Analysis of Brain Imaging Data With Machine Learning Methods

Francisco Pereira, Tom Mitchell

Computer Science Department
Center for the Neural Basis of Cognition
Carnegie Mellon University

Outline

- What is machine learning?
- Decoding of cognitive states
- Other ML research in brain imaging
- Where next?
- Conclusions

Outline

- What is machine learning?
- Decoding of cognitive states
- Other ML research in brain imaging
- Where next?
- Conclusions

- Please, please, please, feel free to ask questions

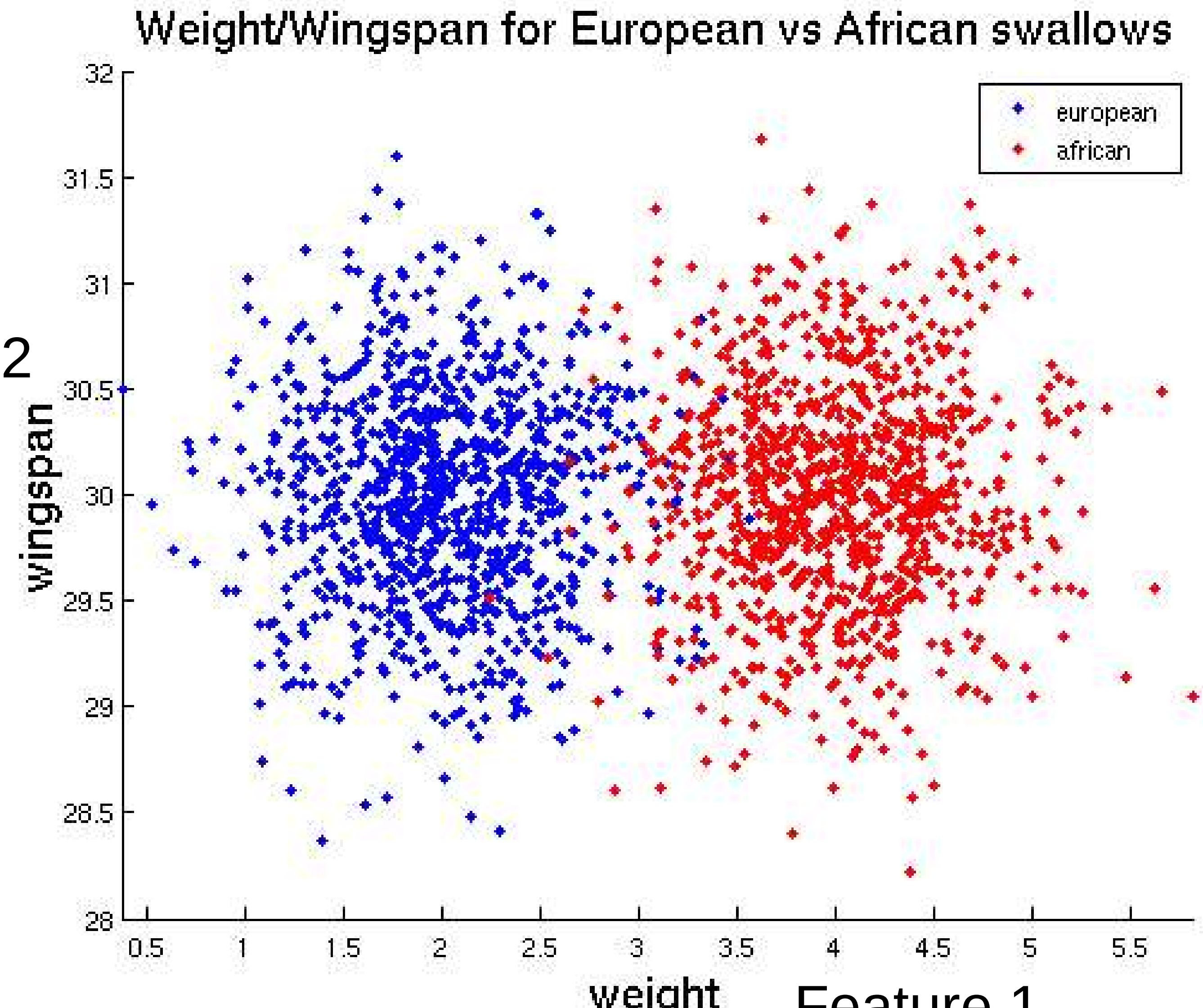
What is Machine Learning?

Computer programs that automatically use experience to improve their performance at some task

What is Machine Learning?

- Experience – examples observed in the past, divided into groups (classes)
- Example – a set of values for several characteristics (features)
- Task – decide which group a new example belongs to (classification)
- Performance - % of classification error

Classification example 1



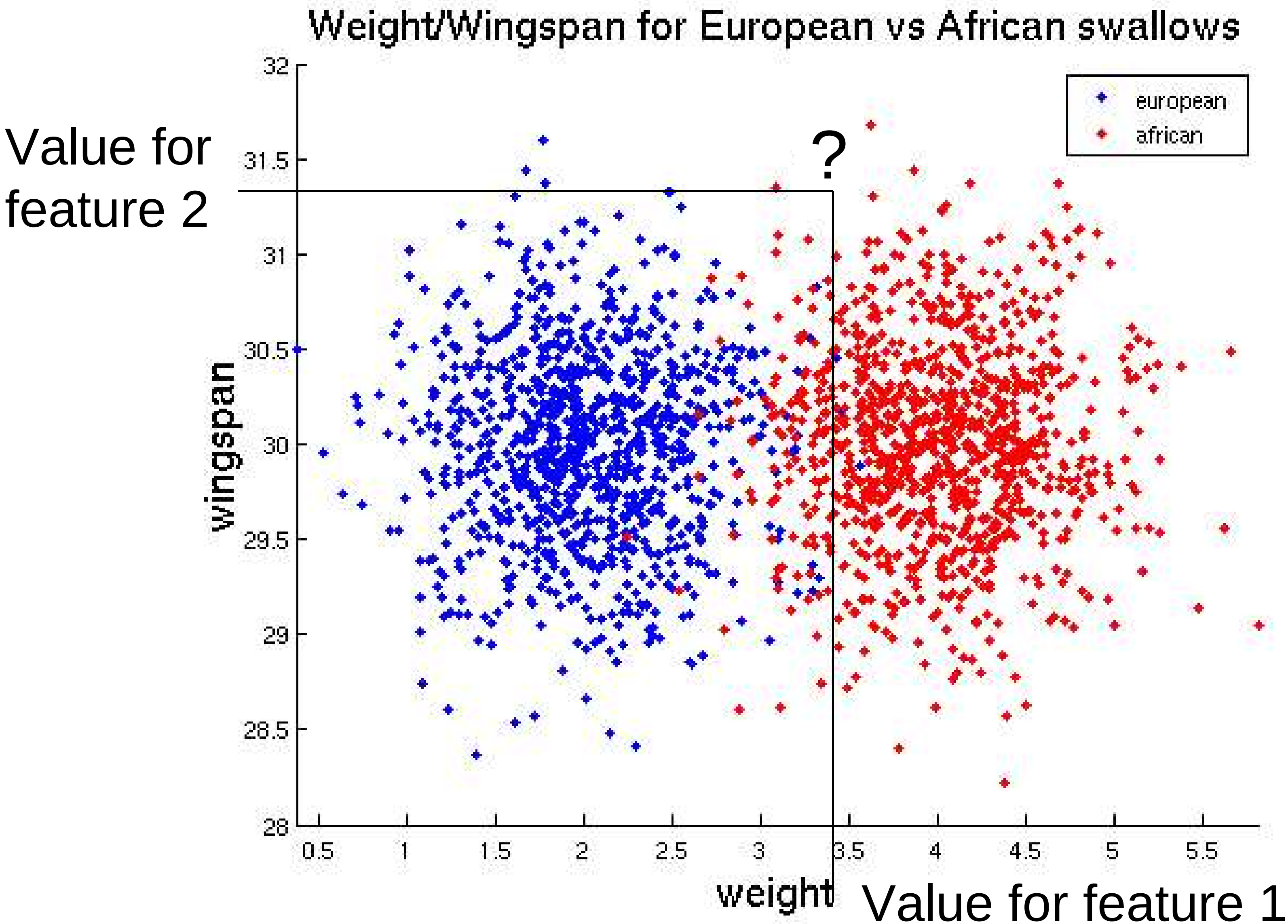
Classes

Feature 2

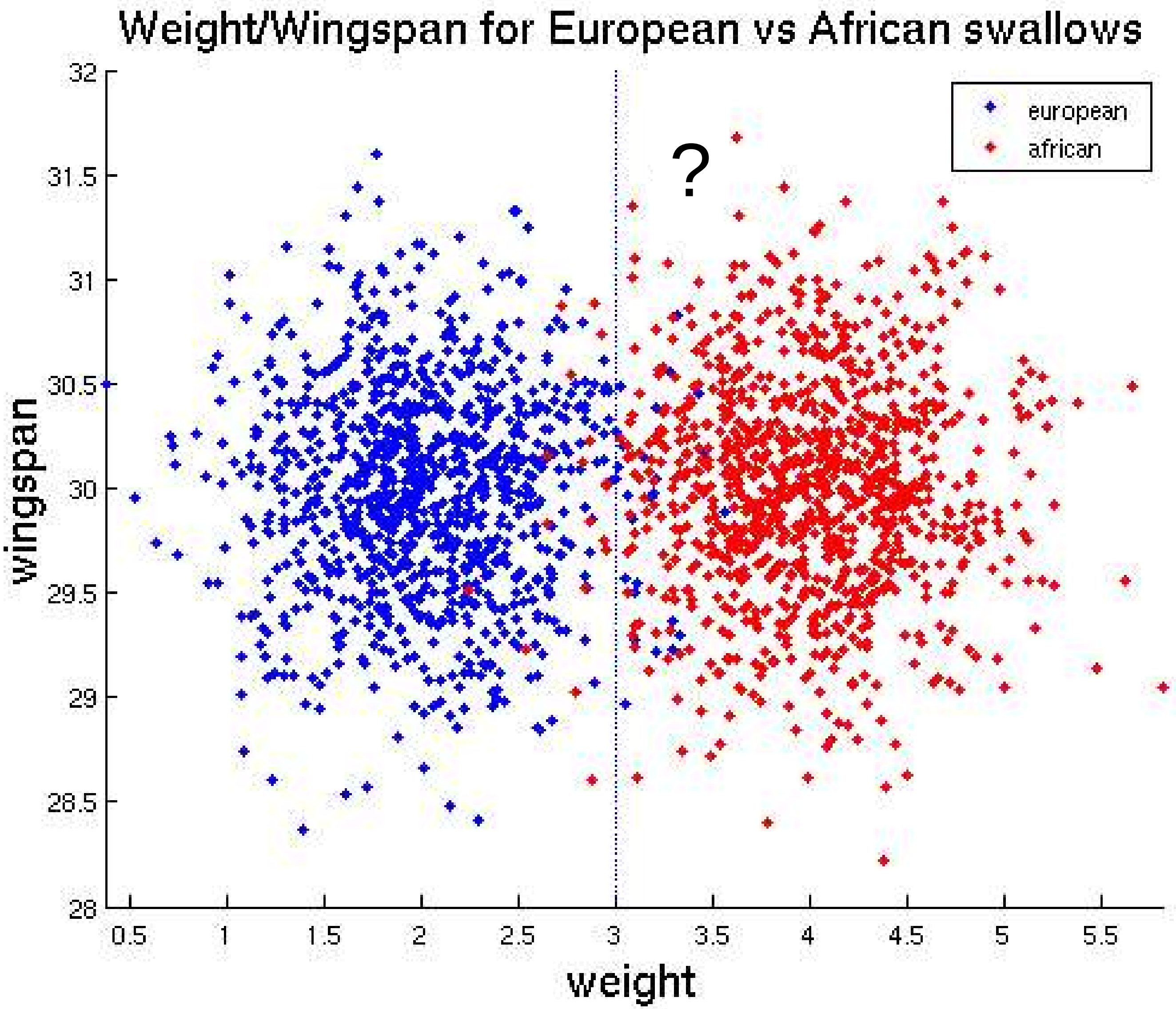
Training examples

Feature 1

Classifying a new example



What does this classifier do?



What does the classifier do?

Gaussian naïve bayes classifier

- learns the means and variances of each cluster of points, assuming it comes from a bidimensional gaussian distribution.
- classifies new point as *european*, if Pr (european|data) > Pr(african|data),

  or as *african* otherwise.

The means and variances of the distribution over each feature, over examples of a class, are the model for that class

Classification example 2

Task: learn to predict whether a patient will need an emergency cesarian

- Experience – past patient records
- Each record contains 215 features

  (e.g. age, history, exam results)
- Hypothesis:
  - which features have anything to do with the likelihood of a c-section being needed?

C-section prediction example

```
Patient103
Age                         23
First Pregnancy             no
Anemia                      no
Diabetes                    YES
PreviousPrematureBirth      no
Ultrasound                  abnormal
Elective c-section          no
...
Emergency c-section         ?
```

9714 patient records (examples)

each with 215 fields

(features)

```
IF
  no previous vaginal delivery AND
  abnormal 2nd trimester ultrasound AND
  malpresentation at admission
THEN
  probability of emergency c-section is 0.6

Over training data: 0.63
Over test data:     0.6
```

Regularities

condensed into

rules

to predict a

label (yes/no)

with a probability

What does the classifier do?

