Statistical Machine Learning for Data Science

Zaid Harchaoui (Instructor)
Corinne Jones (Teaching assistant)

Spring 2019

UNIVERSITY of WASHINGTON

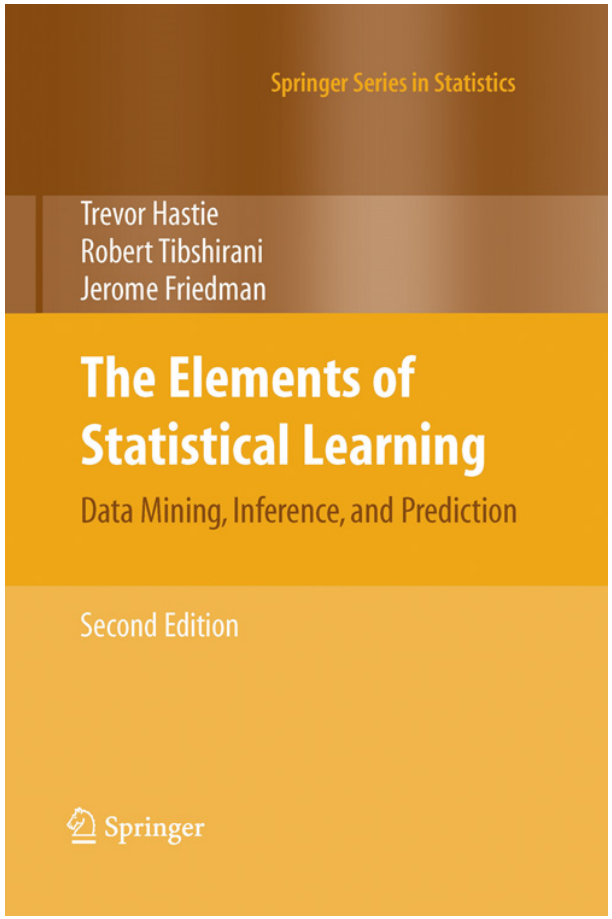
Outline

Outline

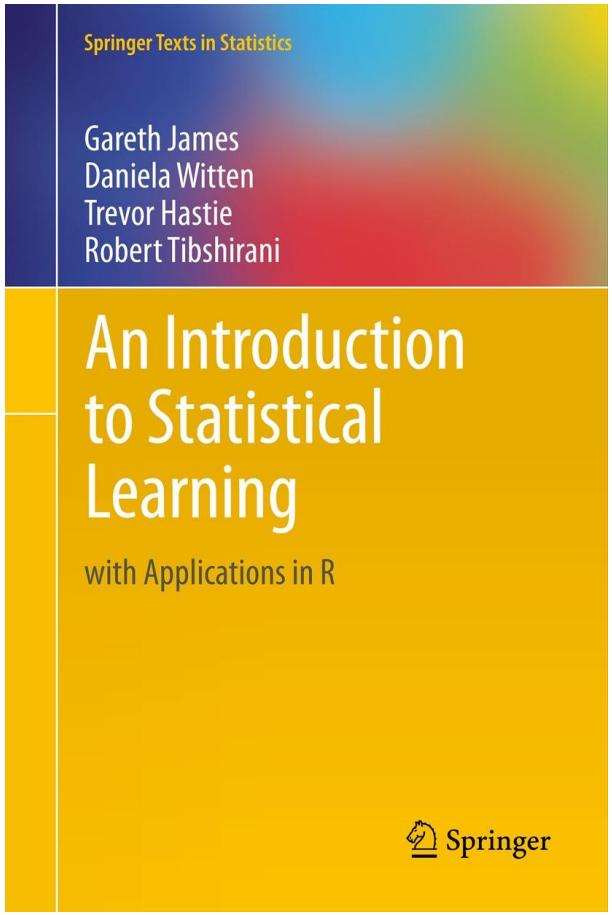
Statistical Machine Learning for Data Science, DATA 558

- Canvas page :
  `https://canvas.uw.edu/courses/1325582`
- Syllabus :
  `https://canvas.uw.edu/courses/1325582/assignments`
- Discussions :
  `https://canvas.uw.edu/courses/1325582/discussion_topics`
- No regular emails please.
  Please use Canvas discussion or email $\rightarrow$ faster response.

Textbook : The Elements of Statistical Learning : Data Mining, Inference, and Prediction



Textbook :
An Introduction to Statistical Learning, with applications in R



Tentative schedule

- Week 1 - Introduction. What is statistical learning ?
- Week 2 - L2-regularized least-squares. Gradient Descent
- Week 3 - L2-regularized logistic regression. Accelerated Gradient Method. L1-regularized models. Coordinate descent.
- Week 4 - L1-regularized models. Coordinate descent. Variable selection. Dimension reduction. Clustering.
- Week 5 - Variable selection. Dimension reduction. Clustering.

Tentative schedule

- Week 6 - Midterm : In-class (math) + Take-home (coding).
- Week 7 - Dimensions of learning. Large-scale optimization. Guest lecture.
- Week 8 - Generalized additive models. Performance metrics. Guest lecture.
- Week 9 - Kernel-based methods. Outlier detection. Change detection.
- Week 10 - Tree-based models. Matrix factorization. Deep learning.

Evaluation

- About 8 Homework assignments.
- Class participation : public release of polished codes.
- Midterm Exam (week 6 or week 7). In-class (exercises) + Take-home (programming exercises).
- Kaggle competition 1 : small-scale image categorization.
- Kaggle competition 2 : large-scale image categorization.
- No final exam.

Homework assignments

- Assignment out after each lecture on Wednesday.
- Due the following Thursday midnight.
- Submit through Canvas.
- Late homework will receive 0 points.

Philosophy of the course

- Write your own codes from scratch in R and Python.
- Reach out for help on Canvas is you encounter pure scripting/programming issues.

General Philosophy

- Mastery vs Performance
- Cheerful learning and do it yourself vs "getting the grade"

Questions ?

- What are you looking to get out of the course ?
- Questions for me ?

Outline

Statistical learning : a tentative big picture

Unsupervised learning (*learning without a teacher*)

- Find structure of $\mathbf{x} \in \mathcal{X}$, given observations $\mathbf{x}_i,\ i = 1, ..., n$

Supervised learning (*learning with a teacher*)

- Predict $y \in \mathcal{Y}$ from $\mathbf{x} \in \mathcal{X}$, given observations $(\mathbf{x}_i, y_i),\ i = 1, ..., n$

Statistical machine learning : a tentative big picture

Applications in many fields

- Biostatistics and bioinformatics
- Computer vision
- Bioinformatics
- Audio/speech processing
- Text mining
- Astrostatistics
- etc.

