CS 182 Lecture 1: Intro

"low-resource"
↓

Multilingual Translation Model: Translating in/out of rare languages works better after training on more common languages.

Zero-shot: unseen pairs

English → French   English   French   "thought" is
  standard   "thought" 2 ────────→ "thought"   language-agnostic
  translation model   representation.
                                                             semantic meaning.

Representation Learning?   complex
                                            ↙ input
  ↳ how do we meaningfully represent the x in "predict y from x"?
  ↳ power of DL: learn such representations

What are we learning?
→ NN   → CV, NLP   - other cool stuff
→ optim. algs.   → RL

ML: define program to acquire input-output relationship from data
  $f(x, \theta) = y$ or $f_\theta(x) = y$.   learn θ. ← feature-based, 'shallow' learning
                                            - coming up w/ good features is hard.

DL: learned features (representations), layered: multiple layers.

  input → lower-level rep. → higher-level rep. → output   ⎤ usually a
                                                          ⎦ deep neural net.
  → params for each layer are trained w.r.t. overall task objective.
    (end-to-end learning)

                              → big models (layers)
What makes DL work? ─┬─   large datasets   ─→ enough compute

train model on Tрф (specialized for training)

Model capacity: how many diff. functions a particular model class can
                                                              represent

Major themes: Automation, end-to-end, learning vs. inductive bias.
                                                  (built-in knowledge)
              algorithms that scale w/ data/capacity/compute

ReLU: Rectified linear unit (nonlinear/nonlinear)