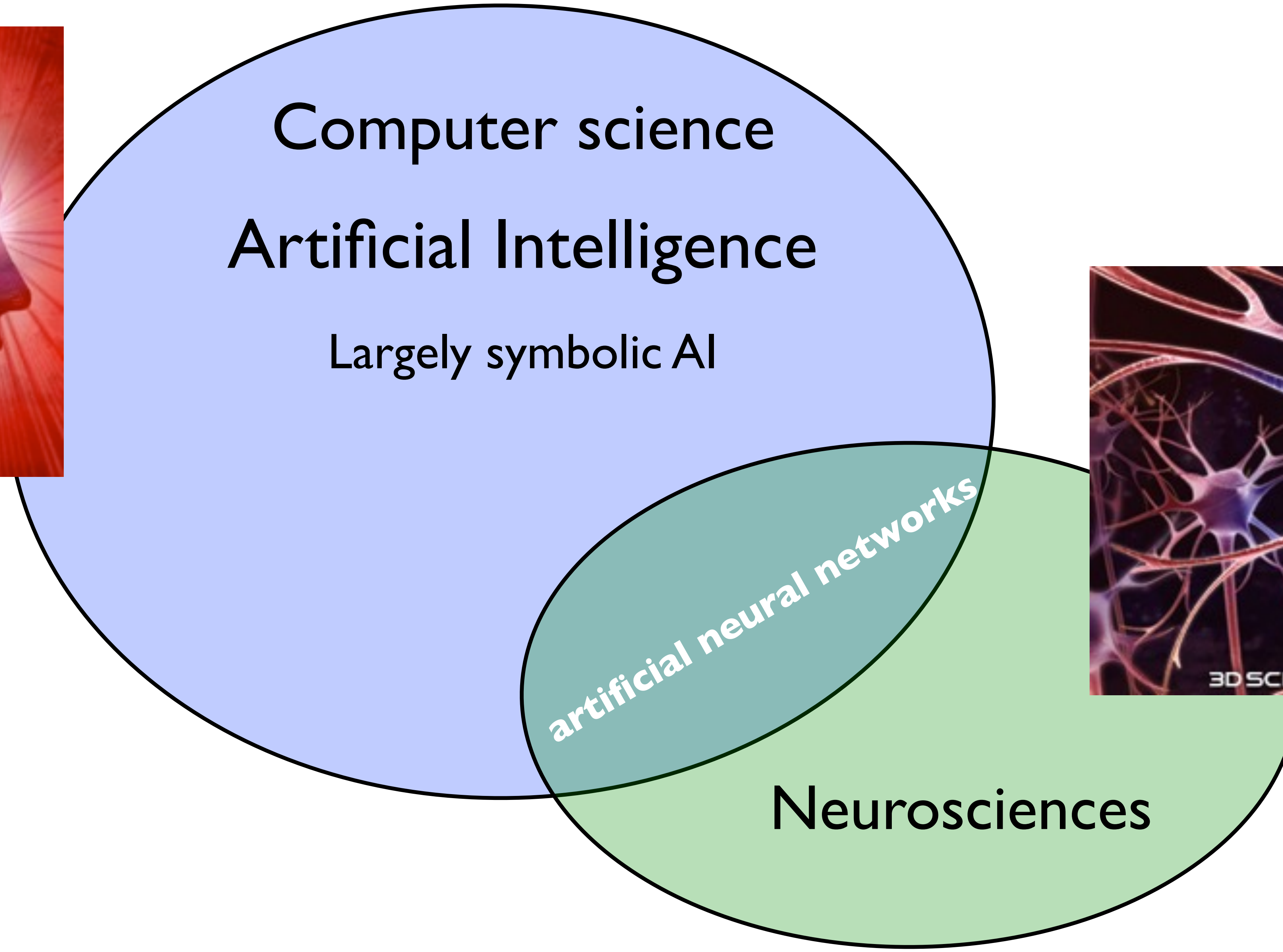
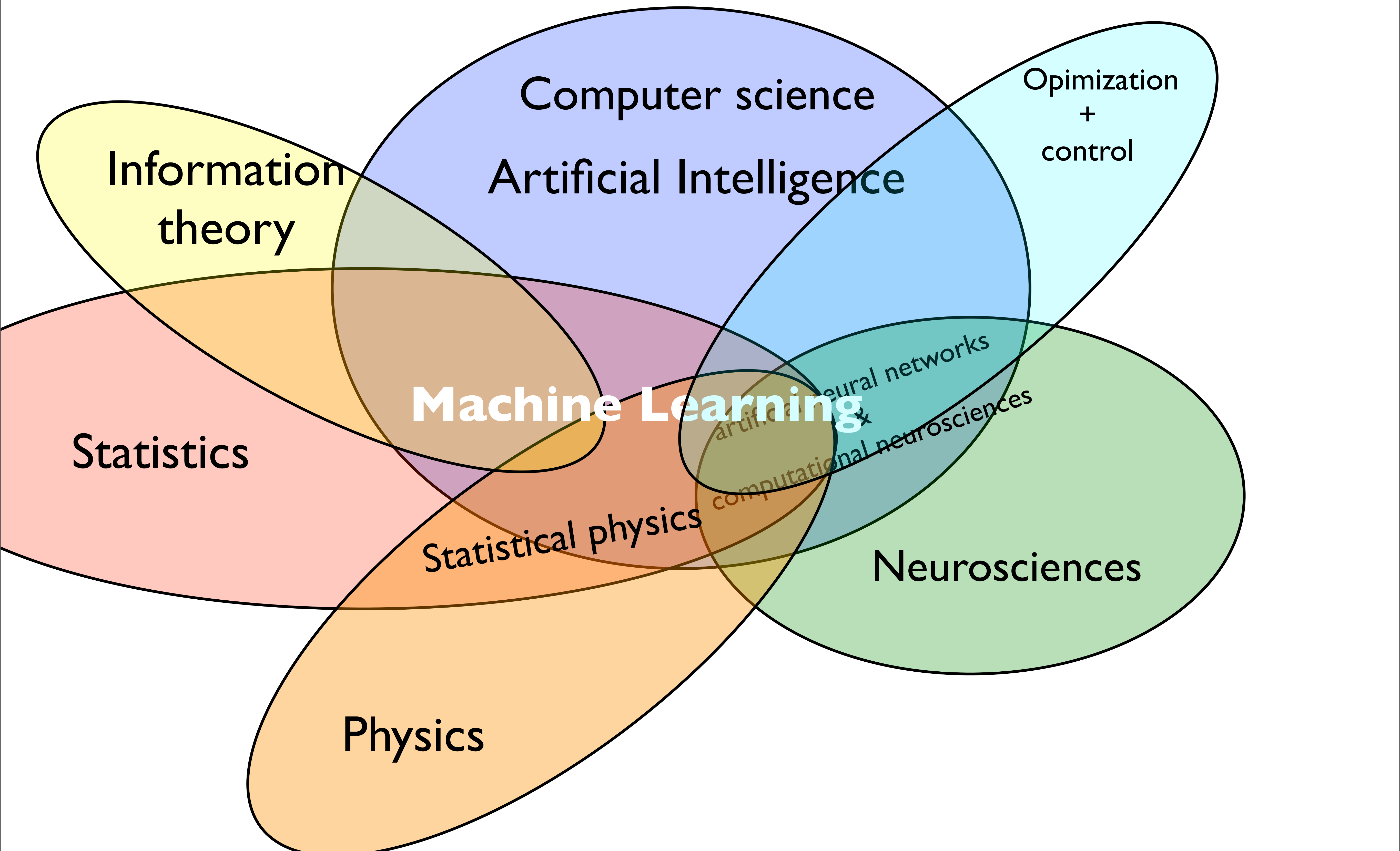
Artificial Intelligence in the 60s

