Machine Learning:
An Applied Econometric Approach

Jann Spiess

based on work with Sendhil Mullainathan

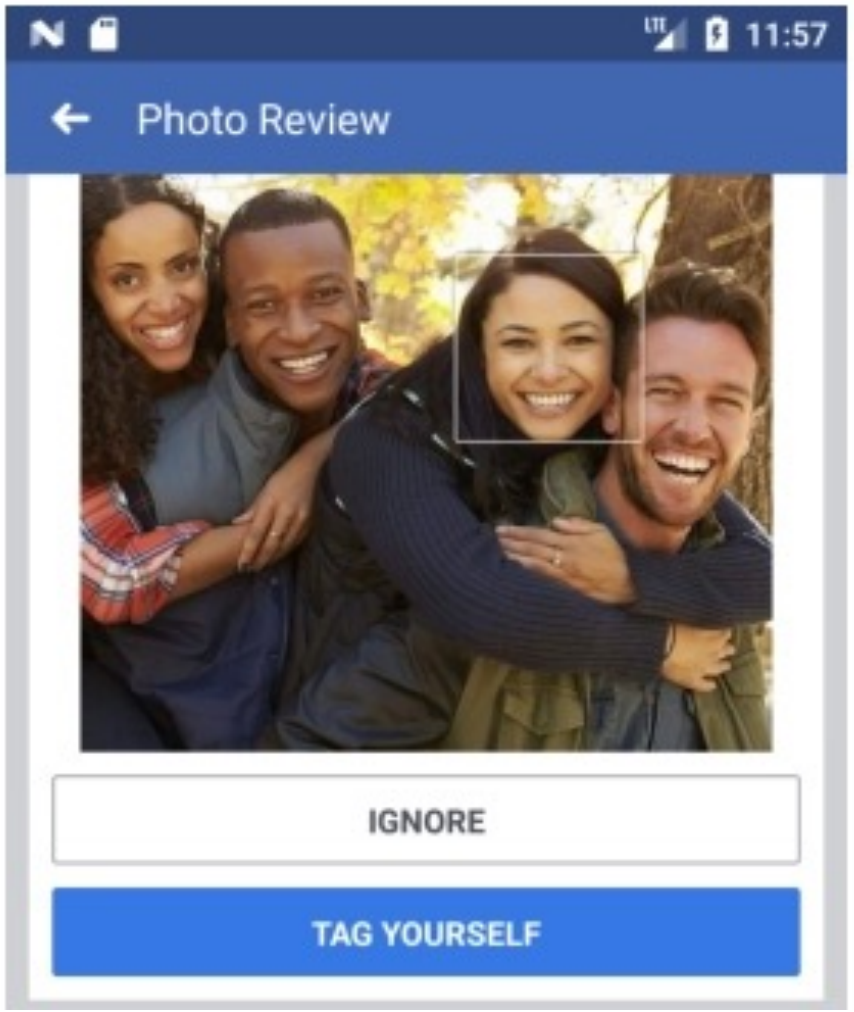
in collaboration with Susan Athey and Niall Keleher

