Semantics. IJCV 2014.

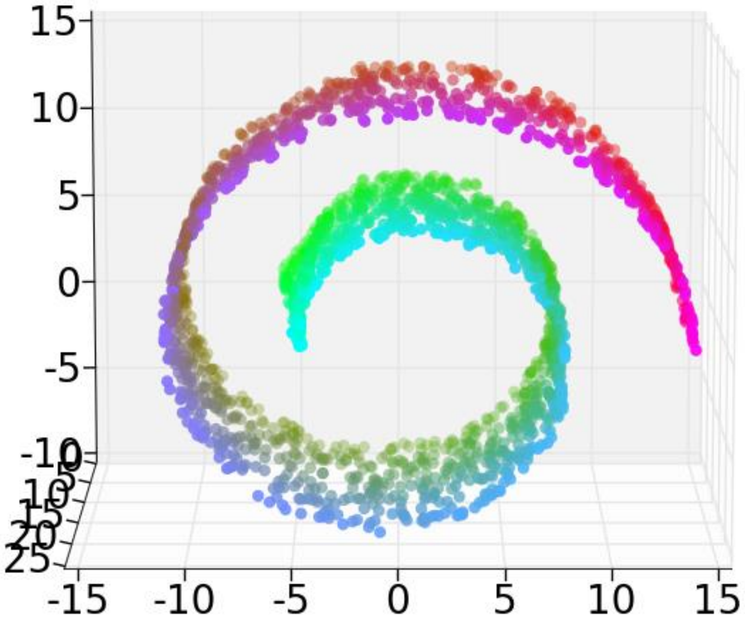
Unsupervised Learning

- **Quantization** or **data compression**
  - Encode the data into a more compact form



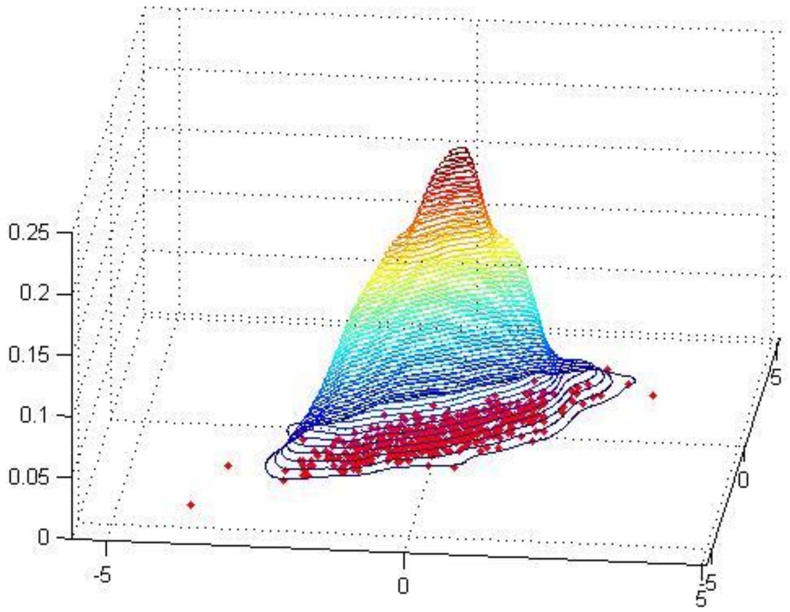
Unsupervised Learning

- **Dimensionality reduction, manifold learning**
  - Discover a lower-dimensional surface on which the data lives