Interplays

- interplay between computer science and data science, towards AI
- interplay between theory, algorithms, and real-world applications

A blend of scientific disciplines

A blend of scientific disciplines

- data science : statistics, machine learning
- computer science : algorithmics, complexity, information theory
- applied mathematics : matrix analysis, convex optimization, probability

Application in Domain Sciences

- biology
- astrophysics
- oceanography
- neuroscience

Unsupervised learning

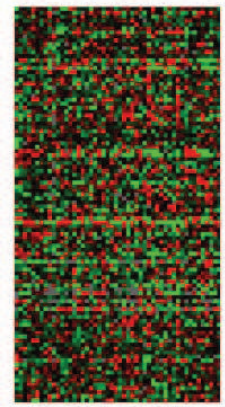
Dimension reduction

face images

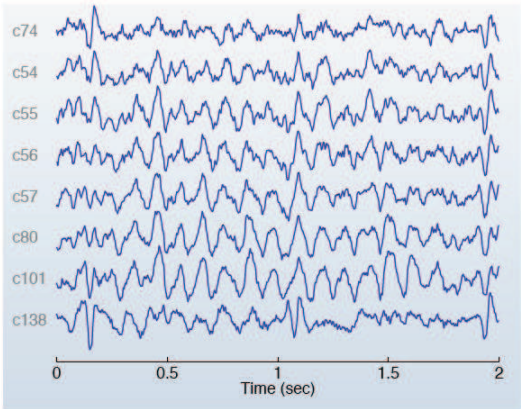
Zambian President Levy Mwanawasa has won a second term in office in an election his challenger Michael Sata accused him of rigging, official results showed on Monday.

According to media reports, a pair of hackers said on Saturday that the Firefox Web browser, commonly perceived as the safer and more customizable alternative to market leader Internet Explorer, is critically flawed. A presentation on the flaw was shown during the ToorCon hacker conference in San Diego.

documents



gene expression data



MEG readings

Unsupervised learning

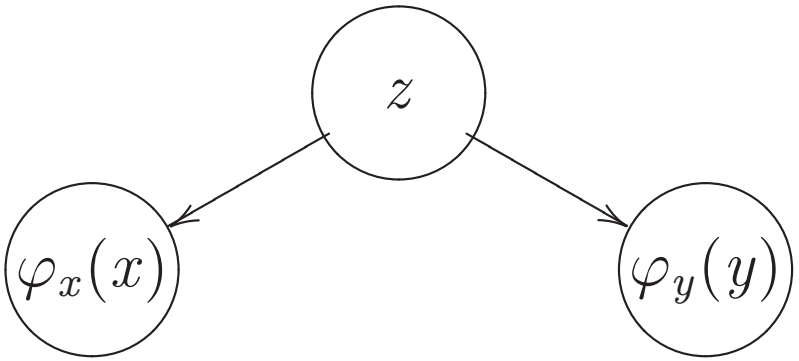
Dimension reduction

- Computational efficiency : space and time savings
- Statistical performance : fewer dimensions $\rightarrow$ regularization
- Visualization : discover underlying structure of the data

$\rightarrow$ Principal Component Analysis, Sparse Principal Component Analysis, Non-negative Matrix Factorization

Unsupervised learning

Feature extraction



x:

y: "A view from Idyllwild, California, with pine trees and snow capped Marion Mountain under a blue sky."

Unsupervised learning

Feature extraction

- Multimodality : leverage the correlation between the modalities
- Statistical performance : take advantage of both views of the data
- Putting in relation : discover underlying relations between the modalities

$\rightarrow$ Canonical Correlation Analysis, Sparse Canonical Correlation Analysis

