343H: Honors AI

Lecture 26:

More applications

4/29/2014

Kristen Grauman

UT Austin

This week

- Tournament Wed night (tomorrow) 7 pm
  - We'll meet here
  - Submit final agent by tonight
  - Otherwise we'll take your last qualifying entry

- Class Thursday
  - Course wrap-up, exam details, tournament recap/awards, surveys

Last time

- Neural networks
- Visual recognition
  - Face detection
  - Gender recognition
  - Boosting
  - Multi-class SVMs
  - Classifier cascades

Today

- Deep learning for image recognition
- Body pose estimation from decision forests
- Non-parametric scene recognition

How many computers to identify a cat?

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | U.S. Edition

Subscribe: Digital / Home Delivery | Log In | Register Now | Help

The New York Times

Business Day

Technology

Search All NYTimes.com

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | SPORTS | OPINION | ARTS | STYLE | TRAVEL | JOBS | REAL ESTATE | AUTOS

How Many Computers to Identify a Cat? 16,000

Jim Wilson/The New York Times

An image of a cat that a neural network taught itself to recognize.

By JOHN MARKOFF
Published: June 25, 2012

MOUNTAIN VIEW, Calif. — Inside Google's secretive X laboratory, known for inventing self-driving cars and augmented reality glasses, a small group of researchers began working several years ago on a simulation of the human brain.

There Google scientists created one of the largest neural networks for machine learning by connecting 16,000 computer processors, which they turned loose on the Internet to learn on its own.

Presented with 10 million digital images found in YouTube videos, what did Google's brain do? What millions of humans do with YouTube: looked for cats.

FACEBOOK | TWITTER | GOOGLE+ | E-MAIL | SHARE | PRINT | REPRINTS

Multimedia

Business Day Live | Google's Brain

Log in to see what your friends are sharing on nytimes.com. Privacy Policy | What's This?

Log In With Facebook

What's Popular Now

Restaurant Review: Guy's American Kitchen & Bar in Times Square

Occupy Offshoot Aims to Erase People's Debts

Today's Headlines Daily E-Mail

Sign up for a roundup of the day's top stories, sent every morning.

Sign Up

See Sample | Privacy Policy

Subscribe to Technology RSS Feeds

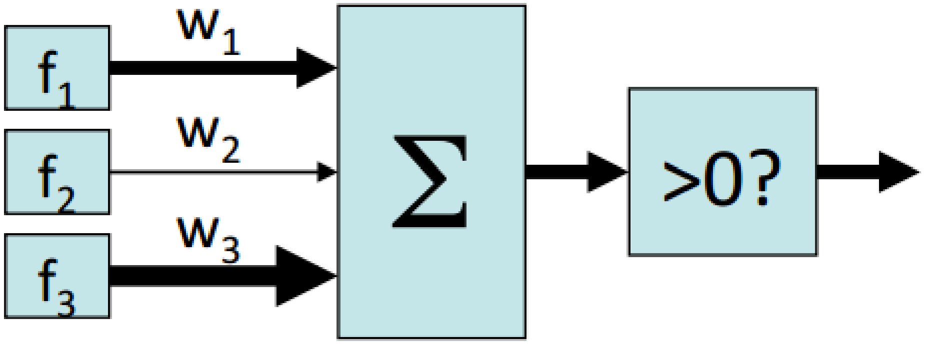
Technology News | Internet | Start-Ups | Business Computing | Companies

Bits Blog | Personal Tech | Pogue's Posts

MOST E-MAILED | MOST VIEWED

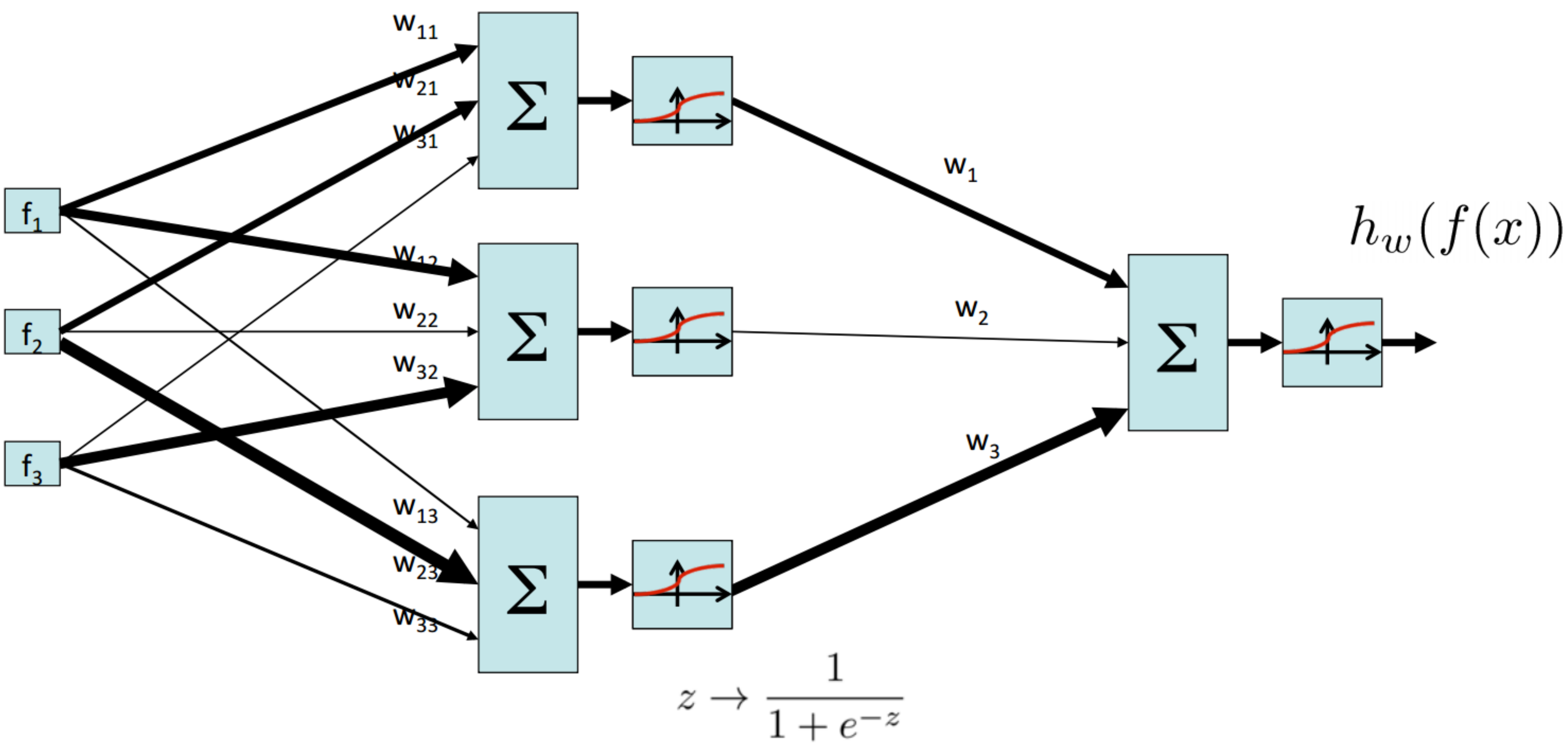
[Le, Ng, Dean, et al. 2012]

