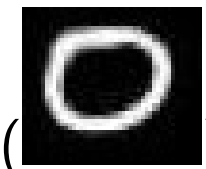
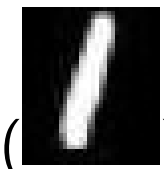
CS499/599
Unsupervised Learning
Lecture 1: applications of machine learning

Toby Dylan Hocking
toby.hocking@nau.edu

23 Aug 2021

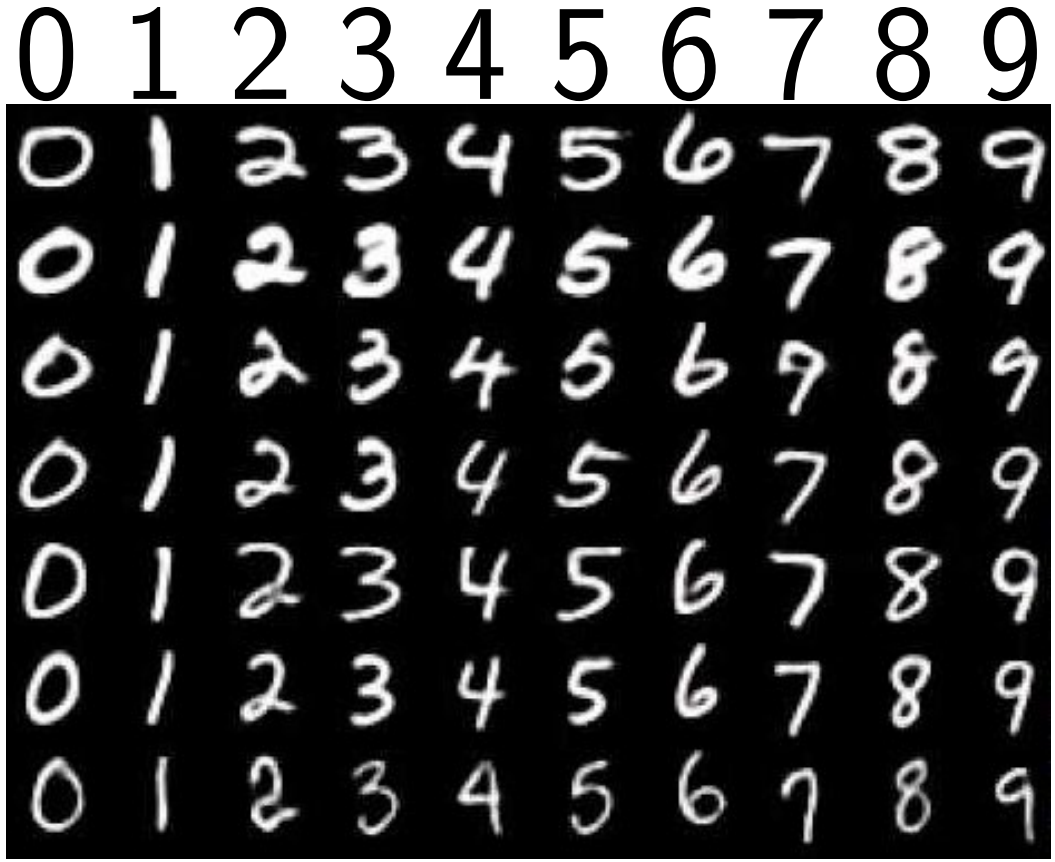
Machine learning intro: image classification example

ML is all about learning predictive functions $f(x) \approx y$, where

- Inputs/features x can be easily computed using traditional algorithms, e.g. matrix of pixel intensities in an image.
- Outputs/labels y are what we want to predict, easy to get by asking a human, but hard to compute using traditional algorithms, e.g. image class.
- Input x = image of digit, output $y \in \{0, 1, \ldots, 9\}$,
  – this is a classification problem with 10 classes.
  $f($  $) = 0$, $f($  $) = 1$
- Traditional algorithm: I give you a pixel intensity matrix $x \in \mathbb{R}^{16 \times 16}$, you code a function f that returns one of the 10 possible digits. Q: how to do that?

Supervised machine learning algorithms

I give you a training data set with paired inputs/outputs, e.g.