Unsupervised learning

Clustering

Unsupervised learning

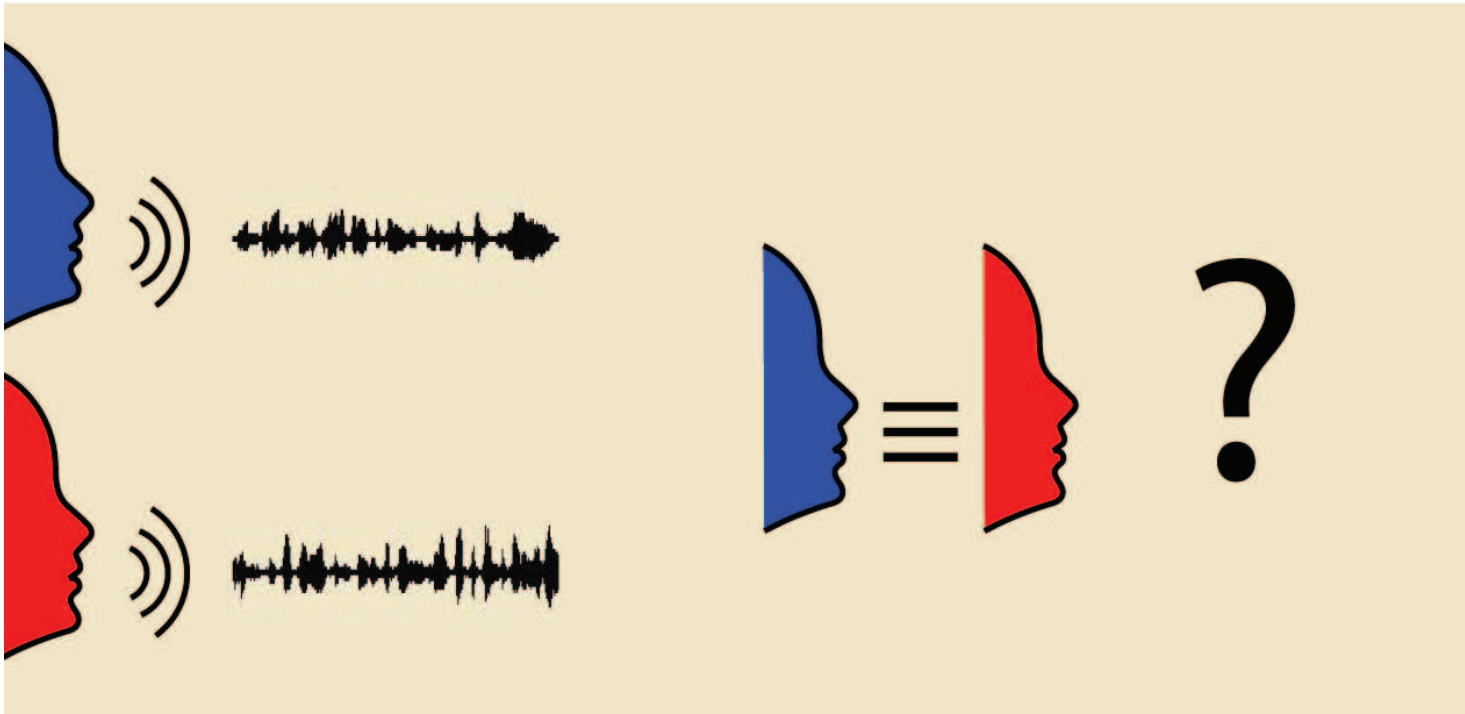
Clustering

- Semantics : grouping datapoints in meaningful clusters
- Statistical performance : intrinsic degrees of freedom of the data
- Visualization : discover groupings between datapoints

$\rightarrow$ k-means, spectral clustering

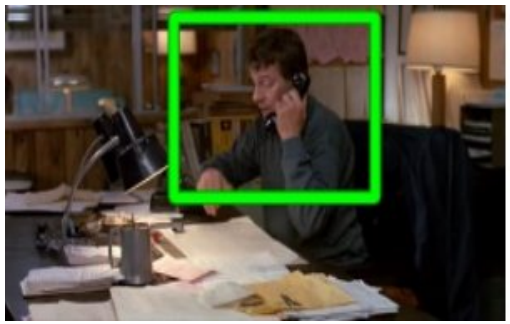
Unsupervised learning

Detection problems



Supervised learning

Human action recognition



Supervised learning

Image classification and scene understanding

- Variance : high intra-class variability
- Structure : spatial and temporal structure
- Unknowns : unknown localization of the object of interest

$\rightarrow$ ridge regression (regularized least-squares), regularized logistic regression, support vector machine

Image categorization/classification

Hierarchy of classes:

Deng, Dong, Socher, Li, Li and Fei-Fei, "Imagenet: a large-scale hierarchical image database", CVPR'09.

Fine-grained subsets: generally more practical problems

→ Fungus: 134 classes, 90K images

Image categorization/classification

Hierarchy of classes:

Deng, Dong, Socher, Li, Li and Fei-Fei, "Imagenet: a large-scale hierarchical image database", CVPR'09.

Fine-grained subsets: generally more practical problems

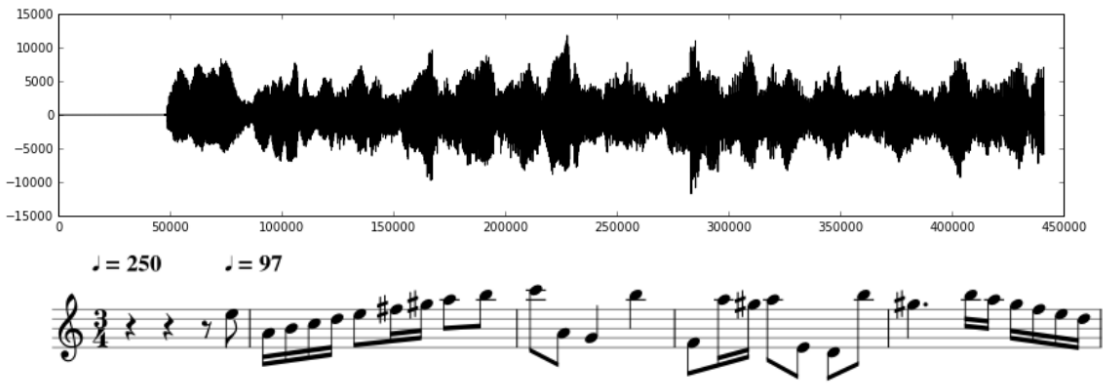
→ Vehicle: 262 classes, 226K images

Music

Music transcription

1. Which note is being played ?
2. When is it played ?
3. By what instruments ?

MusicNet



Music

Music transcription

1. Which note is being played ?
2. When is it played ?
3. By what instruments ?

Challenges

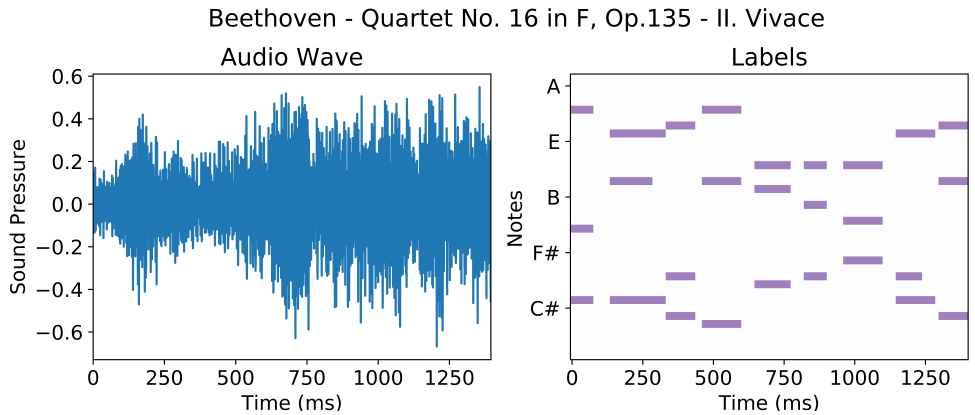
1. Difficult for “musically rich” music pieces
2. Metric, polyphonic, structures

