Applications of machine learning

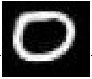
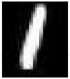
Toby Dylan Hocking
toby.hocking@nau.edu

15 January 2019

Machine learning intro: image classification example

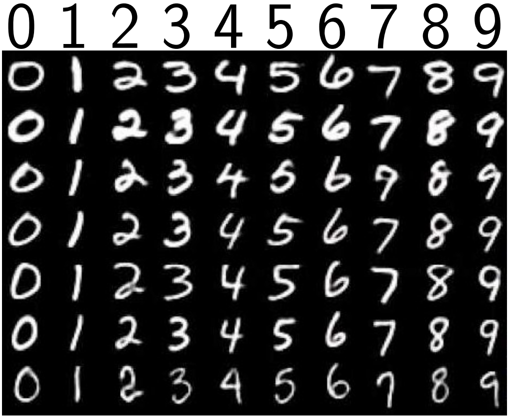
ML is all about learning predictive functions $f(x) \approx y$, where

- 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/unsupervised 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

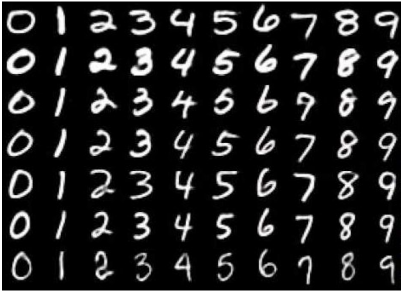
I give you a training data set with paired inputs/outputs, e.g.



Your job is to code an algorithm that learns the function f from the training data. (you don't code f)

Source: github.com/cazala/mnist

Advantages of supervised machine learning

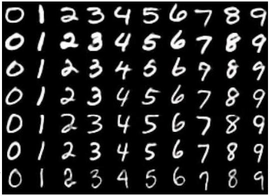


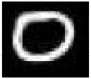
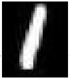
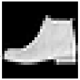
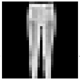
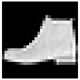
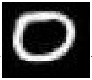
- Input $x \in \mathbb{R}^{16\times 16}$, output $y \in \{0, 1, \dots, 9\}$ types the same!
- Can use same learning algorithm regardless of pattern.
- Pattern encoded in the labels (not the algorithm).
- Useful if there are many un-labeled data, but few labeled data (or getting labels is long/costly).
- State-of-the-art accuracy (if there is enough training data).

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

Learning two different functions

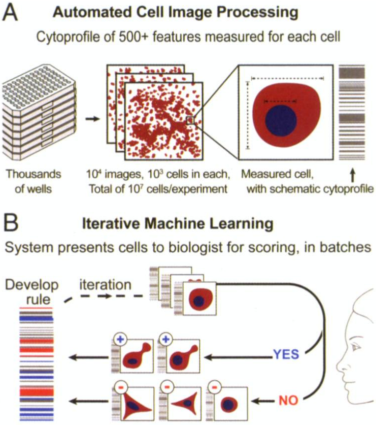
Say $\textsc{Learn}$ is a learning algorithm you have coded.

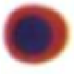
$\textsc{Learn}($$) \rightarrow f$, $\textsc{Learn}($$) \rightarrow g$

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

Machine learning for cell image classification

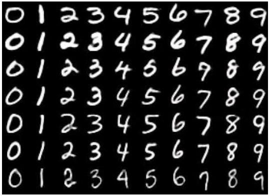
Jones *et al.* PNAS 2009. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning.



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

Advantages of supervised machine learning

- Can use same learning algorithm regardless of pattern.
- Pattern encoded in the labels (not the algorithm).
- State-of-the-art accuracy (if there is enough training data).



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