Machine Learning

Intro to AI
Bert Huang
Virginia Tech

Machine Learning

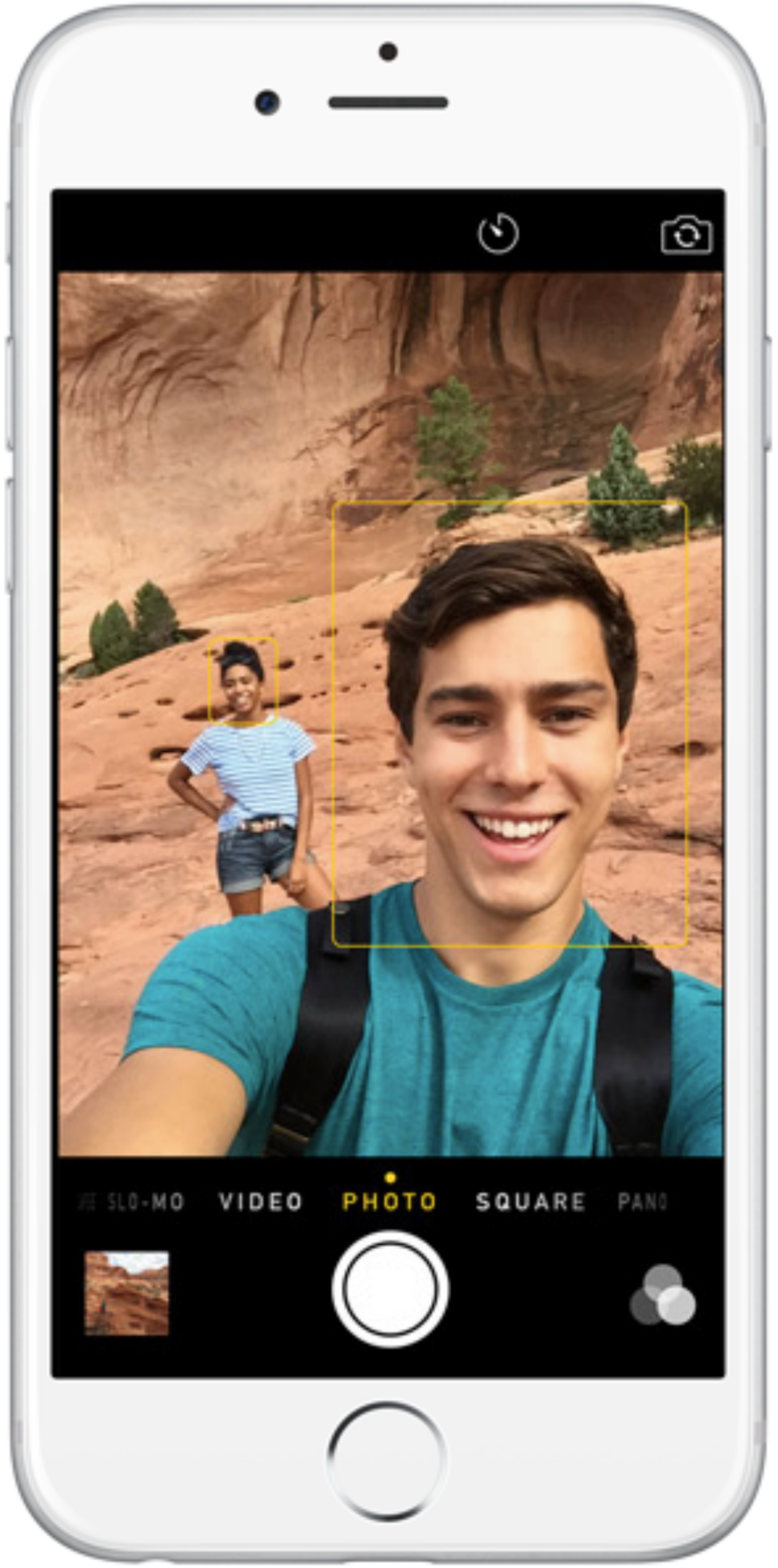
- Learning: improving with experience at some task
    - Improve over **task**
    - with respect to some **performance measure**
    - based on some **experience**
- Writing computer programs that write computer programs