Music

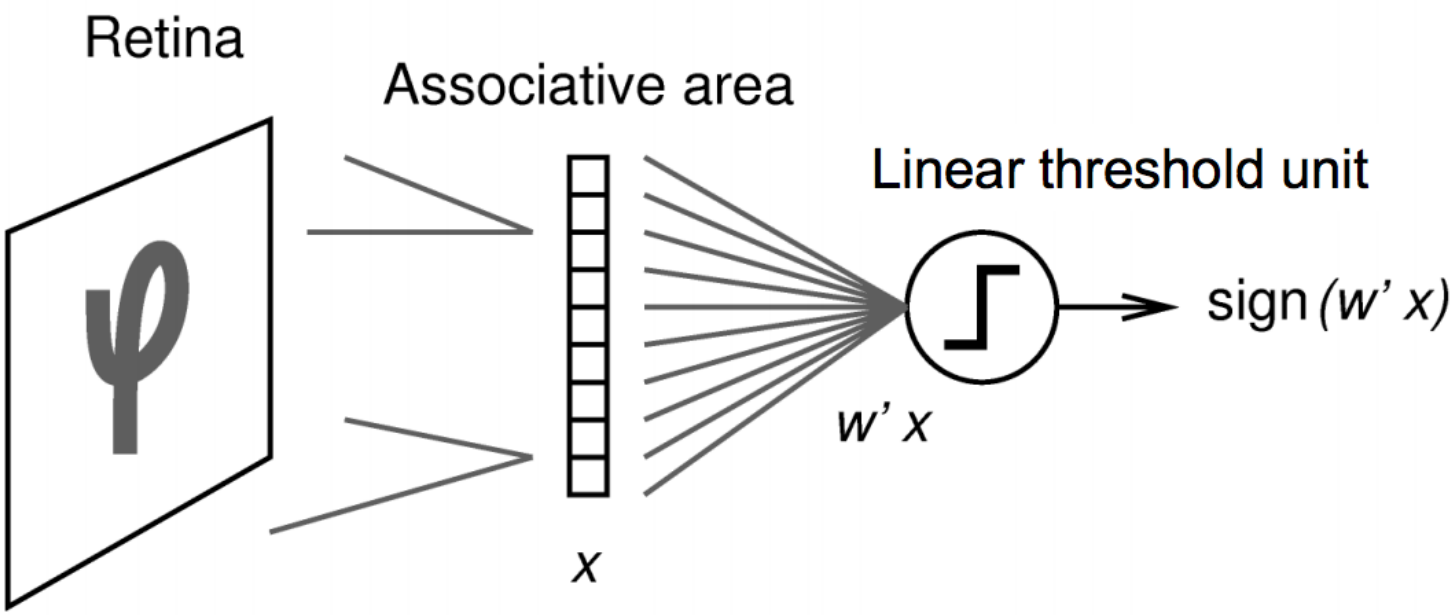
Music transcription

1. Which note is being played ?
2. When is it played ?
3. By what instruments ?

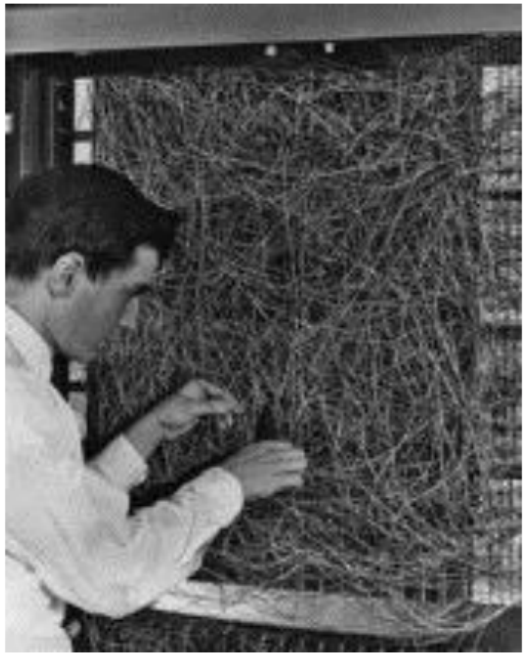
MusicNet



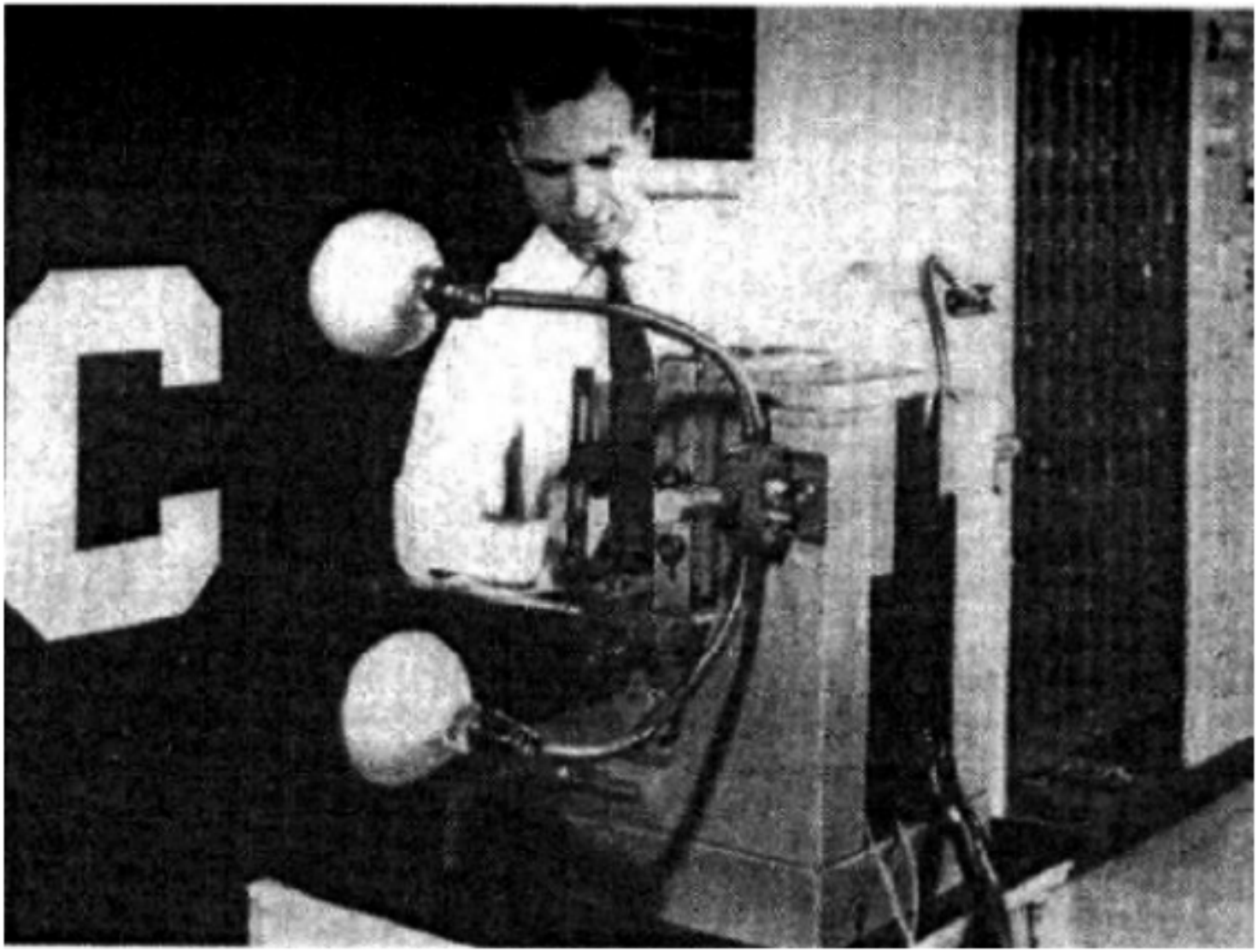
Artificial Learning Machines (1940-1950s)