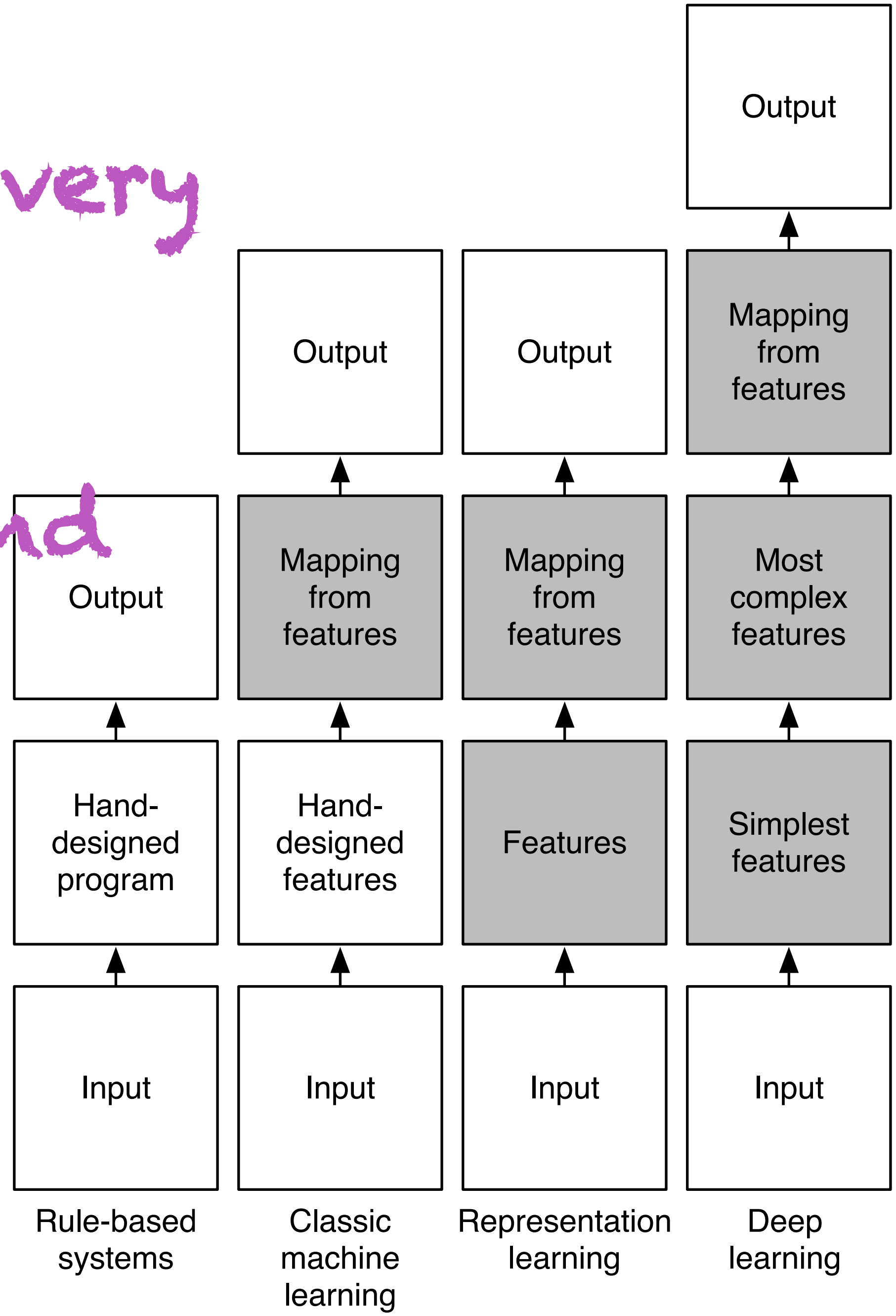


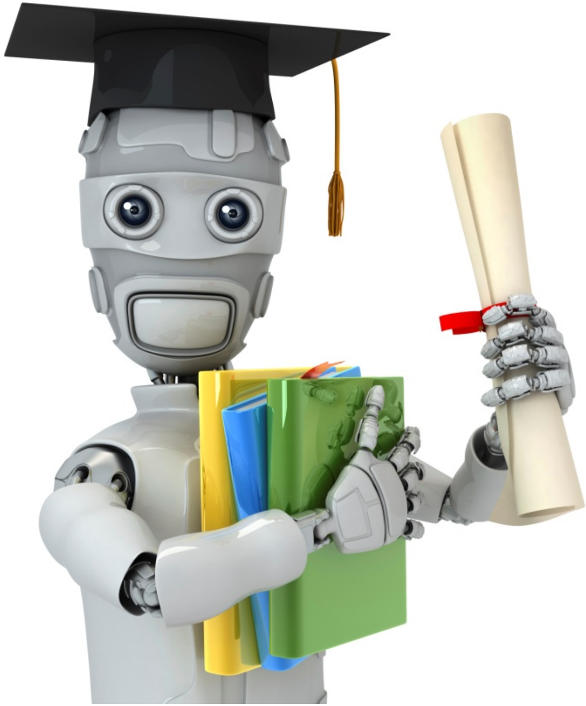
Current view of ML founding disciplines



Automating Feature Discovery

Discovering and representing higher-level abstractions





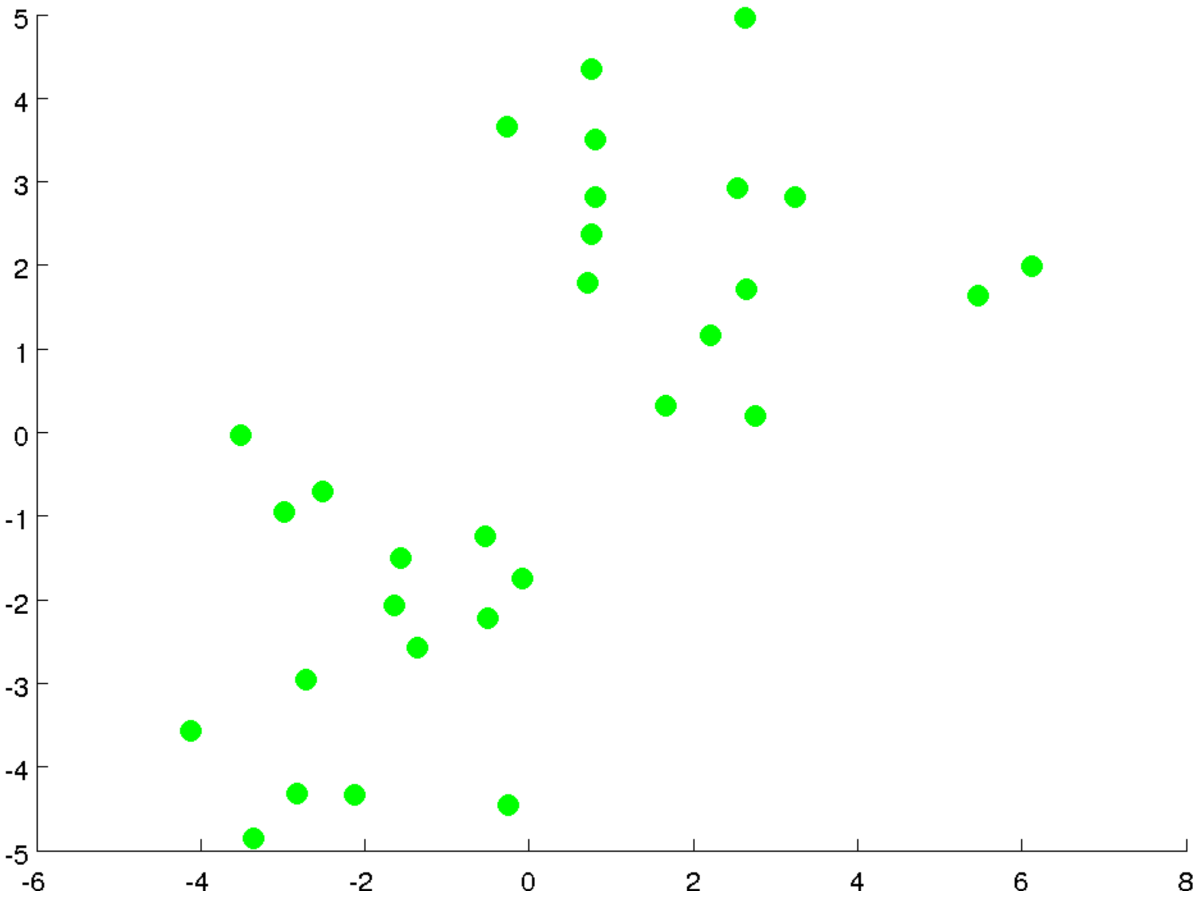
Machine Learning

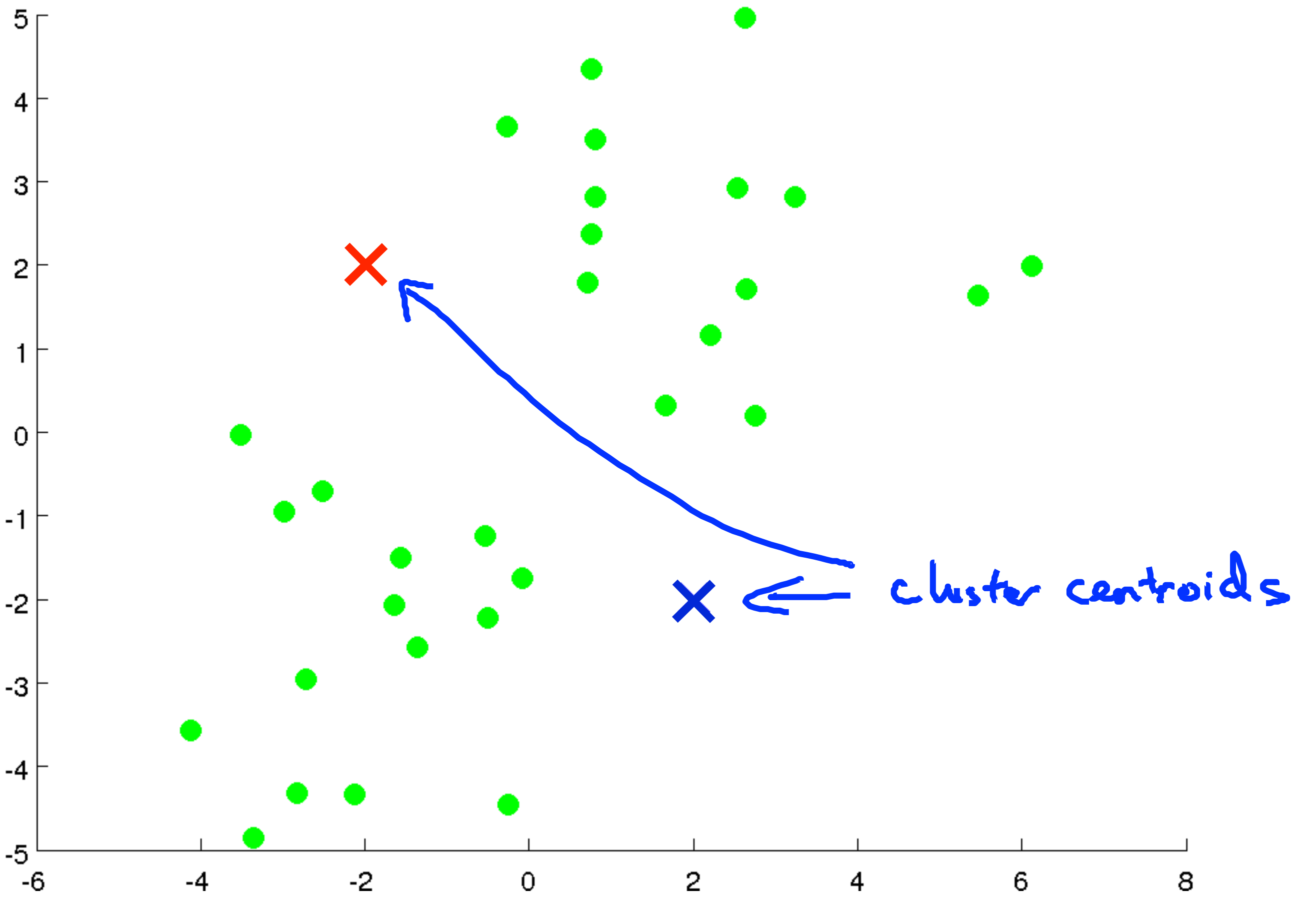
Clustering

K-means algorithm

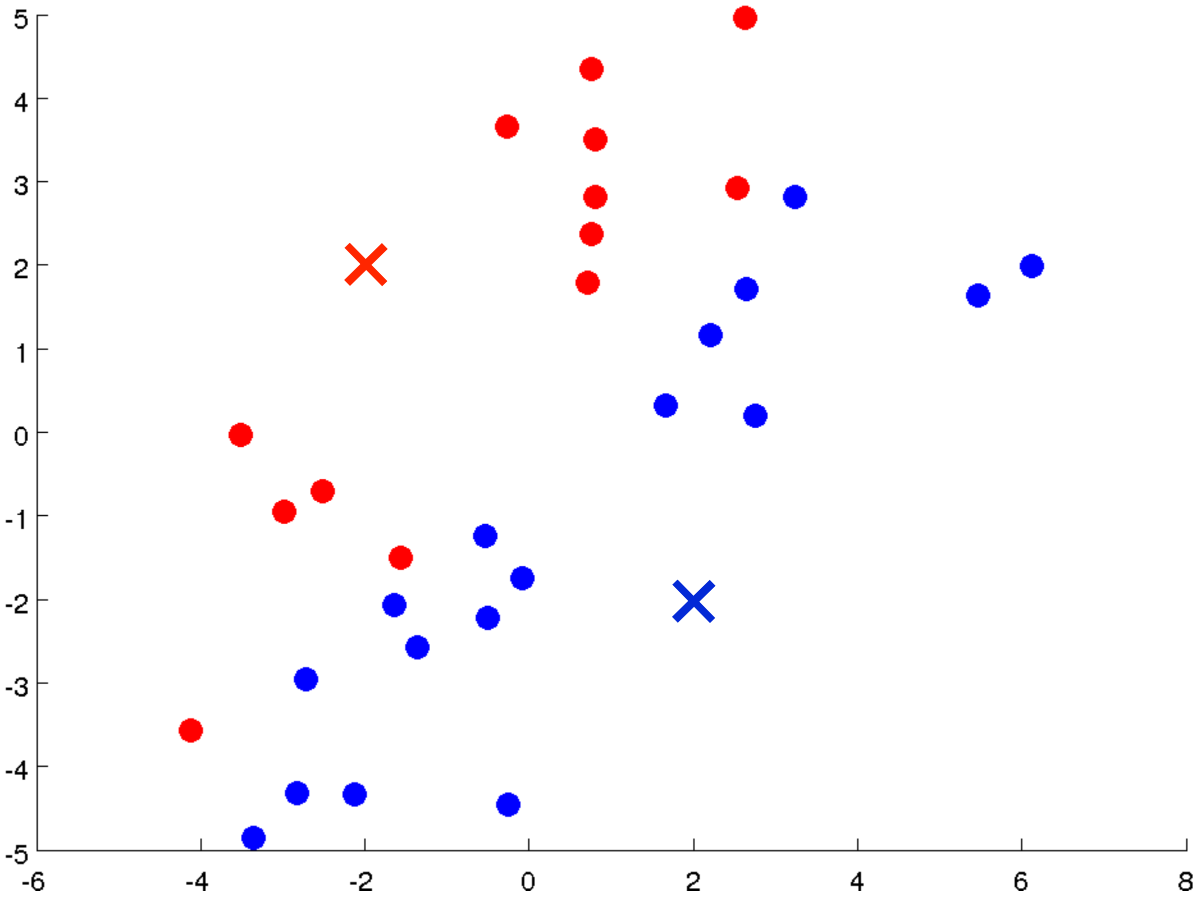
The *K-Means* Clustering Method

- *K-Means* (MacQueen'67, Lloyd'57/'82)
  - Each cluster is represented by the center of the cluster
- Given K, the number of clusters, the *K-Means* clustering algorithm is outlined as follows
  - Select *K* points as initial centroids
  - **Repeat**
    - Form *K* clusters by assigning each point to its closest centroid
    - Re-compute the centroids (i.e., *mean point*) of each cluster
  - **Until** convergence criterion is satisfied
- Different kinds of measures can be used
  - Manhattan distance (L_1 norm), Euclidean distance (L_2 norm), Cosine similarity