Decision tree classifier

- finds classification rules:
  - involving as few features as possible
  - that apply to as many examples as possible
- the model learnt is the set of rules

What does any classifier do?

In summary:

- build a model that captures the essential regularities in the data (generative) or the difference between the classes (discriminative)
- … and only those – in the limit the training data is a perfect model of itself
- a good model generalizes to unseen examples

What is Machine Learning?

Typical tasks:

- predicting labels

  (e.g. does a patient have disease X?)
- learning a function

  (e.g. how high will this marker be in 3 months?)
- grouping data

  (e.g. are there natural subsets of patients, according to some indicator?)
- recognizing structures in images

  (e.g. abnormal structure or pattern of activation)

What is Machine Learning?

More technical names:

- predicting labels

  classification

- learning a function

  regression (does not have to be linear)

- grouping data

  clustering

- recognizing structures in images

  detection (e.g. “is this patch a X?”)

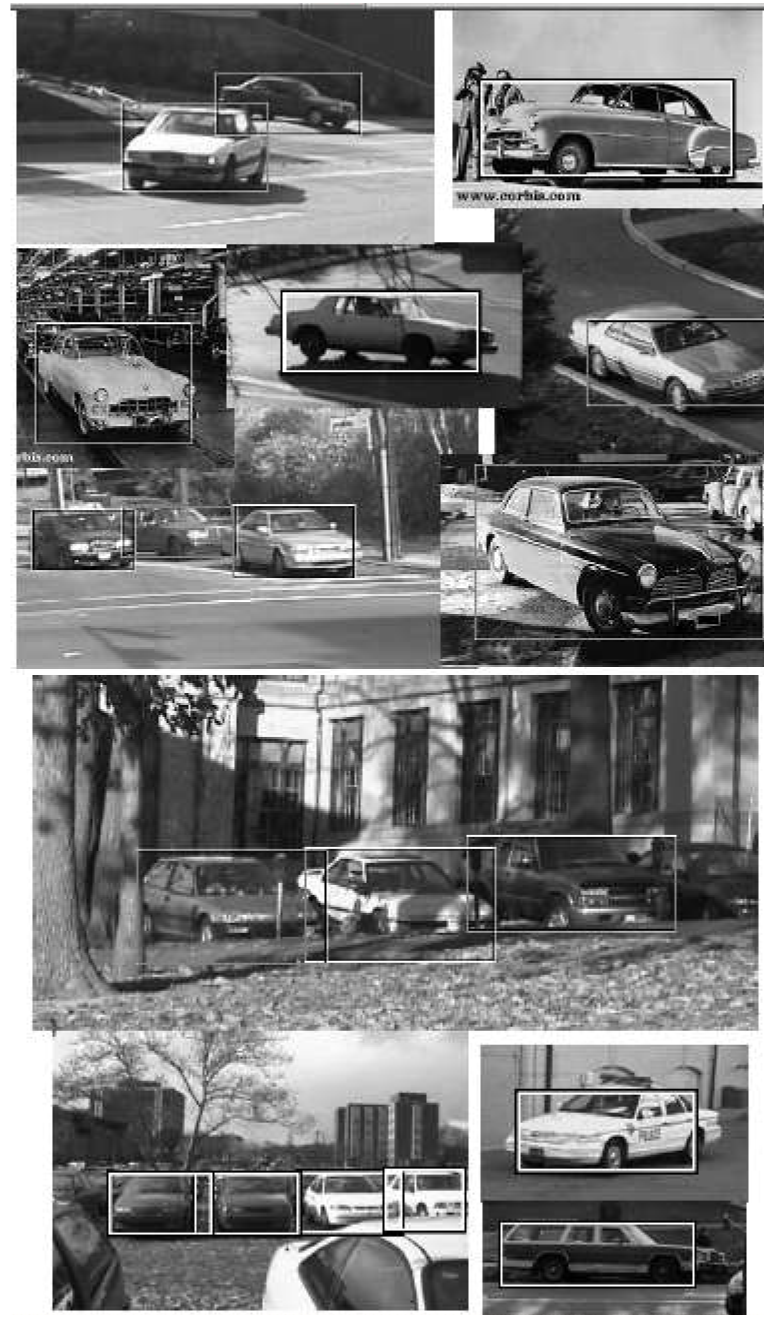
  segmentation (e.g. parcellation of a structural image)

Object Detection

(Henry Schneiderman, CMU)

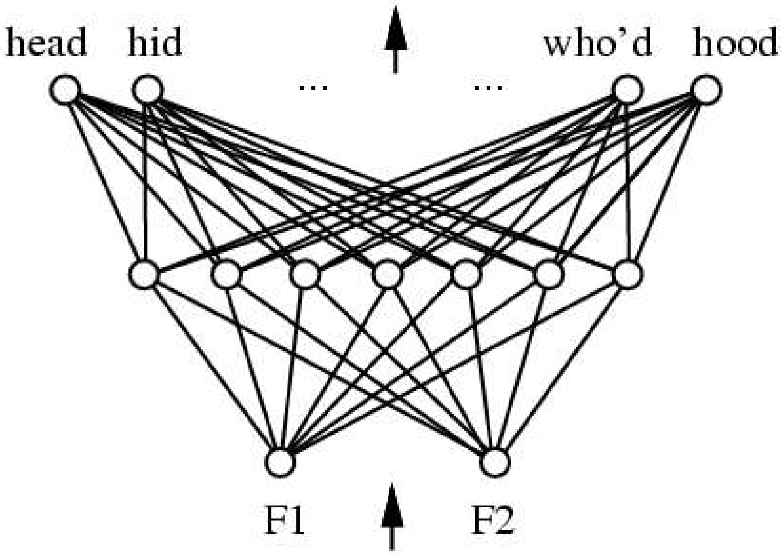
Figure 2. Examples of training images for each car orientation

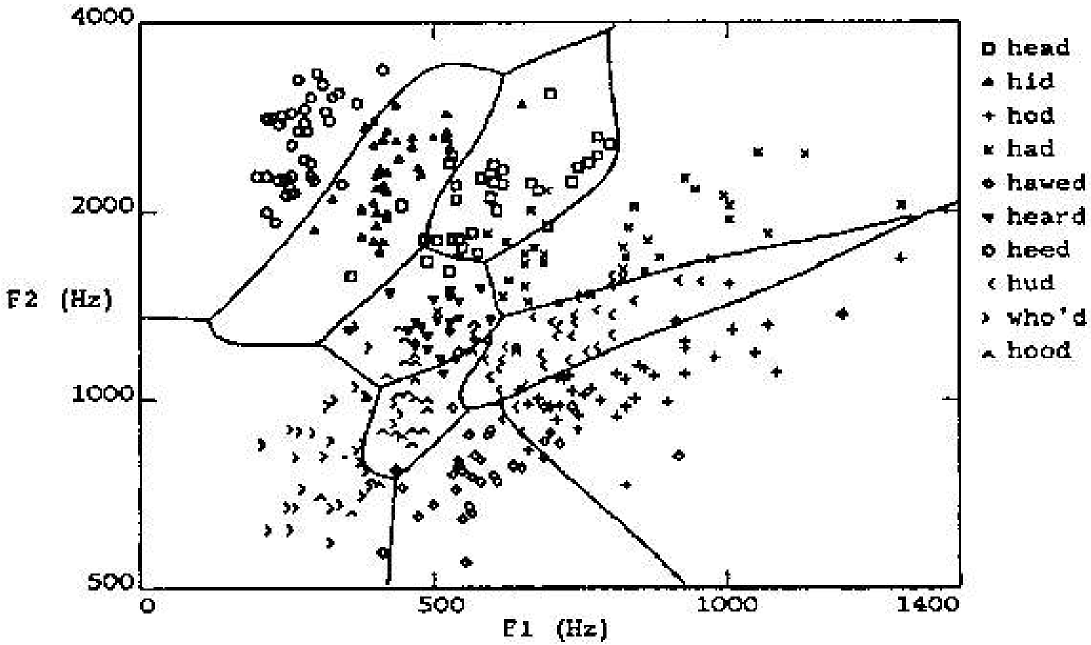
Example training images for each orientation, the features are pixels on the image.



Classification/Regression

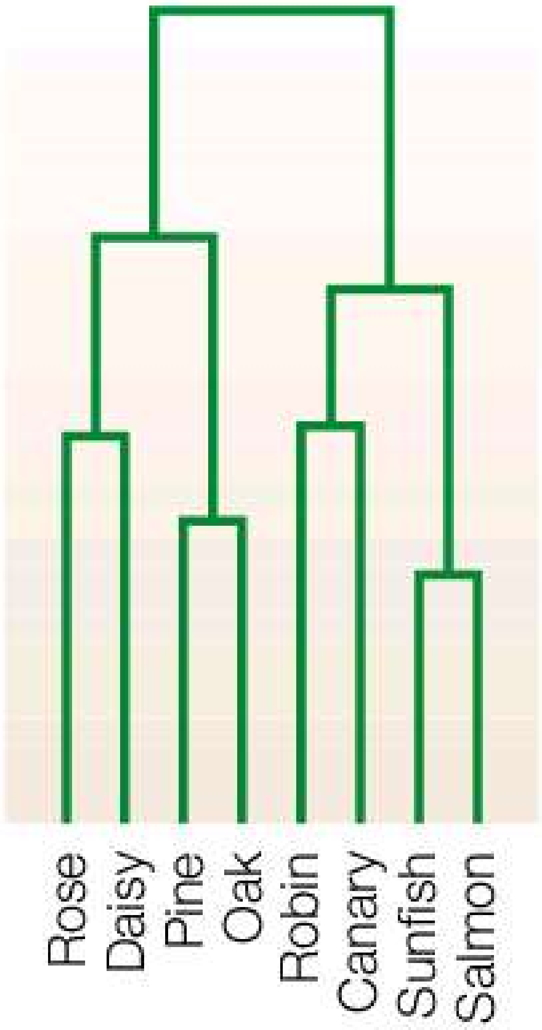
Artificial neural network for speech recognition





The features are power of a frequency in the speech signal

Clustering



[McClelland, Nature Neuroscience 2003]

- Groups items with their neighbours according to their similarity under some measure
- Does it hierarchically, joining groups with other groups
- Can be applied to anything that supports a similarity measure (e.g. images or time series)