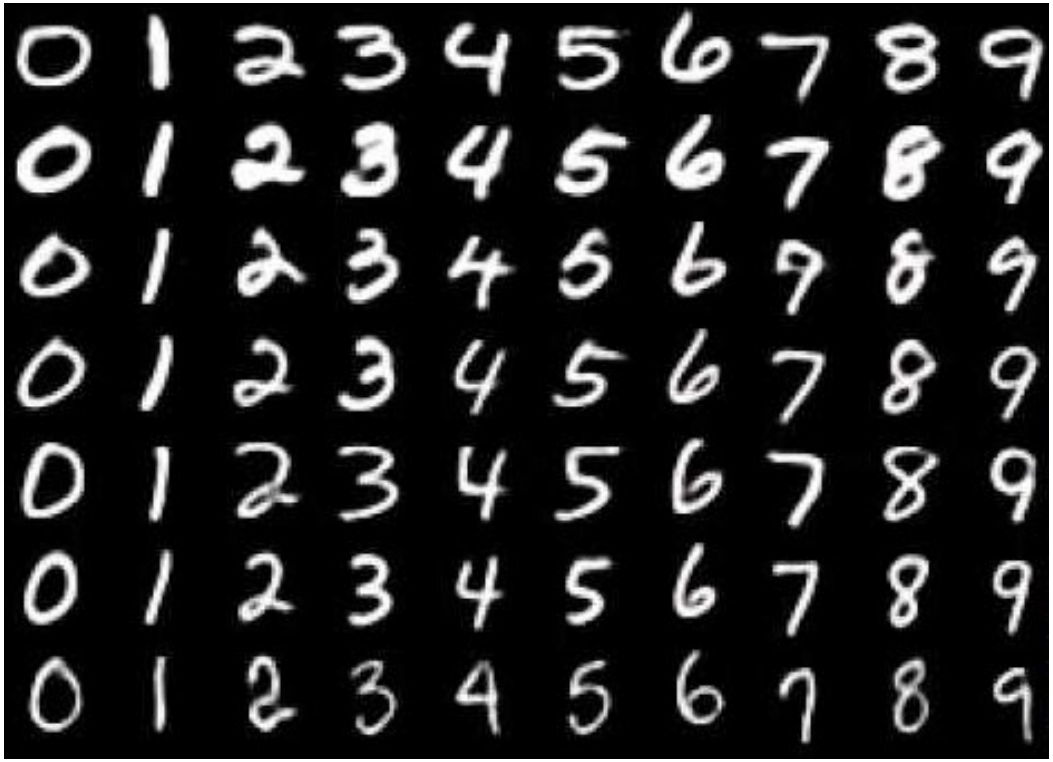


Your job is to learn f from the training set (including expected outputs/labels).

Source: github.com/cazala/mnist

Unsupervised machine learning algorithms

I give you a training data set with only inputs (no outputs), e.g.

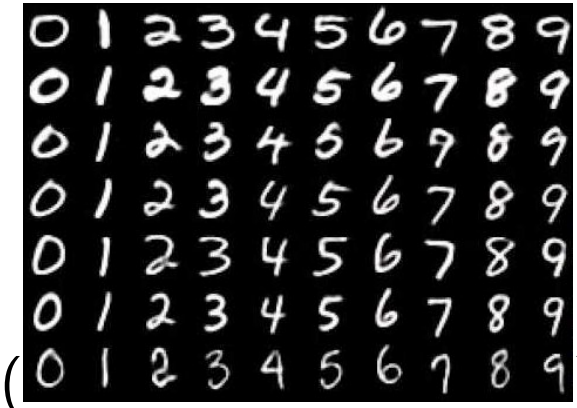
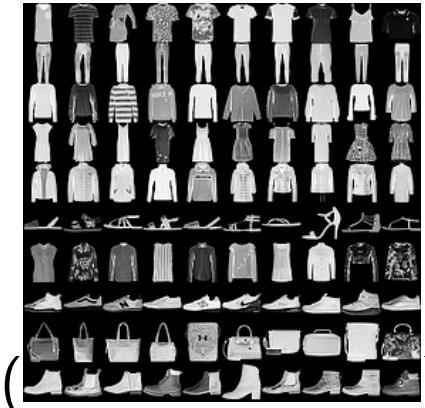