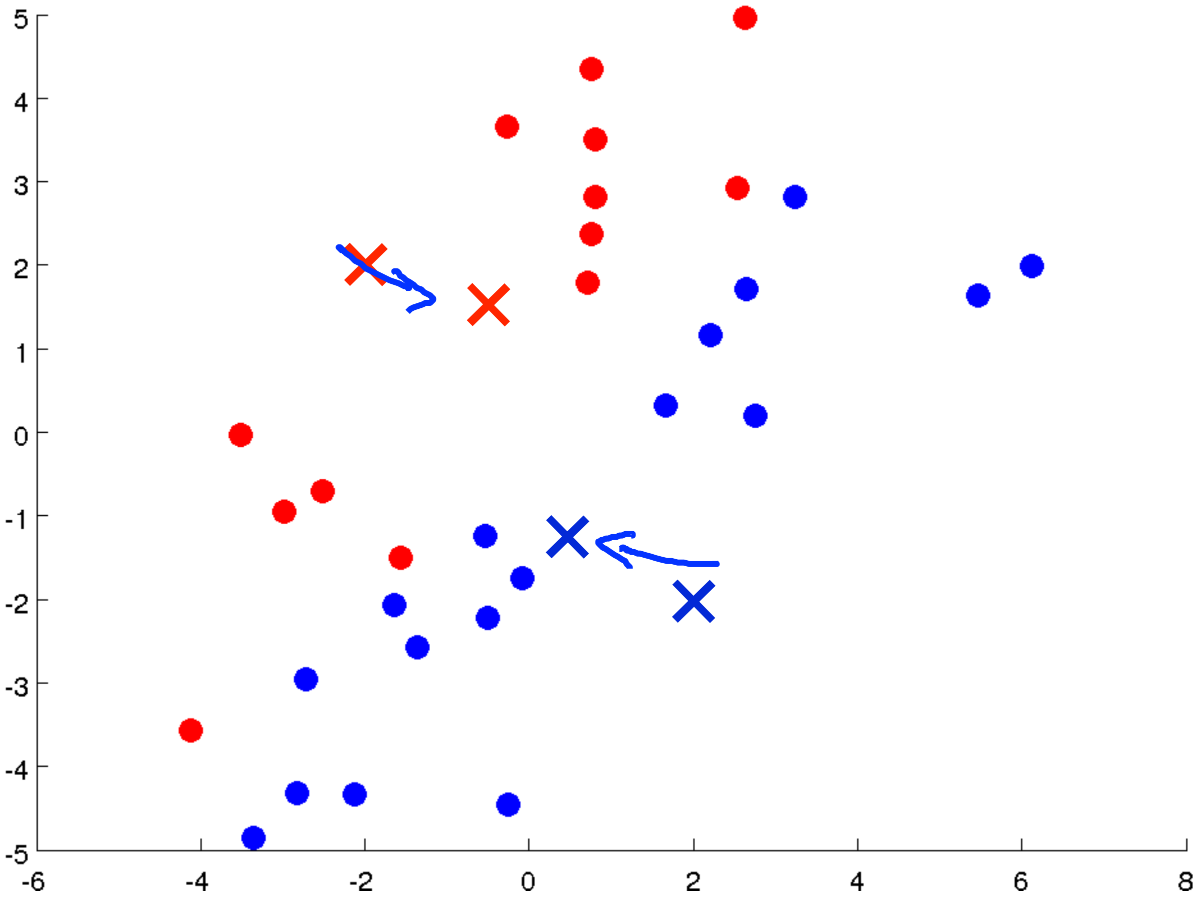

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8



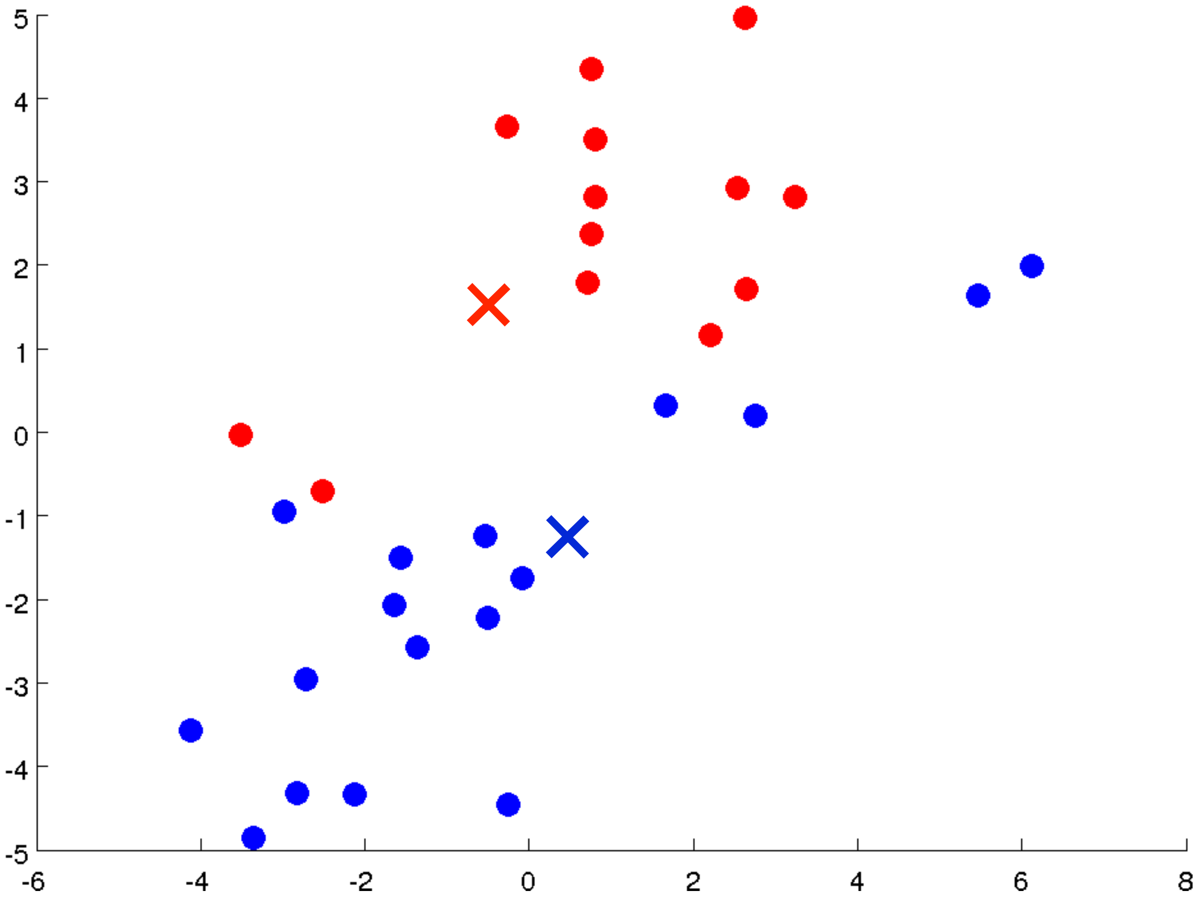
cluster centroids

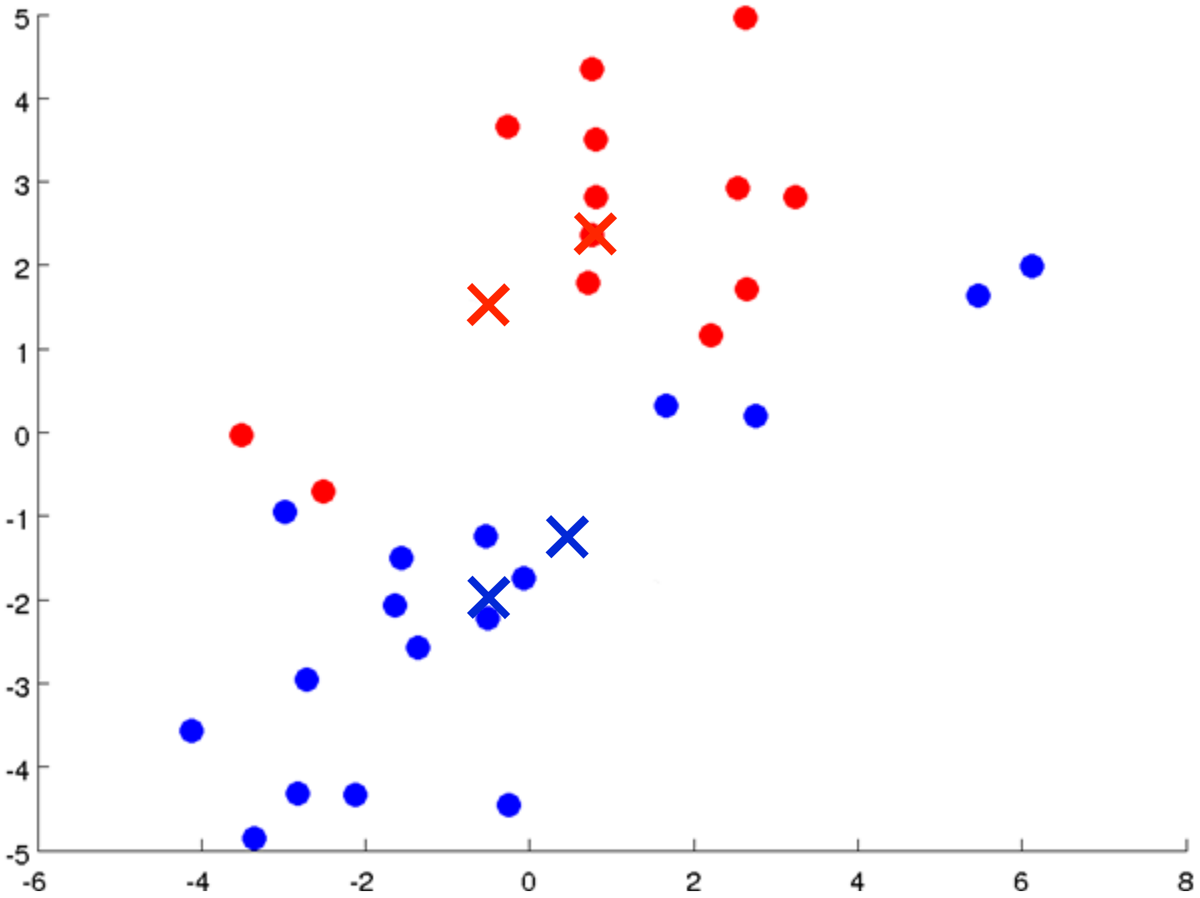


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

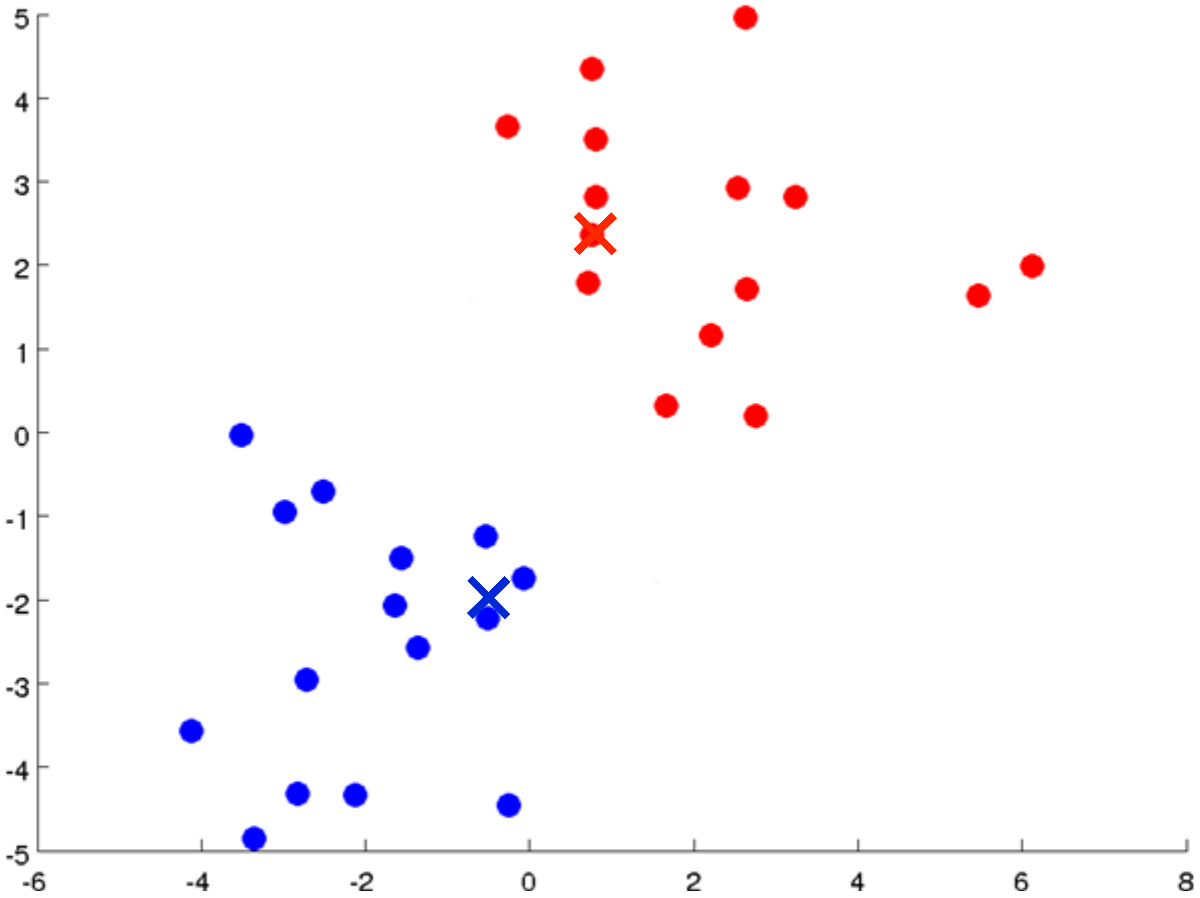


Andrew Ng

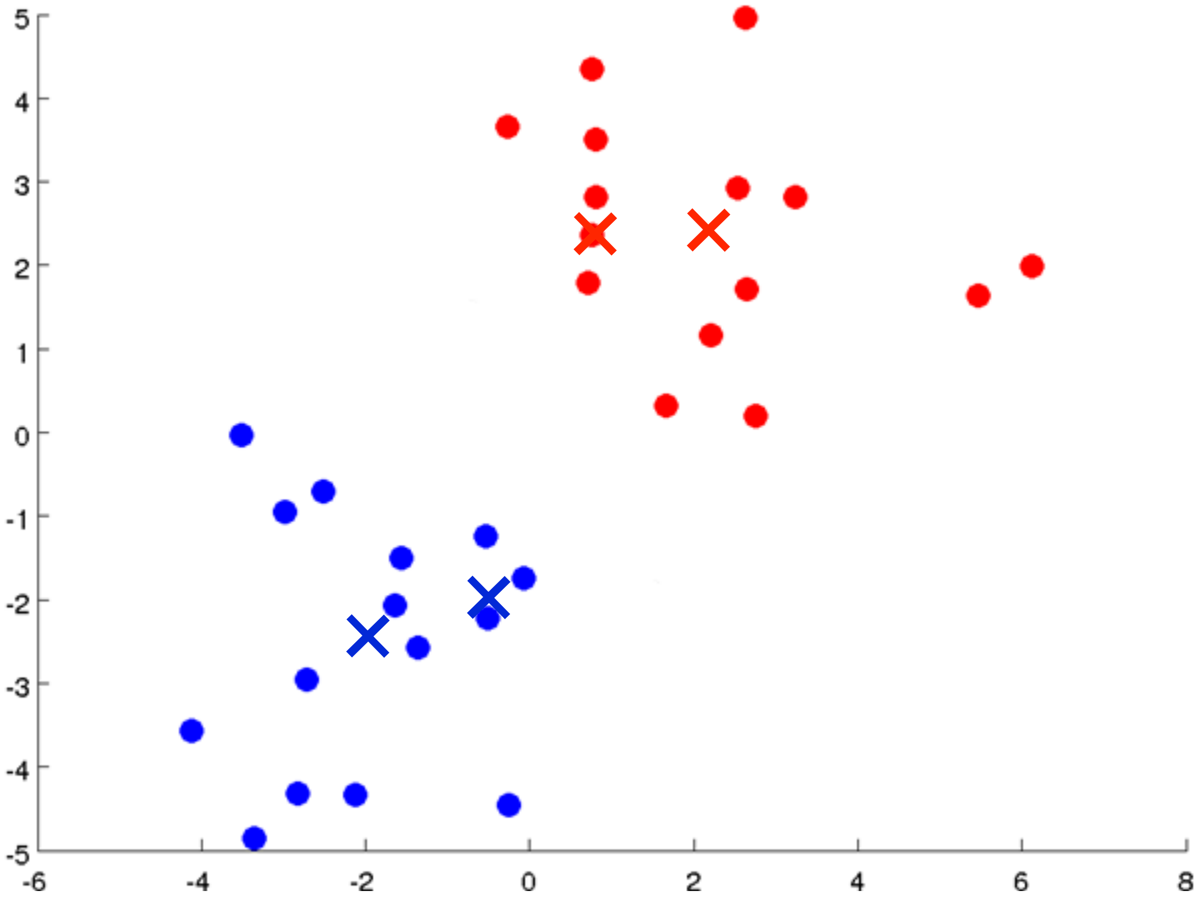

5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

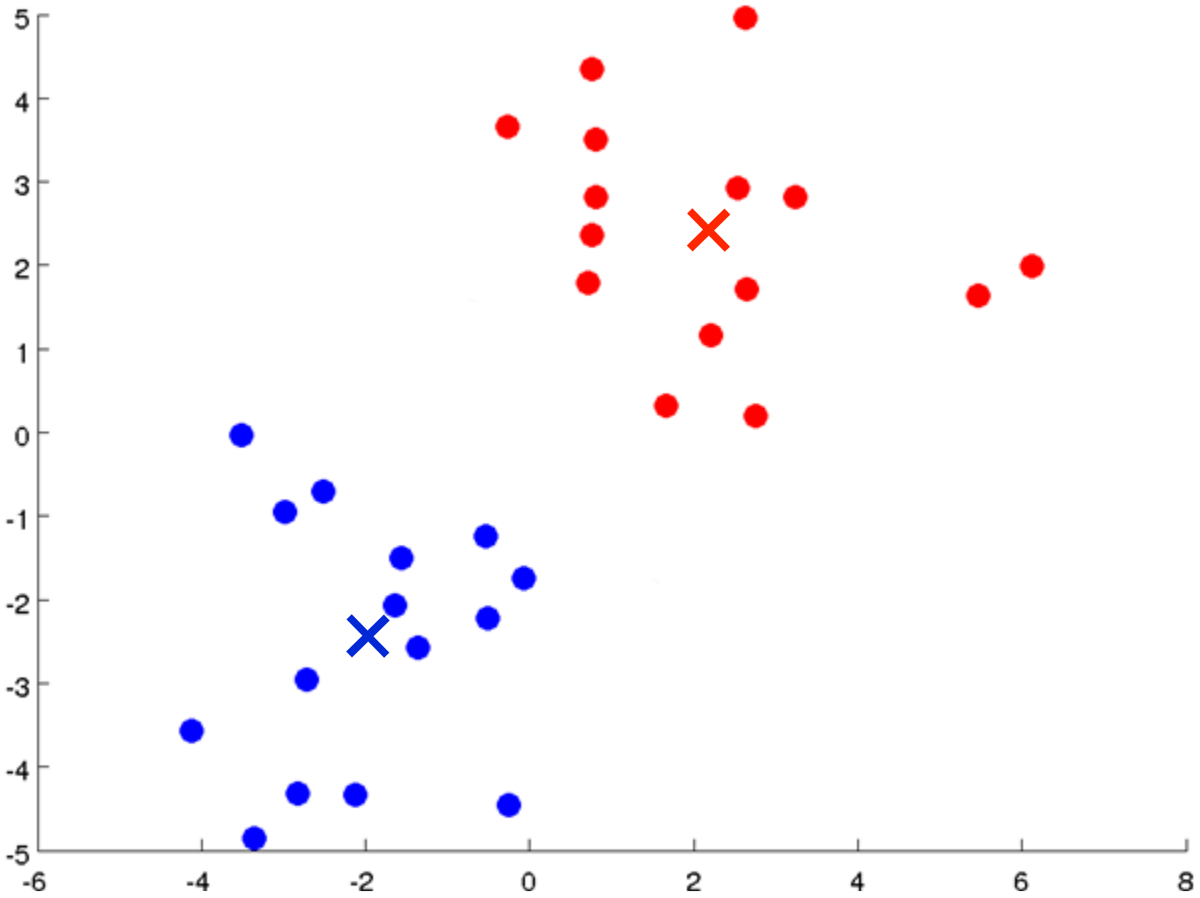


Andrew Ng



Andrew Ng


5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-4
-2
0
2
4
6
8

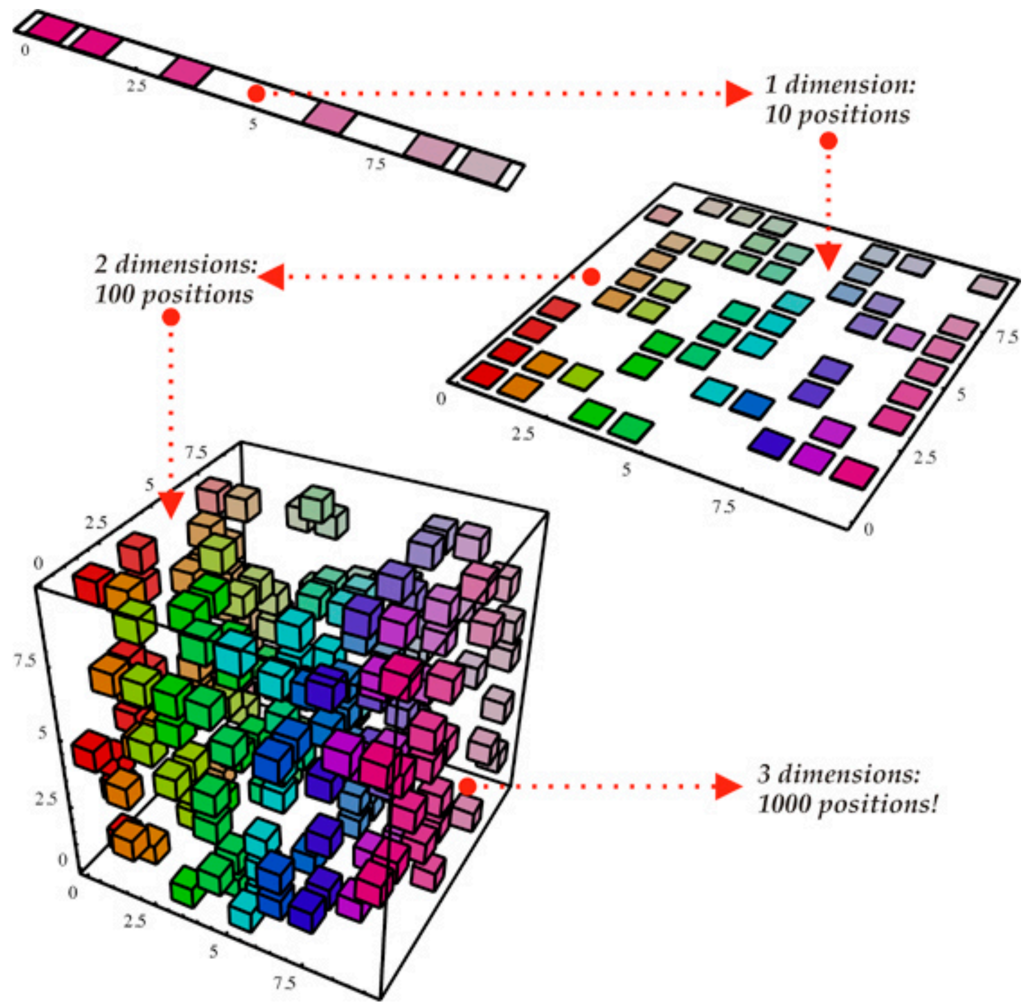


Andrew Ng

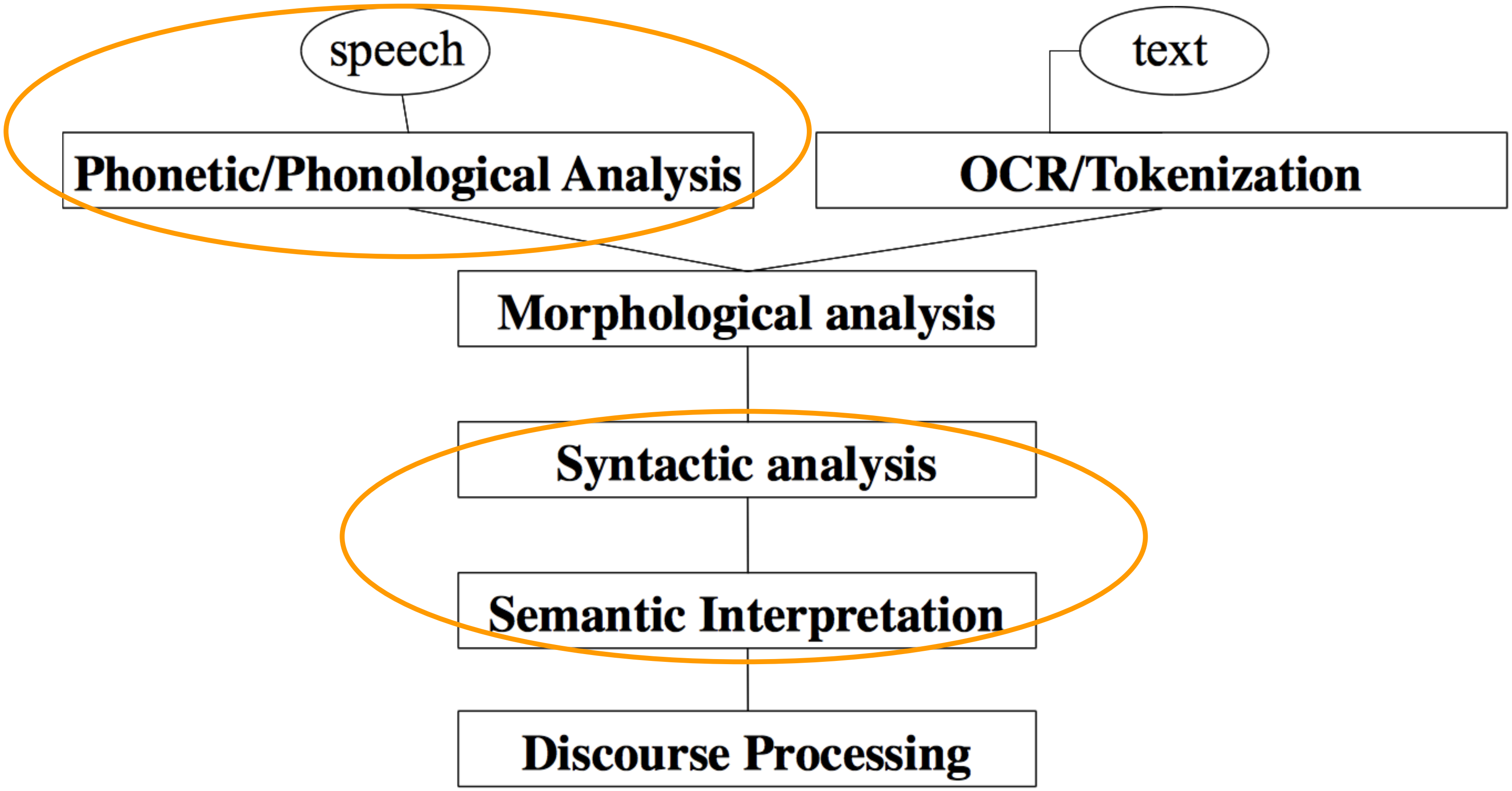
ML 101. What We Are Fighting Against: The Curse of Dimensionality

To generalize locally, need representative examples for all relevant variations!