Markov models

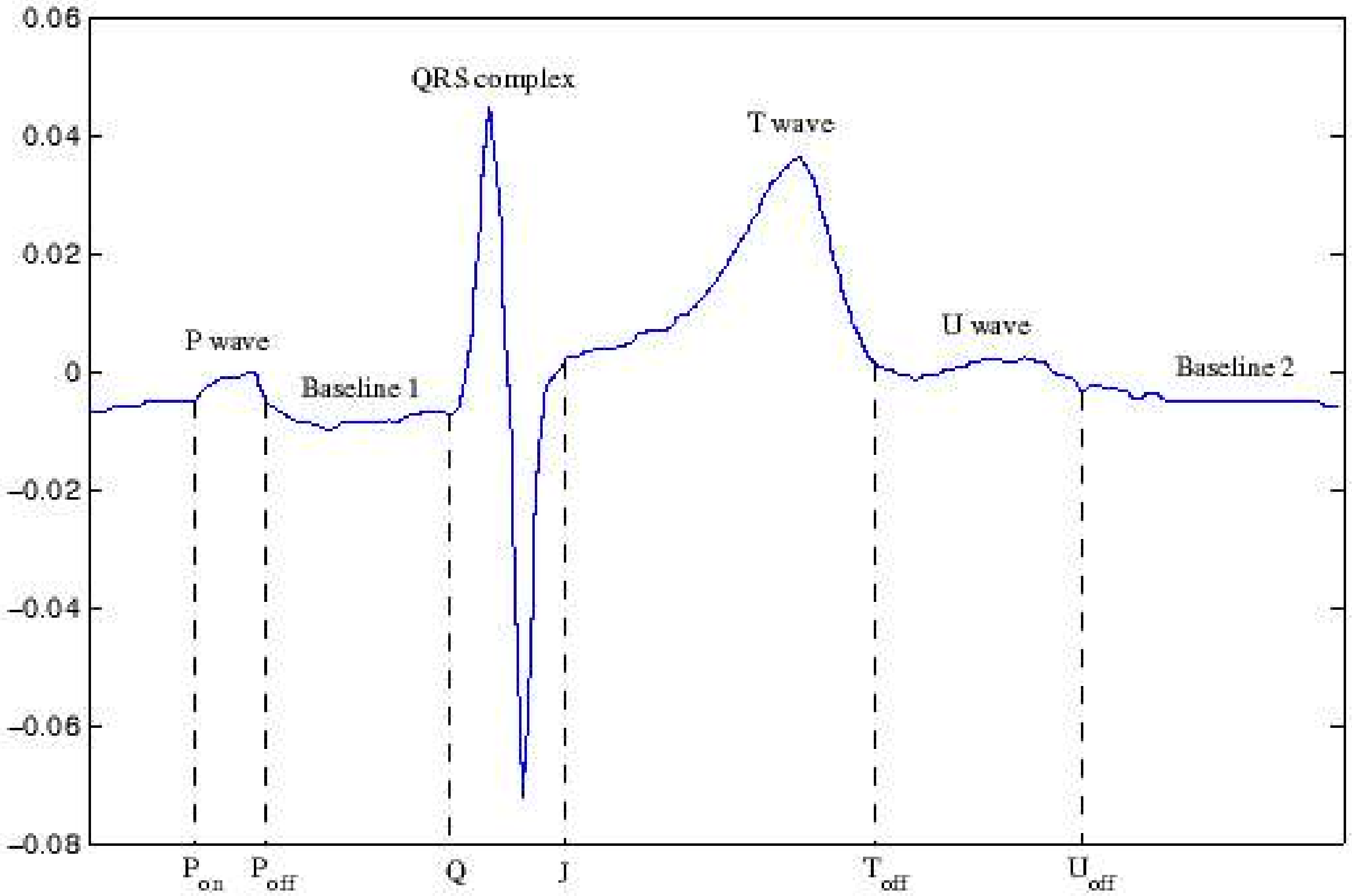
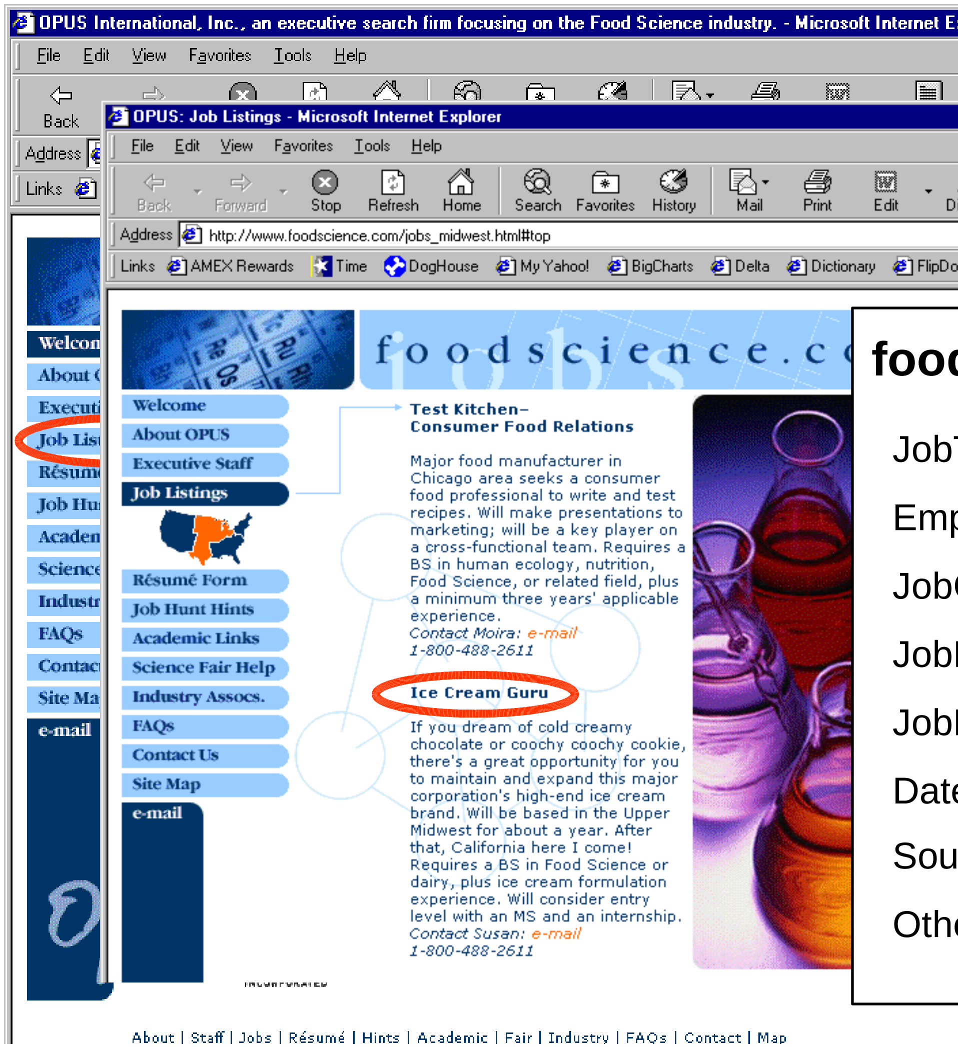


Figure 1: A human ECG waveform.

[Nicholas Hughes, NIPS 2003]

Text mining

Extraction of formatted information from text on web pages



These are job postings, see www.flipdog.com

foodscience.com-Job2

JobTitle: Ice Cream Guru

Employer: foodscience.com

JobCategory: Travel/Hospitality

JobFunction: Food Services

JobLocation: FL-Deerfield Beach

DateExtracted: January 8, 2001

Source: www.foodscience.com/jobs_midwest.html#top

OtherCompanyJobs: foodscience.com-Job1

Outline

- What is machine learning?
- Decoding of cognitive states
- Other ML research in brain imaging
- Where next?
- Conclusions

Decoding of cognitive states

Cognitive psychology studies

[Marcel Just and colleagues, CMU]

- Language and spatial position judgements
- Semantic categorization

Goals:

- Train "virtual sensors" of cognitive states, classifiers of the form

  fMRImage([time interval]) => state

[AMIA2003, also to appear in the *Machine Learning Journal*]

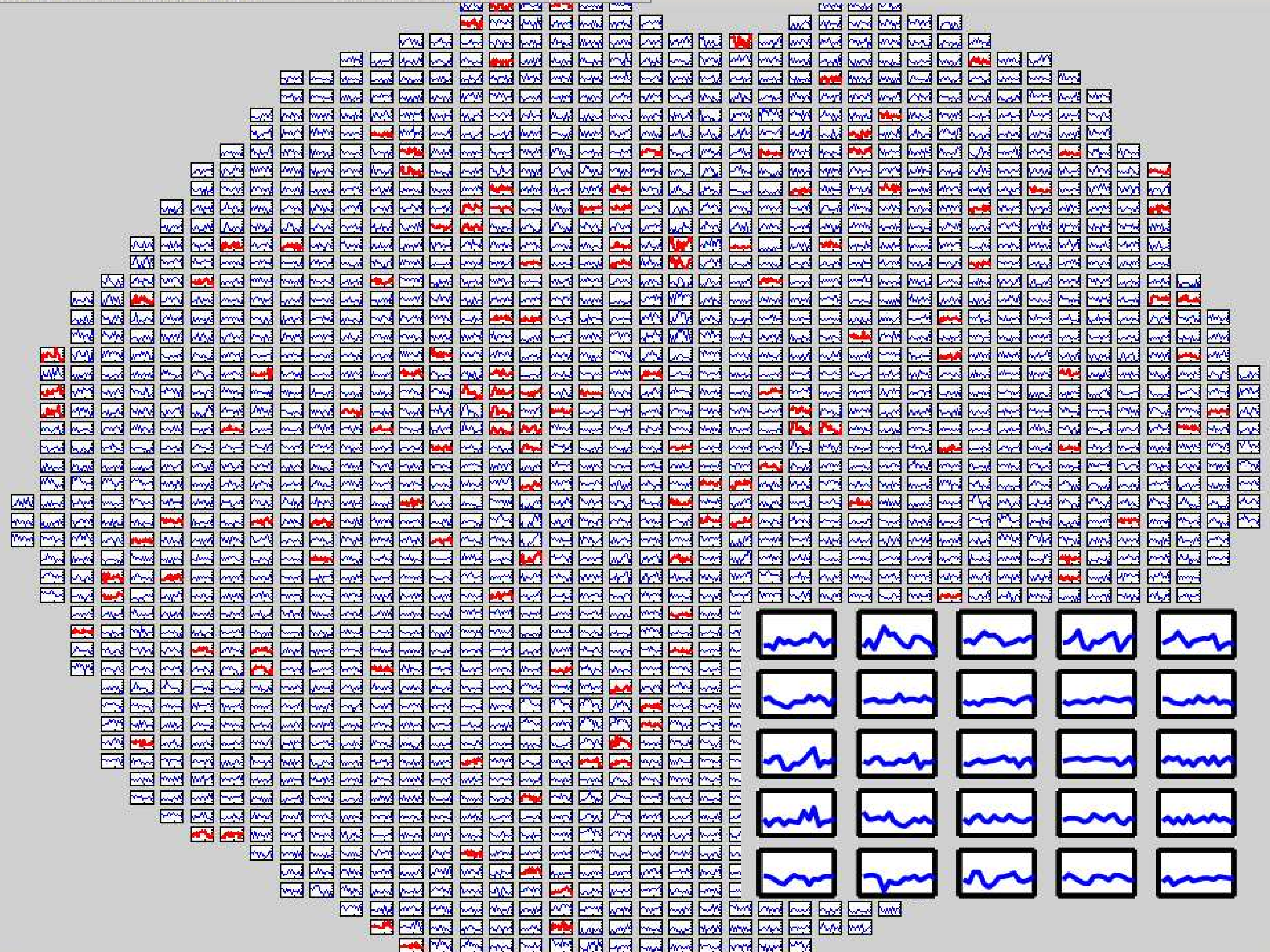
Approach

Classifiers considered:

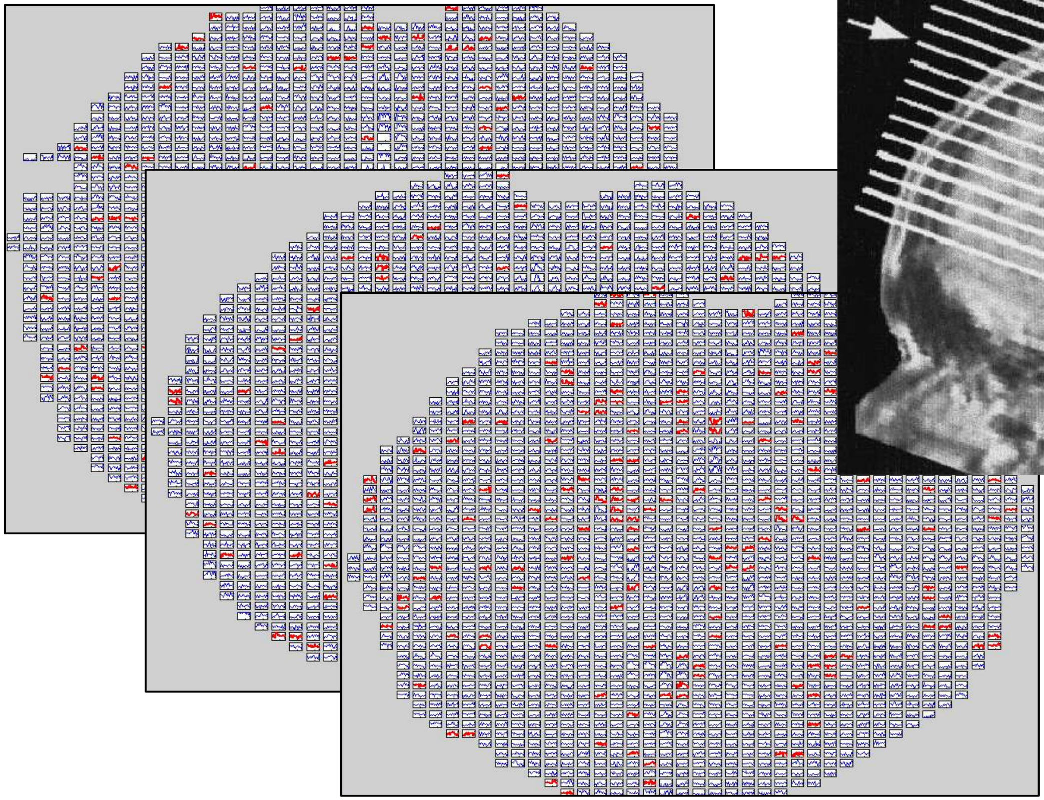
- Gaussian Naïve Bayes (GNB)
- Linear kernel Support Vector Machine (SVM)

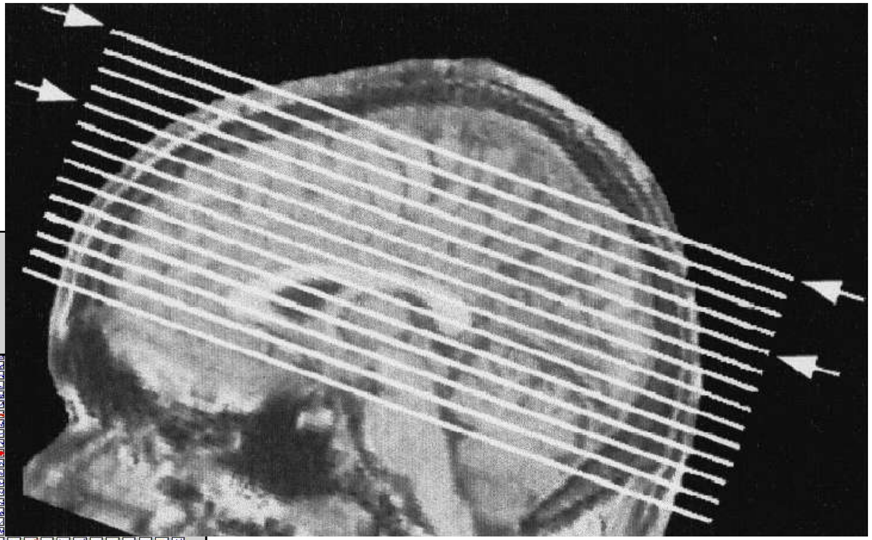
Feature selection/abstraction

- Select subset of voxels (by mean signal, by anatomy)
- Select subinterval of time
- Normalization of voxel activities



From Images to Examples





~10,000 voxels per 3D image

~8 3D images in each trial

→ 80,000 feature (voxel at time t) vector as an example

Study 1 – Pictures and Sentences

It is true that the star is above the plus?