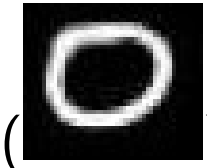
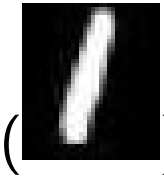
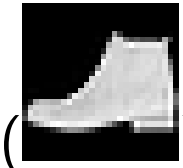
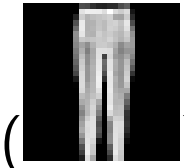
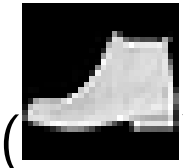
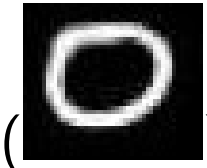
Your job is to learn f from the training set (without expected outputs/labels).

Source: github.com/cazala/mnist

Learning two different functions

Say LEARN is a learning algorithm you have coded.

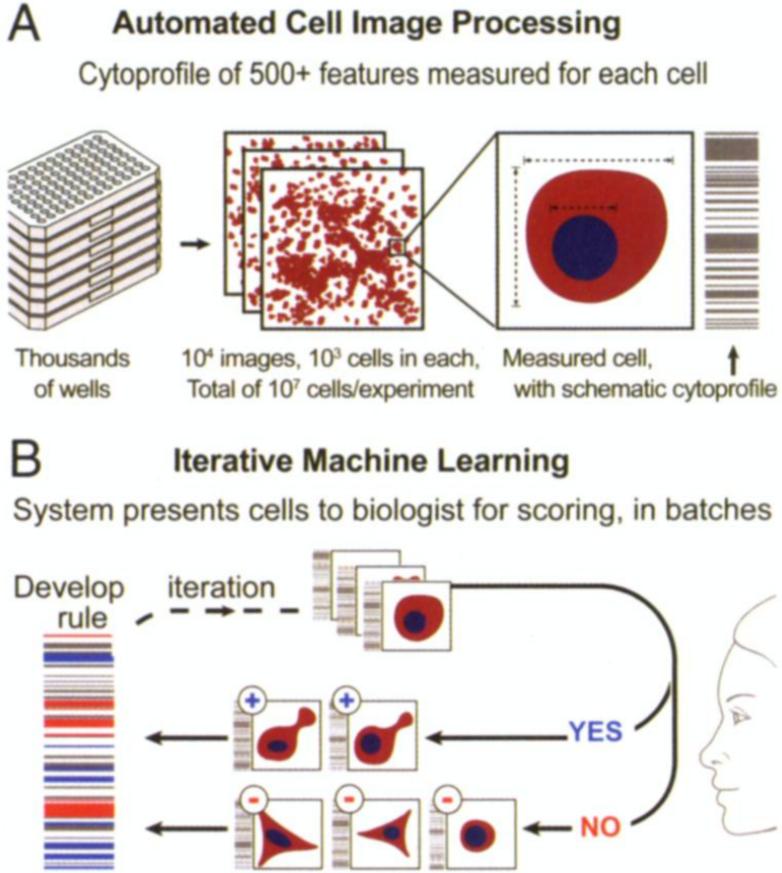
LEARN() $\rightarrow f$, LEARN() $\rightarrow g$

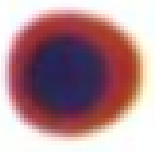
- Then we would expect $f($  $) = 0$, $f($  $) = 1$
- $g($  $) =$ shoe$/0$, $g($  $) =$ pants$/1$
- Q: what happens if you do $f($  $)$, or $g($  $)$?

Sources: github.com/cazala/mnist, github.com/zalandoresearch/fashion-mnist

Machine learning for cell image classification (CellProfiler)

Jones *et al.* PNAS 2009. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning.



- Input x = image of cell,
- Output $y \in \{\text{yes}, \text{no}\}$ (binary classification),
- $f($  $) = \text{yes}$,
- $f($  $) = \text{no}$.