Classical solution: hope for a smooth enough target function, or make it smooth by handcrafting good features / kernel



NLP Levels


speech
text
Phonetic/Phonological Analysis
OCR/Tokenization
Morphological analysis
Syntactic analysis
Semantic Interpretation
Discourse Processing

Why is NLP hard?

- Complexity in representing, learning and using linguistic/situational/world/visual knowledge

- Jane hit June and then **she** [fell/ran].

- Ambiguity: “I made her duck”

Reasons for Exploring Deep Learning

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- **Learned Features** are easy to adapt, fast to learn

- Deep learning provides a very flexible, (almost?) universal, learnable framework for **representing** world, visual and linguistic information.

- Deep learning can learn **unsupervised** (from raw text) and **supervised** (with specific labels like positive/negative)

Reasons for Exploring Deep Learning

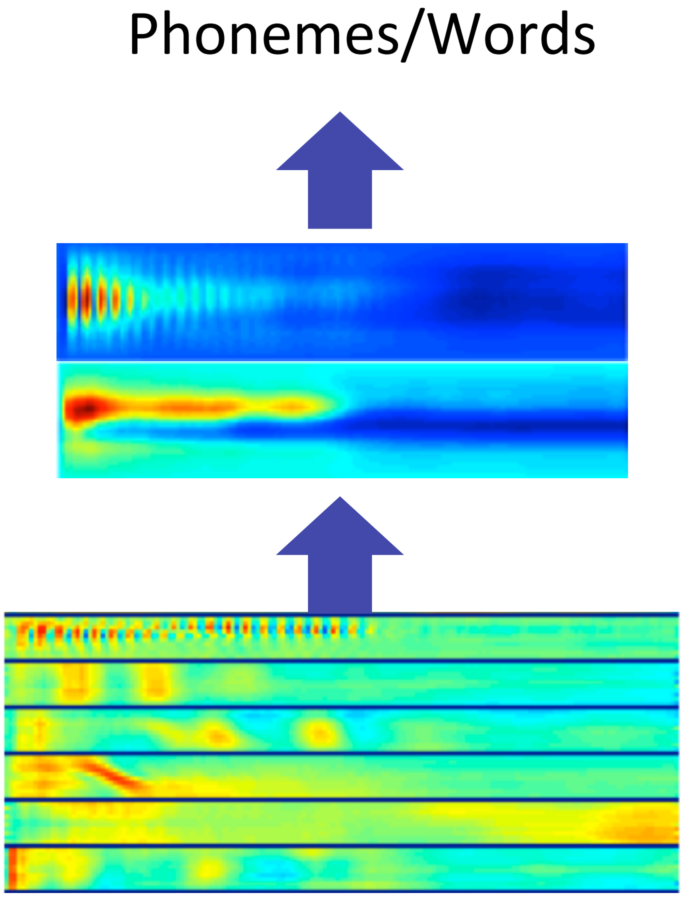
- In 2006 **deep** learning techniques started outperforming other machine learning techniques. Why now?

- DL techniques benefit more from a lot of data
- Faster machines and multicore CPU/GPU help DL
- New models, algorithms, ideas

→ **Improved performance** (first in speech and vision, then NLP)

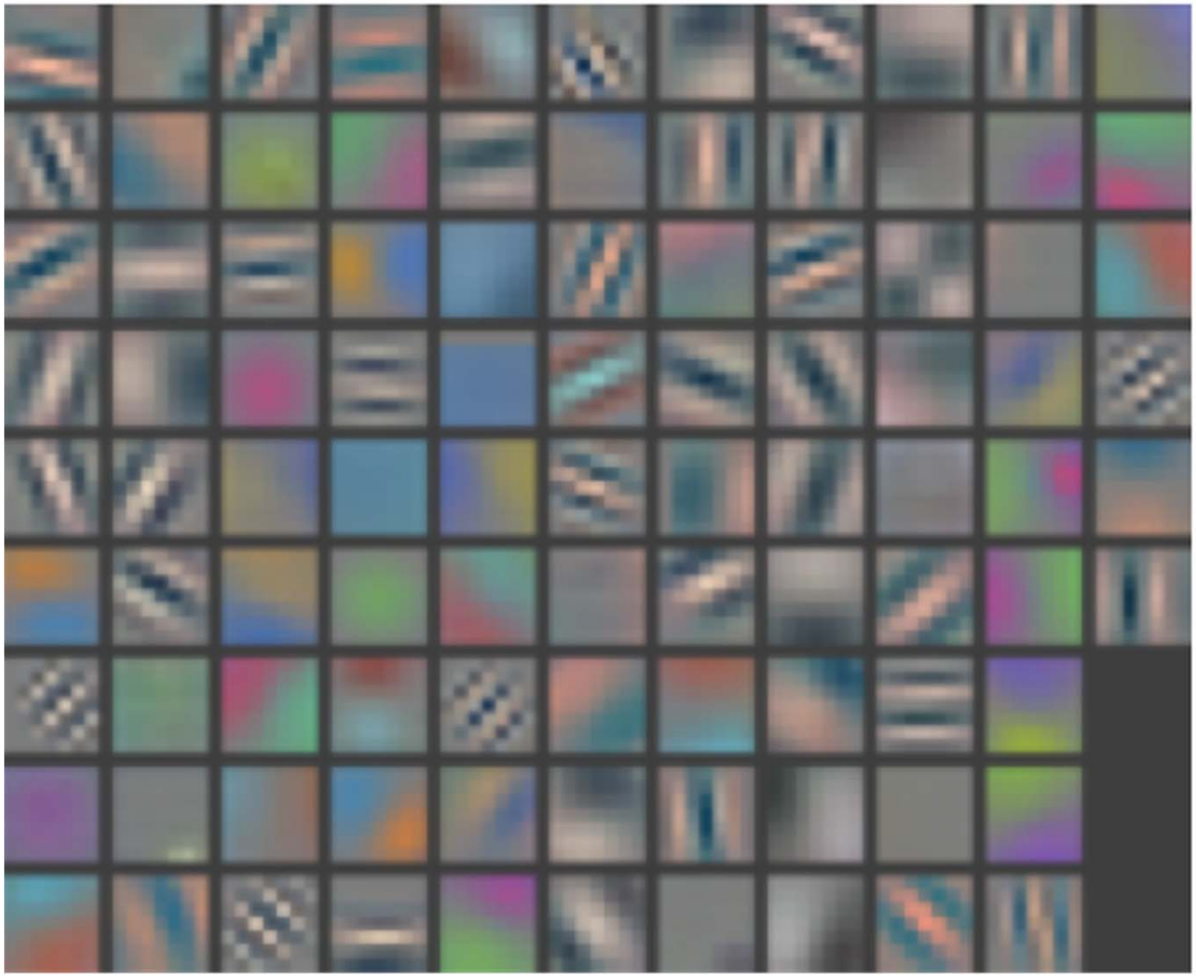
Deep Learning for Speech

- The first breakthrough results of "deep learning" on large datasets happened in speech recognition
- Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition Dahl et al. (2010)



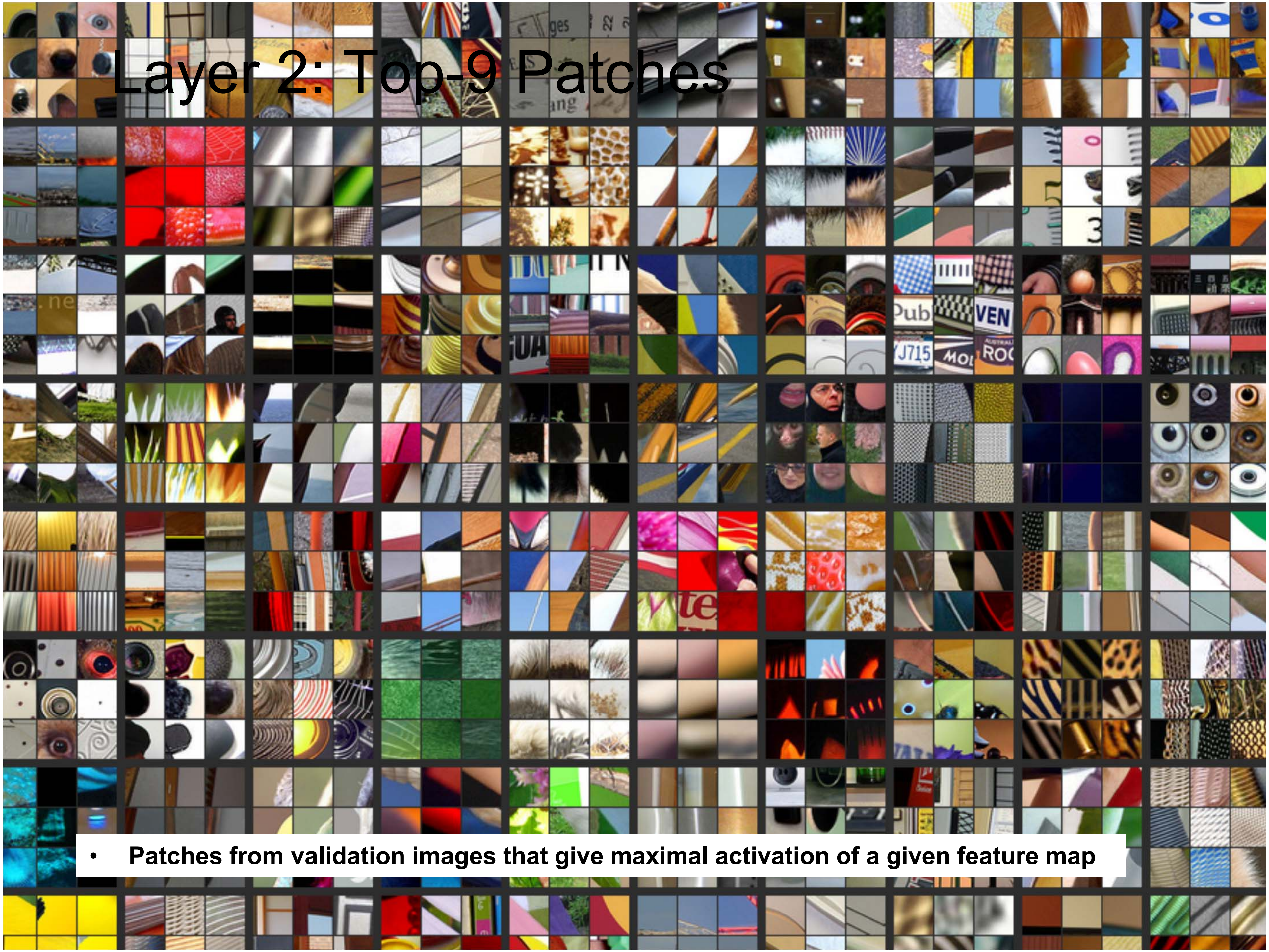
Acoustic model	Recog \ WER	RT03S FSH	Hub5 SWB
Traditional features	1-pass −adapt	**27.4**	**23.6**
Deep Learning	1-pass −adapt	**18.5** (−33%)	**16.1** (−32%)