Unsupervised Learning

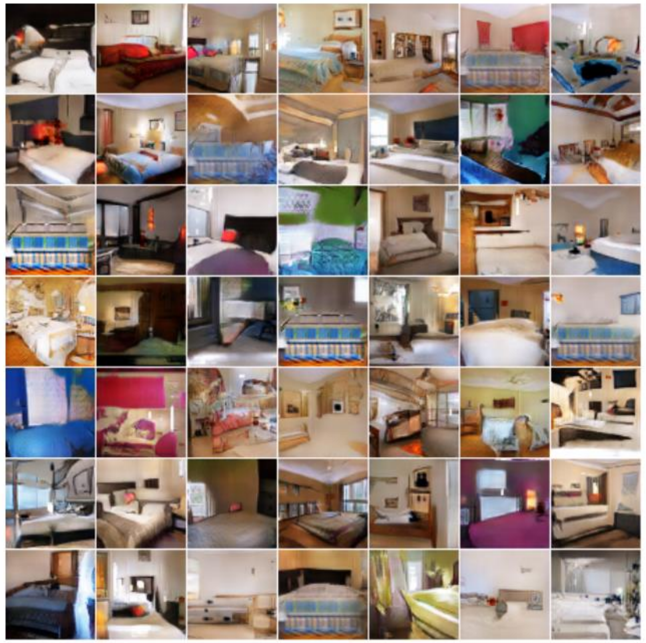
- **Learning the data distribution**
  - Density estimation: Find a function that approximates the probability density of the data (i.e., value of the function is high for "typical" points and low for "atypical" points)
  - Can be used for anomaly detection



Unsupervised Learning

- **Learning the data distribution**
  – Learning to sample: Produce samples from a data distribution that mimics the training set

“Bedroom”



“Face”

Generative adversarial networks

Beyond classification and supervised learning

- Other prediction scenarios (output types)
  - Regression
  - Structured prediction
- Other supervision scenarios
  - Unsupervised learning
    - Clustering and quantization
    - Dimensionality reduction, manifold learning
    - Density estimation
    - Learning to sample

Between "unsupervised" and "fully supervised"

Semi-supervised
(labels for a small portion of training data)

Unsupervised
(no labels)

Weakly supervised
(noisy labels, labels not exactly for the task of interest)

Supervised
(clean, complete training labels for the task of interest)

Beyond classification and supervised learning

- Other prediction scenarios (output types)
  - Regression
  - Structured prediction
- Other supervision scenarios
  - Unsupervised learning
  - Self-supervised or predictive learning
  - Reinforcement learning
  - Active learning
  - Lifelong learning

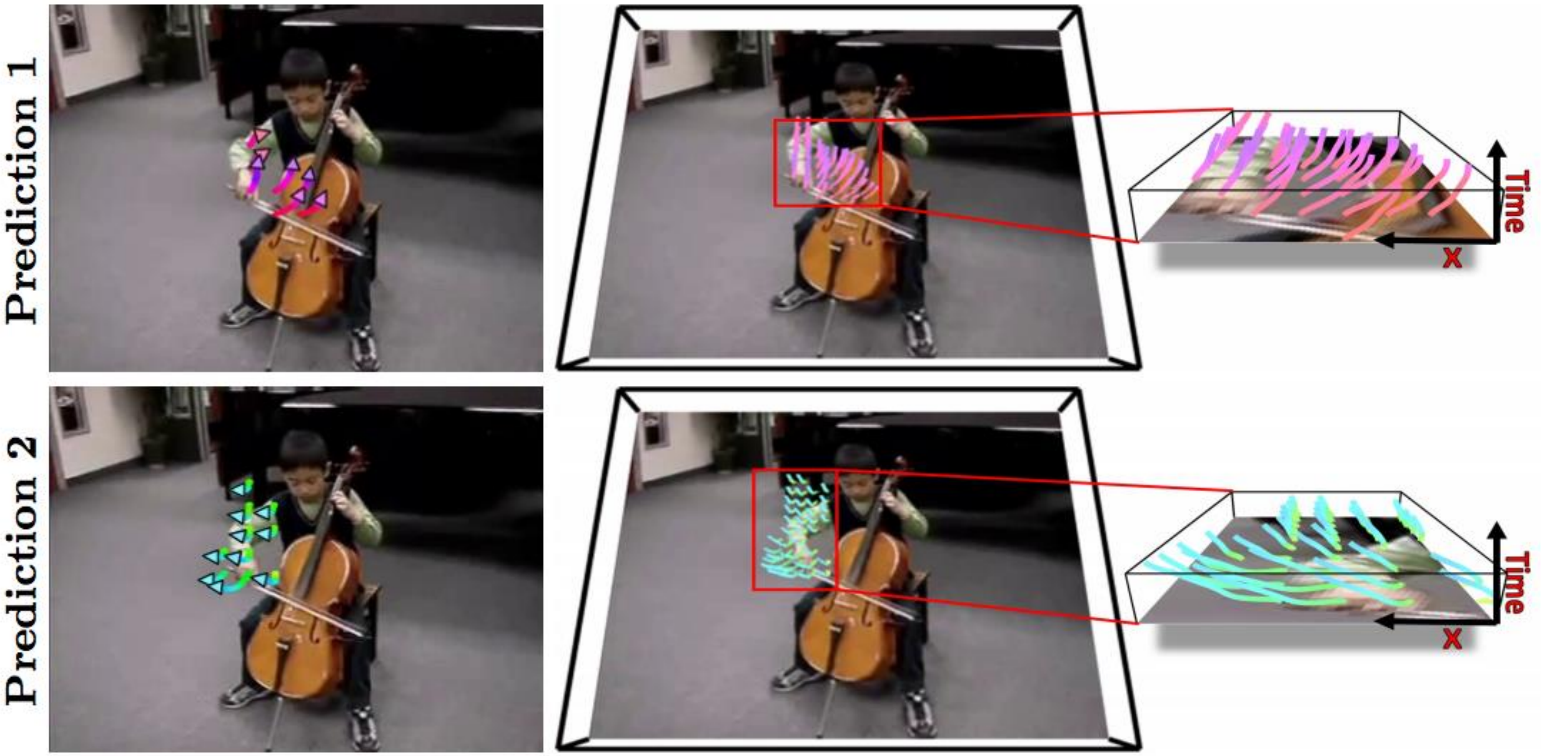
Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Image colorization