Machine learning for image segmentation (LabelMe)

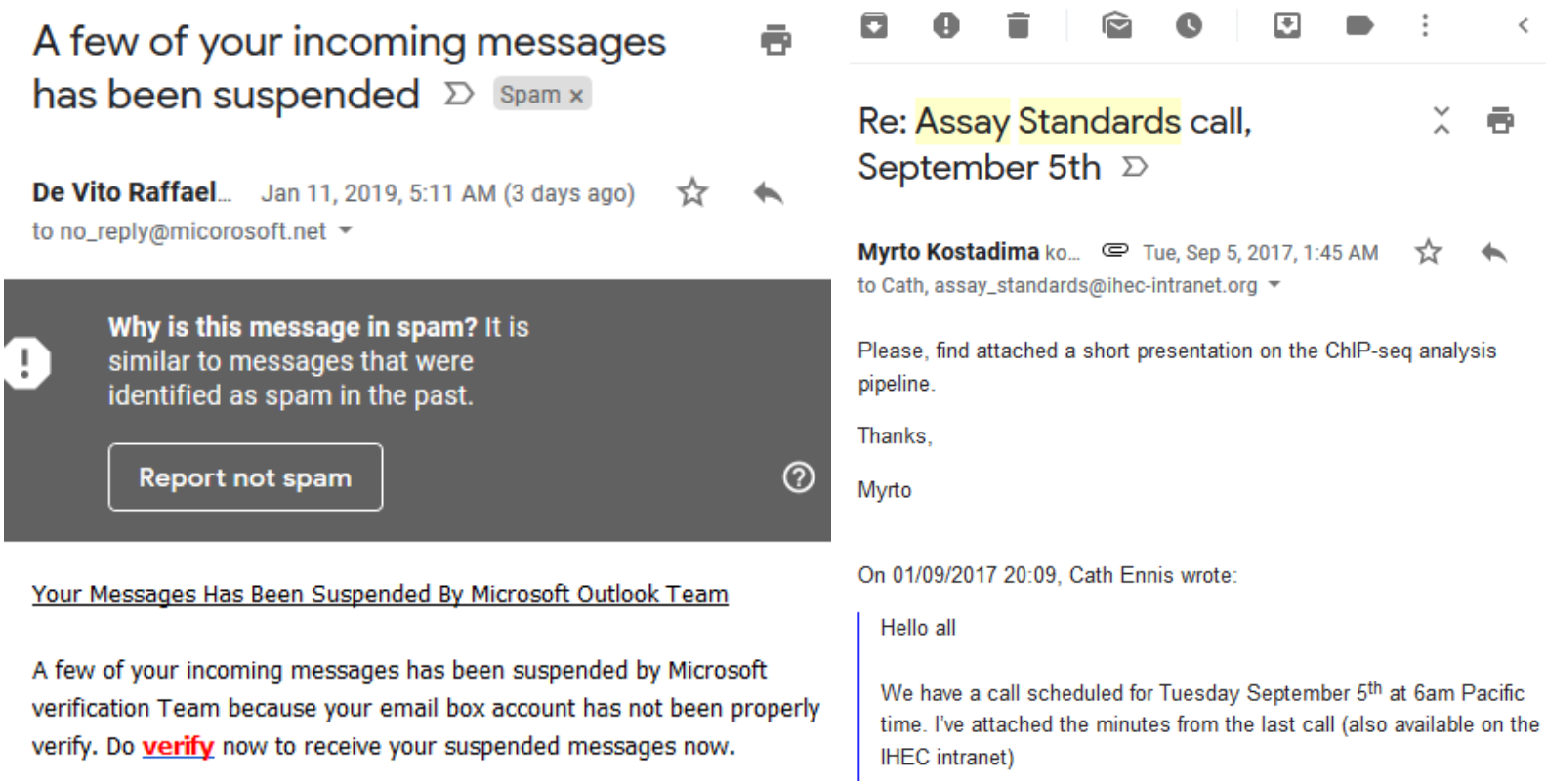
Figure 1. A screenshot of the labeling tool in use. The user is shown an image along with possibly one or more existing annotations, which are drawn on the image. The user has the option of annotating a new object by clicking along the boundary of the desired object and

Machine learning for image segmentation

Russell *et al.* 2007.

Q: What are the types/dimensions of x, y, f in this example?

Machine learning for spam filtering (Gmail)



Want: $f(\text{email message}) \in \{0, 1\}$ – binary classification, spam=1 or not=0.

Machine learning for translation (google books)

LE COMTE

DE

MONTE-CRISTO

PAR

ALEXANDRE DUMAS

ABRIDGED AND ANNOTATED

BY

EDGAR EWING BRANDON, A.M.