Layer 1 Filters



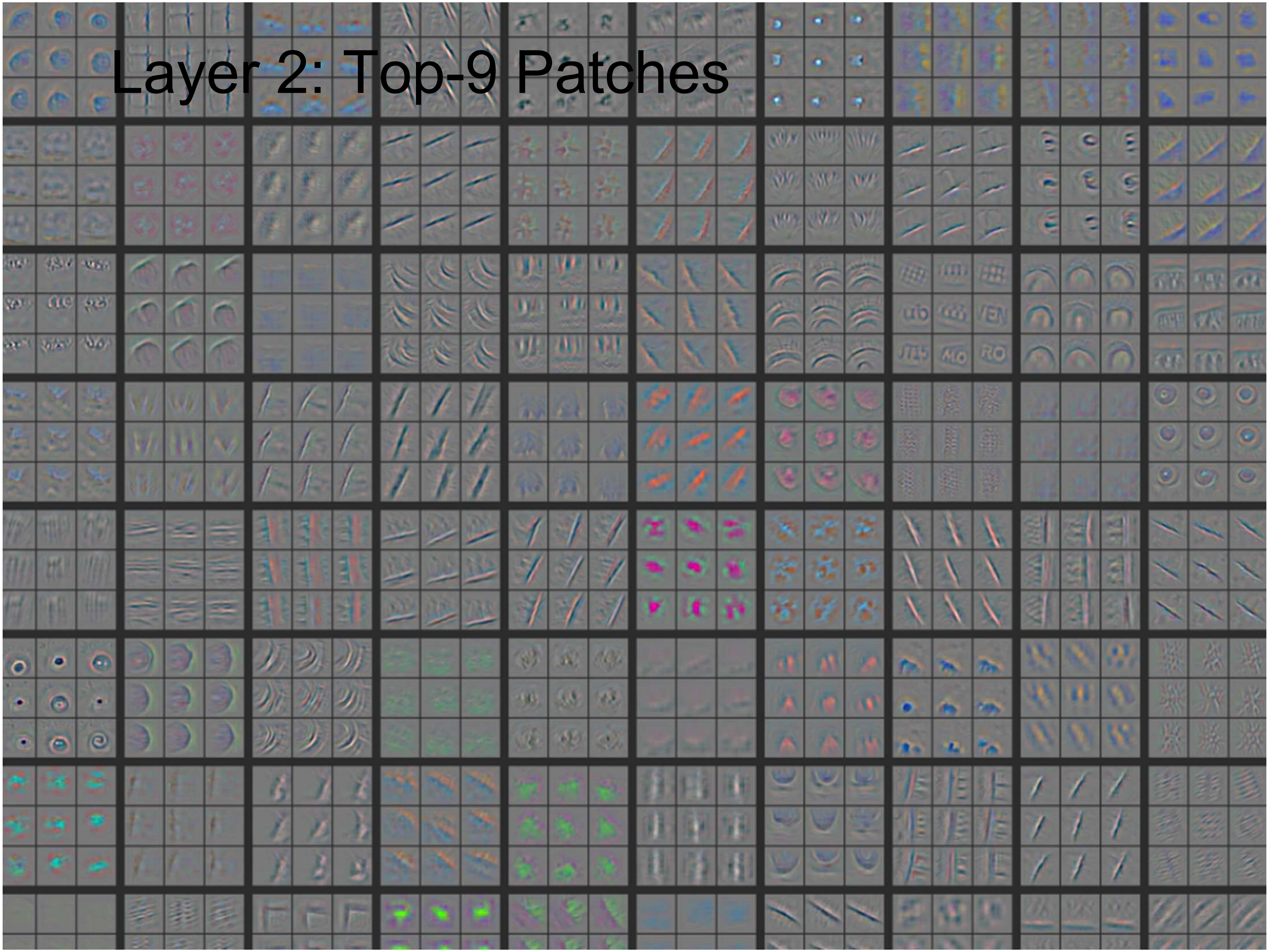
Layer 1: Top-9 Patches


Layer 2: Top-9 Patches

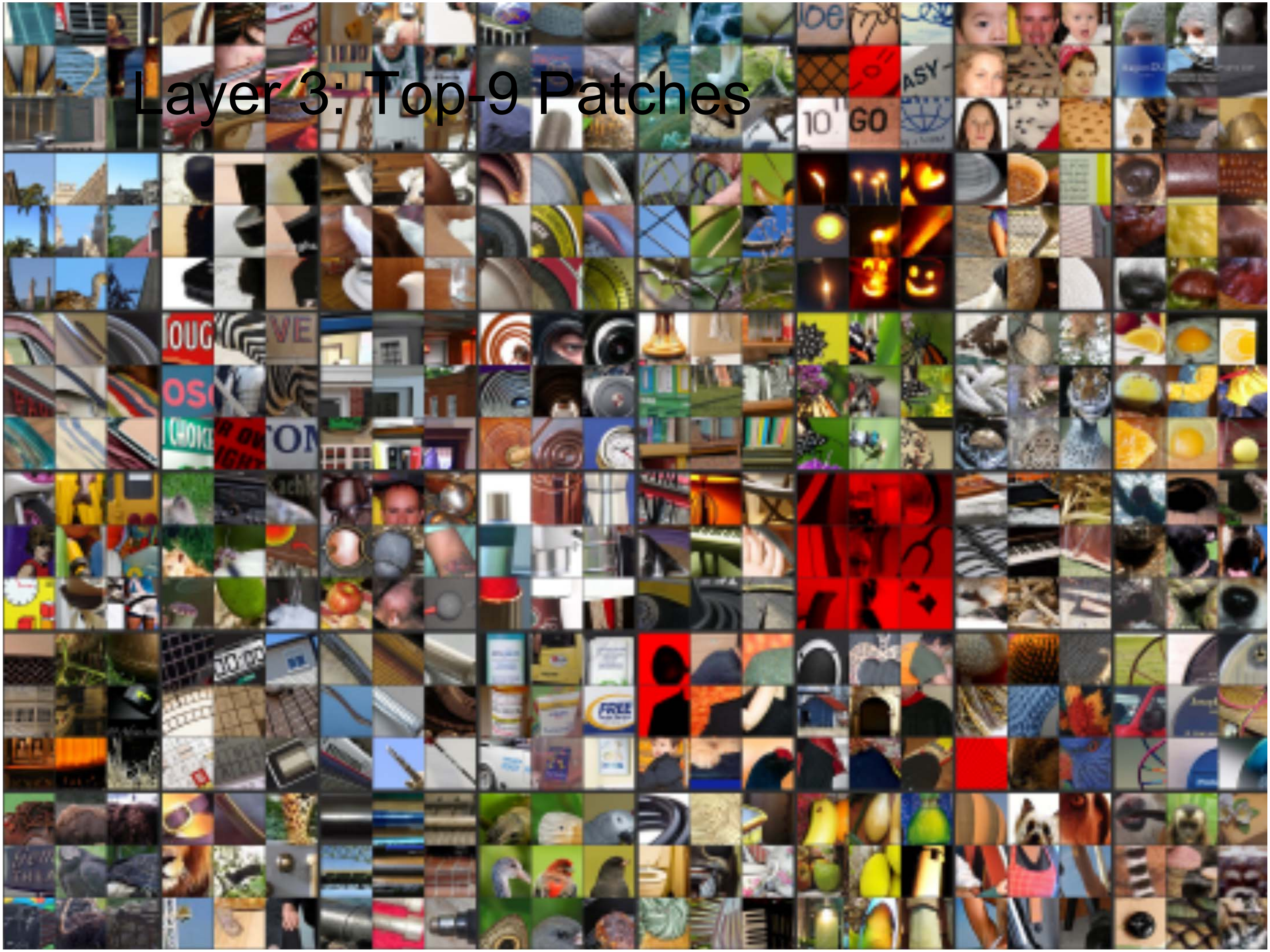


- **Patches from validation images that give maximal activation of a given feature map**