R. Zhang et al., [Colorful Image Colorization](), ECCV 2016

Self-supervised or predictive learning

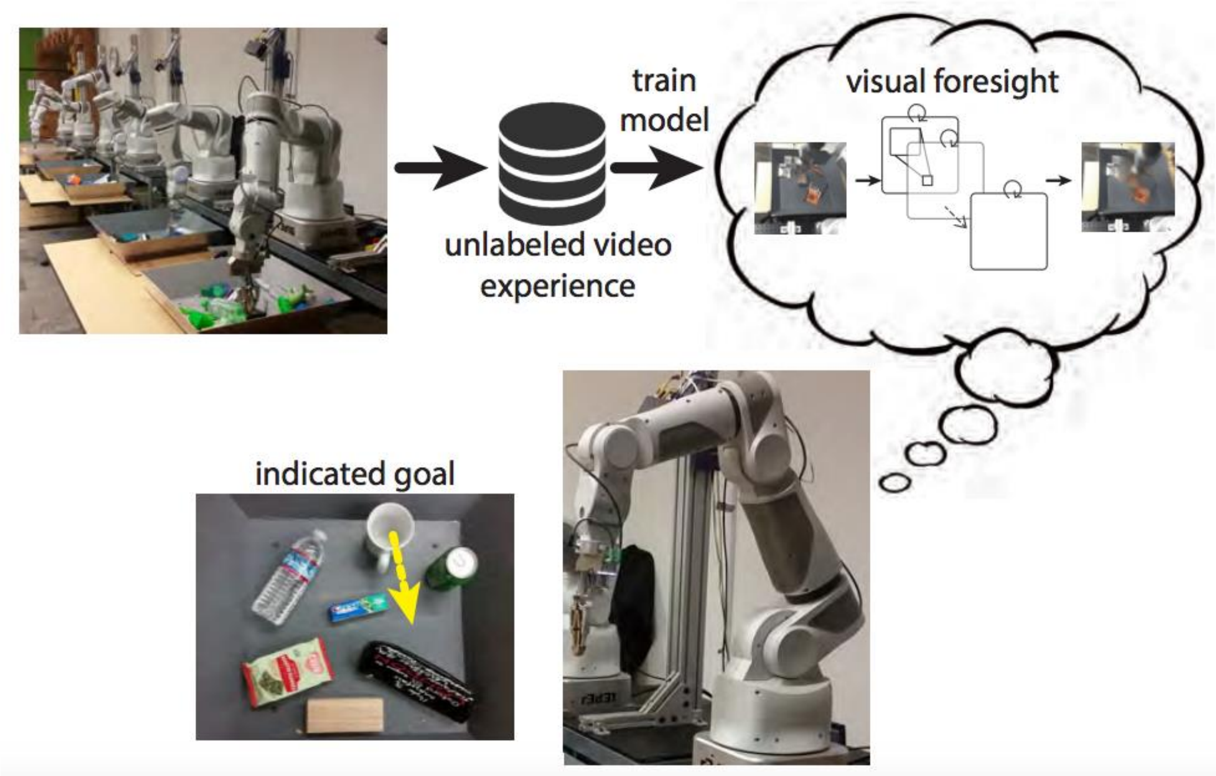
- Use part of the data to predict other parts of the data
  - Example: Future prediction