Perceptron



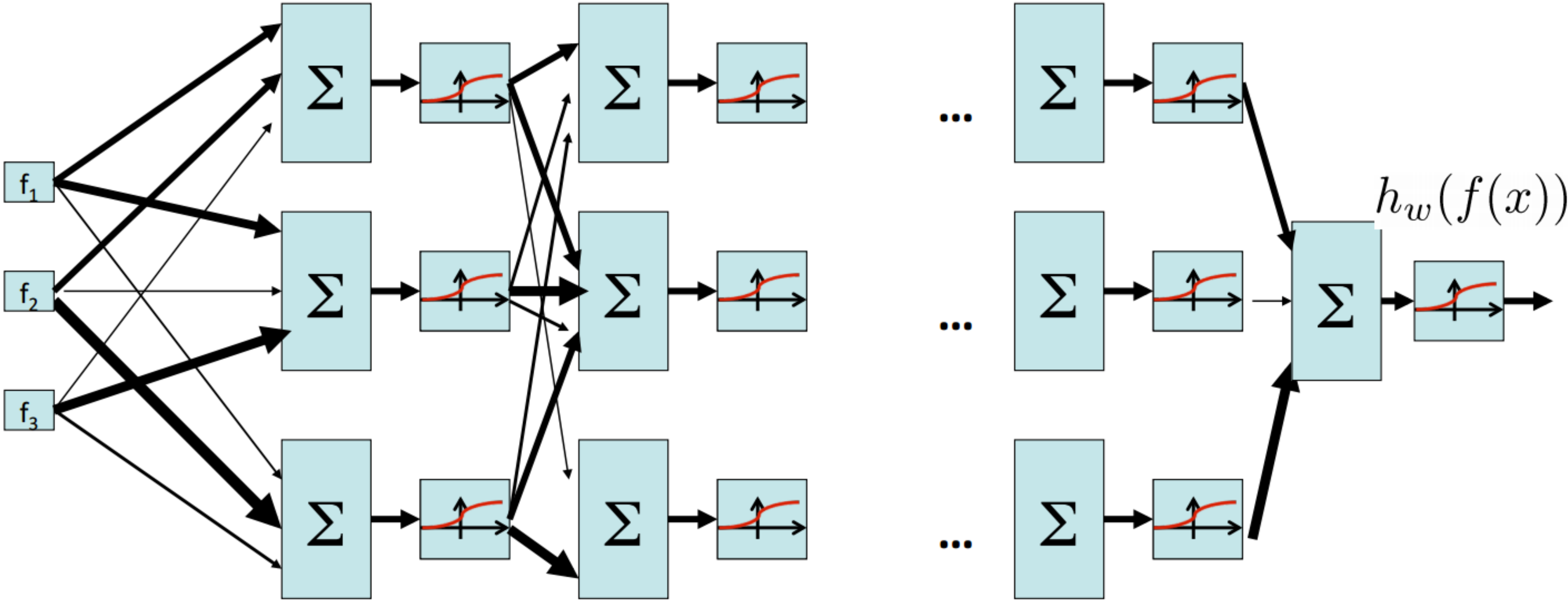
Slide credit: Dan Klein and Pieter Abbeel

Two-layer neural network



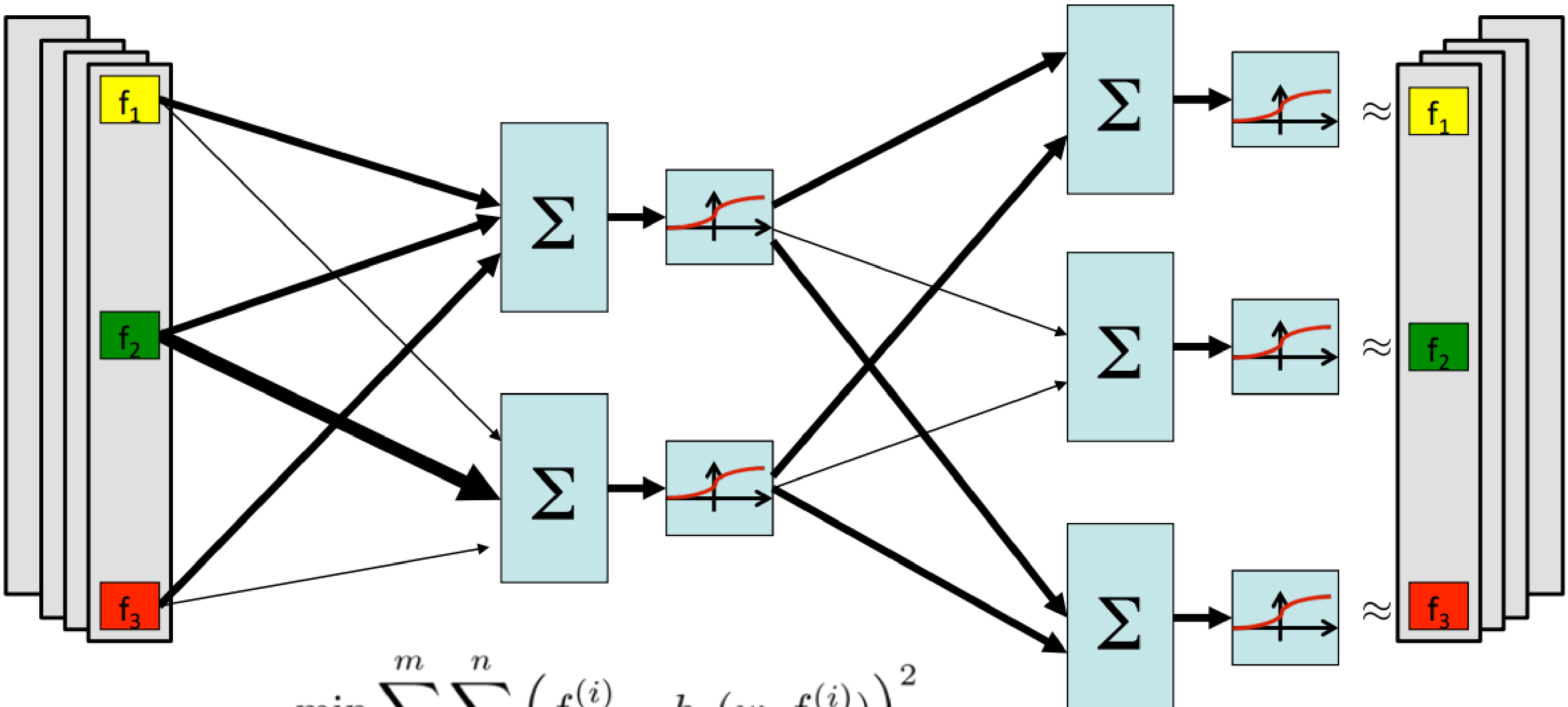
Slide credit: Dan Klein and Pieter Abbeel