Layer 2: Top-9 Patches

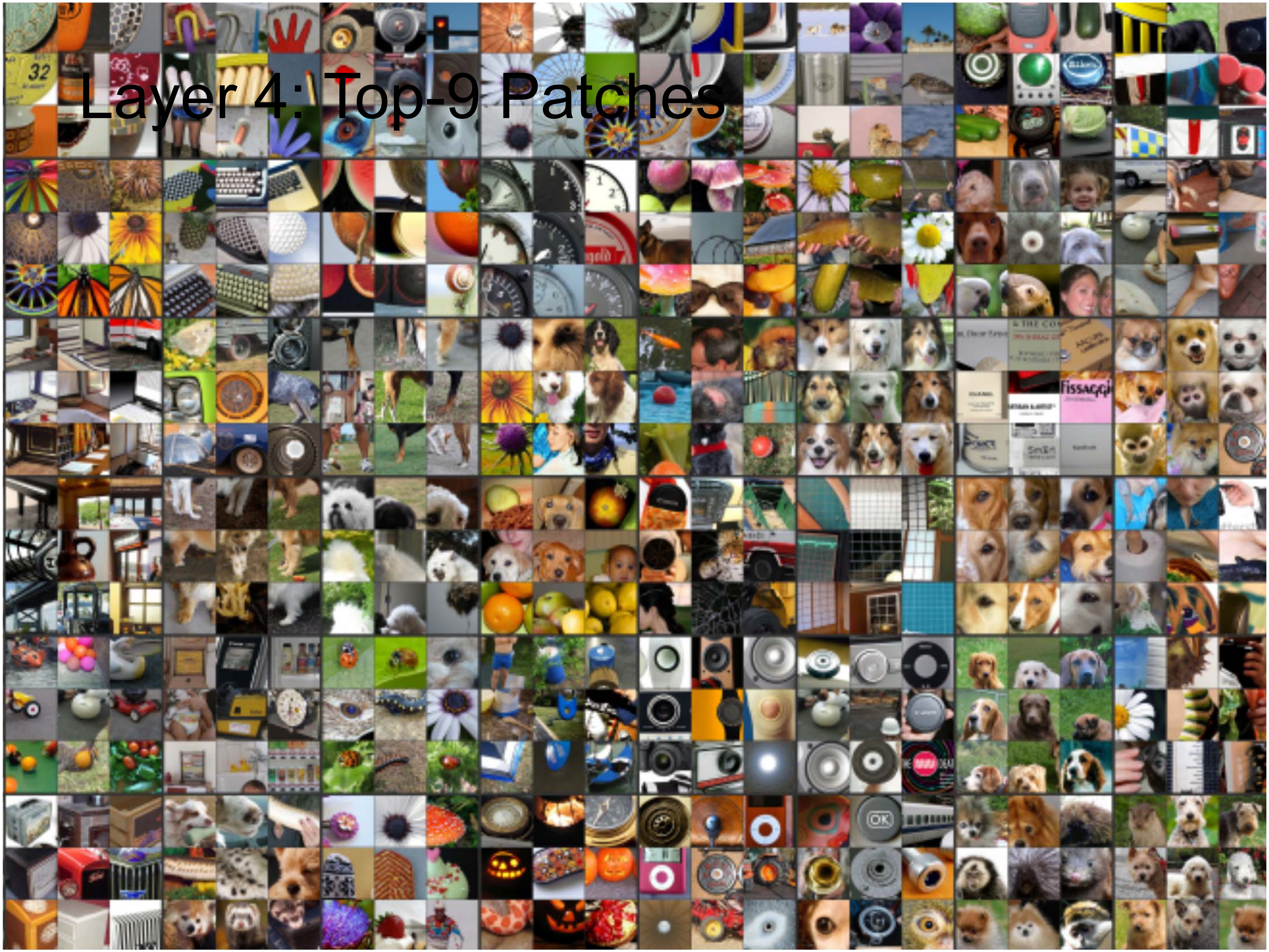


Layer 3: Top-9 Patches



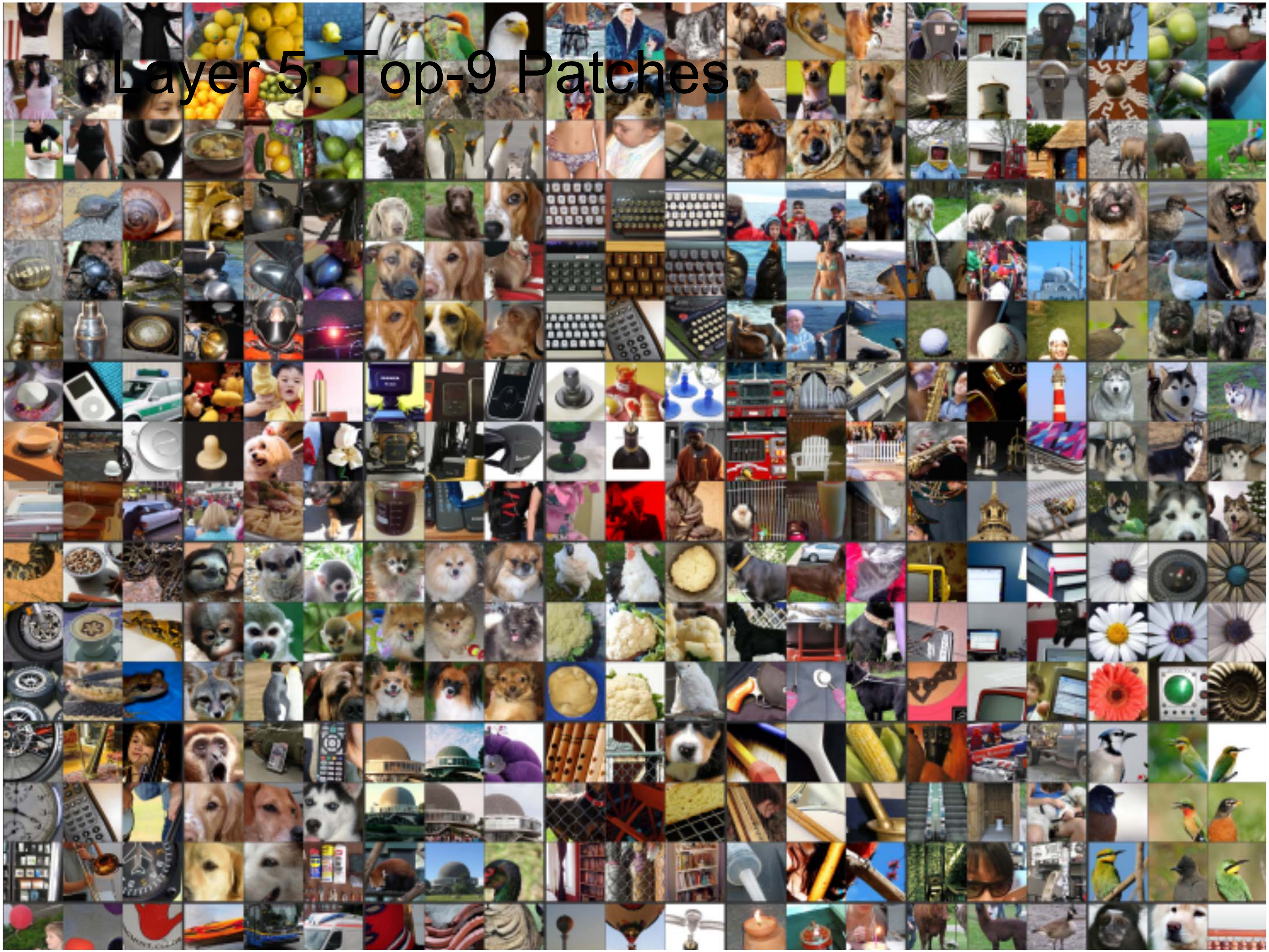
Layer 3: Top-9 Patches

Layer 4: Top-9 Patches



Layer 4: Top-9 Patches

Layer 5: Top-9 Patches



Layer 5: Top-9 Patches

Deep Learning + NLP = Deep NLP

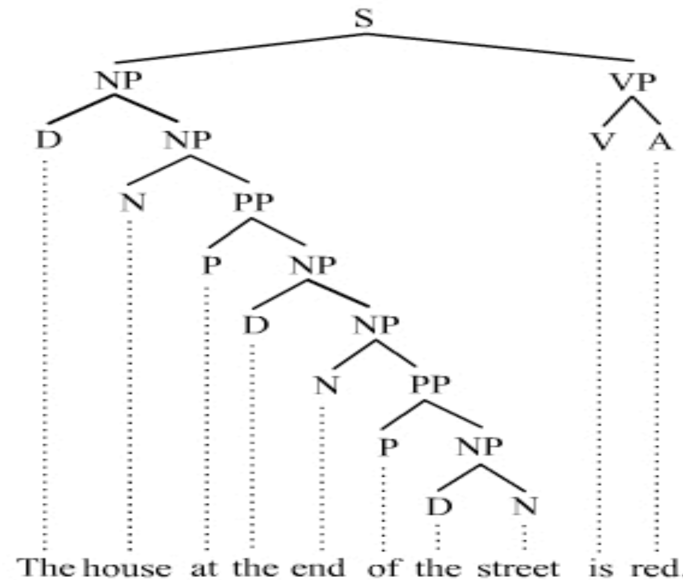
- Combine ideas and goals of NLP and use representation learning and deep learning methods to solve them

- Several big improvements in recent years across different NLP
  - **levels**: speech, morphology, syntax, semantics
  - **applications**: machine translation, sentiment analysis and question answering

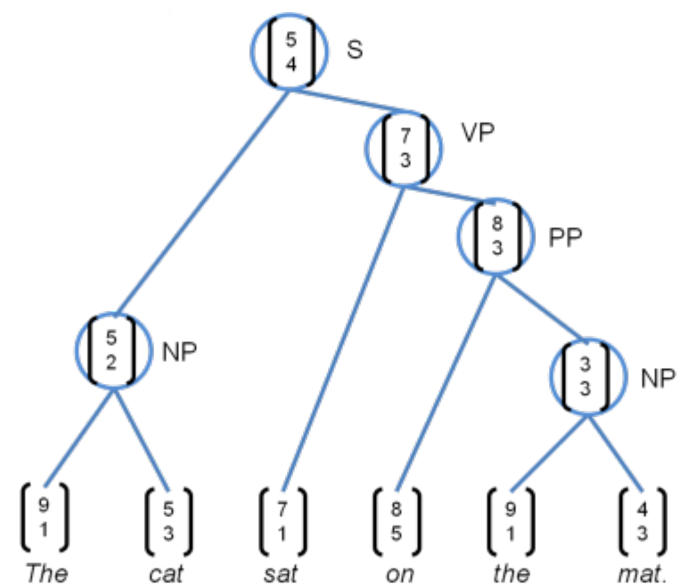
Neural word vectors - visualization

Representations at NLP Levels: Syntax

- Traditional: Phrases
  Discrete categories like NP, VP

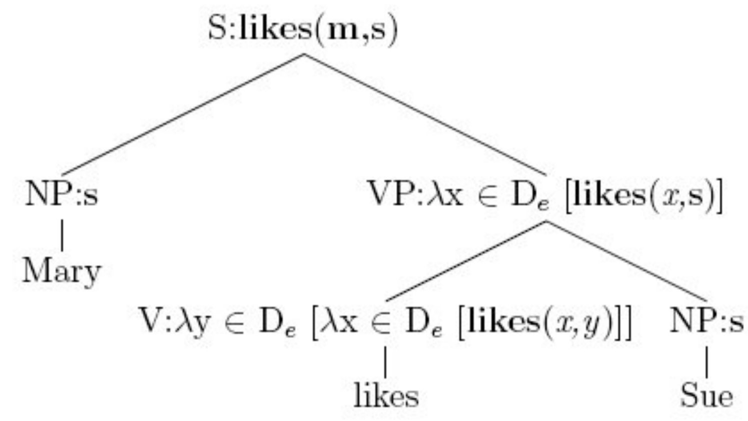


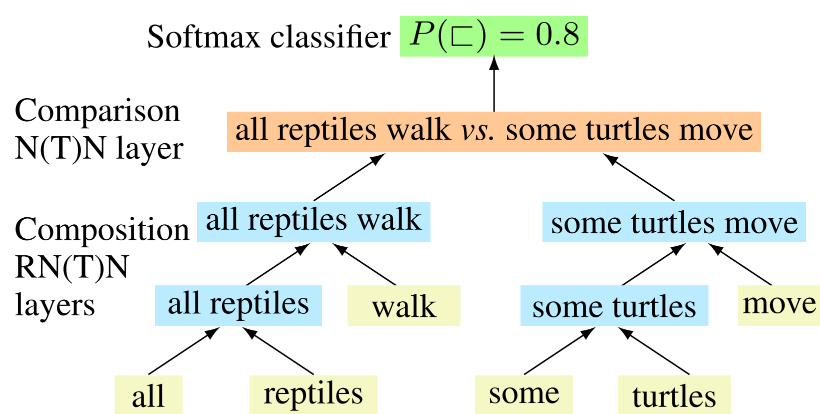
- DL:
  - Every word and every phrase is a vector
  - a neural network combines two vectors into one vector
  - Socher et al. 2011



Representations at NLP Levels: Semantics

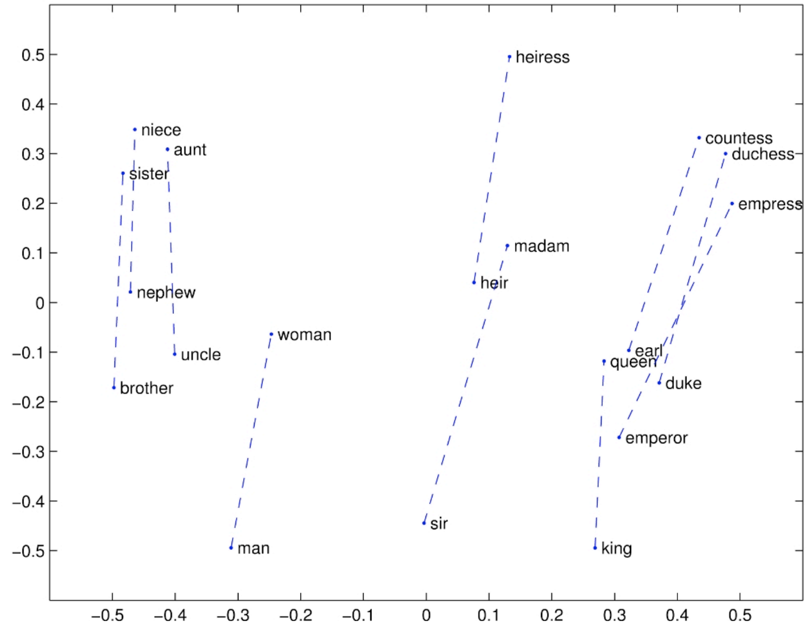
- Traditional: Lambda calculus
  - Carefully engineered functions
  - Take as inputs specific other functions
  - No notion of similarity or fuzziness of language
- DL:
  - Every word and every phrase and every logical expression is a vector
  - a neural network combines two vectors into one vector
  - Bowman et al. 2014





Representation for all levels: Vectors

- We will learn in the next lecture how we can learn vector representations for words and what they actually **represent.**



- Next week: neural networks and how they can use these vectors for all NLP levels and many different applications

