What is machine learning?

Machine Learning - 10601

Geoff Gordon, Miroslav Dudík
http://www.cs.cmu.edu/~ggordon/10601/

August 24, 2009

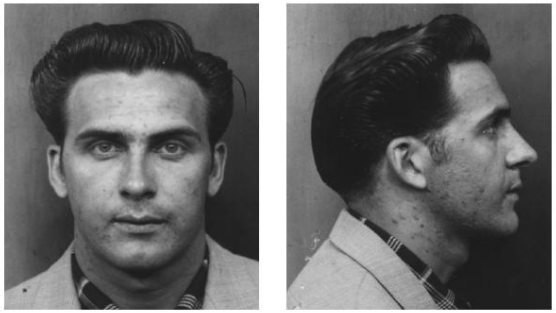
Machine learning

Study of algorithms that

- improve their **performance**
- at some **task**
- with **experience**

Object detection

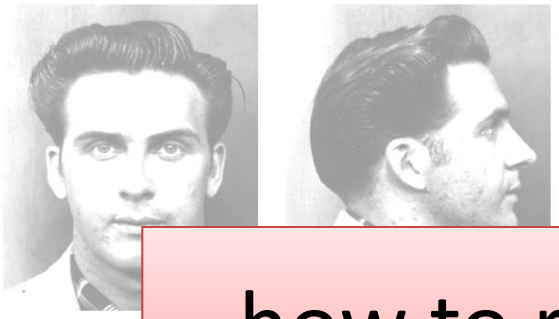
[Schneiderman, Kanade 2002]



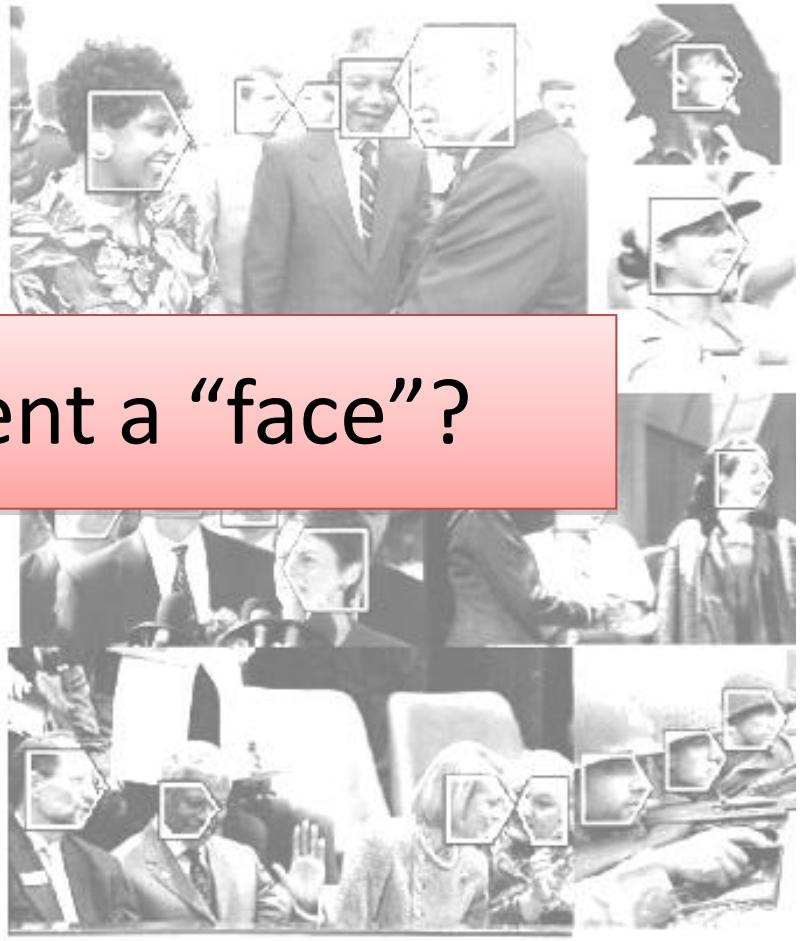
Example training images
for each orientation

Object detection

[Schneiderman, Kanade 2002]



Example training images
for each orientation



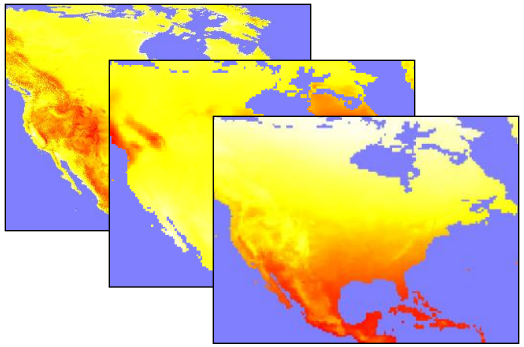
how to represent a “face”?