+

---

*

Study 1 - Pictures and Sentences

- Picture presented first in half of trials, sentence first in other half
- One image every 500 msec
- 12 normal subjects
- Three possible objects: star, dollar, plus

Study 1 – Pictures and Sentences

Learn

fMRImages(t, …, t+8) $\rightarrow$ {Picture, Sentence}

- 80 examples

  (40 pictures and 40 sentences, first and second)
- ~10,000 voxels

Classifiers:

- GNB and SVM
- Using <n> voxels most active in at least one class

Study 1 - Results

Results obtained via cross-validation

- pick 1 example, train on the others, test on it
- repeat for all examples, average results

Results (leave one out)

- Random guess = 50% accuracy
- Gaussian Naïve Bayes: 82% accuracy
- Support Vector Machine: 89% accuracy
  (average across subjects)
- feature selection important for both,
  worked best using 240 most active voxels

Study 2: Word Categories

Subjects decide whether words belong to one of the following semantic categories

- Family members
- Occupations
- Tools
- Kitchen items
- Dwellings
- Building parts
- 4 legged animals
- Fish
- Trees
- Flowers
- Fruits
- Vegetables

Study 2: Word Categories

Stimulus:

- 12 blocks of words:
  - Category name (2 sec)
  - Word (400 msec), Blank screen (1200 msec); answer
  - Word (400 msec), Blank screen (1200 msec); answer
  - …
- Subject answers whether each word in category
- 32 words per block, nearly all in category
- Category blocks interspersed with 5 fixation blocks
- 10 normal subjects

Study 2: Word Categories

Learn

fMRImage(t) $\rightarrow$ word-category(t)

Classifier:

- Gaussian Naïve Bayes
- Select <n> voxels most active in at least one category

Study 2 - Results

Results obtained via cross-validation

- pick 1 example, train on the others, test on it
- repeat for all examples, average results

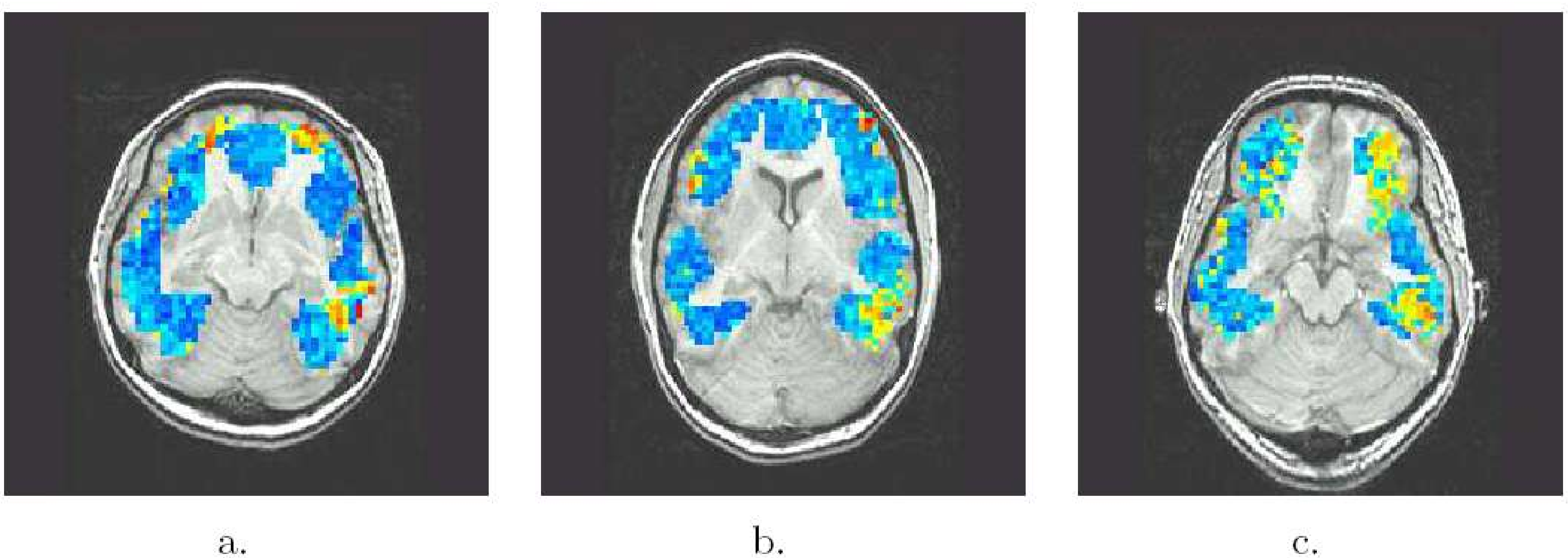
Results (leave one out)

- Classifier outputs a ranked list of classes
- Error: fraction of classes ranked ahead of true
- Gaussian Naïve Bayes: 91% accuracy
  (on average across subjects)
- feature selection important, worked best using 1600 most active voxels

OK, but does this have any use?

- Discover what parts of the brain contain information relevant to a classifier that works, by looking at the model.
- Ask questions other than “where is the activation?”
- Figure outif a certain transformation of the data (e.g. averaging across time) still contains enough information to drive a classifier.

Where is the information?



Voxelwise predictability in an inferior slice (3 subjects).

All of these get added up in the classifier's decision.

Outline

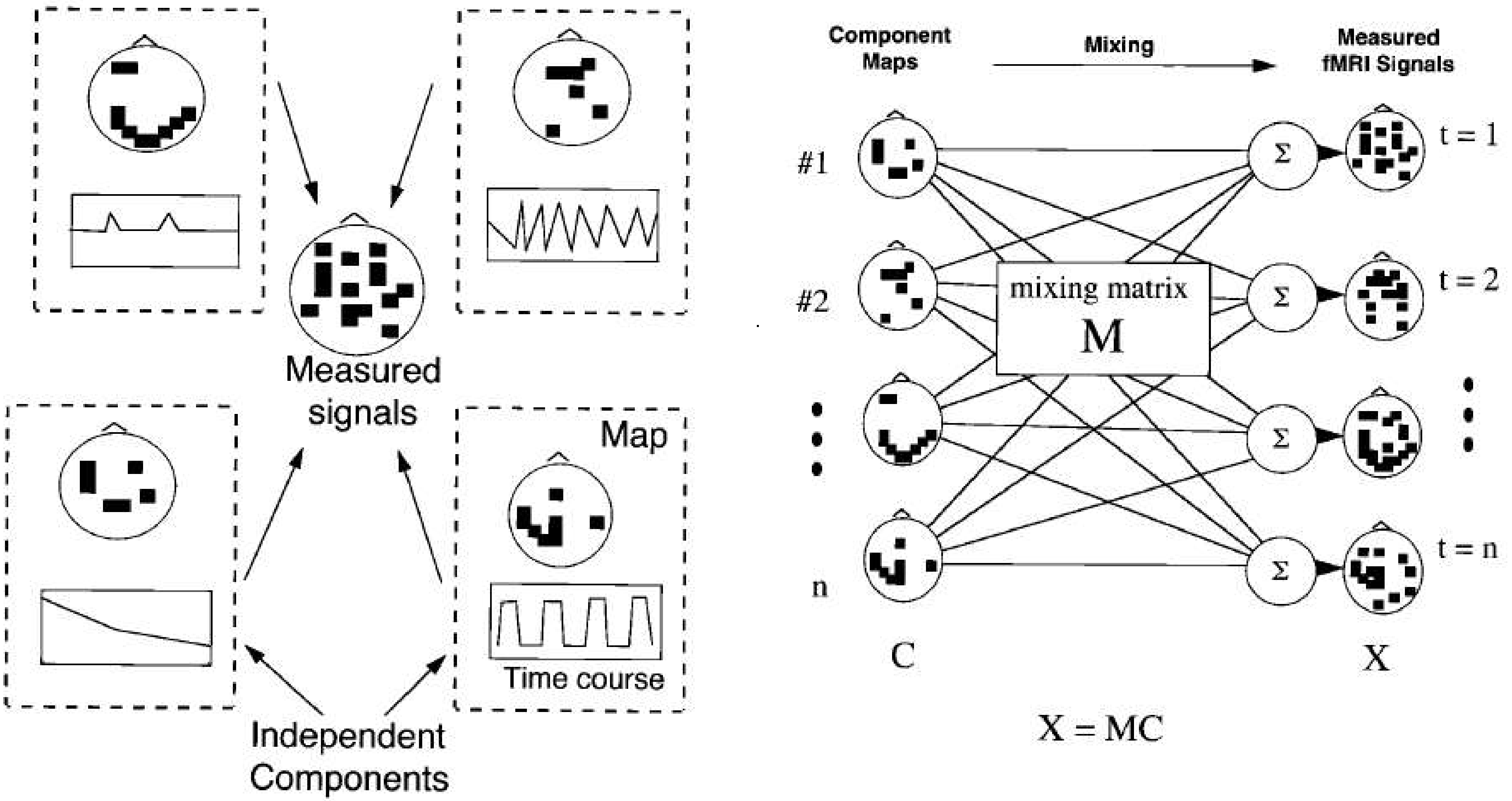
- What is machine learning?
- Decoding of cognitive states
- Other ML research in brain imaging
- Where next?
- Conclusions

Independent Component Analysis

- ICA discovers statistically independent component images that combine to form the observed fMRI signal
- ICA is a data-driven approach, complementary to hypothesis-driven methods (e.g. GLM)
- Requires no a-priori knowledge about hemodynamics, noise models, time-courses of subject stimuli,…

Spatial ICA decomposition

[McKeown et al. Human Brain Mapping 98]



Spatially independent components, different time courses

Sample ICA decomposition

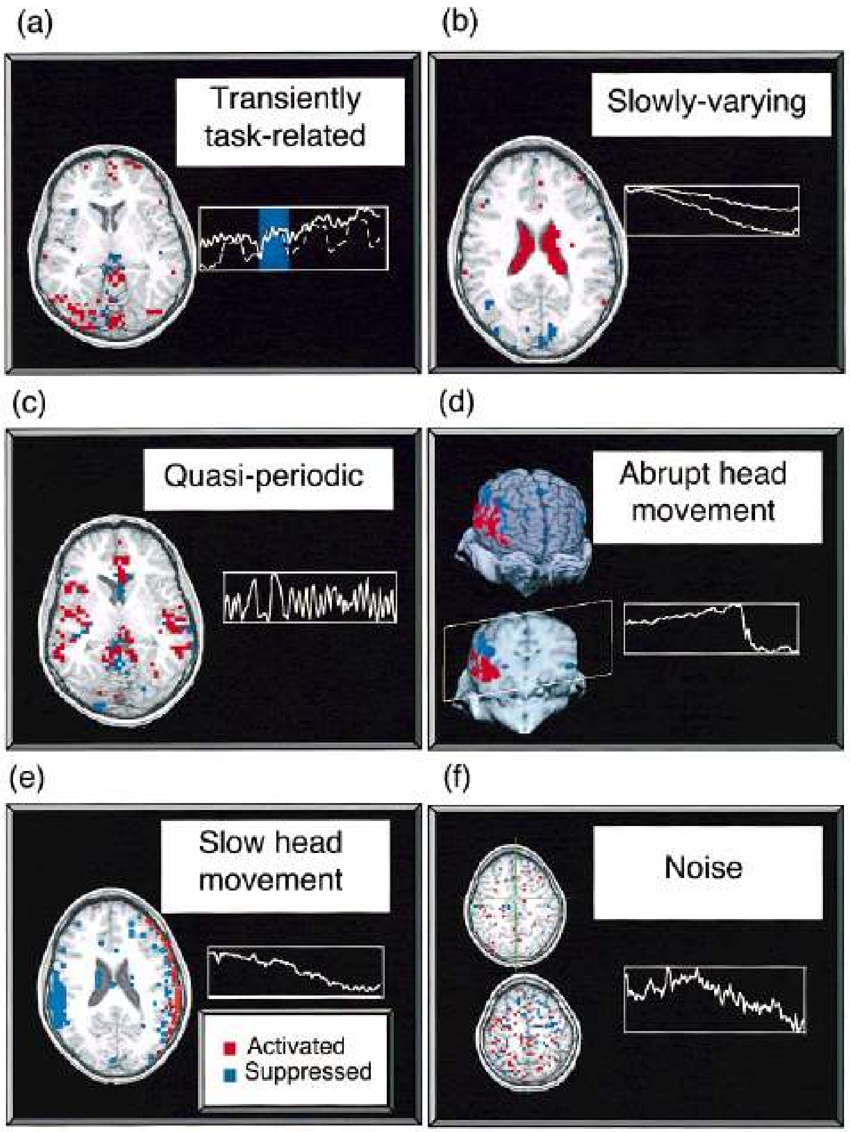
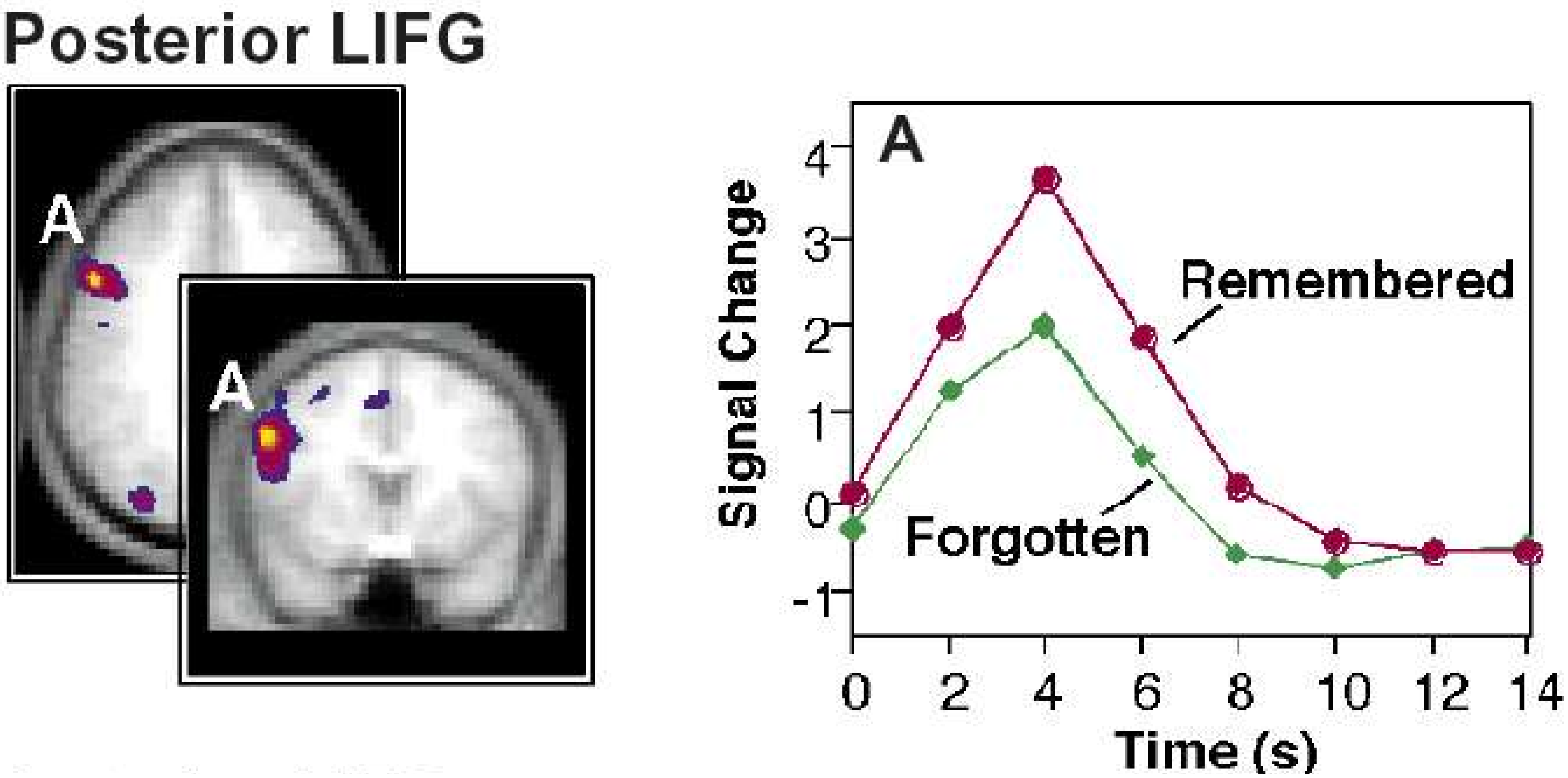