1. Introduction

Machine Intelligence

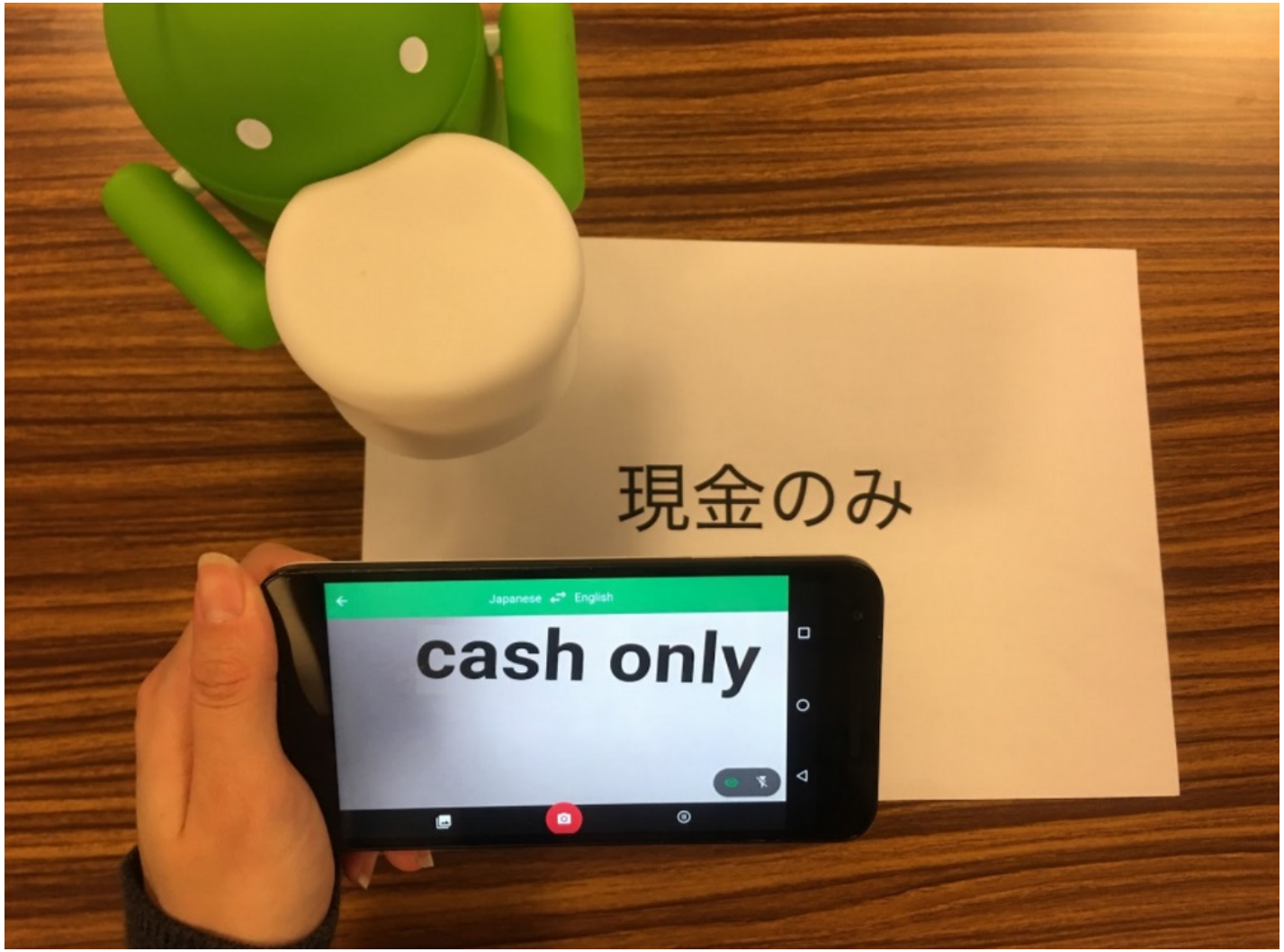
Picture source: Ex Machina / A24

Advances in Machine Intelligence



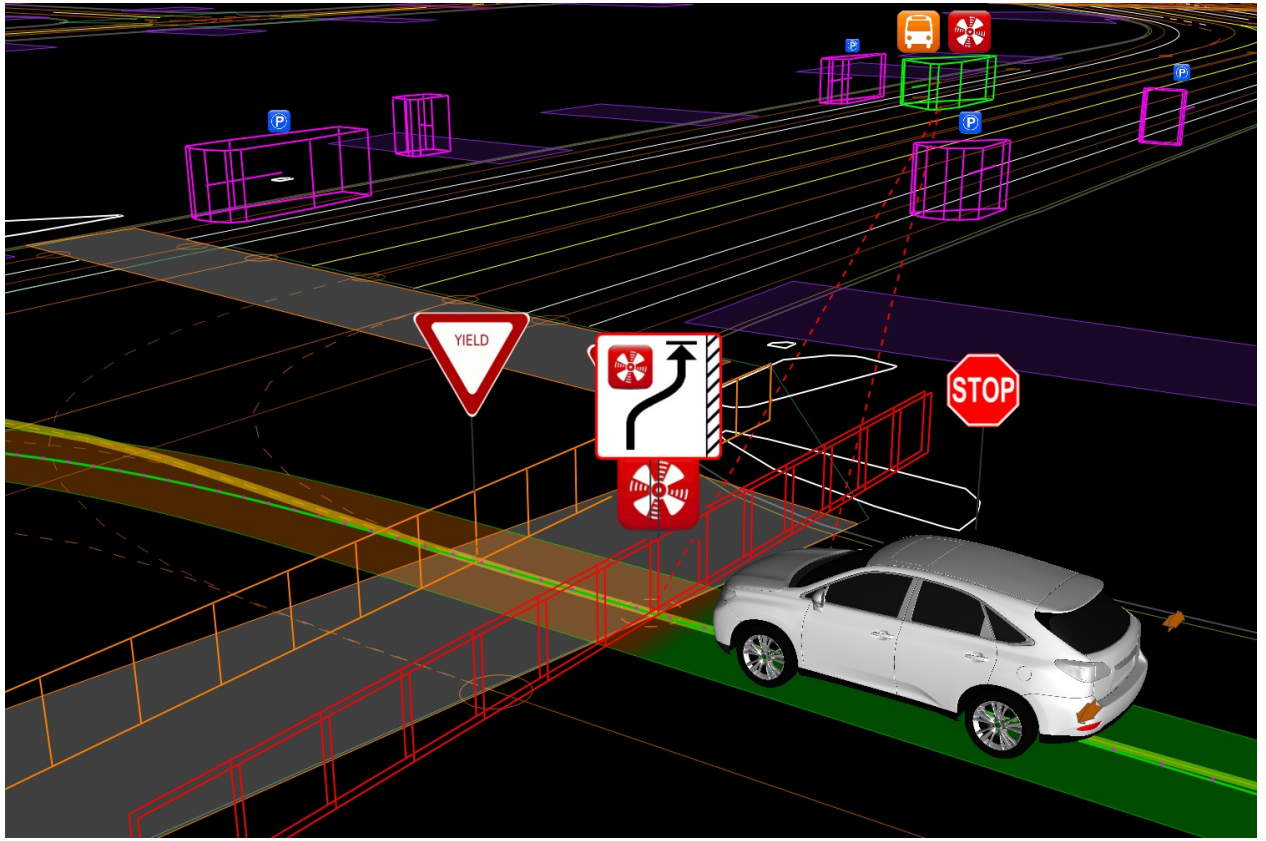
Picture source: Facebook

Advances in Machine Intelligence



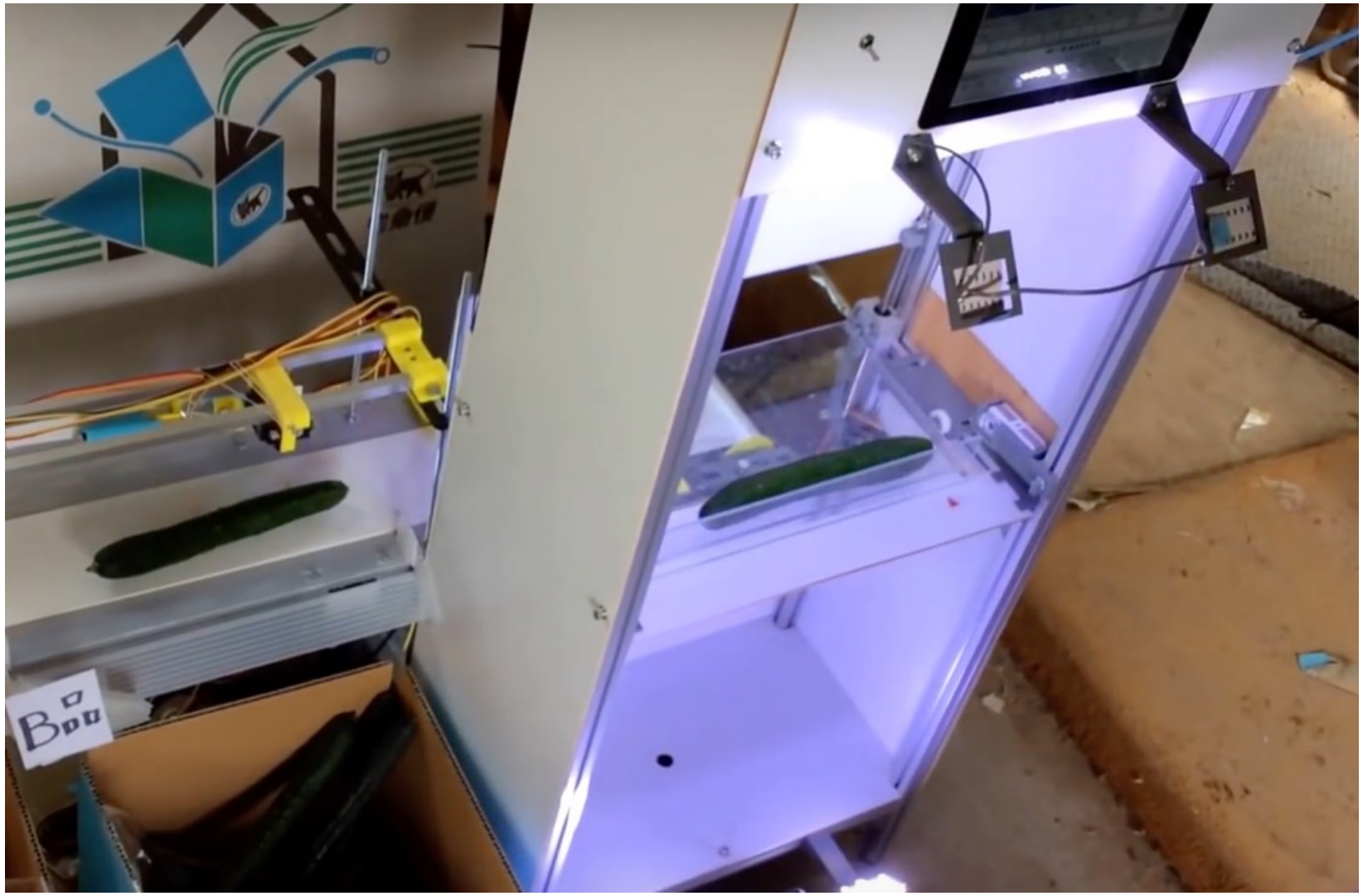