J. Walker et al. An Uncertain Future: Forecasting from Static Images Using Variational Autoencoders. ECCV 2016.

Self-supervised or predictive learning

- Use part of the data to predict other parts of the data
  - Example: Future prediction



C. Finn and S. Levine. [Deep Visual Foresight for Planning Robot Motion](). ICRA 2017. [YouTube video]()

Reinforcement learning

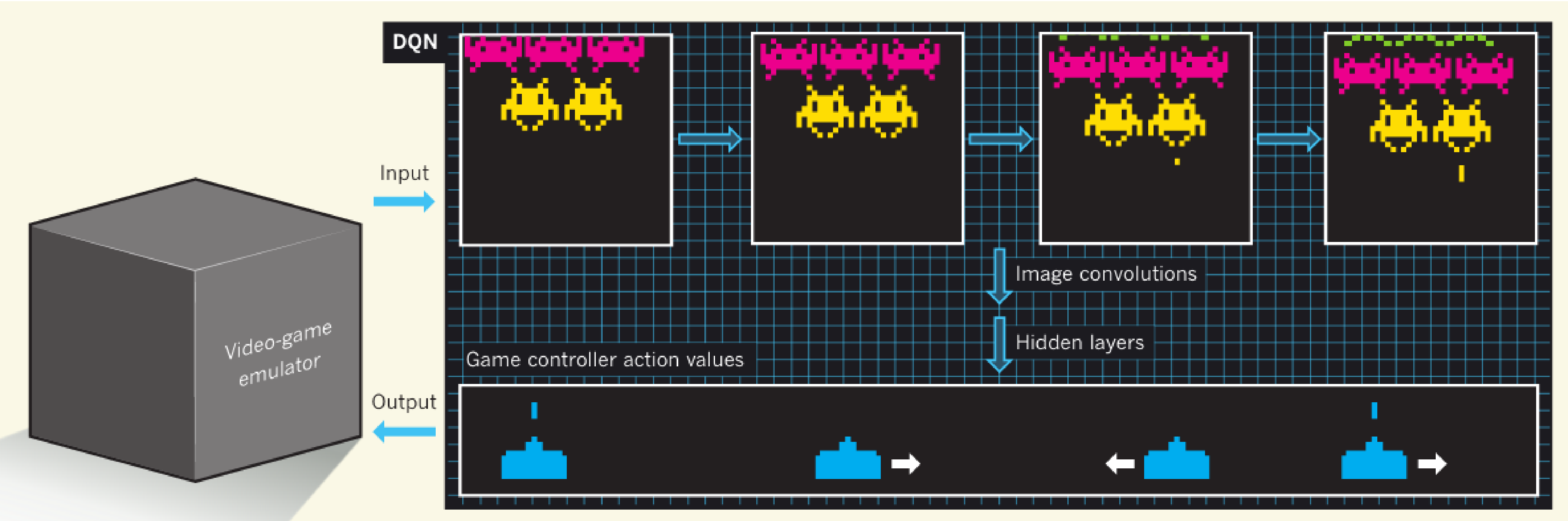
- Learn from rewards in a *sequential* environment