Professor of French in Miami University

NEW YORK

HENRY HOLT AND COMPANY

1900

THE

COUNT OF MONTE-CRISTO.

BY

ALEXANDRE DUMAS.

WITH

Twenty Illustrations,

DRAWN ON WOOD BY M. VALENTIN,

AND EXECUTED BY THE MOST EMINENT ENGLISH ENGRAVERS,
UNDER THE SUPERINTENDENCE OF MR. CHARLES HEATH.

IN TWO VOLUMES.

VOL. II.

LONDON:

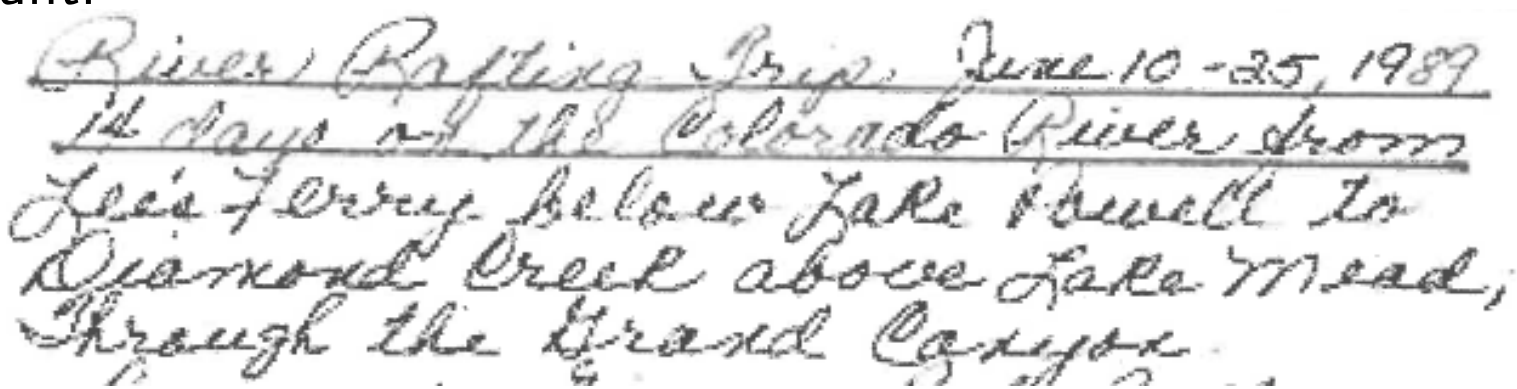
CHAPMAN AND HALL, 186 STRAND.

1846.

Machine learning for recognizing cursive handwriting

Optical/intelligent character/word recognition

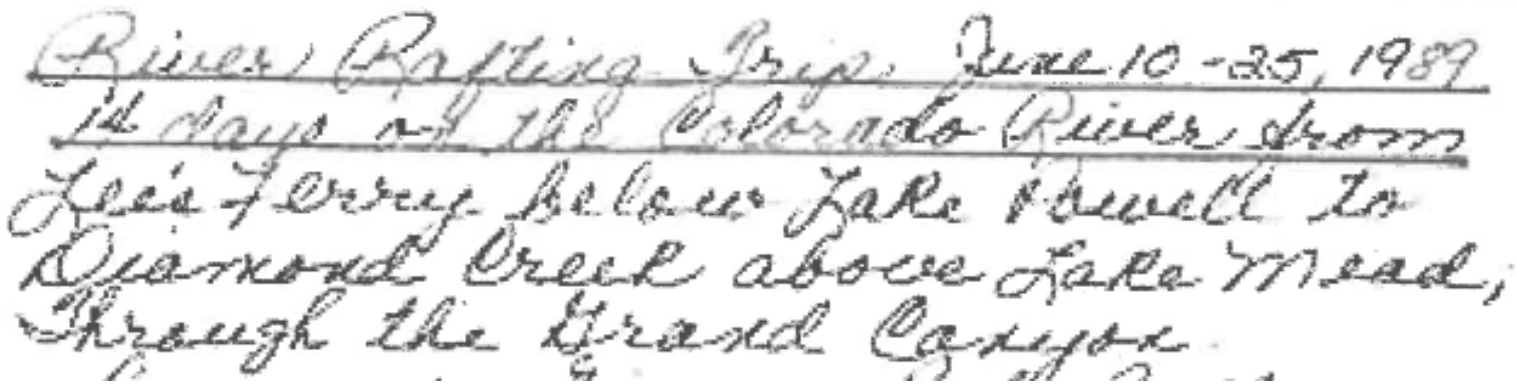
Want:

$f($  $)=$

Machine learning for recognizing cursive handwriting

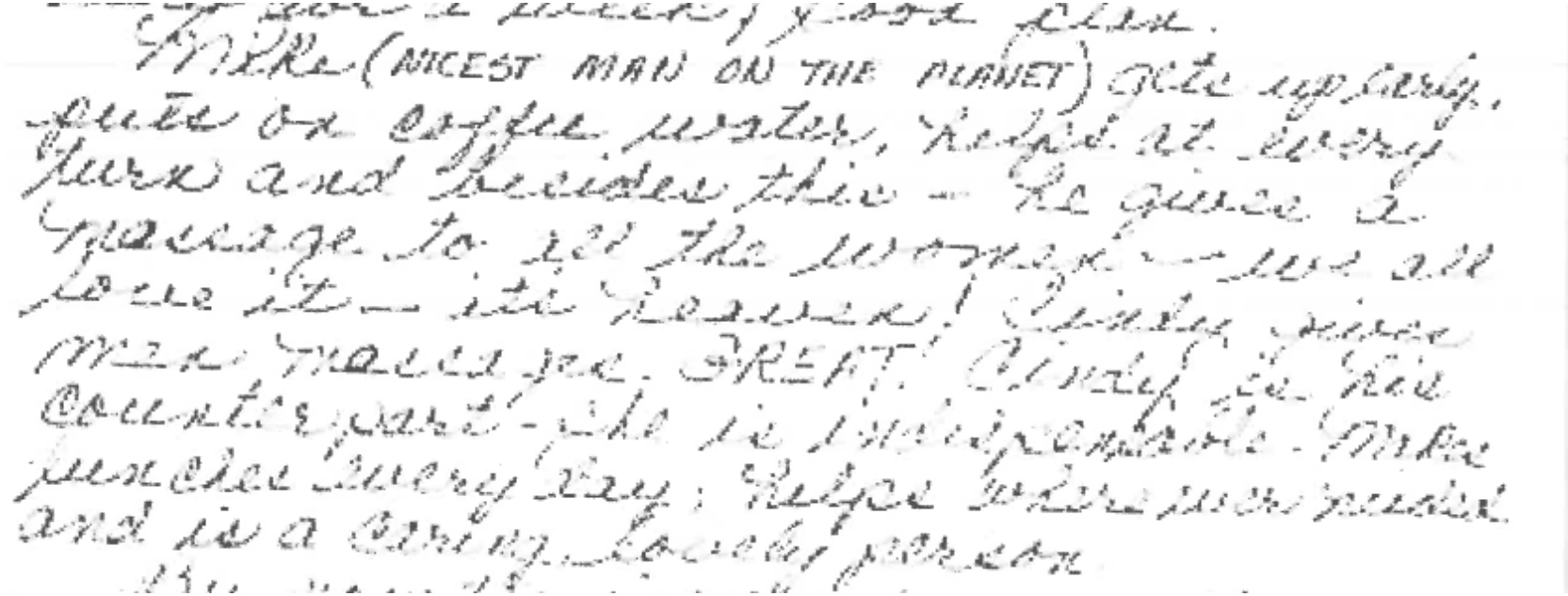
Optical/intelligent character/word recognition

Want:

$f($  $)=$

River Rafting Trip June 10–25, 1989
14 days on the Colorado River from
Lee's Ferry below Lake Powell to
Diamond Creek above Lake Mead,
Through the Grand Canyon

Machine learning for recognizing cursive handwriting



Q: sometimes you hear about "AI more accurate than human" – what does that mean, if the human is defining the labels?

Q: Can AI be more accurate at reading this than my grandmother, who wrote it?