Picture source: Google

Advances in Machine Intelligence



Picture source: Waymo

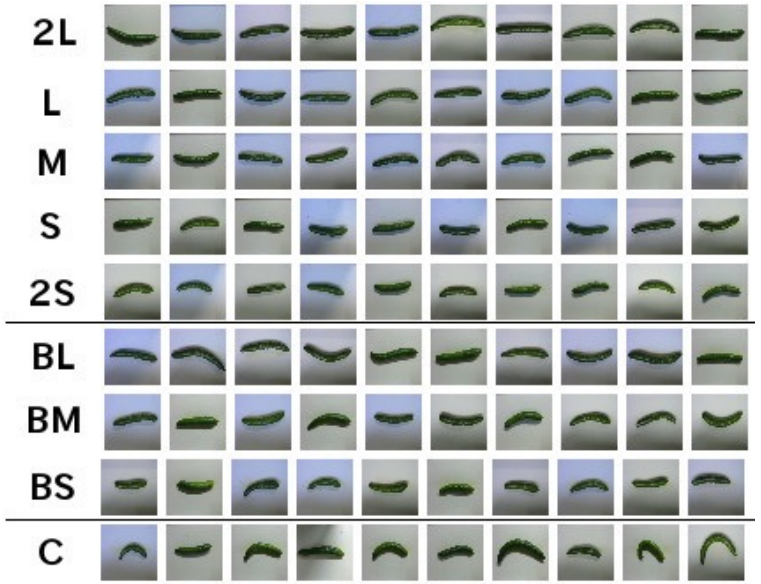
Machine Intelligence in the Field



Images: Makoto Koike (via The New Yorker), Kazunori Sato (via YouTube)

Cucumber Classification Problem

$$\underbrace{\hat{y}}_{\text{cucumber quality}} = \hat{f}(\overbrace{\tilde{x}}^{\text{pixels}})$$



- Old-style AI: deduce from human intuition, introspection
- New-style ML: induce from training data
  - Take "labelled" data
  - Fit a function $\hat{f}$ in the training sample