Arthur Samuel

AlphaGo

Reinforcement learning

- Playing Atari with deep reinforcement learning



Breakout video

V. Mnih et al., *Nature*, February 2015

Reinforcement learning

- Learn from rewards in a *sequential* environment

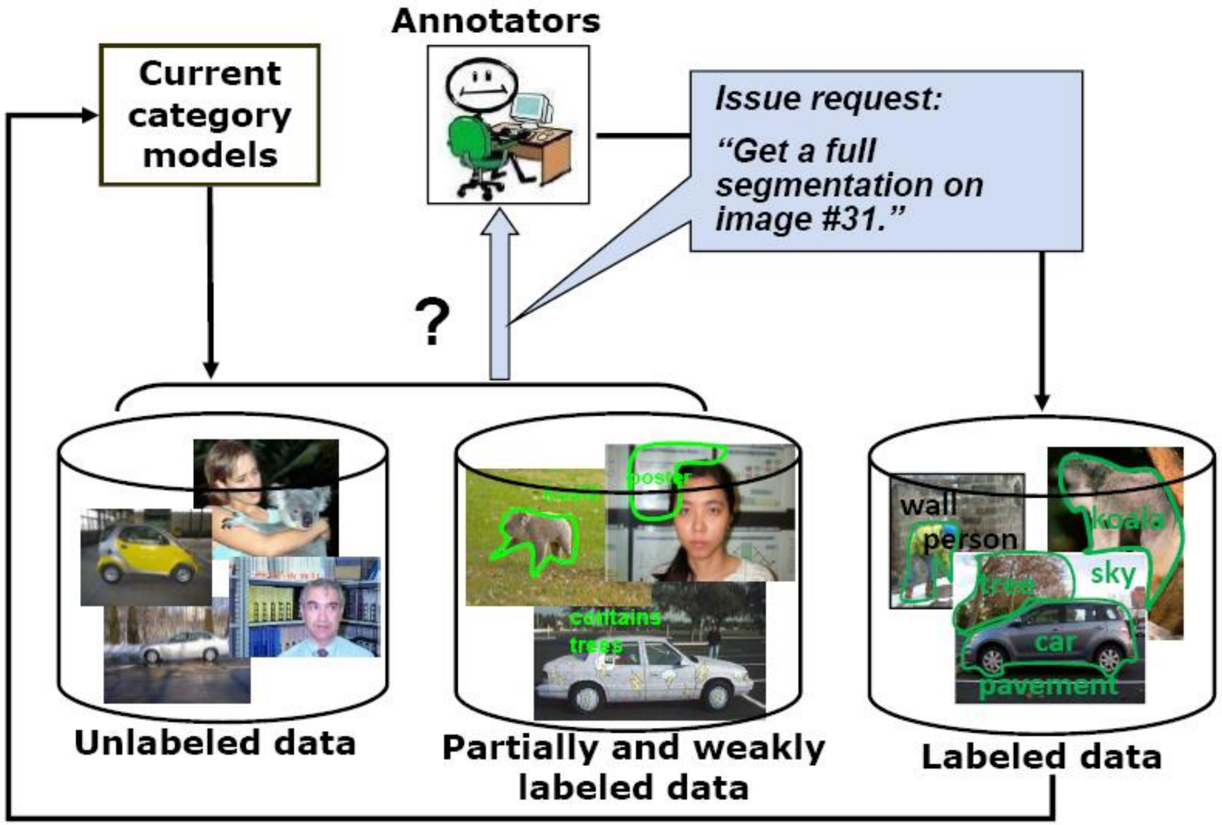
Initial gait

Learned gait

N. Kohl and P. Stone. Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion. ICRA 2004

Active learning

- The learning algorithm can choose its own training examples, or ask a “teacher” for an answer on selected inputs



S. Vijayanarasimhan and K. Grauman. Cost-Sensitive Active Visual Category Learning. IJCV 2010

Lifelong learning

Read the Web

Research Project at Carnegie Mellon University

Home | Project Overview | Resources & Data | Publications | People

NELL: Never-Ending Language Learning

Can computers learn to read? We think so. "Read the Web" is a research project that attempts to create a computer system that learns over time to read the web. Since January 2010, our computer system called NELL (Never-Ending Language Learner) has been running continuously, attempting to perform two tasks each day:

- First, it attempts to "read," or extract facts from text found in hundreds of millions of web pages (e.g., `playsInstrument(George_Harrison, guitar)`).
- Second, it attempts to improve its reading competence, so that tomorrow it can extract more facts from the web, more accurately.