Figure 15.

The time course of component a) is the only one that shows the same periodicity as the task trials.

Verbal Remembering and Forgetting Prediction

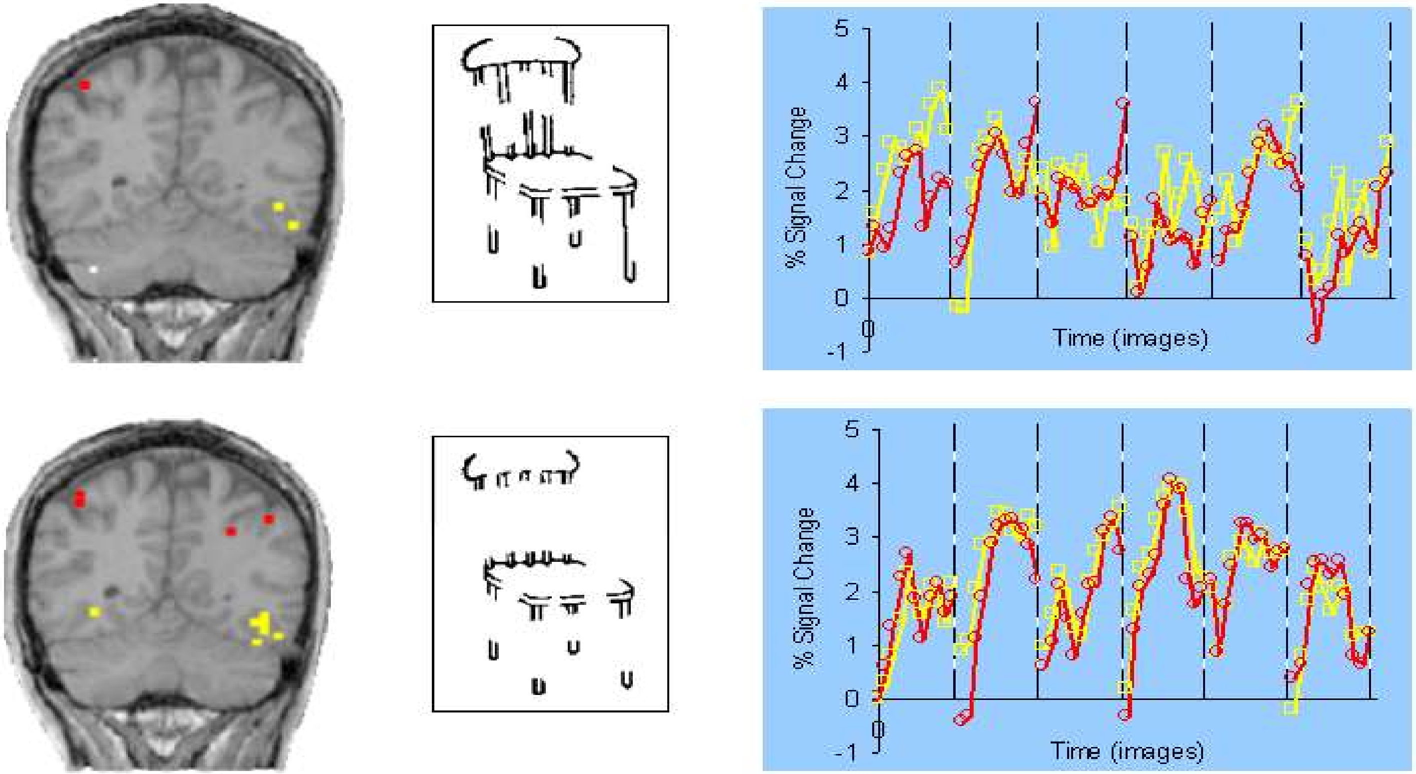
[Wagner et al., *Science*, 1998]



Activation in left inferior frontal gyrus predicts remembering

Functional connectivity

(from Diwadkar, Carpenter, & Just, 2001)

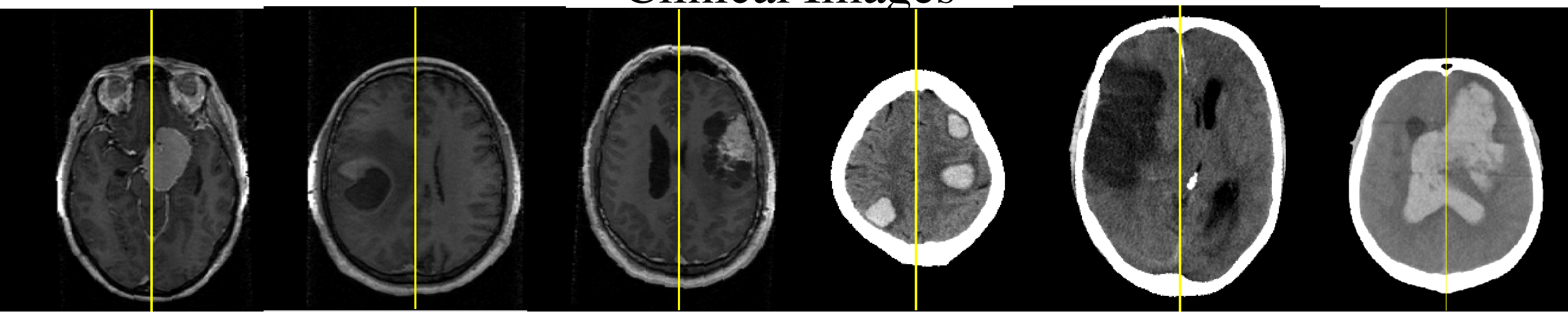


The average signal in two cortical areas (**parietal/dorsal** and **inferior temporal/ventral**) becomes more synchronized as the object recognition task becomes more difficult.

Robust Midsagittal Plane Extraction from Coarse, Pathological 3D Images

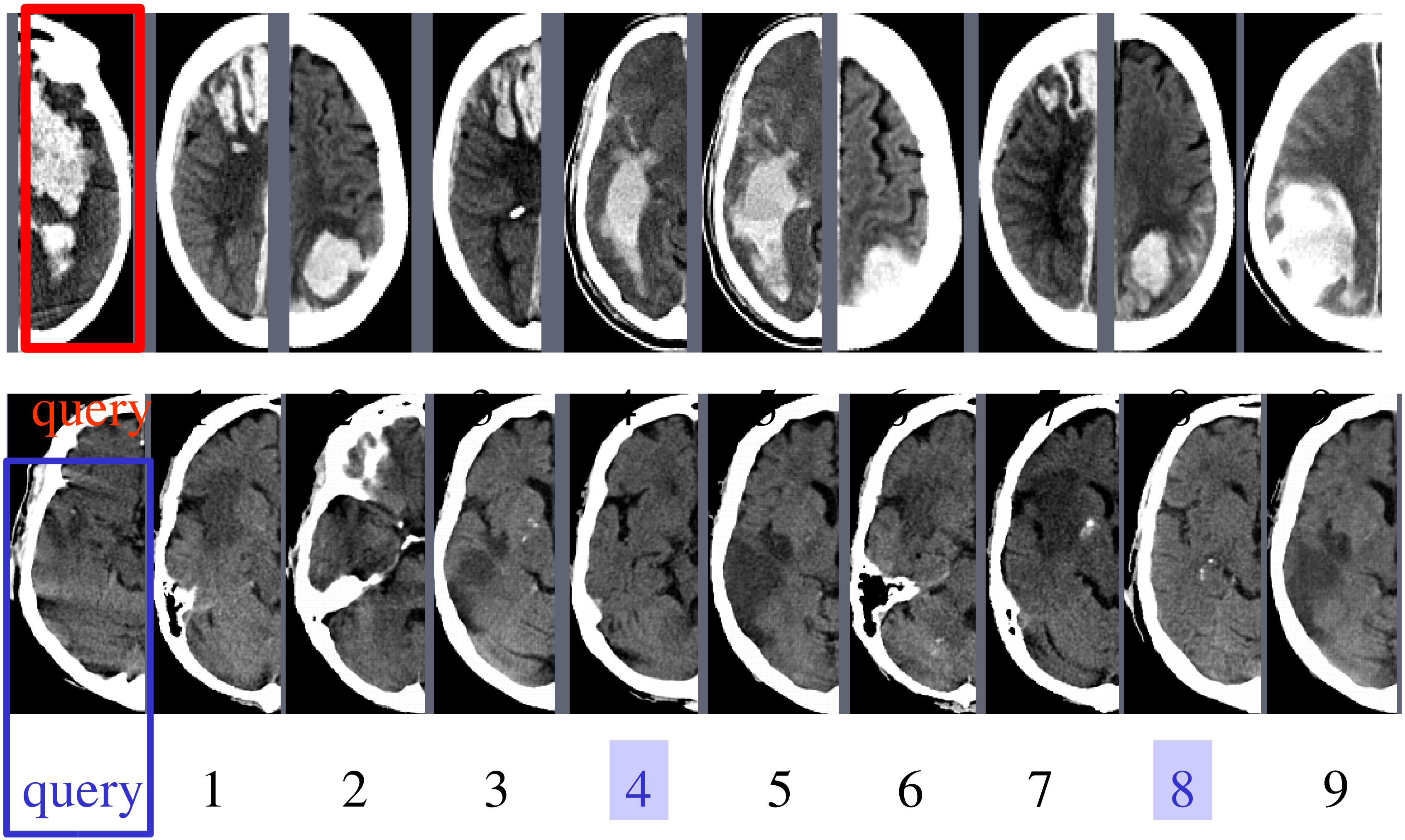
Liu et al (2001) Robust Midsagittal Plane Extraction From Normal and Pathological 3D Neuroradiology Images IEEE Transactions on Medical Imaging 20(3)