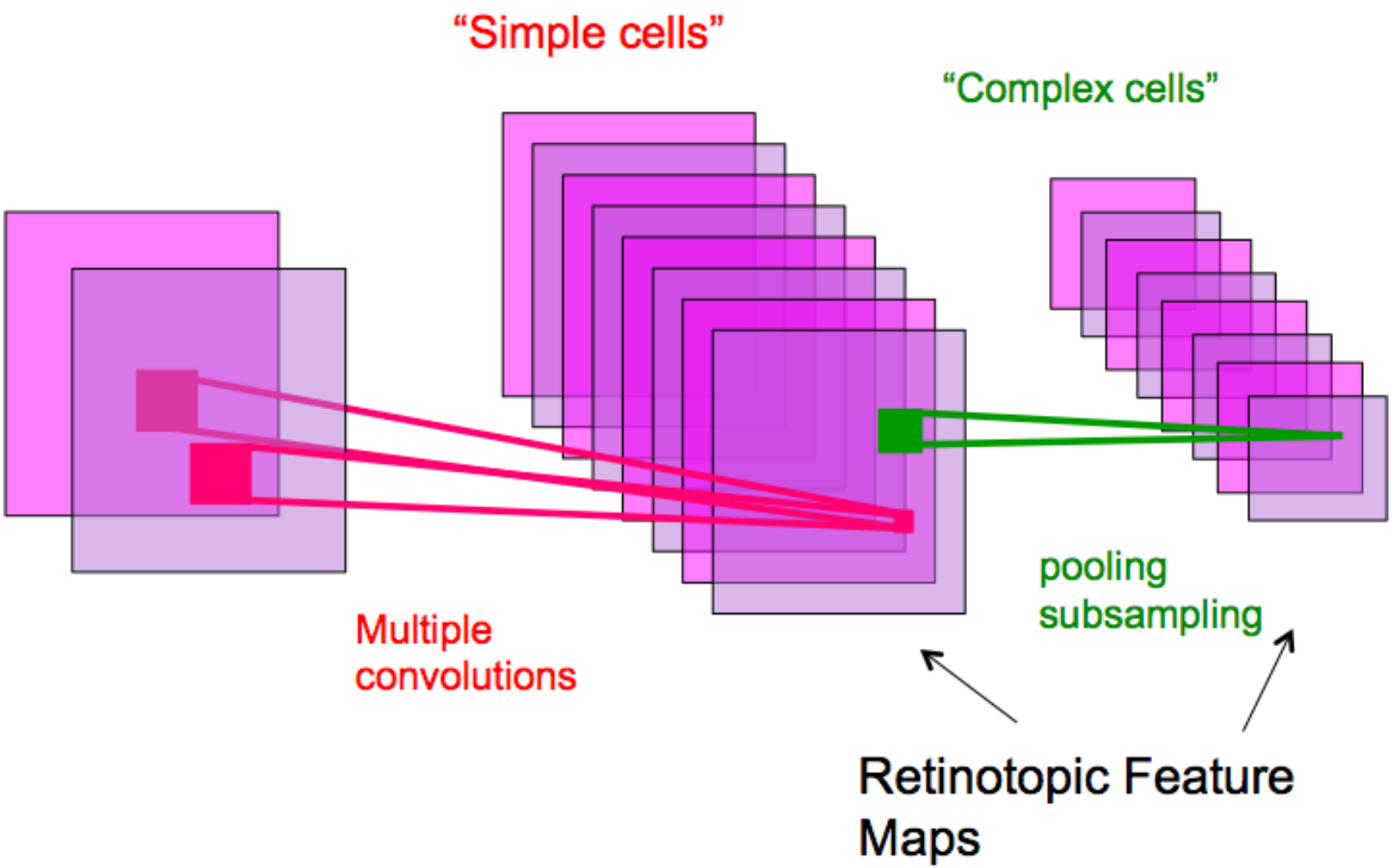
Perceptron (Rosenblatt, 1957)