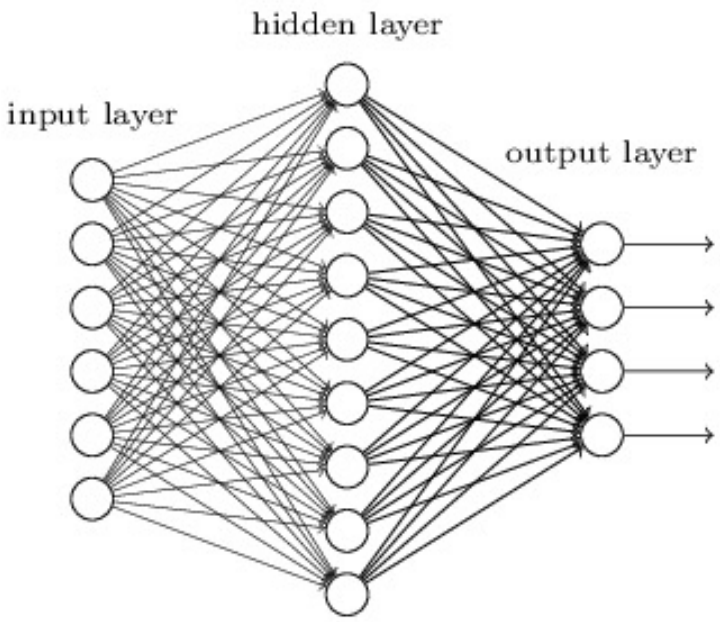
Image source: Google

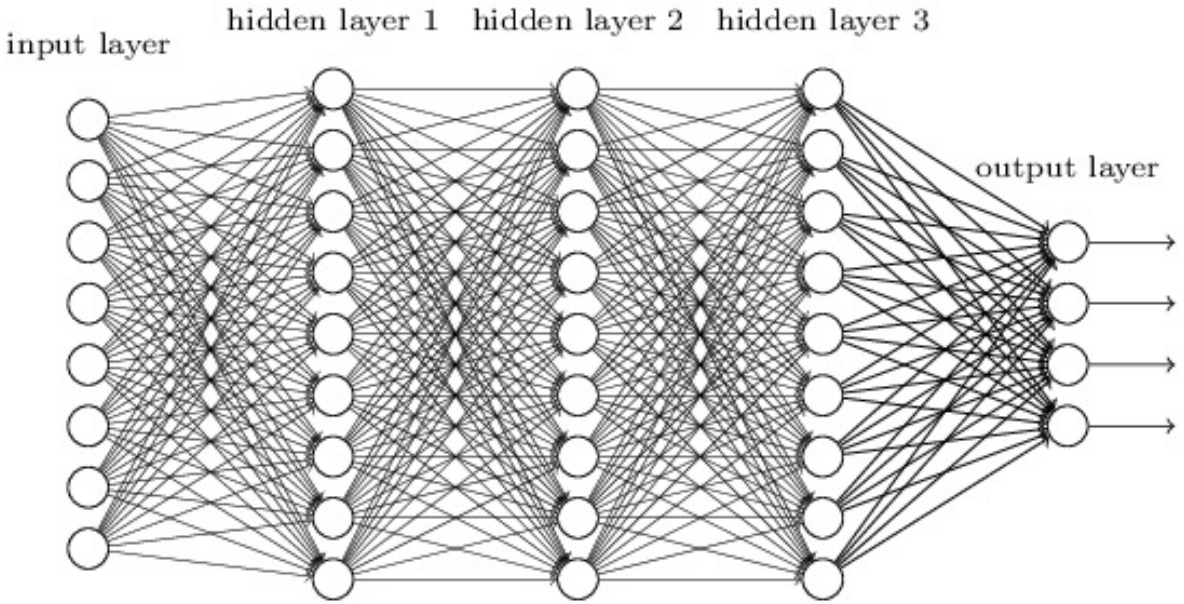
Isn't This Just Statistics?

Training data
(y, x)

$\hat{f}$

Application data
$(\hat{y} = \hat{f}(x), x)$





Some Features of Machine Learning (ML)

- Flexible, rich, data-driven models
- Can work with very high-dimensional data
- Limit expressiveness to avoid overfit (**regularization**)
- Learn how much expressiveness to allow (**tuning**)
- Industry-strength tools readily available

- Supervised learning: focus on **prediction**

→ Idea: turn intelligence task into supervised-learning problems

  - Bank decides who to give credit to
  - Tax authority decides which returns to audit
  - Image recognition
  - Self-driving cars