Clinical Images



Testing Images

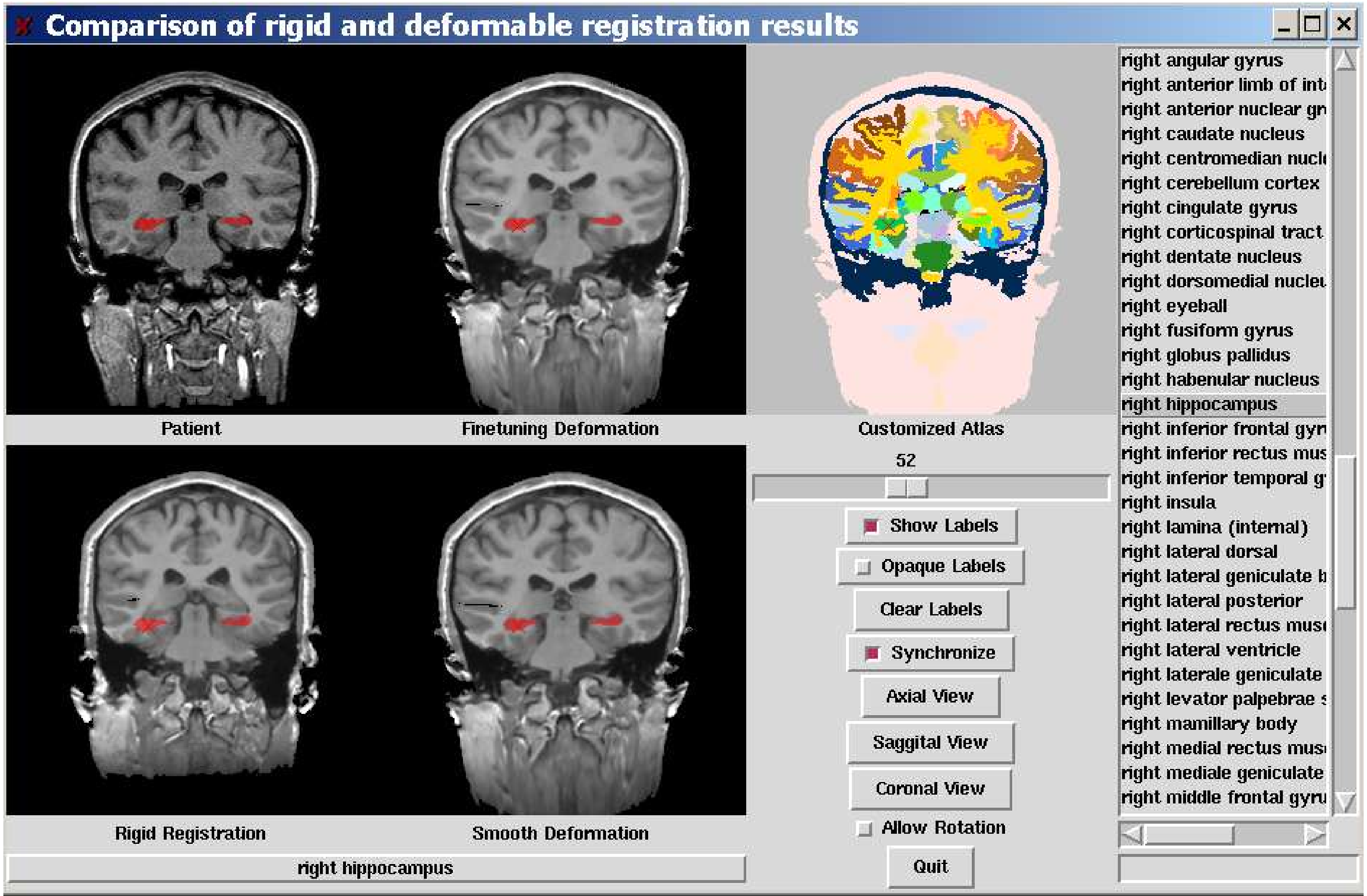
Image Retrieval



Liu et al (2001) Classification-Driven Pathological Neuroimage Retrieval Using Statistical Asymmetry Measures

Deformable Registration for Segmentation using Brain Atlas

[Yanxi Liu, CMU]



Outline

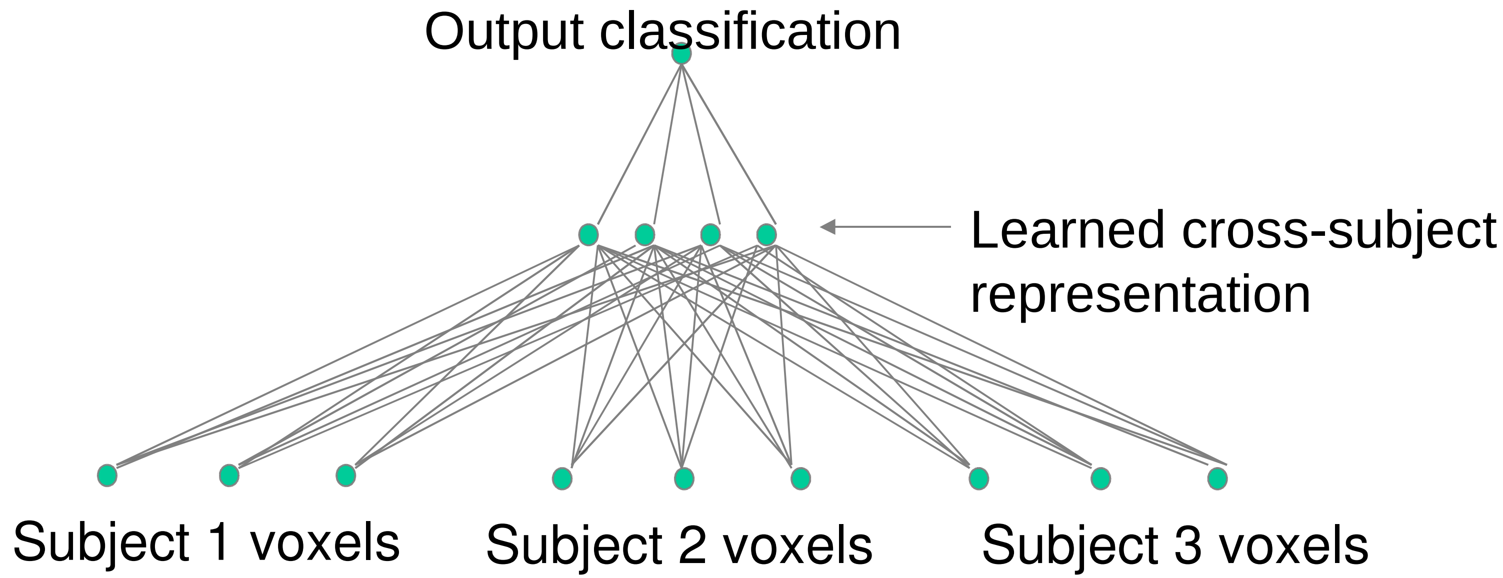
- What is machine learning?
- Decoding of cognitive states
- Other ML research in brain imaging
- Where next?
- Conclusions

What next?

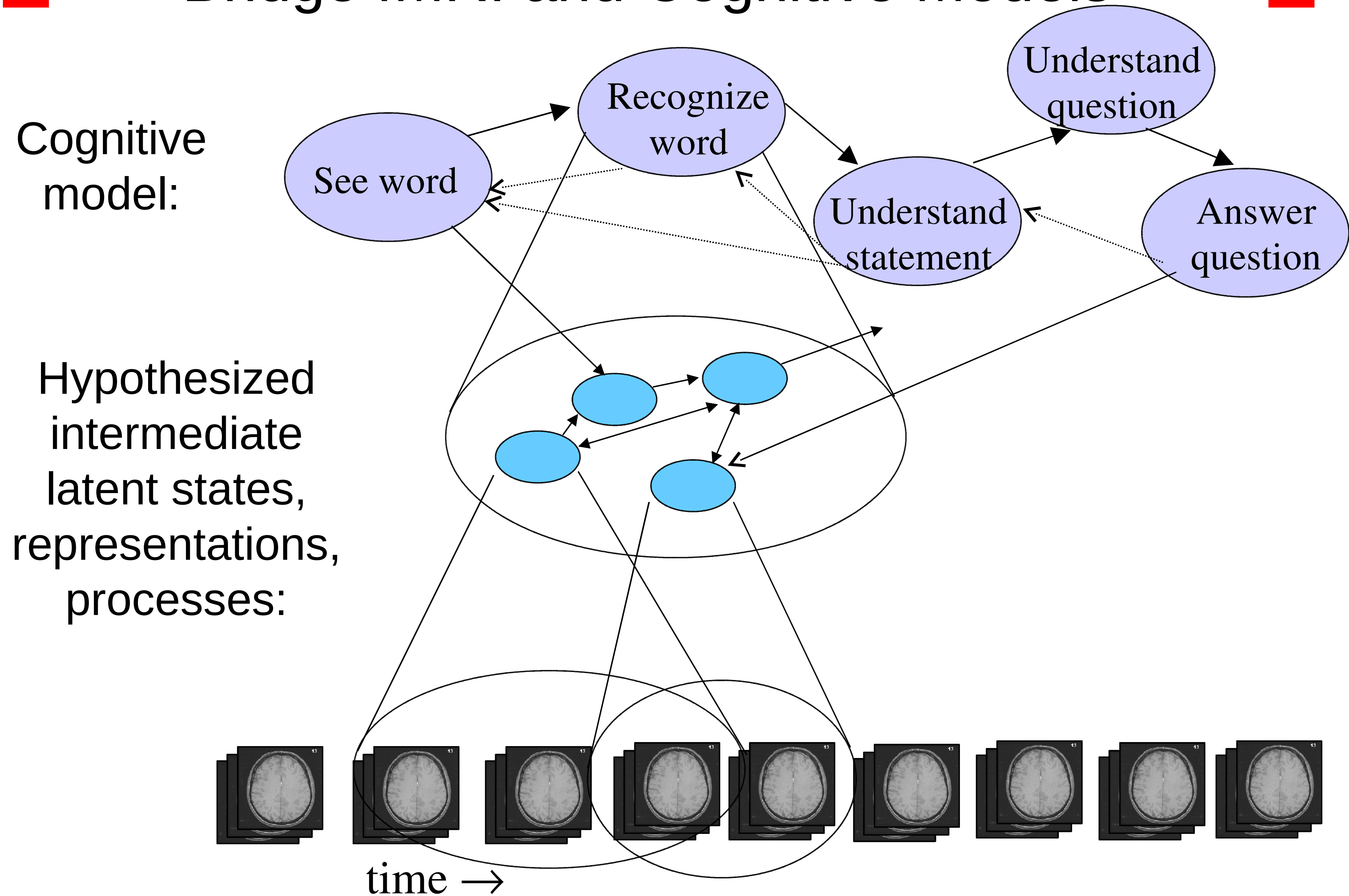
- Better feature selection strategies
- Understand how to use prior knowledge from cognitive neuroscience to help train classifiers
- Learn classifiers that work across multiple subjects
- Learn to track *sequences* of cognitive states and cognitive *processes* over time

Cross Subject Classifiers

Train a neural network to discover intermediate data abstraction across multiple subjects.

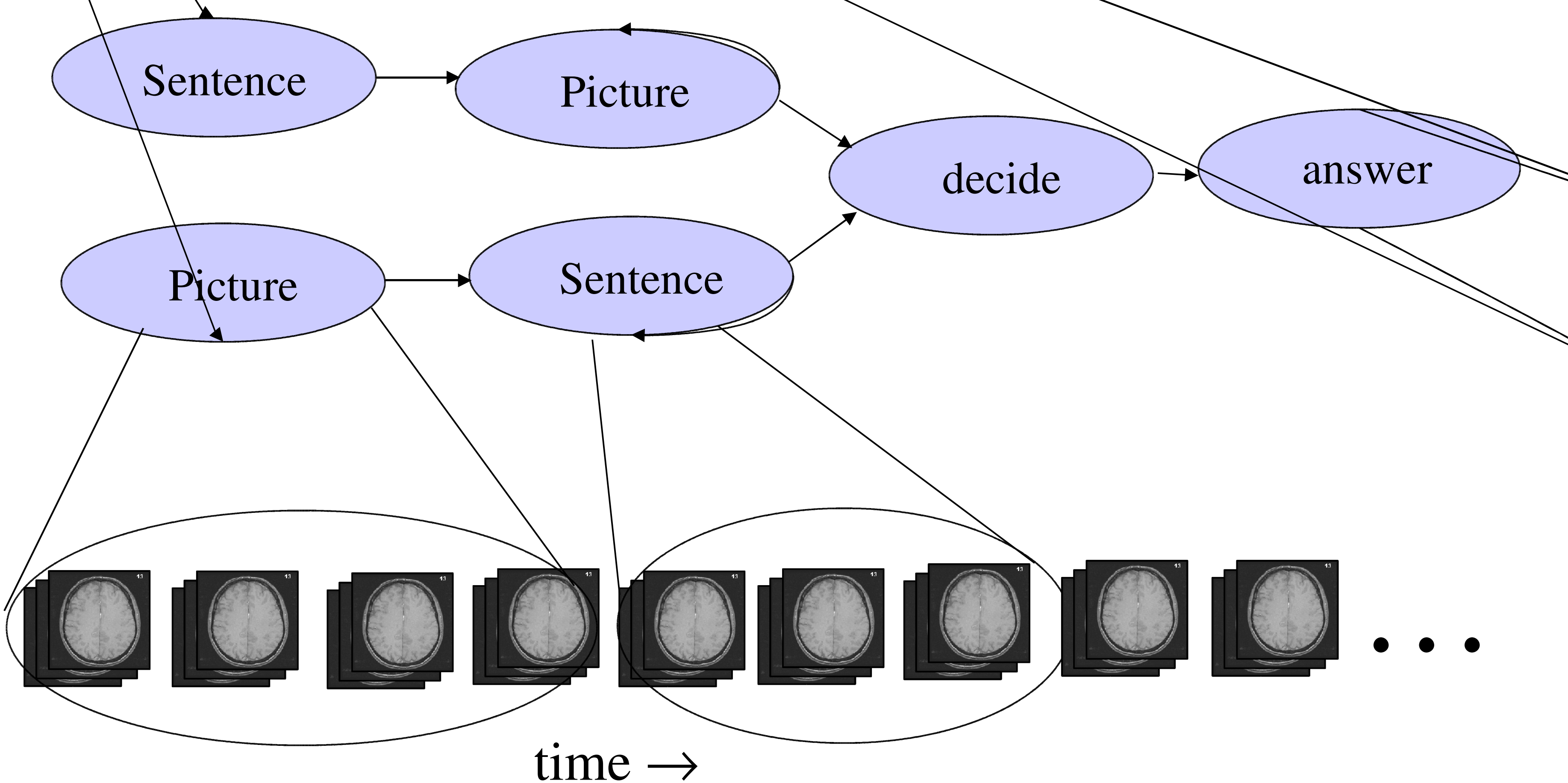


Bridge fMRI and Cognitive Models



Hidden Markov Models to track sequences of cognitive states

[Rebecca Hutchinson,CMU]



The End

If you'd like references on any of this, please email fpereira@cs.cmu.edu

Summary

- Successful training of classifiers for distinguishing cognitive states in four studies
- Feature selection and abstraction are essential
  - Introduced new “zero-signal” learning setting
- Cross-subject classifiers trained by abstracting to anatomically defined ROIs
- Key question: how broad a context, and how broad a range of subjects, can we train classifiers for?