Modeling distributions of species

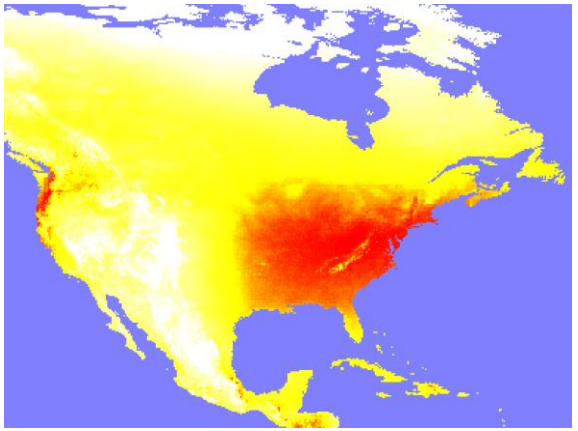
[Phillips, Dudík, Schapire 2004]

occurrence localities

Yellow-throated Vireo



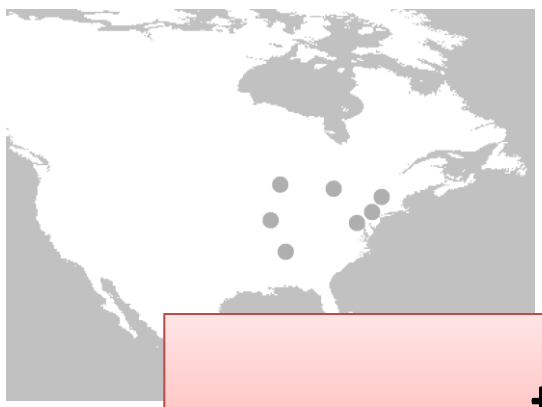
environmental variables



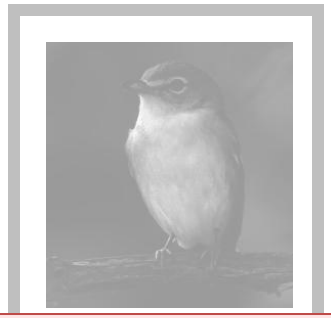
predicted geographic distribution

Modeling distributions of species

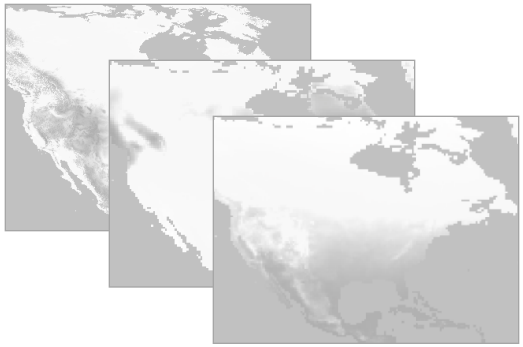
[Phillips, Dudík, Schapire 2004]



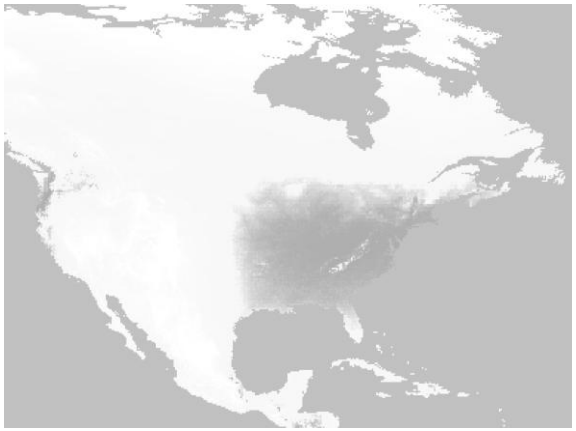
occur



too little data!



environmental variables



predicted geographic distribution

Other machine learning applications?

Divide into groups of 4-5 people.

- **think of 3 (or more) ML applications**
  commercial, research, future promise
- **think of challenges**
  what prevents you from solving them right now?

You have 5mins. Start now...

ML applications & challenges

Application	Challenge
object detection	representing “objects”
species-distribution modeling	scarce data

Machine learning challenges

Representation of

- observations, assumptions, solutions

Generalization to

- present (but unobserved) data, future data

Computation

Machine learning challenges

Representation

- computer science (AI), physics, biology, ...

Generalization

- probability, statistics

Computation

- computer science (algorithms, complexity)

This course

- from **basics** to **state of the art**
- both **theory** and **applications**

Syllabus:

- graphical models, naïve Bayes, logistic regression, decision trees, boosting, neural nets, regularization, dimensionality reduction, PCA, mistake bounds, VC dimension, SVMs, kernels, margin bounds, k-means, EM, HMMs, reinforcement learning, ...

Logistics

Web-page:
http://www.cs.cmu.edu/~ggordon/10601/

Mailing list:
10601-09f-announce@cs

Staff:

- instructors: **Geoff Gordon**, **Miro Dudík**
- TAs: **Oznur Tastan**, **Joseph Gonzales**
- administrative assistant: **Michelle Martin**
- **check webpage for office hours**

Logistics

Recitations

Logistics

Auditing:

- students must register
- can submit homeworks, but indicate as “AUDITING”

Grading

- Homeworks (40%)
  - first **out 8/26**, **due 9/2 at 10:30am in class**
  - total of five late days, **no exceptions**
  - collaboration: **write alone**, **list collaborators**
- Final project (25%)
  - teams of 2-3 students; more info mid-September
- "Mid"-term (30%)
  - Monday **11/2**, **5pm-7pm**, location TBD
- Class participation (5%)

It’s going to be hard work!

Remember to have fun! ☺