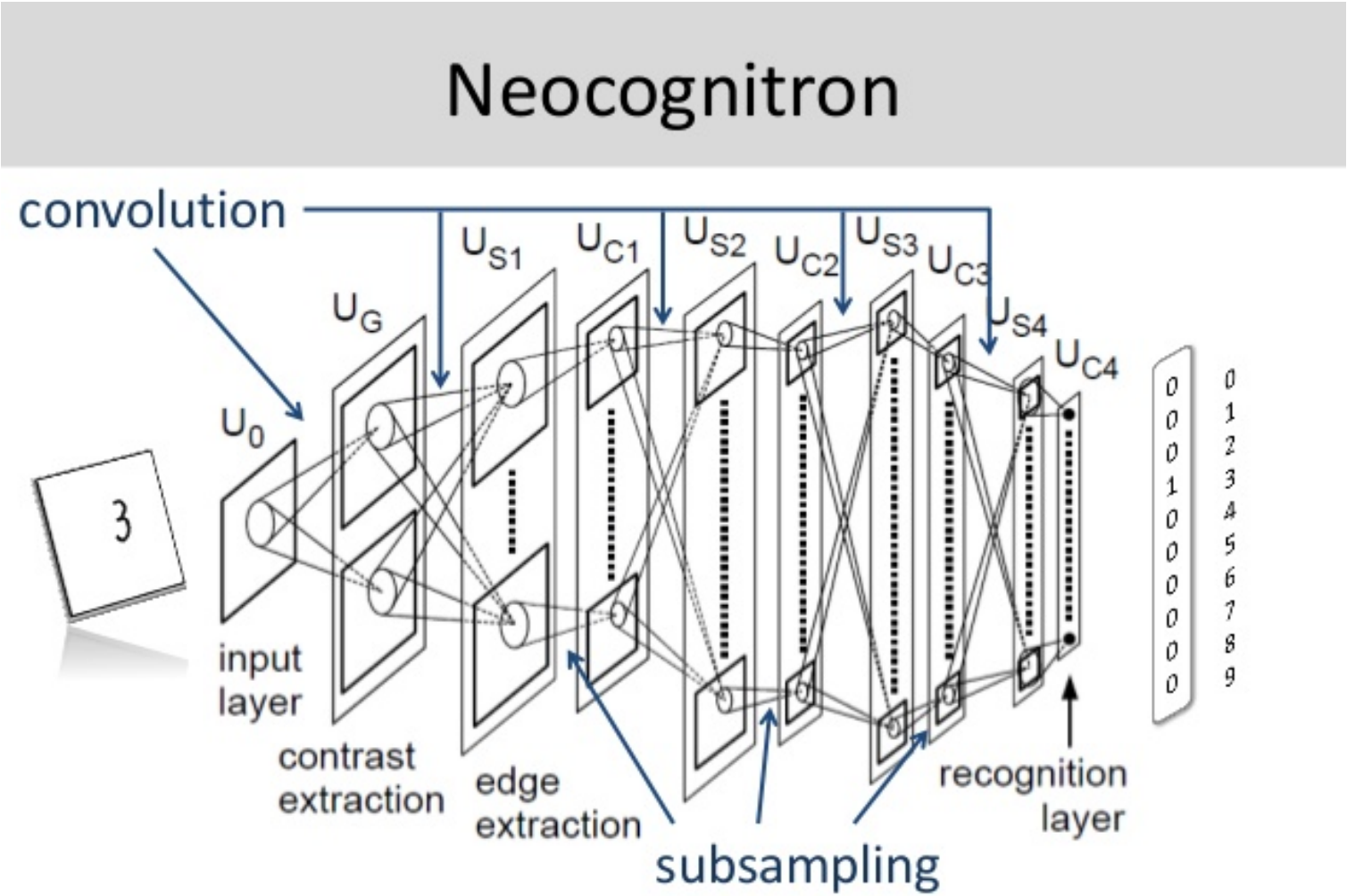
Perceptron (Rosenblatt, 1957)



Convolutional Neural Nets : origins



Neocognitron (Fukushima, 1980)



Convolutional Neural Net (LeCun, 1988)

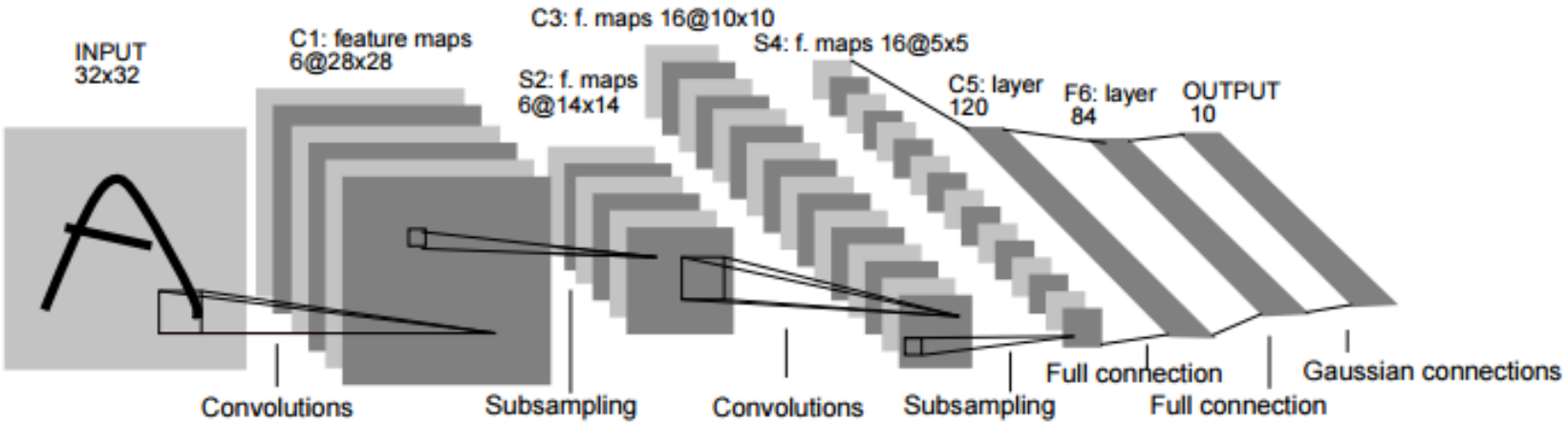
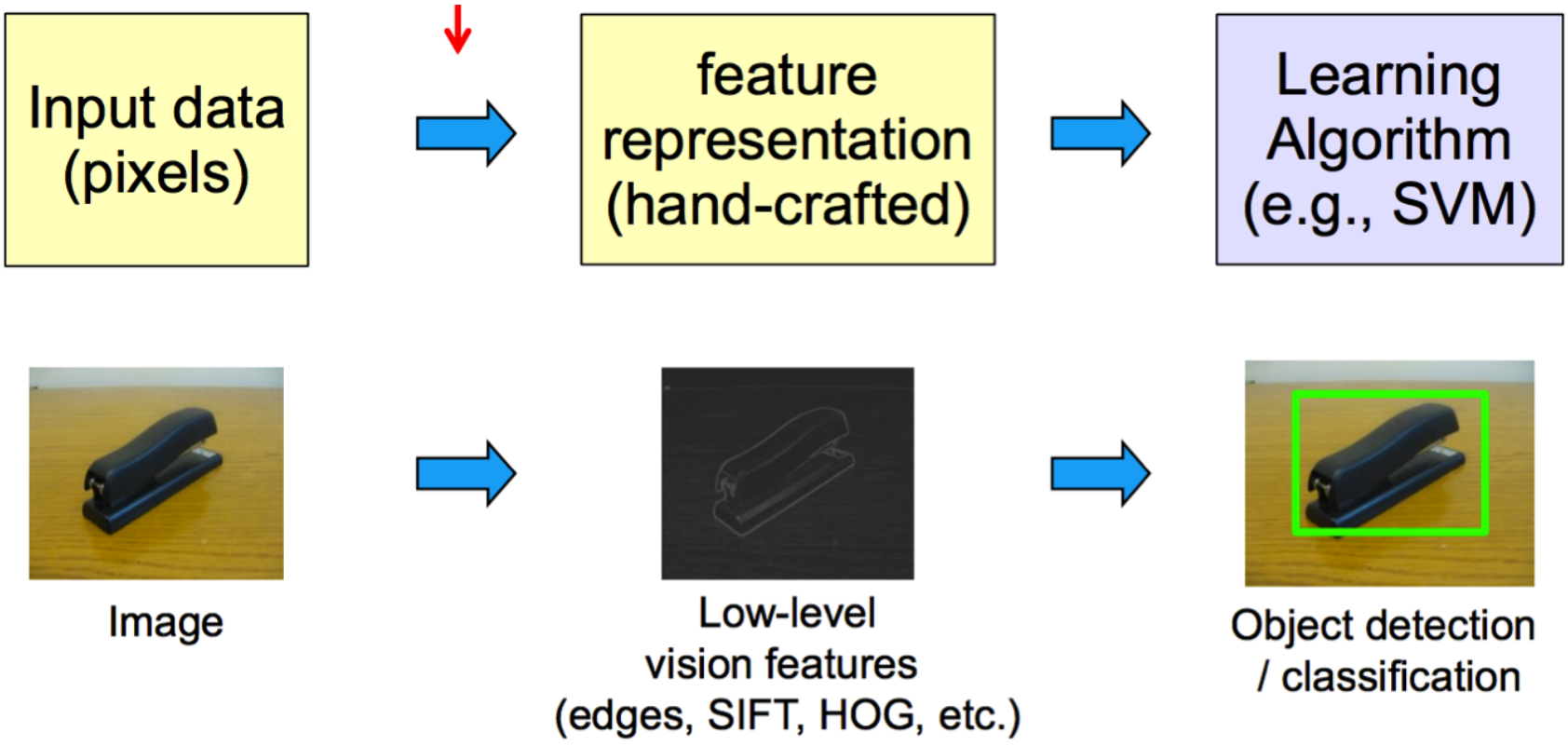