Lessons Learnt

Classifiers work

- Picture vs. sentence, Semantic categories of nouns, Ambiguous vs. unambiguous sentences.
- For picture vs sentence, it even works across subjects.

Can we devise learning algorithms to construct such “virtual sensors”?

- High dimension, sparse data → difficult
- New, more accurate methods possible

The End

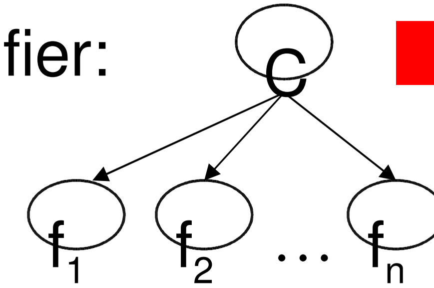
Lessons Learned

Yes, one can train machine learning classifiers to distinguish a variety of mental states

- Picture vs. Sentence
- Noun semantic categories
- Ambiguous sentence vs. unambiguous
- Noun vs. Verb

Failures too:

- True vs. false sentences
- Negative vs. affirmative sentences



Learning a Gaussian Naïve Bayes (GNB) classifier:

For each class value, c_i

3. Estimate $\widehat{P}(c_i)$

4. For each feature f_i and possible value v_k estimate

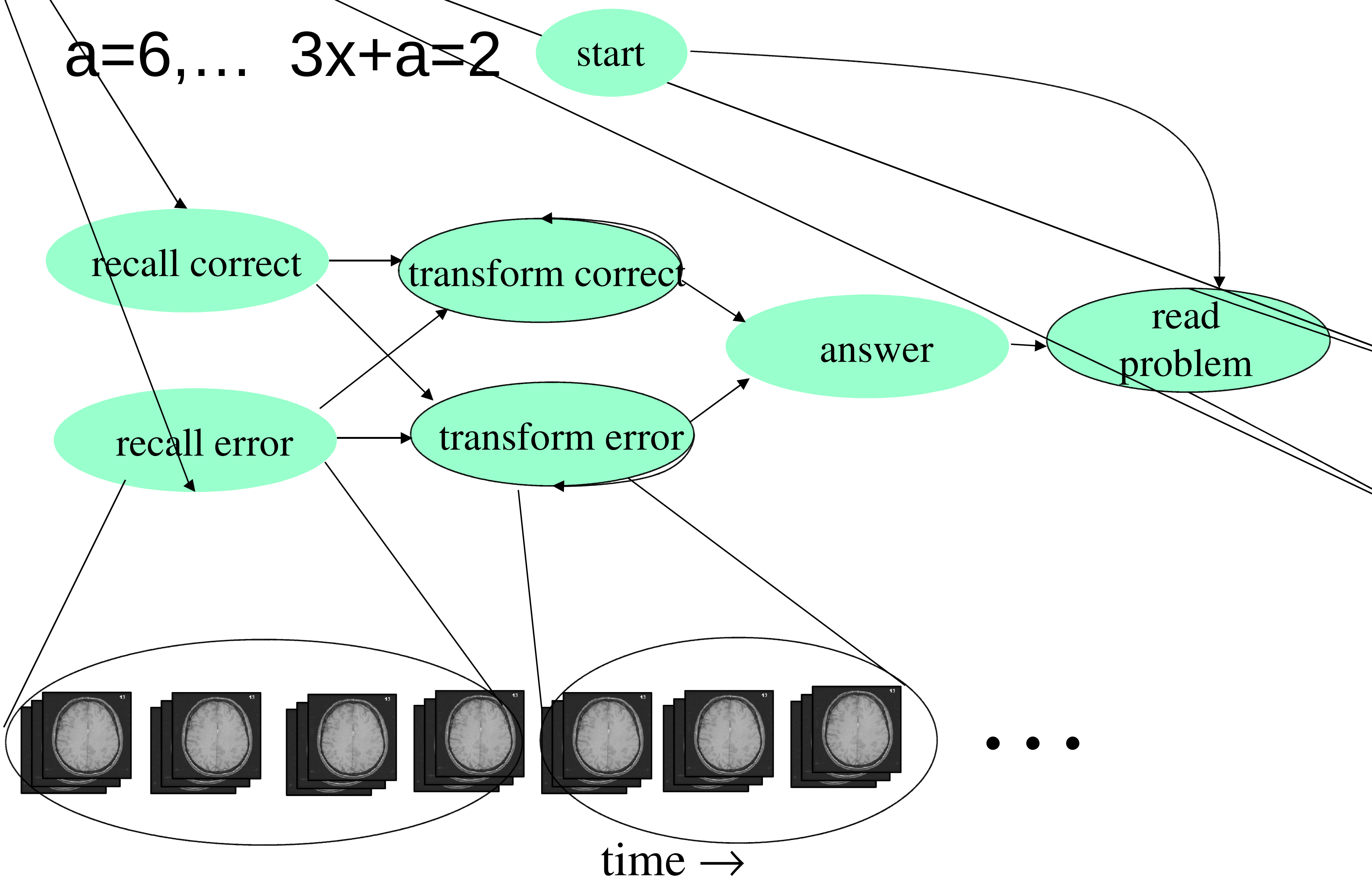
$$\widehat{P}(f_j = v_k | c_i)$$

modeling distrib. for each c_i , f_j, as Gaussian, $N(\widehat{\mu}_{ij}, \widehat{\sigma}_{ij}^2)$

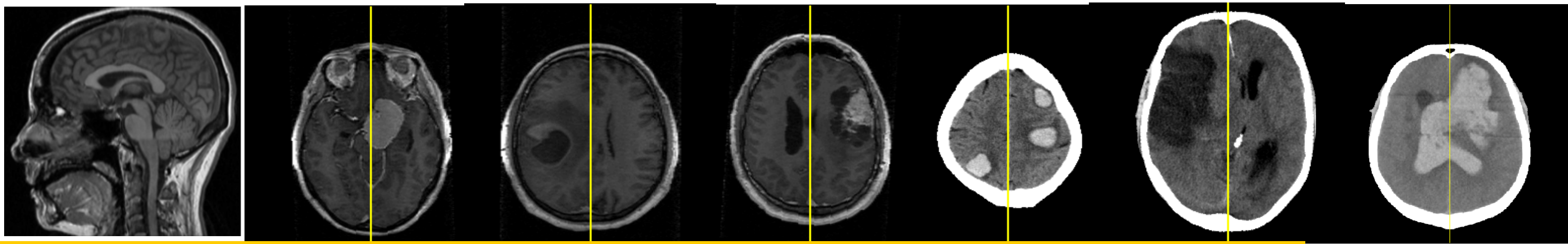
Applying GNB classifier to new instance $\langle v_1, v_2, ... v_n \rangle$

$$c \leftarrow \arg\max_{c_i} \widehat{P}(c_i) \prod_j \widehat{P}(f_j = v_k | c_i)$$

Example: HMMs to track algebra solving



Classification-Driven Pathological Neuroimage Retrieval Using Statistical Asymmetry Measures



3D image registration and image feature extraction/selection

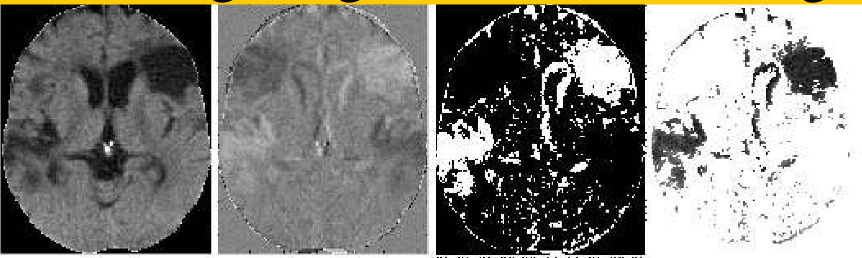
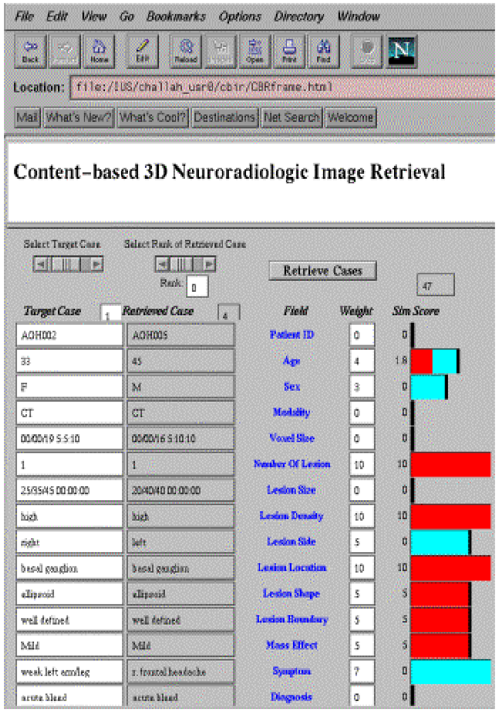
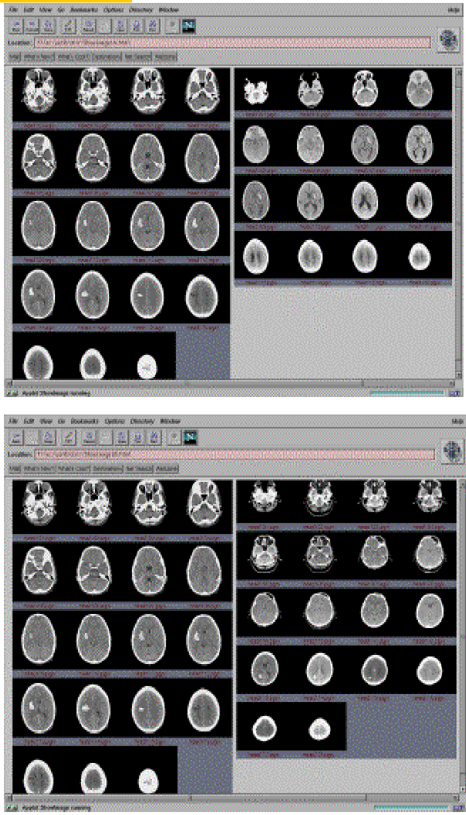