Advances in Supervised Learning

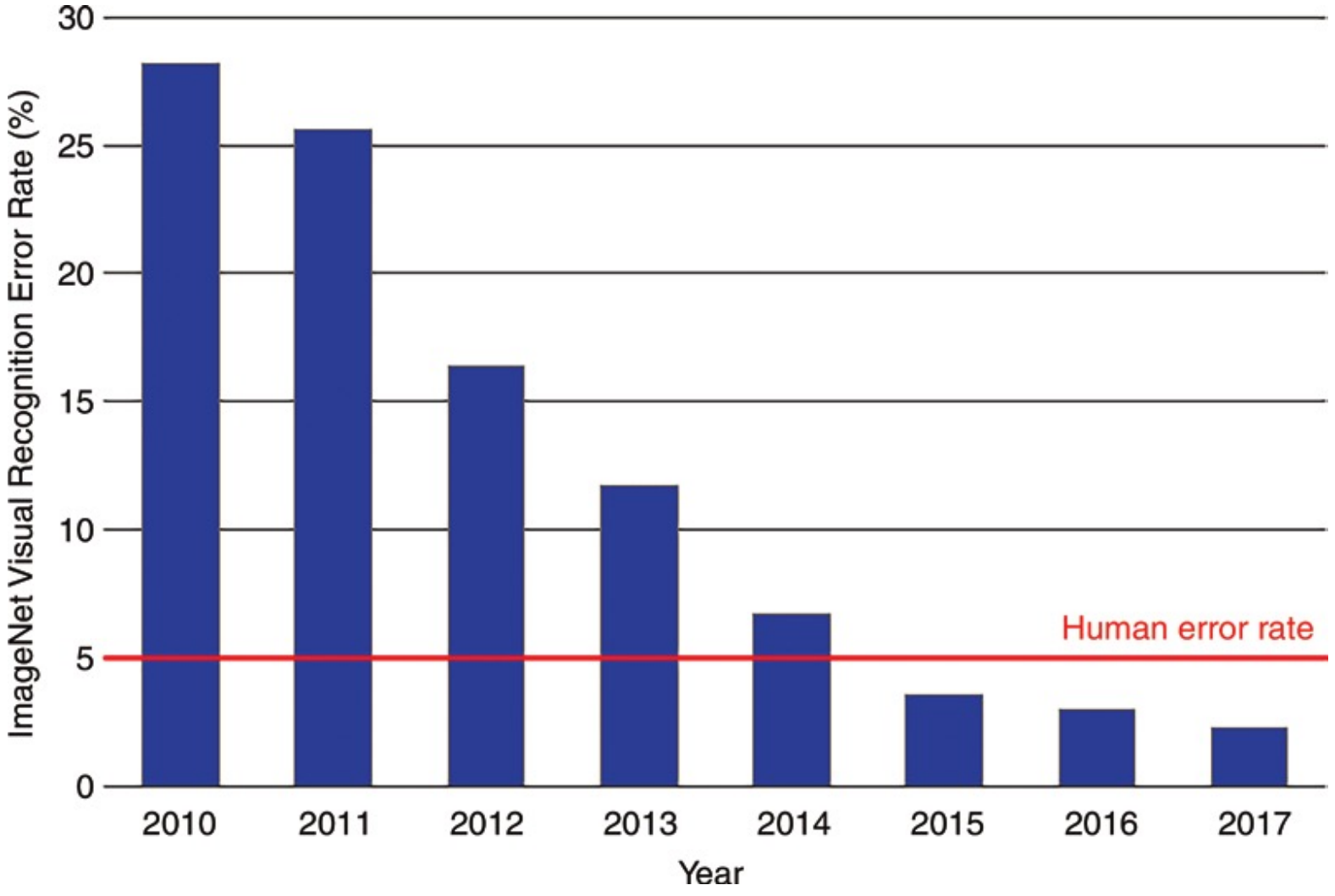
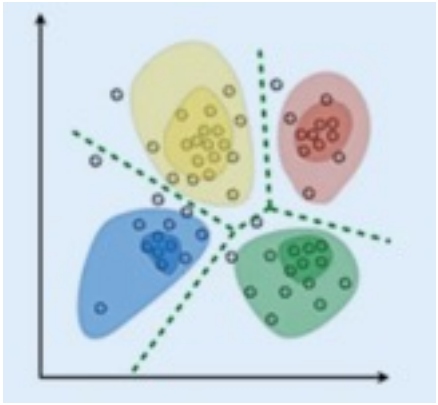


Image sources: Krizhevsky et al. (2012), Langlotz et al. (2019)