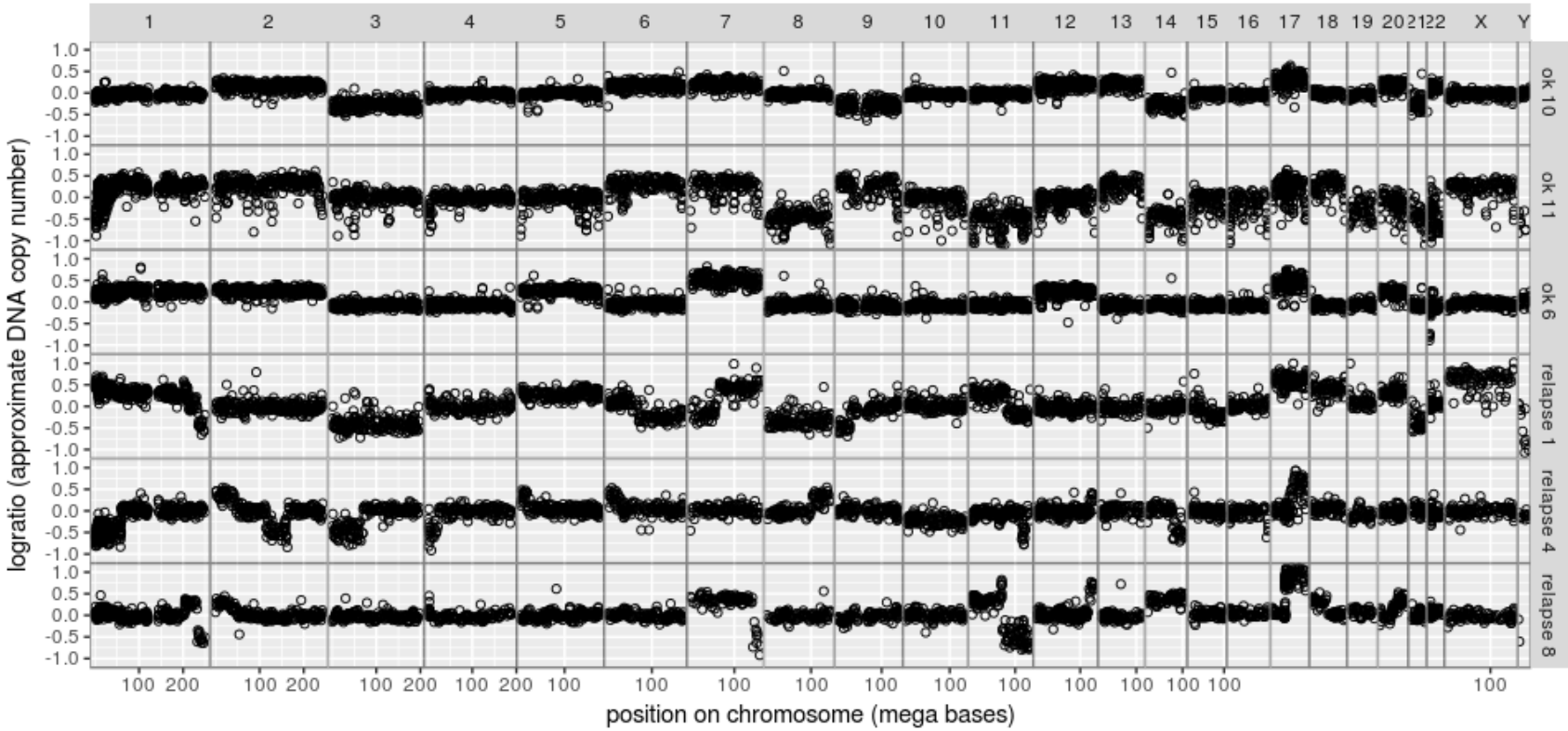
Machine learning for medical diagnosis

Original

f( ) =

Thanks to artificial intelligence and machine learning, diagnostic tools can be trained to read tissue samples and radiologic scans. Google researchers fed more than a quarter-million patients' retinal scans into algorithms that recognize patterns—and the technology "learned" to spot which patterns predict a patient has high blood pressure or is at increased risk for heart attack or stroke. In some comparisons, digital tools produced more accurate analyses than did human pathologists, dermatolo-

Machine learning for medical diagnosis 2



- Each row is a genomic profile from a cancer patient.
- Each column is a different chromosome.
- Approximate copy number plotted against chrom position.
- Want $f(\text{profile}) \in \{\text{ok}, \text{relapse}\}$.
- Can you determine an f by visual inspection?