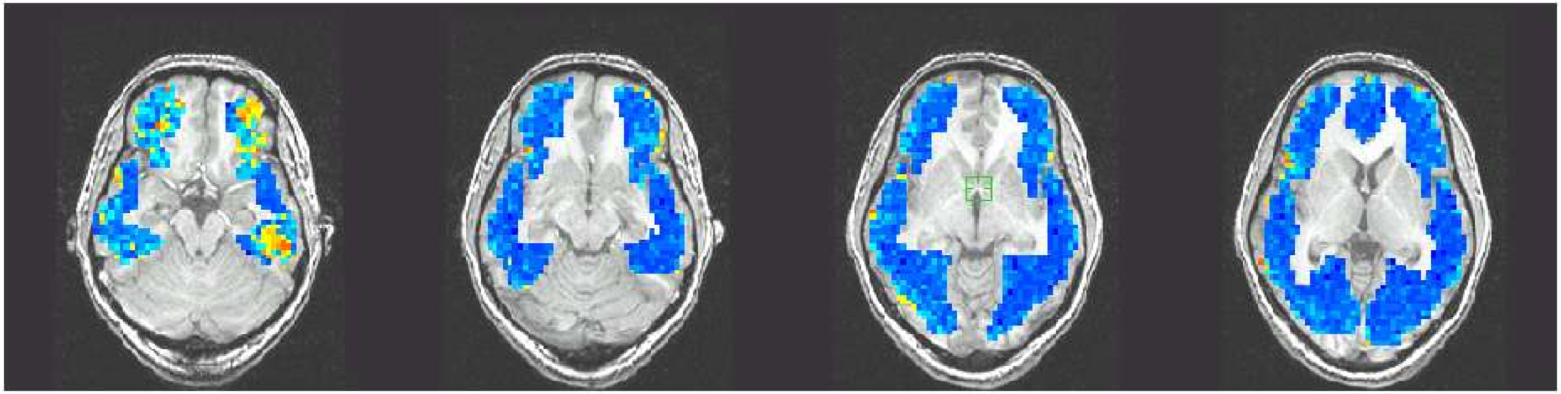
Image classification and learning





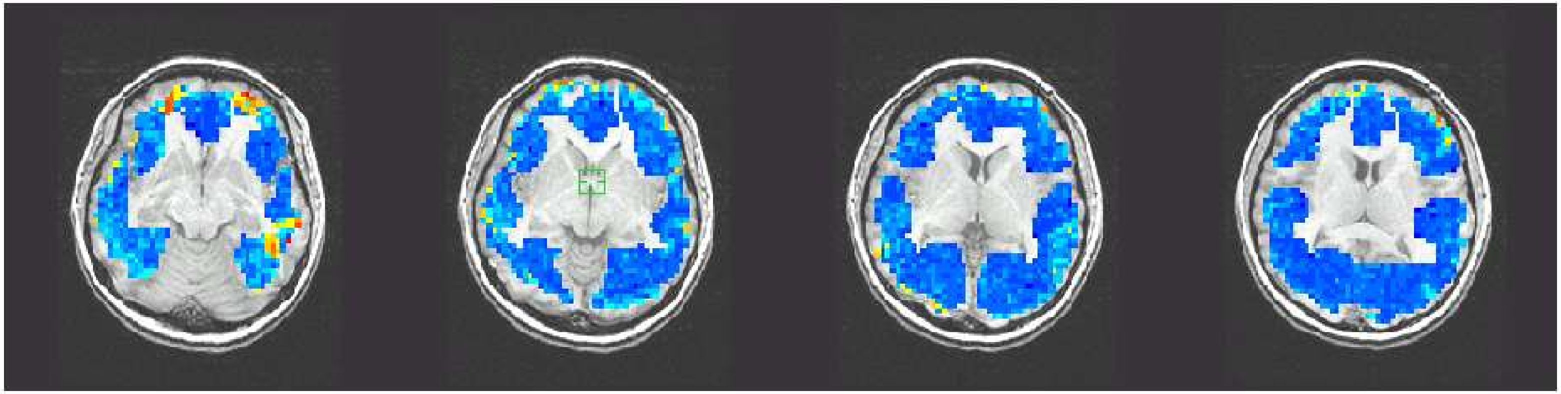
Remote image retrieval

Subject J:



(c) Distribution of discriminating voxels for subject J (the more red the better)

Subject I:



(a) Distribution of discriminating voxels for subject I (the more red the better)

Six-Category Study: Pairwise Classification Errors

* Worst   * Best

	Fish	Vegetables	Tools	Dwellings	Trees	Bldg Parts
Subj1	.20	.55 *	.20	.15	.15	.05 *
Sub2	.10 *	.55 *	.35	.20	.10 *	.30
Sub3	.20	.35 *	.15 *	.20	.20	.20
Sub4	.15	.45 *	.15	.15	.25	.05 *
Sub5	.60 *	.55	.25	.20	.15 *	.15 *
Sub6	.20	.25	.00 *	.30 *	.30 *	.05
Sub7	.15	.55 *	.15	.25	.15	.05 *
Mean	.23	.46	.18	.21	.19	.12

Question:

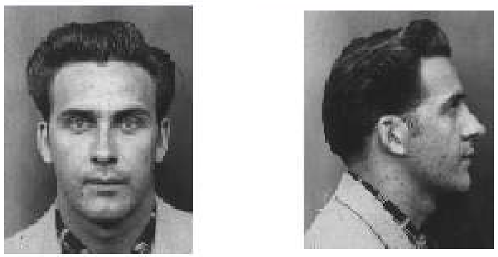
Do different people's brains 'encode' semantic categories using the same spatial patterns?

Yes at a coarse grain, no in detail.

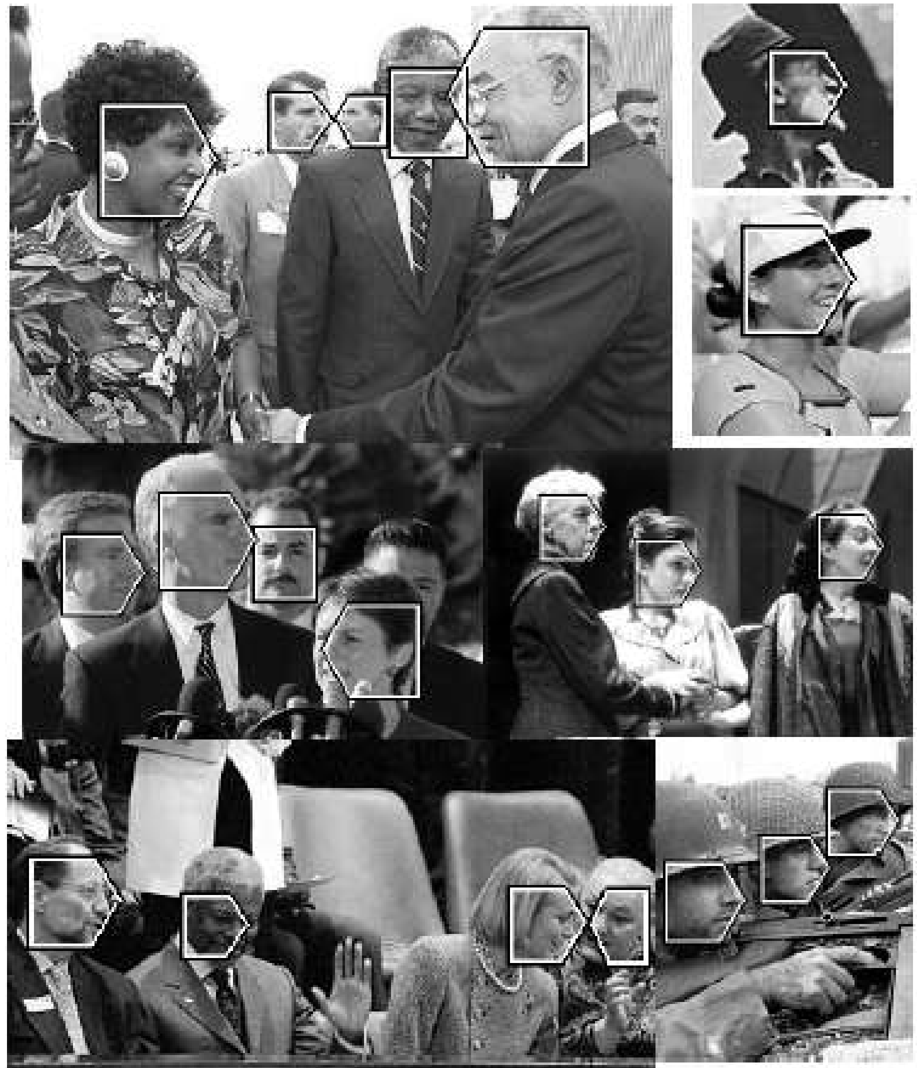
But, there are cross-subject regularities in "distances" between categories, as measured by classifier error rates.

Object Detection 1

(Henry Schneiderman)



Example training images for each orientation, the features are pixels on the image.



Rise of Machine Learning in CS

Machine learning is already the best approach to

- Computer vision
- Speech recognition, Natural language processing
- Medical outcomes analysis
- Robot control
- …

This trend is accelerating

- Improved machine learning algorithms
- Improved data capture, networking, faster computers
- Software too complex to write by hand
- New sensors / IO devices
- Need for self-customization to user, environment

Study 3: Syntactic Ambiguity

Is subject reading ambiguous or unambiguous sentence?

"The <u>experienced soldiers</u> <u>warned</u> about the dangers conducted the midnight raid."

"The <u>experienced soldiers</u> <u>spoke</u> about the dangers before the midnight raid."

Study 3: Results

GNB classifier accuracy:

- Average accuracy over five subjects: .79
- Best subject: .90

Which Machine Learning Method Works Best?

GNB and SVM tend to outperform KNN

Feature selection important

Study	Feature Selection	GNB	SVM	1NN	3NN	5NN	7NN	9NN
Picture vs Sentence	No	0.29	0.32	0.43	0.41	0.37	0.37	0.33
	Yes	0.16	0.09	0.20	0.18	0.19	0.18	0.17
Semantic Categories	No	0.10	N/A	0.40	0.40	0.40	0.40	0.25
	Yes	0.08	N/A	0.30	0.20	0.16	0.14	0.13
Syntactic Ambiguity	No	0.43	0.38	0.50	0.46	0.47	0.39	0.43
	Yes	0.25	0.23	0.29	0.29	0.28	0.29	0.26
Noun vs Verb	No	0.36	0.39	0.44	0.45	0.39	0.44	0.41
	Yes	0.23	0.28	0.38	0.38	0.33	0.28	0.31

Which Feature Selection Works Best?

Conventional wisdom: pick features that best distinguish between target classes

- Information gain, single-feature classifier accuracy, etc.

Surprise: better to pick features that best distinguish between fixation (rest) and any of the target classes

- *Active*: choose voxels with highest t-statistic contrasting any target class against fixation
- *ROI Active*: as above, but choose per anatomical ROI
- *ROI Active Avg*: use mean of *ROI Active* voxels

GNB Classifier Errors: Feature Selection

fMRI study

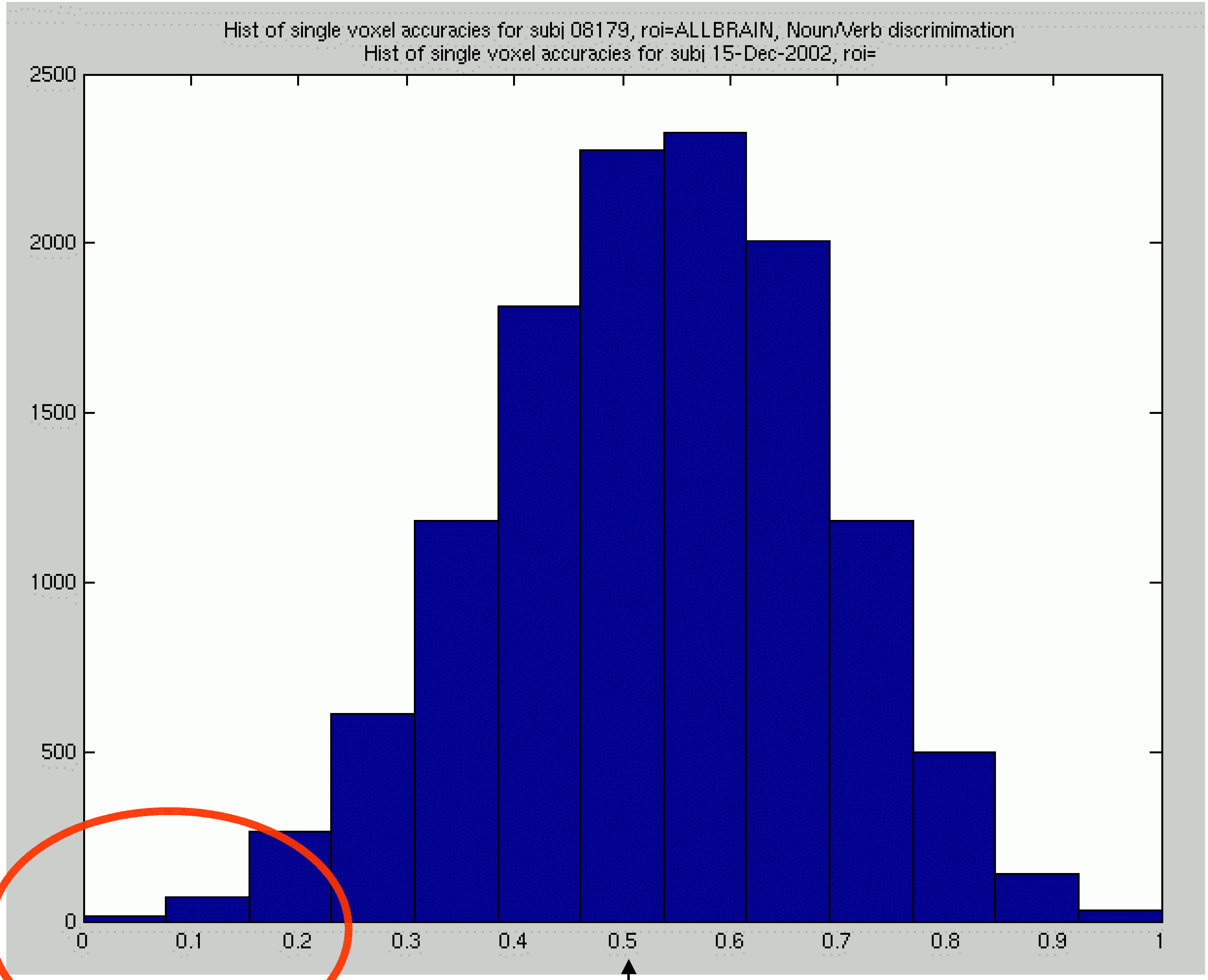
feature selection

	Picture Sentence	Syntactic Ambiguity	Nouns vs. Verbs	Word Categories
All features	.29	.43	.36	.10
Discriminate target classes	.26	.34	.36	.10
Active	.16	.25	.34	.08
ROI Active	.18	.27	.31	.09
ROI Active Average	.21	.27	.23	NA

What's wrong with selecting features based on their ability to discriminate target classes??

10^4 features, 10^1 examples, much noise

Test Set LOO Accuracies of Single-Voxel Noun/Verb Classifiers



Uh oh…

Random guessing

8 training examples per class, test accuracy based on leave-out-one-per-class