N-layer neural network



Slide credit: Dan Klein and Pieter Abbeel

Auto-encoder (sketch)



$$\min_w \sum_{i=1}^{m} \sum_{j=1}^{n} \left(f_j^{(i)} - h_j(w, f^{(i)}) \right)^2$$

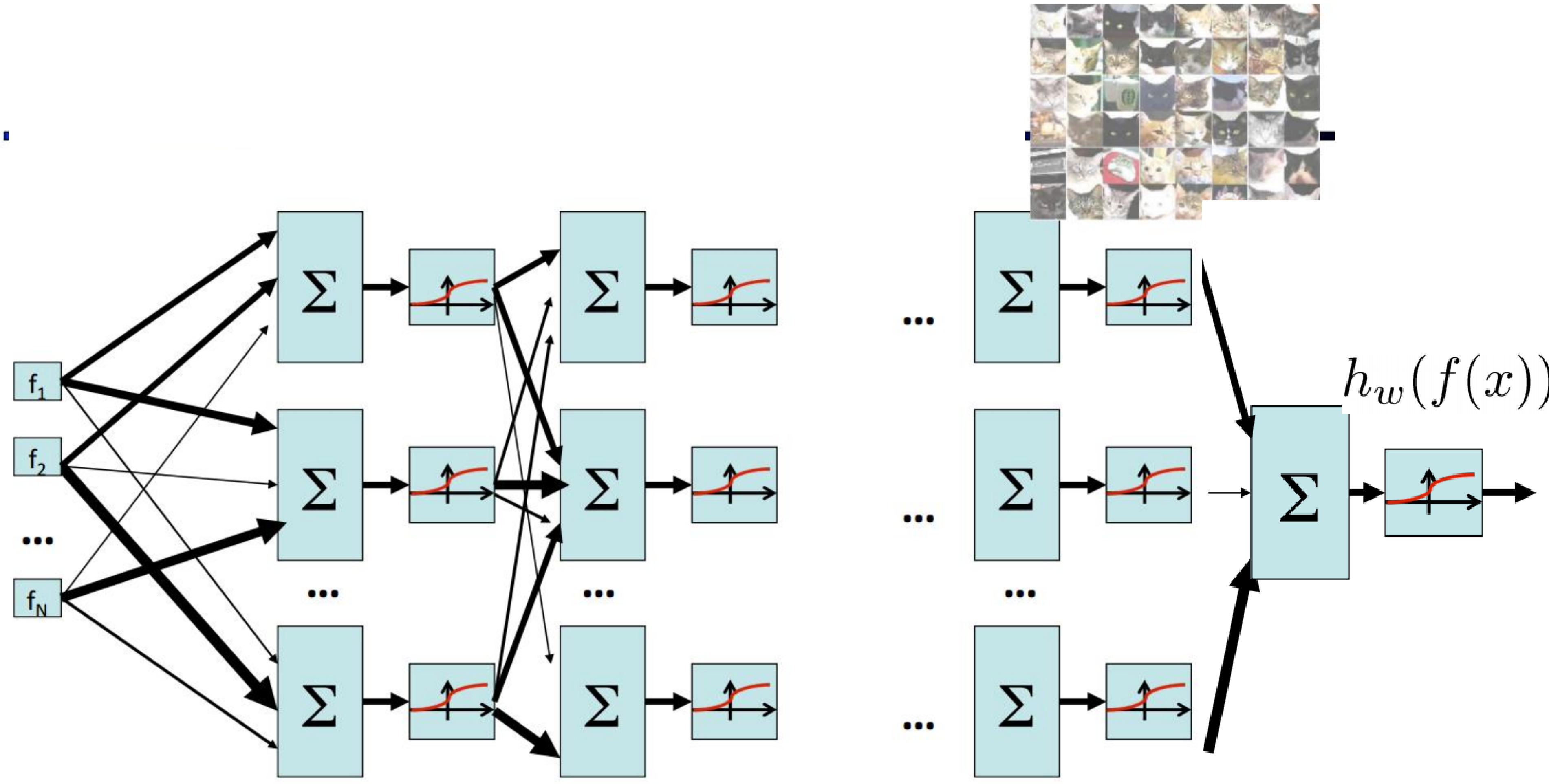
Slide credit: Dan Klein and Pieter Abbeel

Training procedure: stacked auto-encoder

- Auto-encoder
  - Layer 1 = “compressed” version of input layer
- Stacked auto-encoder
  - For every image, make a compressed image (=layer 1 response to image)
  - Learn Layer 2 by using compressed images as input, and as output to be predicted
  - Repeat similarly for Layer 3, 4, etc.
- Some details left out
  - Typically in between layers responses get agglomerated from several neurons (“pooling” / “complex cells”)

Slide credit: Dan Klein and Pieter Abbeel

Final result: trained neural network



Slide credit: Dan Klein and Pieter Abbeel

Real-Time Human Pose Recognition in Parts from Single Depth Images

Jamie Shotton, Andrew Fitzgibbon, Mat Cook,
Toby Sharp, Mark Finocchio, Richard Moore,
Alex Kipman, Andrew Blake