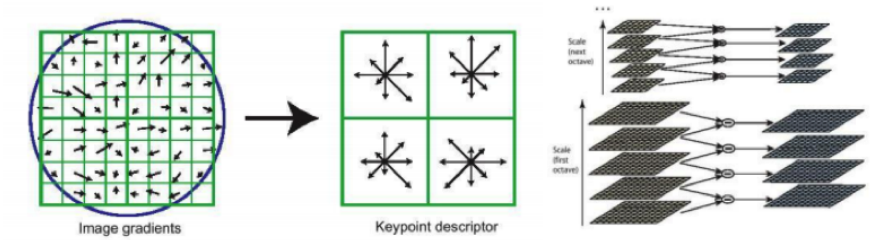
Figure: Picture from (LeCun, 1998)

- CNNs perform "simple" operations such as convolutions, point-wise non-linearities and subsampling.
- for most successful applications of CNNs, training is supervised.

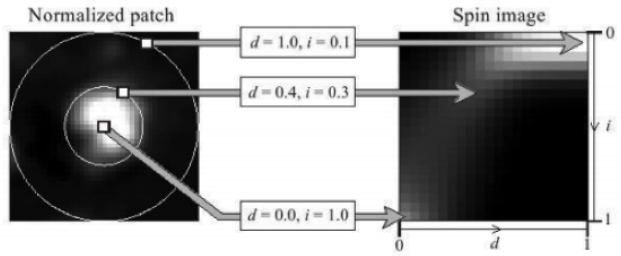
Visual recognition : traditional approach



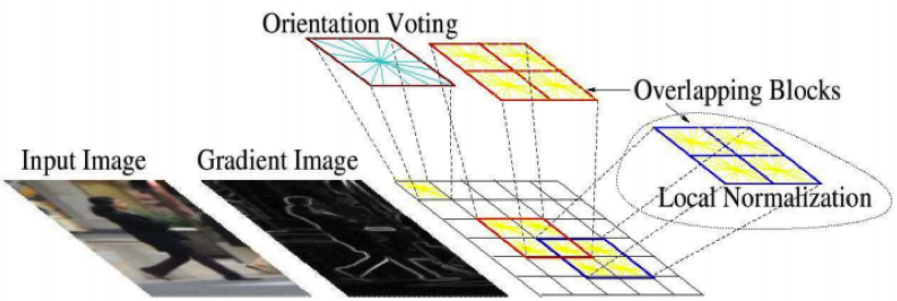
Visual recognition : traditional features