So far, NELL has accumulated over 50 million candidate beliefs by reading the web, and it is considering these at different levels of confidence. NELL has high confidence in 2,033,557 of these beliefs — these are displayed on this website. It is not perfect, but NELL is learning. You can track NELL's progress below or @cmunell on Twitter, browse and download its knowledge base, read more about our technical approach, or join the discussion group.

http://rtw.ml.cmu.edu/rtw/

NEIL: Never Ending Image Learner

I Crawl, I See, I Learn.

WHAT COMMON SENSE FACTS HAVE NEIL LEARNED?

Here are a few examples:

Airbus_330 can be a kind of / look similar to Airplane.

Deer can be a kind of / look similar to Antelope.

Car can have a part Wheel.

Airbus_330 can have a part Airplane_nose.

Leaning_tower can be found in Pisa.

Zebra can be found in Savanna.

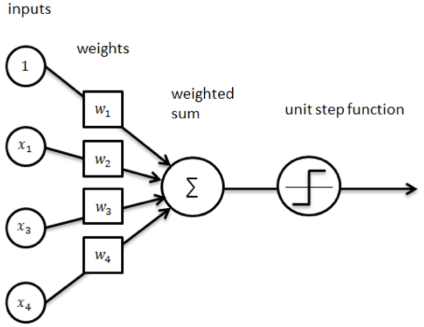
Xinlei Chen, Abhinav Shrivastava and Abhinav Gupta. [NEIL: Extracting Visual Knowledge from Web Data](). In ICCV 2013

Review: Beyond classification and supervised learning

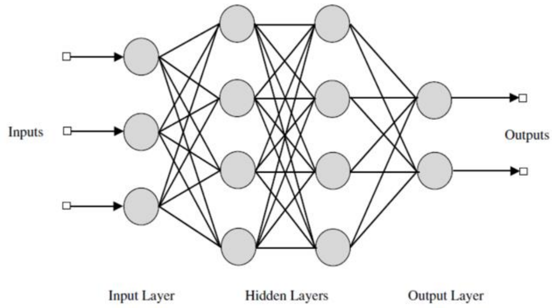
- Other prediction scenarios
  - Regression
  - Structured prediction
- Other supervision scenarios
  - Unsupervised learning
  - Self-supervised or predictive learning
  - Reinforcement learning
  - Active learning
  - Lifelong learning