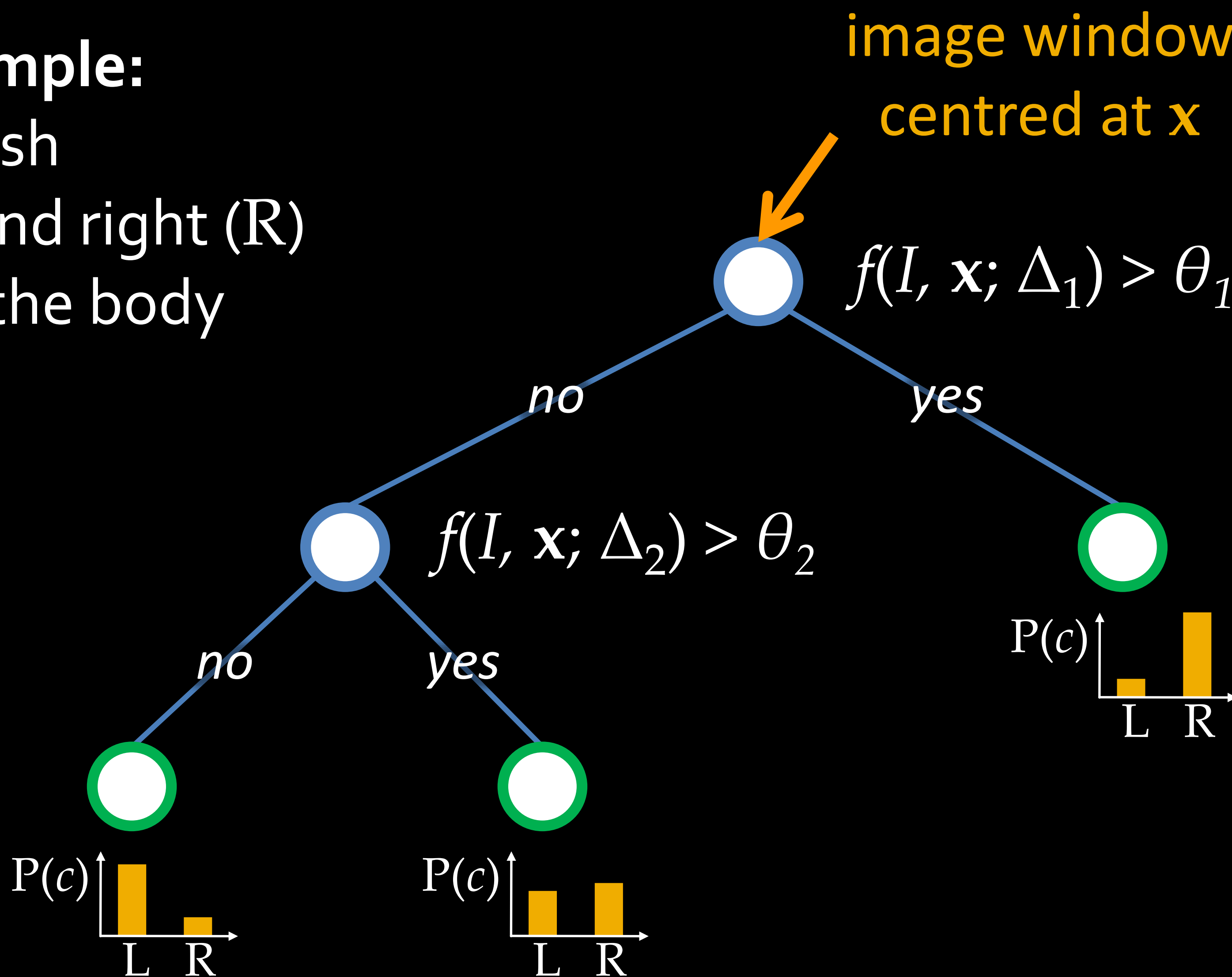
CVPR 2011

Microsoft Research

XBOX 360

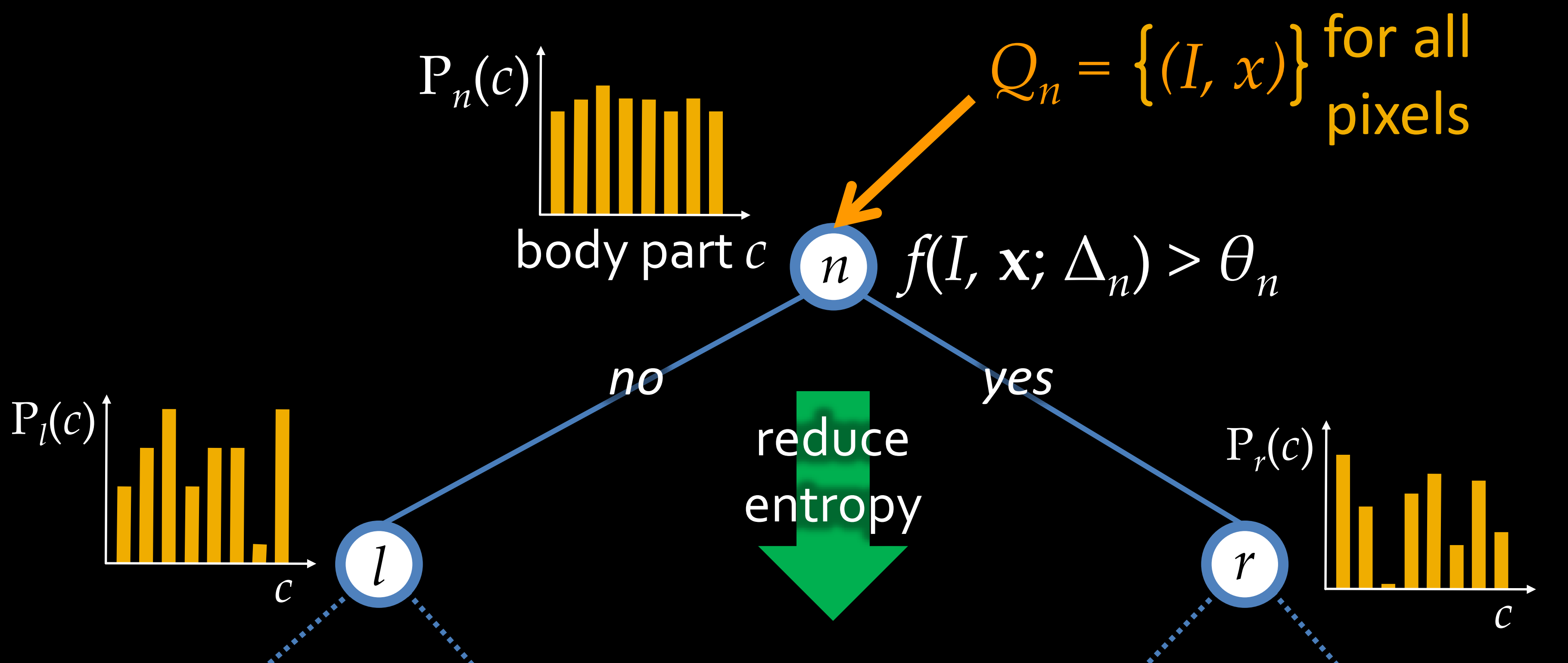
Decision tree classification

Toy example:
distinguish
left (L) and right (R)
sides of the body



Training decision trees

[Breiman *et al.* 84]



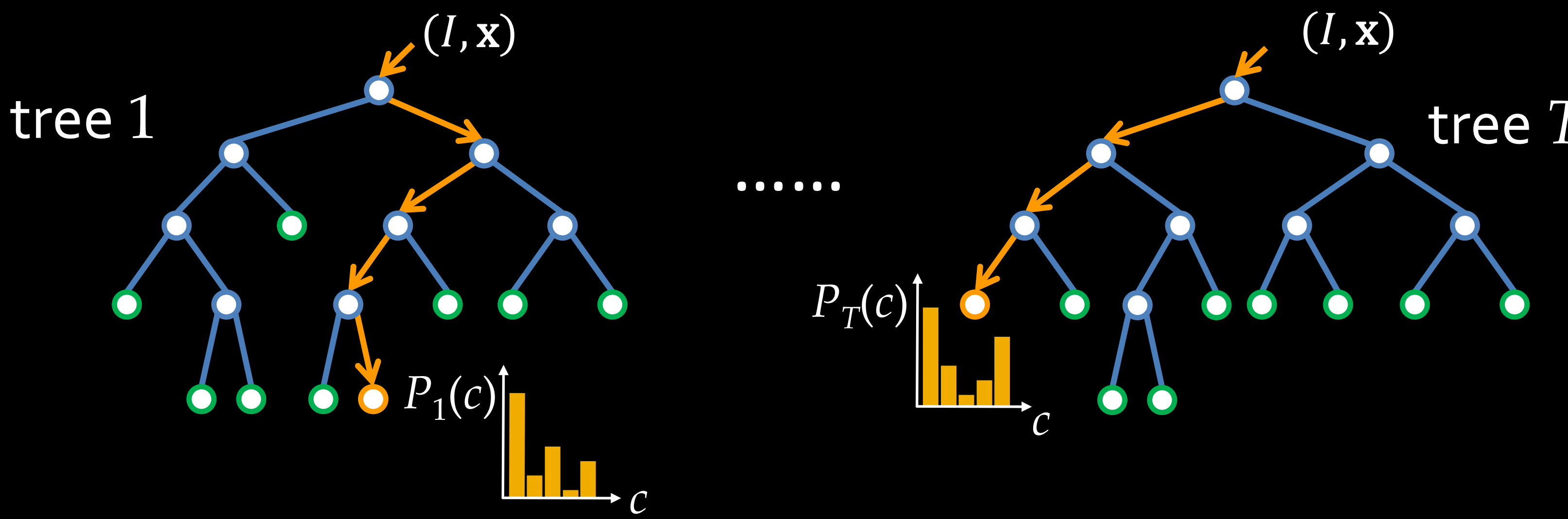
Take (Δ, θ) that maximises information gain:

$$\Delta E = -\frac{|Q_\mathrm{l}|}{|Q_n|} E(\mathrm{Q_l}) - \frac{|Q_\mathrm{r}|}{|Q_n|} E(\mathrm{Q_r})$$

Goal: drive entropy at leaf nodes to zero

Decision forest classifier

[Amit & Geman 97]
[Breiman 01]
[Geurts *et al.* 06]



- Trained on different random subset of images
  - "bagging" helps avoid over-fitting
- Average tree posteriors $P(c|I,\mathbf{x}) = \frac{1}{T}\sum_{t=1}^{T} P_t(c|I,\mathbf{x})$

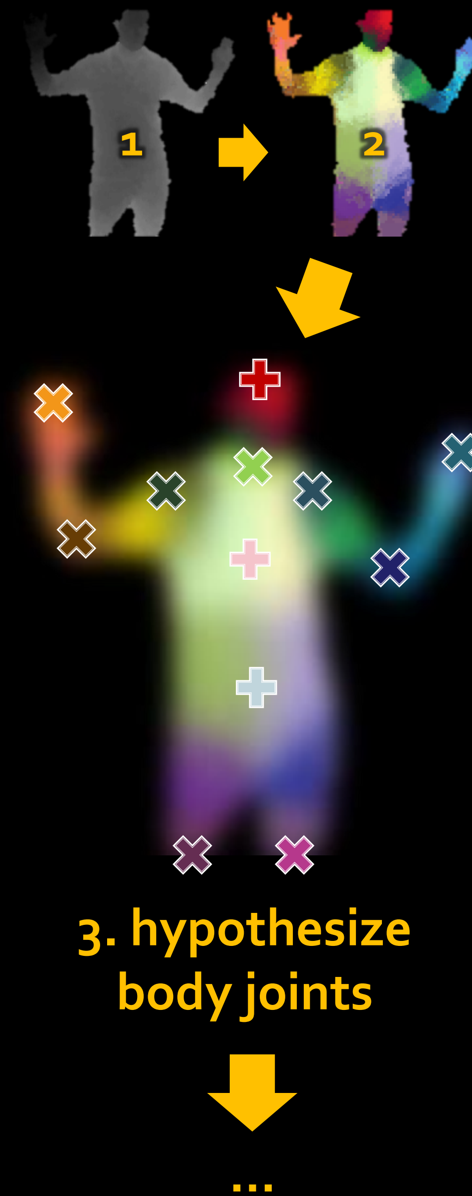
Body parts to joint hypotheses

- Define 3D world space density:

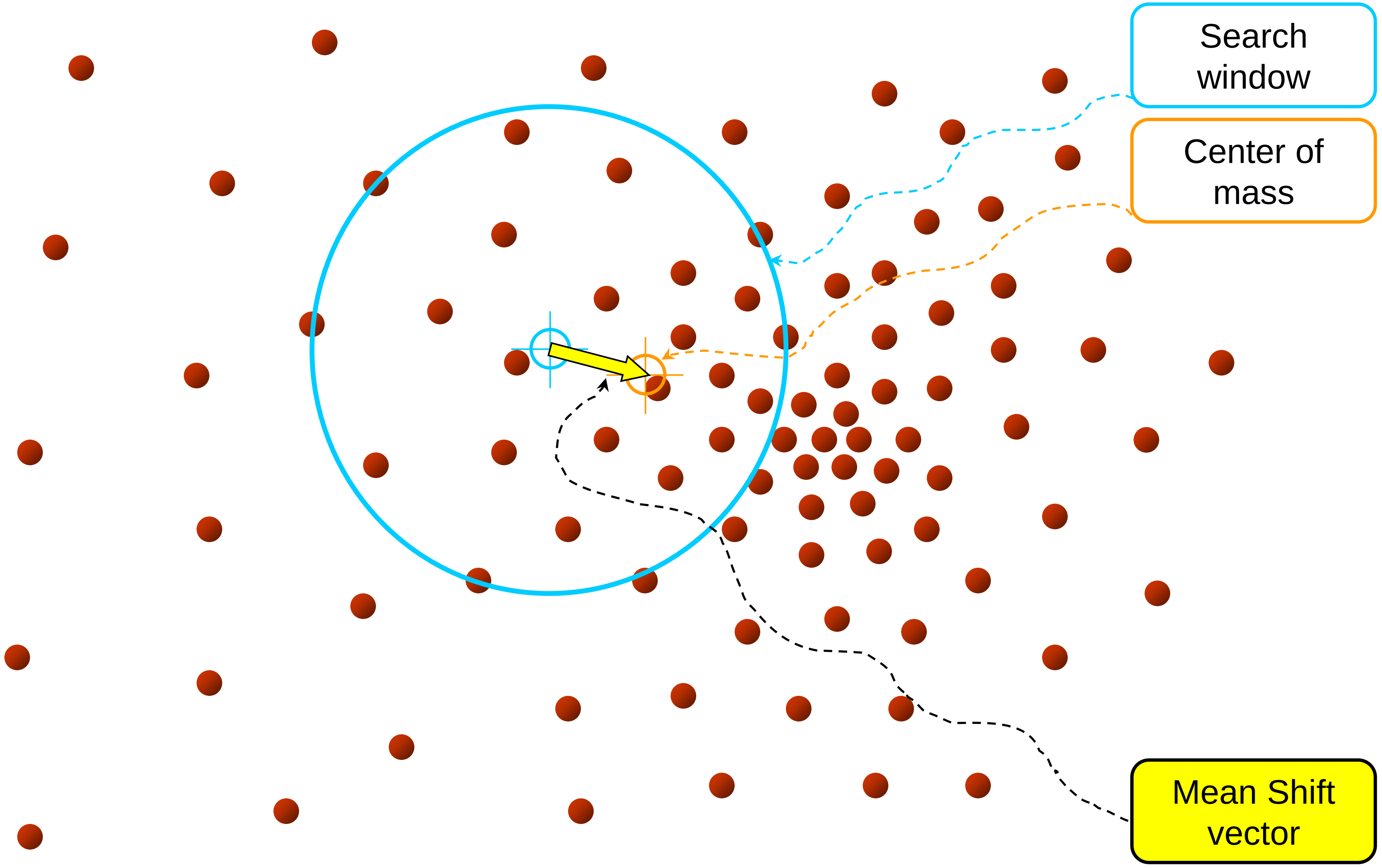
$$f_c(\overbrace{\hat{\mathbf{x}}}^{\text{3D coord}}) \propto \underbrace{\sum_{i=1}^{N}}_{\text{pixel index } i} \overbrace{w_{ic}}^{\text{pixel weight}} \exp\left(-\left\|\frac{\hat{\mathbf{x}} - \overbrace{\hat{\mathbf{x}}_i}^{\text{3D coord of } i^{\text{th}} \text{ pixel}}}{\underbrace{b_c}_{\text{bandwidth}}}\right\|^2\right)$$