Advances Beyond Supervised Learning



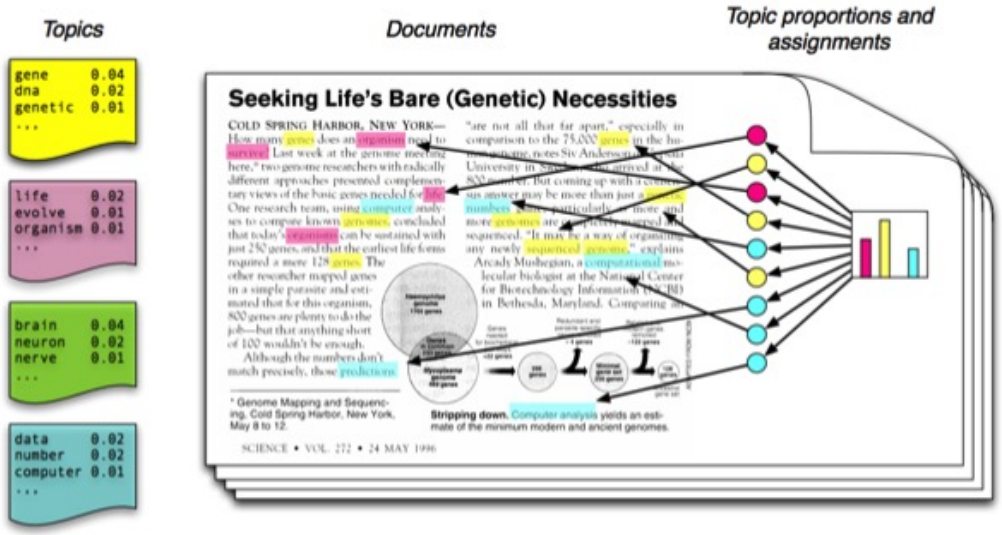
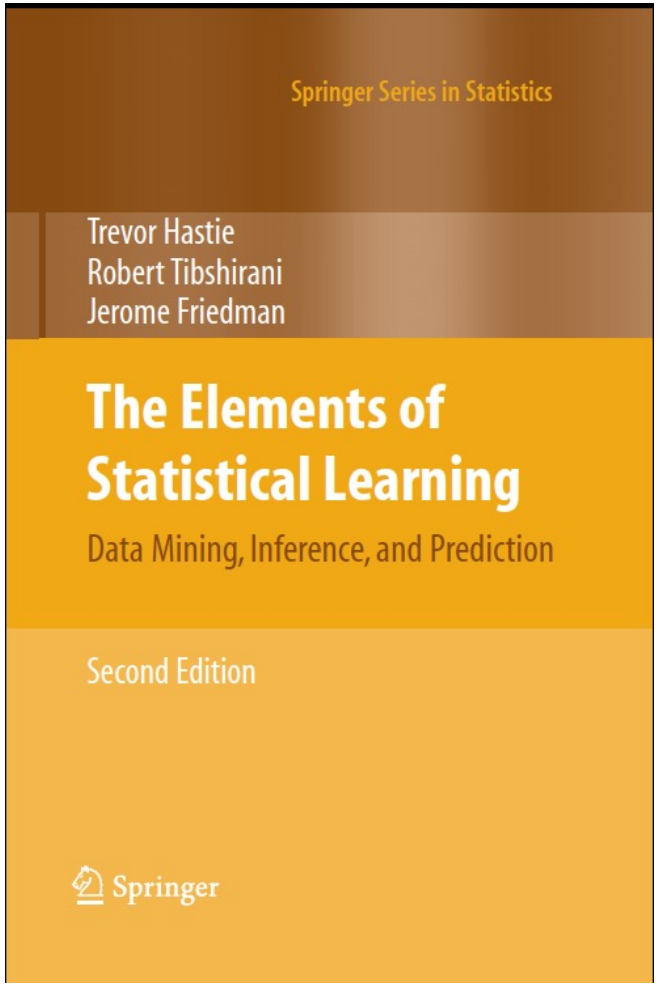


Image sources: Langs et al. (2018), Blei (2012), DeepMind

Goal of Webinar

- Plenty of resources to learn ML, statistics behind it – but not necessarily how to use ML conceptually and in program evaluation
- Framework for using (and critically evaluating) ML in applied econometrics
- Concrete tools in which ML can enhance applied work
- Will focus on supervised learning on social-science type data
- Will only touch on technical challenges in implementation
- Work in progress on an evolving agenda!



Structure of Part on Machine Learning Basics

1. Introduction

2. The Secret Sauce of Machine Learning

3. Prediction vs Estimation