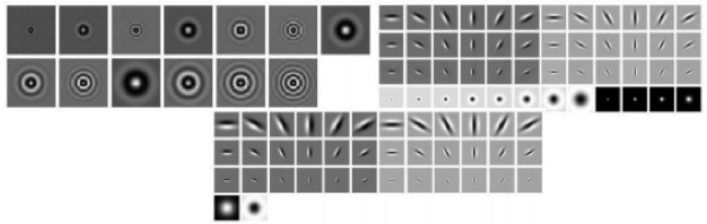
SIFT



Spin image



HoG

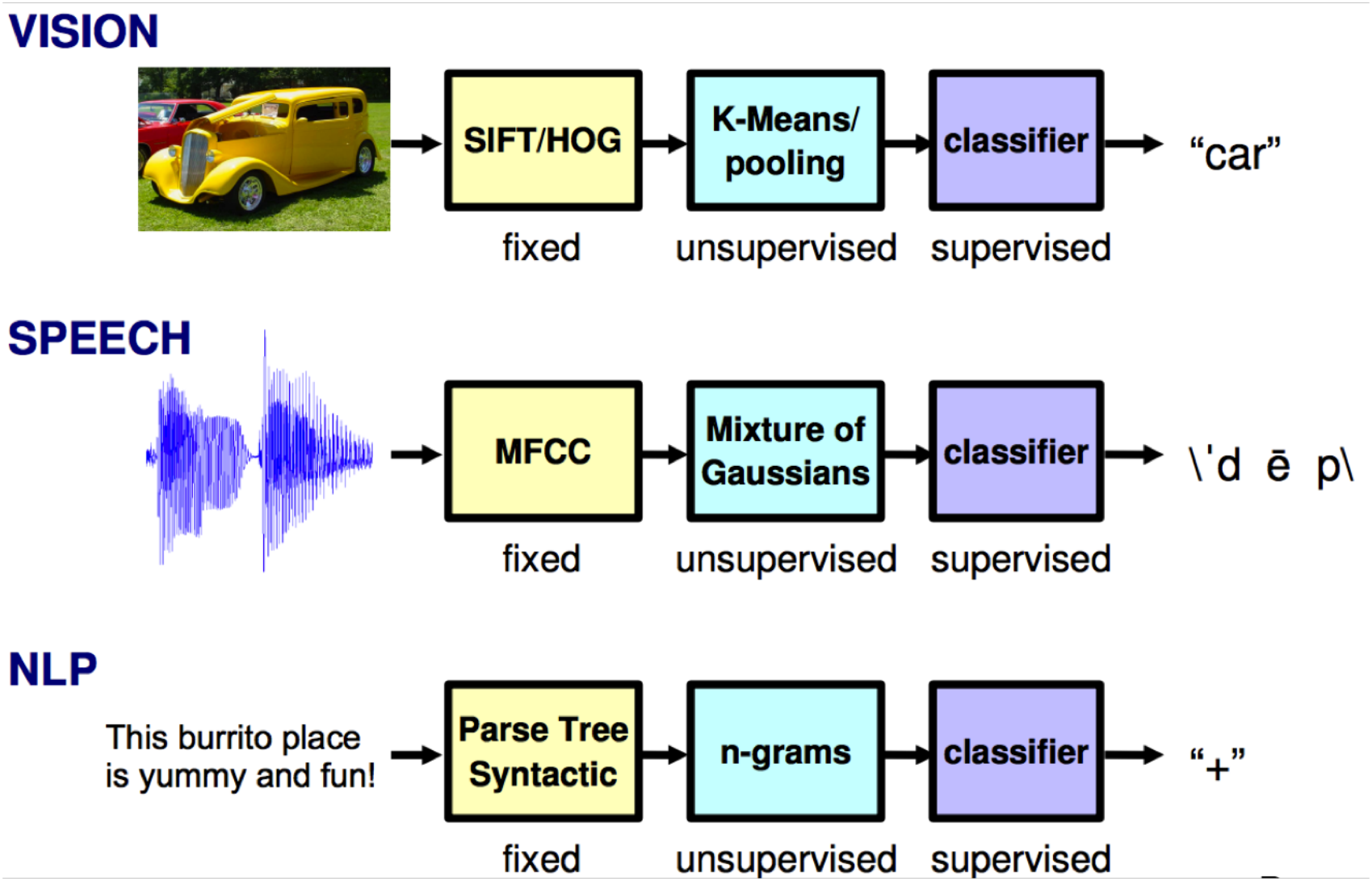


Textons

and many others:

SURF, MSER, LBP, Color-SIFT, Color histogram, GLOH, ...

Parallel approaches in vision, speech, and NLP



Challenges in vision, speech, and NLP

VISION

pixels → edge → texton → motif → part → object

SPEECH

sample → spectral band → formant → motif → phone → word

NLP

character → word → NP/VP/.. → clause → sentence → story

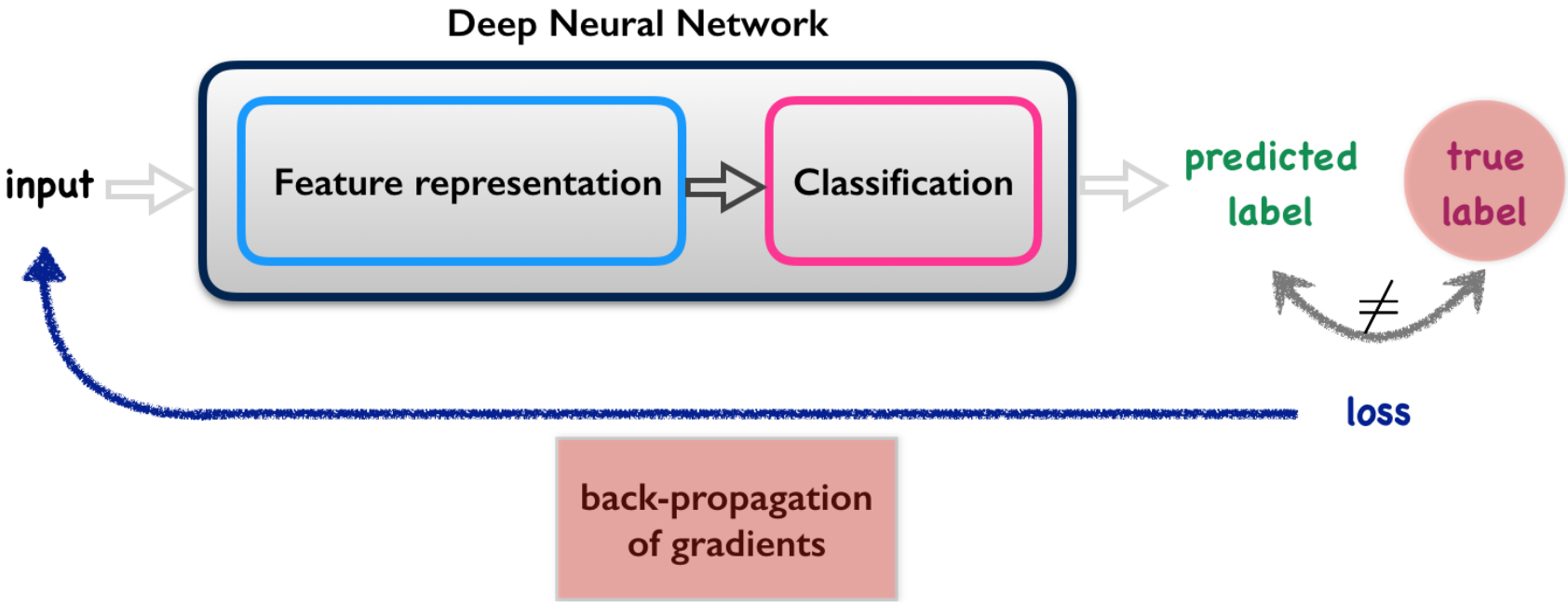
Recent popular strategy : learning all the way through

Recent popular strategy : learning all the way through

1. Low-level feature representation is learnt from raw data
2. Mid-level feature representation is learnt from data
3. High-level feature representation *should be learned* from data
4. Downstream learning classifier is learnt from data

Recent popular strategy : learning all the way through

Deep Convolutional Networks



Recent popular strategy : learning all the way through

Current methodology

1. Frame the task as predicting output **label** from input **example**
2. Collect a **huge training sample**
3. Train using **supervised learning** and **stochastic back-prop**
4. Done