$$w_{ic} = \underbrace{P(c|I, \mathbf{x}_i)}_{\text{inferred probability}} \cdot \underbrace{d_I(\mathbf{x}_i)^2}_{\text{depth at } i^{\text{th}} \text{ pixel}}$$

- Mean shift for mode detection



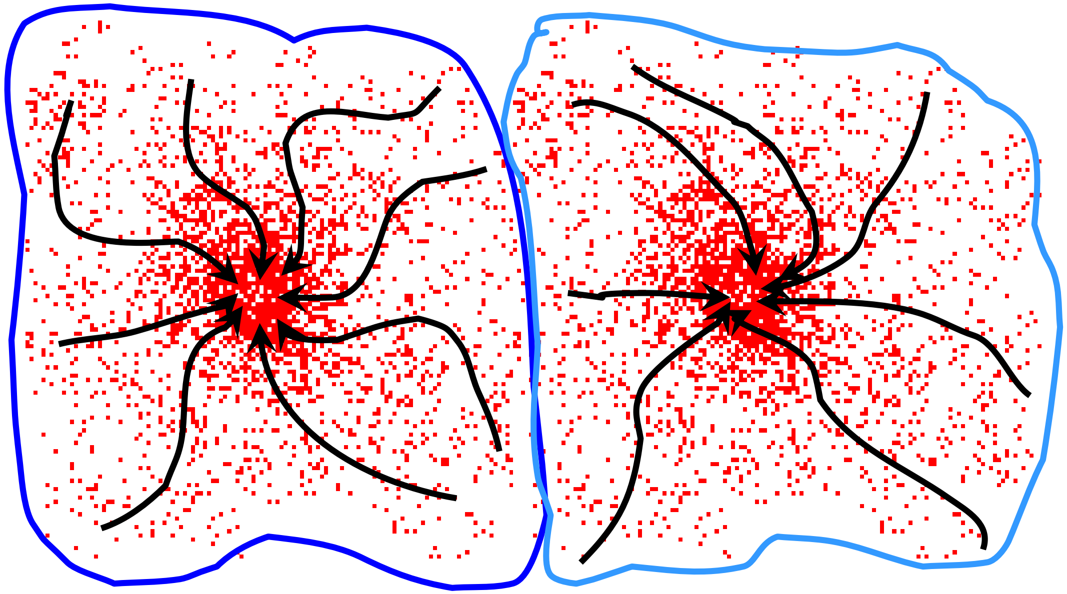
Mean shift



Slide by Y. Ukrainitz & B. Sarel

Mean shift clustering

- Cluster: all data points in the attraction basin of a mode
- Attraction basin: the region for which all trajectories lead to the same mode

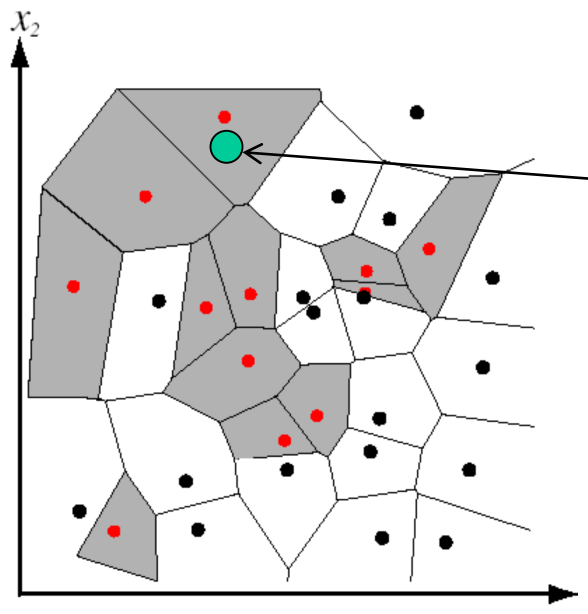


Slide by Y. Ukrainitz & B. Sarel

Nearest Neighbor classification

- Assign label of nearest training data point to each test data point

Black = negative
Red = positive



from Duda *et al.*

Novel test example

Closest to a positive example from the training set, so classify it as positive.

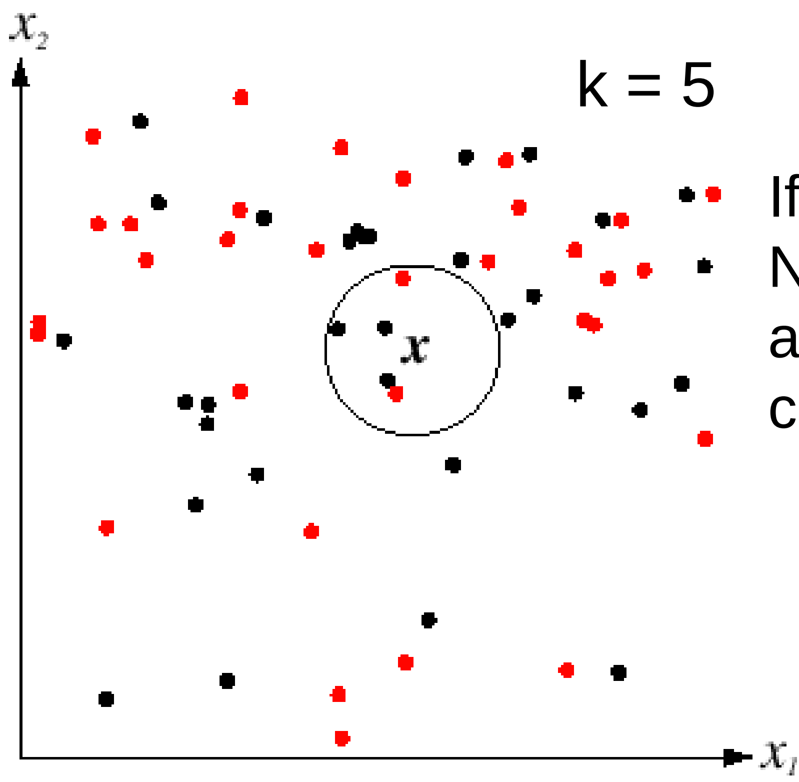
Voronoi partitioning of feature space
for 2-category 2D data