In this class

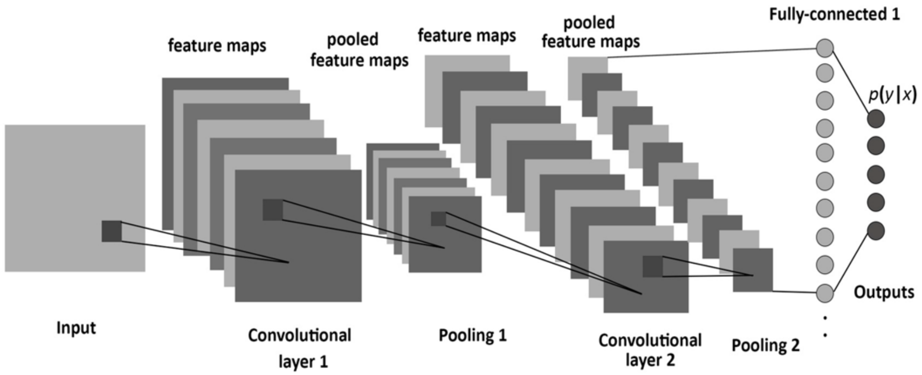
ML basics, linear classifiers



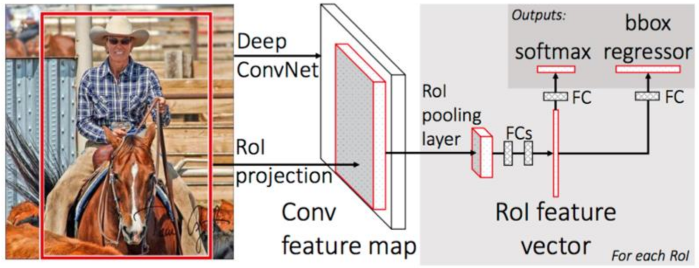
Multilayer neural networks, backpropagation



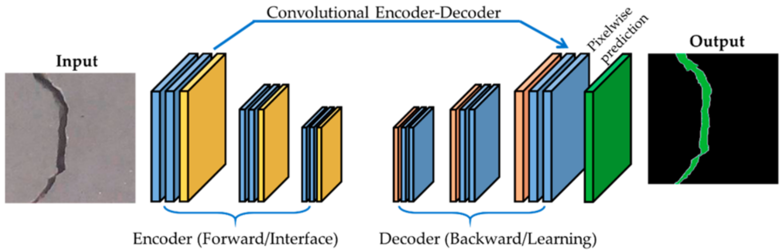
Convolutional networks for classification



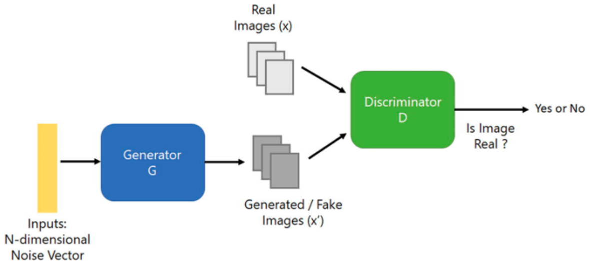
Networks for detection



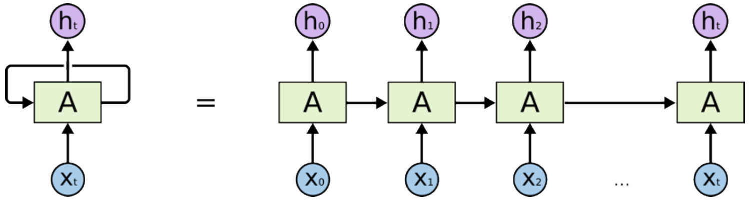
Networks for dense prediction



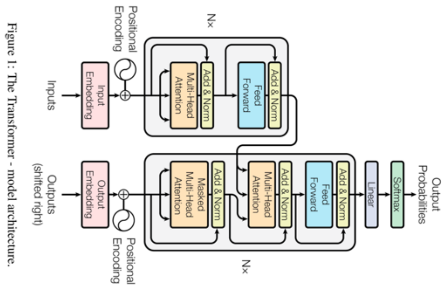
Generative models (GANs, VAEs)



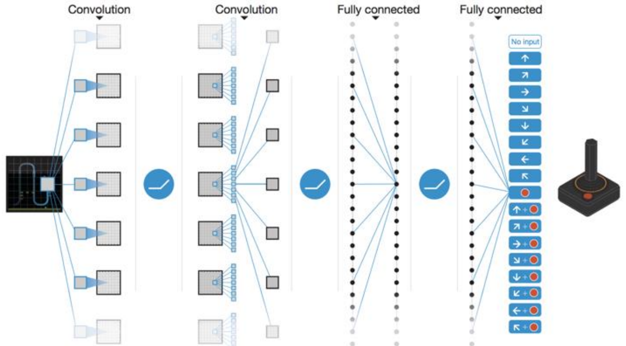
Recurrent models



Transformers



Deep reinforcement learning



Acknowledgement

- Deep Learning, Stanford University
- Introduction to Deep Learning, University of Illinois at Urbana-Champaign
- Introduction to Deep Learning, Carnegie Mellon University
- Convolutional Neural Networks for Visual Recognition, Stanford University
- Natural Language Processing with Deep Learning, Stanford University