Mathematical optimization

(mathematical) optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: constraint functions

optimal solution $x^\star$ has smallest value of f_0 among all vectors that satisfy the constraints

Examples

portfolio optimization

- variables: amounts invested in different assets
- constraints: budget, max./min. investment per asset, minimum return
- objective: overall risk or return variance

device sizing in electronic circuits

- variables: device widths and lengths
- constraints: manufacturing limits, timing requirements, maximum area
- objective: power consumption

data fitting

- variables: model parameters
- constraints: prior information, parameter limits
- objective: measure of misfit or prediction error

Solving optimization problems

general optimization problem

- very difficult to solve
- methods involve some compromise, $e.g.$, very long computation time, or not always finding the solution

exceptions: certain problem classes can be solved efficiently and reliably

- least-squares problems
- linear programming problems
- convex optimization problems

Convex optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- objective and constraint functions are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

  if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$

- includes least-squares problems and linear programs as special cases

solving convex optimization problems

- no analytical solution
- reliable and efficient algorithms
- computation time (roughly) proportional to $\max\{n^3, n^2m, F\}$, where F is cost of evaluating f_i's and their first and second derivatives
- almost a technology

using convex optimization

- often difficult to recognize
- many tricks for transforming problems into convex form
- surprisingly many problems can be solved via convex optimization

A brief history of NLU

- 1960s: Pattern-matching with small rule-sets
- 1970-80s: Linguistically rich, logic-driven, grounded systems; restricted applications
- 1990s: the statistical revolution in NLP leads to a decrease in NLU work
- 2010s: NLU returns to center stage, mixing techniques from previous decades

NLU today and tomorrow

- It's an exciting time to be doing NLU!
- In academia, a resurgence of interest in NLU (after a long winter)
- In industry, an explosion in products & services that rely on NLU (Siri, Google Now, Microsoft Cortana, Amazon Echo, …)
- Systems are impressive, but show their weaknesses quickly
- NLU is far from solved — big breakthroughs lie in the future

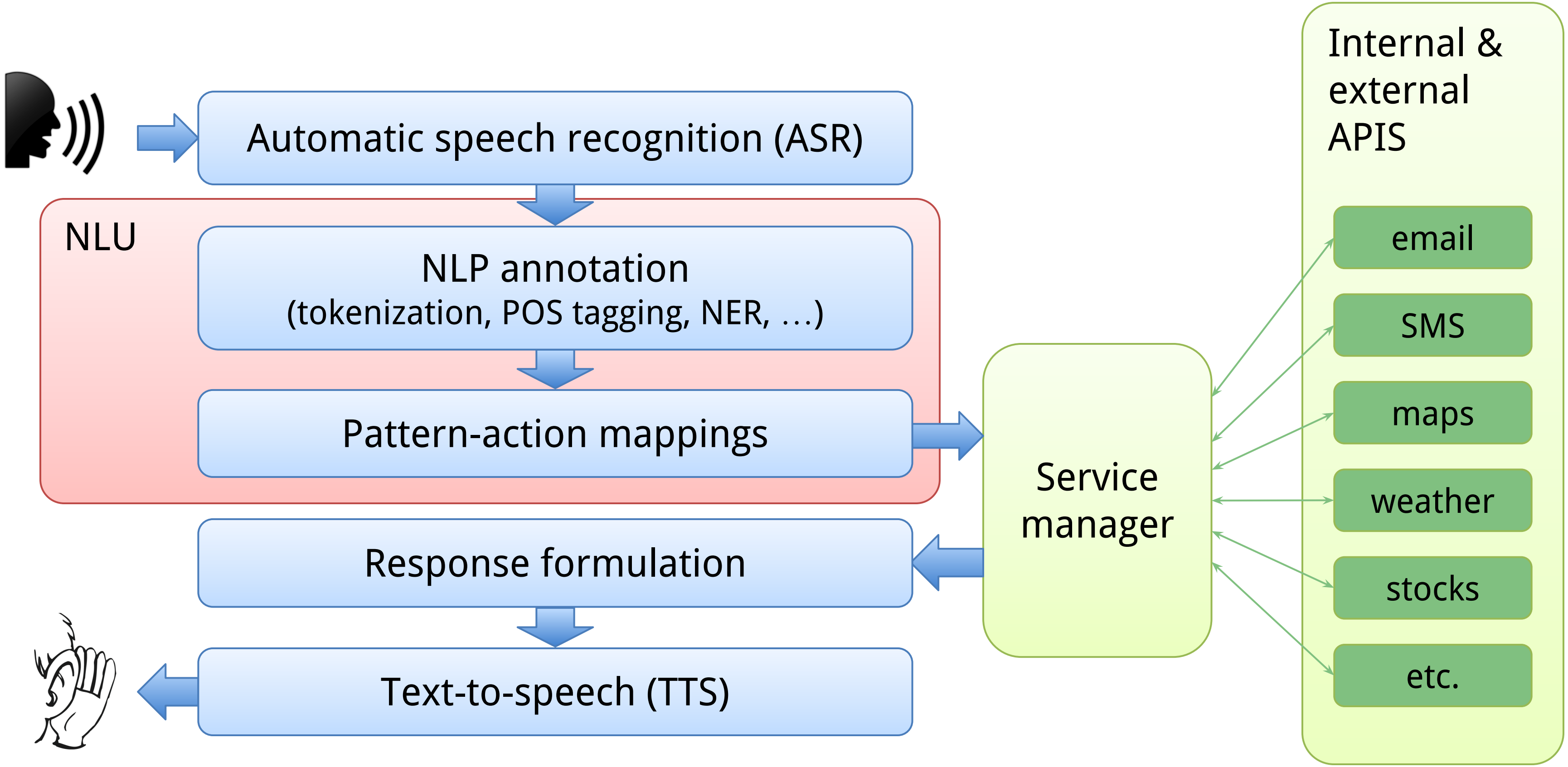
Siri: NLU’s celebrity spokesperson

Susan Bennett, a veteran voice actor, claimed that she is the original voice behind the popular digital assistant. Tight-lipped Apple officials won't confirm or deny the claim, but an audio forensics expert confirmed that the voices are a match. http://nydn.us/1g6E0eR

Siri

- The voice-driven personal assistant on your iPhone
- Perhaps the most visible & exciting application of NLU today
- A major breakthrough in artificial intelligence (AI)??
- The next generation of UI design??

How does Siri work?



The promise of Siri

Where is The Hobbit playing in Mountain View?

The Hobbit is playing at the Century 16 Theater.

When is it playing there?

It's playing at 2pm, 5pm, and 8pm.

OK. I'd like 1 adult and 2 children for the first show.
How much would that cost?

Need domain knowledge, discourse knowledge, world knowledge

The reality of Siri

Colbert: Write the show.

Siri: What would you like to search for?

Colbert: I don't want to search for anything! I want to write the show!

Siri: Searching the Web for "search for anything. I want to write the shuffle."

[...]

Colbert: For the love of God, the cameras are on, give me something?

Siri: What kind of place are you looking for? Camera stores or churches?

http://thecolbertreport.cc.com/videos/h40j2n/talking-iphone-4s

Reminiscent of Eliza (1966)

User: You are like my father in some ways.

ELIZA: What resemblance do you see

User: You are not very aggressive but I think you don’t want me to notice that.

ELIZA: What makes you think I am not aggressive?

User: You don’t argue with me.

ELIZA: Why do you think I don’t argue with you?

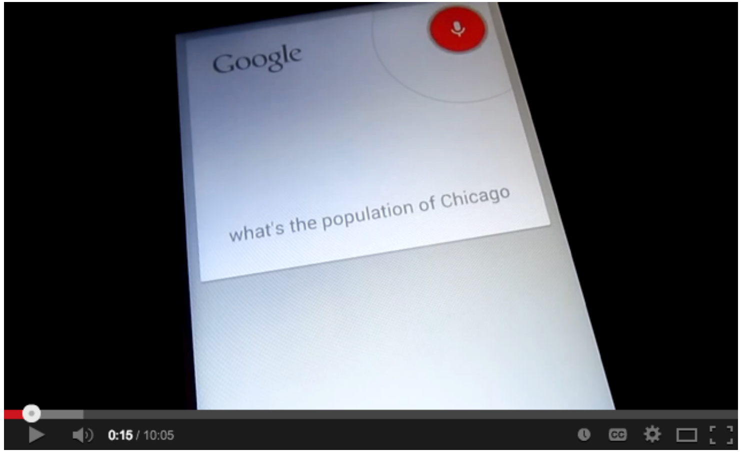
User: You are afraid of me.

ELIZA: Does it please you to believe I am afraid of you?

Conversational search at Google

what's the population of Chicago
who's the mayor
how old is he
who is he married to

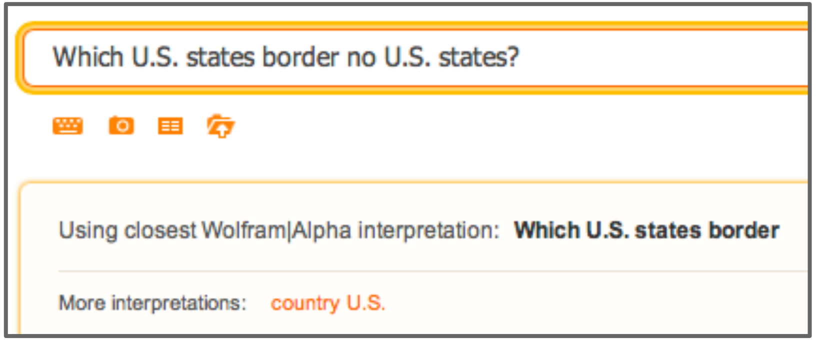
OK Google, where am I
how is traffic in San Diego
show me things to do there
when did the San Diego Zoo open
is it open
how far is it
call them



https://www.youtube.com/watch?v=yiQX-_Y0gms

when is Thanksgiving
I meant the Canadian one

Wolfram Alpha



Watson

Watson and pragmatics

Answer: Grasshoppers eat it.

Watson: What is kosher?

Watson and discourse processing

Watson Almost Sneaks Wrong Response by *Jeopardy*'s Trebek:

Watson also tripped up on an "Olympic Oddities" answer, but so imperceptibly that Alex Trebek didn't notice at first, raising an important point of clarification. After Jennings responded incorrectly that Olympian gymnast George Eyser was "missing a hand", Watson responded, "What is a leg?"

http://www.wired.com/business/2011/02/watson-wrong-answer-trebek/

Twitter prognostication

- Twitter mood predicts the stock market [Bollen et al. 2011]
- “In February 2011 Derwent Capital Markets launched a hedge fund using Twitter for investment direction.” [Wikipedia]
- The junk science behind the ‘Twitter Hedge Fund’
- Derwent closes shop

Hathaway vs. Hathaway

Does Anne Hathaway News Drive Berkshire Hathaway's Stock?

MAR 18 2011, 10:50 AM ET 28 Share 257 Tweet 471 +1 7 Recommend 616

Given the awesome correlating powers of today's stock trading computers, the idea may not be as far-fetched as you think.

Deep problems of sentiment analysis

1. There was an earthquake in LA
2. The team failed the physical challenge. (We win/lose!)
3. They said it would be great. They were right/wrong.
4. Many consider the masterpiece bewildering, boring, slow-moving or annoying.
5. The party fat-cats are sipping their expensive, imported wines.
6. Oh, you’re terrible!

The 2008 United Airlines “bankruptcy”

- Newspaper accidentally republished old bankruptcy story
- Automated trading reacted within seconds
- $1B in market value evaporated within 12 minutes

Read more at http://nyti.ms/1dBzJSK

The 2013 @AP Twitter hack

@AP Twitter feed hacked.

Within seconds,
Dow plunged 140 points.

Recovered in 6 minutes.

S&P 500 temporarily lost
$136B in market cap!

Oops.

The 2013 @AP Twitter hack

The rapid fire trading also highlights the role of computers and algorithmic trading on Wall Street. “That goes to show you how algorithms read headlines and create these automatic orders — you don’t even have time to react as a human being,” said Kenny Polcari of O’Neill Securities, on *Power Lunch*. “I’d imagine the SEC’s going to look into how this happens. It’s not about banning computers, but it’s about protection and securing our markets.”

http://www.cnbc.com/id/100646197

NLU: Traditional organization

- Lexical semantics: meanings of words
- Compositional semantics: meanings of sentences
- Language in context: meanings of dialogues and discourses

Current view of ML founding disciplines