K-Nearest Neighbors classification

- For a new point, find the k closest points from training data
- Labels of the k points "vote" to classify

Black = negative

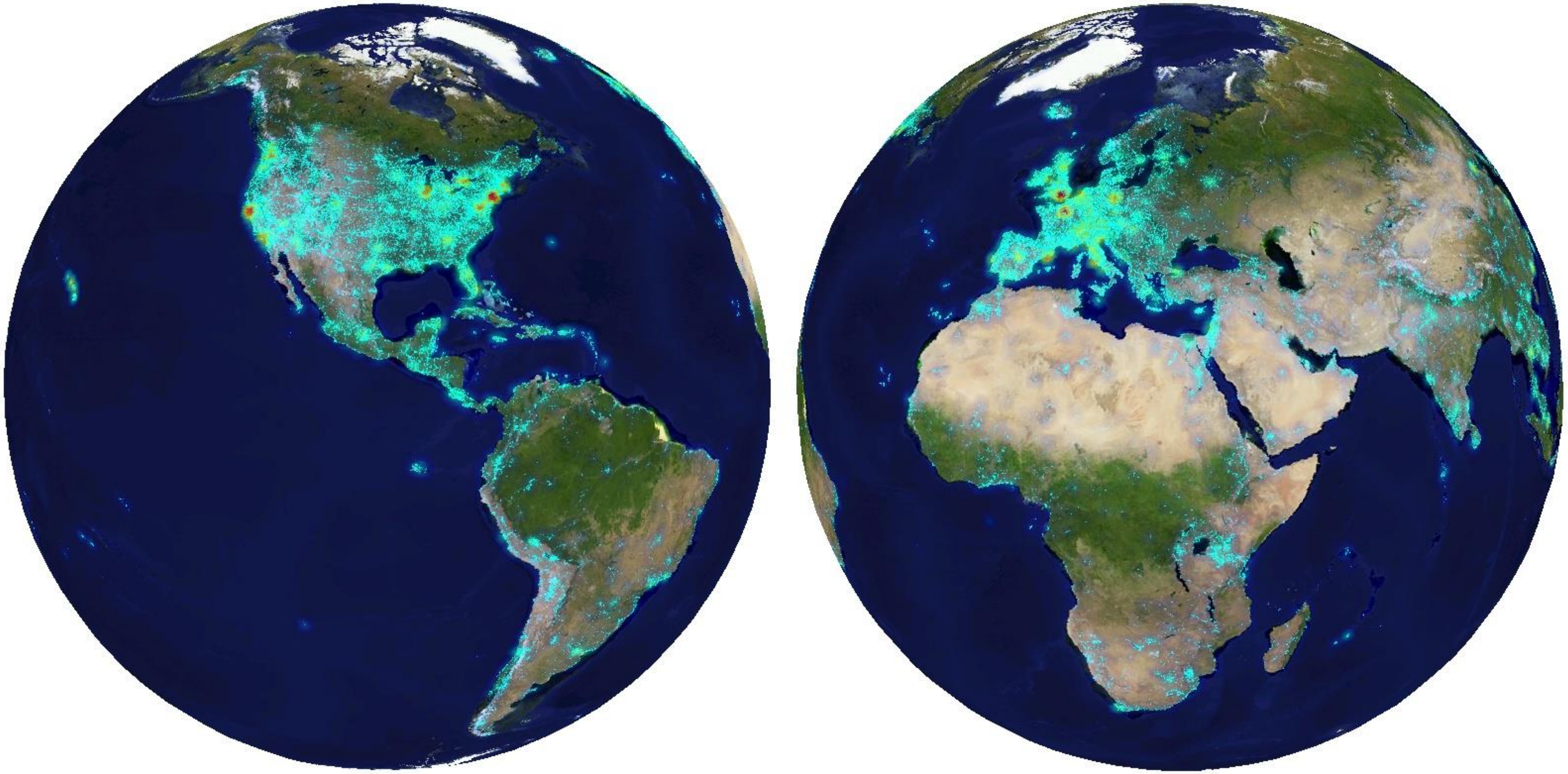
Red = positive



If query lands here, the 5 NN consist of 3 negatives and 2 positives, so we classify it as negative.

Source: D. Lowe

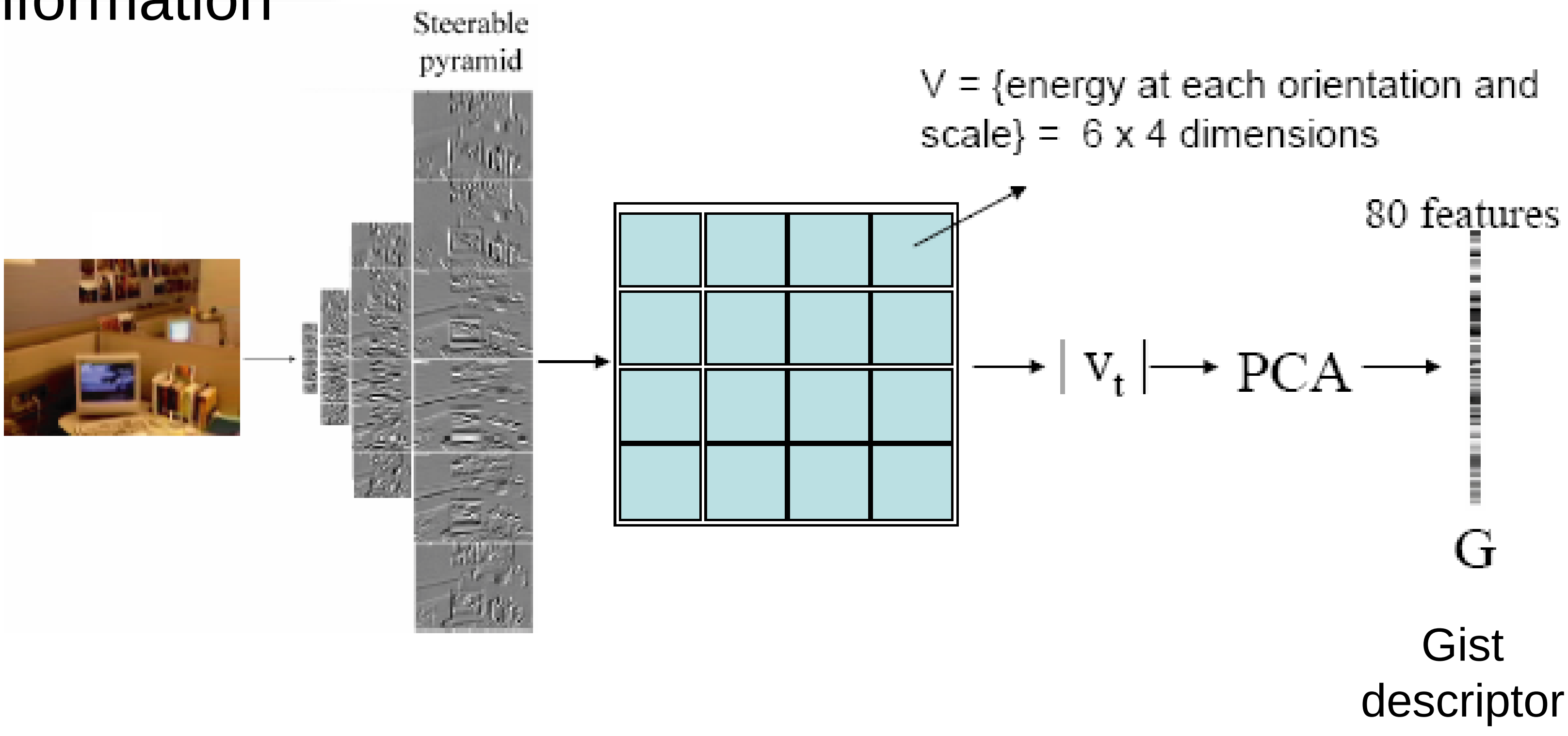
6+ million geotagged photos by 109,788 photographers



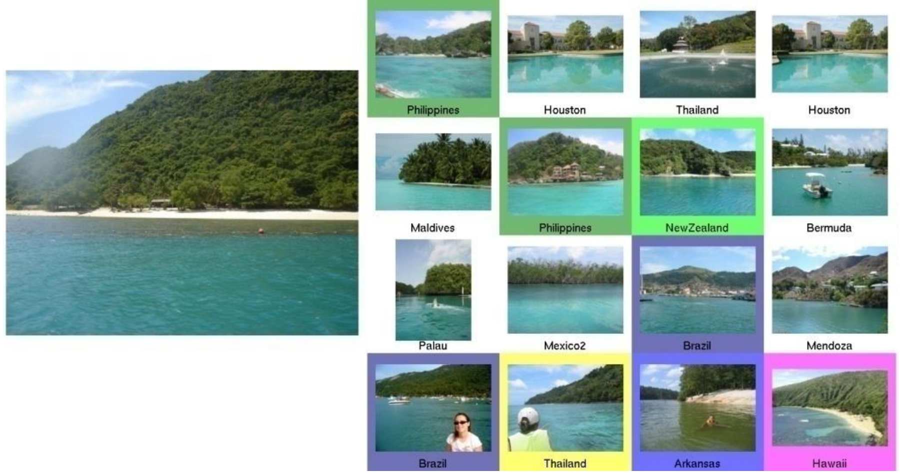
Annotated by Flickr users

Global texture: capturing the "Gist" of the scene

Capture global image properties while keeping some spatial information

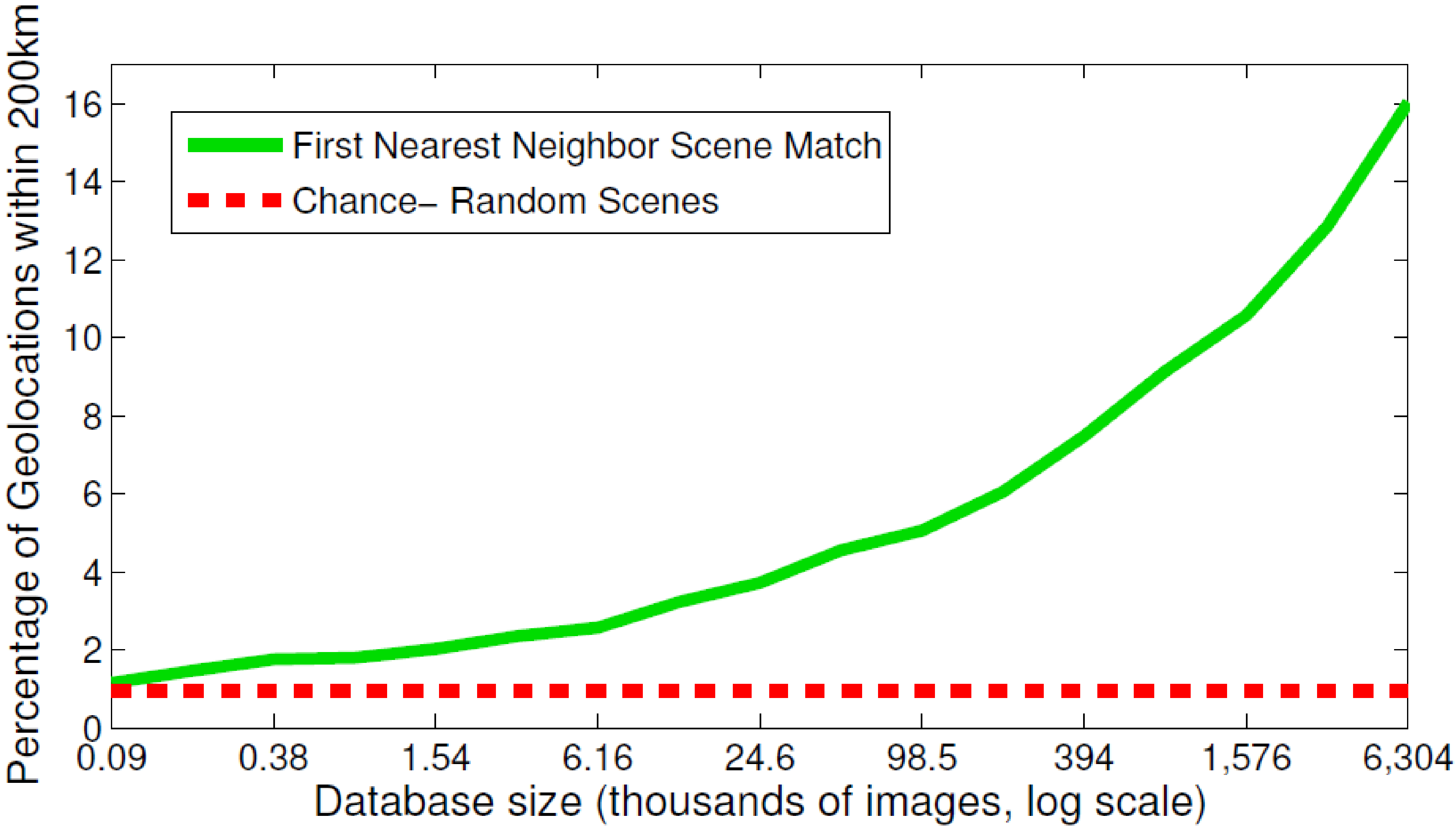


Oliva & Torralba IJCV 2001, Torralba et al. CVPR 2003



[Hays and Efros. **im2gps**: Estimating Geographic Information from a Single Image. CVPR 2008.]

The Importance of Data



[Hays and Efros. **im2gps**: Estimating Geographic Information from a Single Image. CVPR 2008.]

Recap

- Deep learning for image recognition
- Body pose estimation from decision forests
- Non-parametric scene recognition

- Visual recognition tasks with supervised classification
  - Variety of features and models
  - Training data quality and/or quantity essential