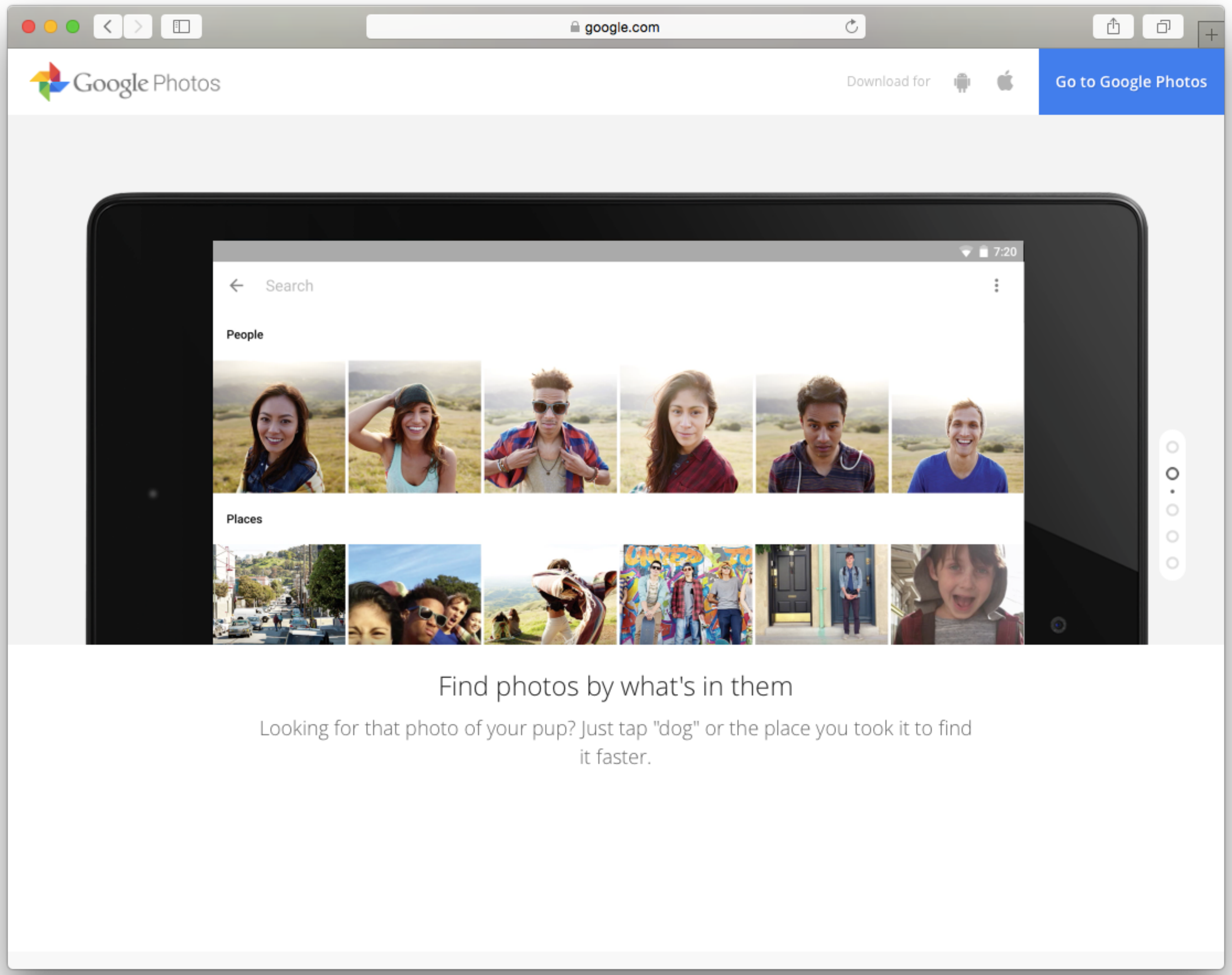
Learning definition by Tom Mitchell

Outline

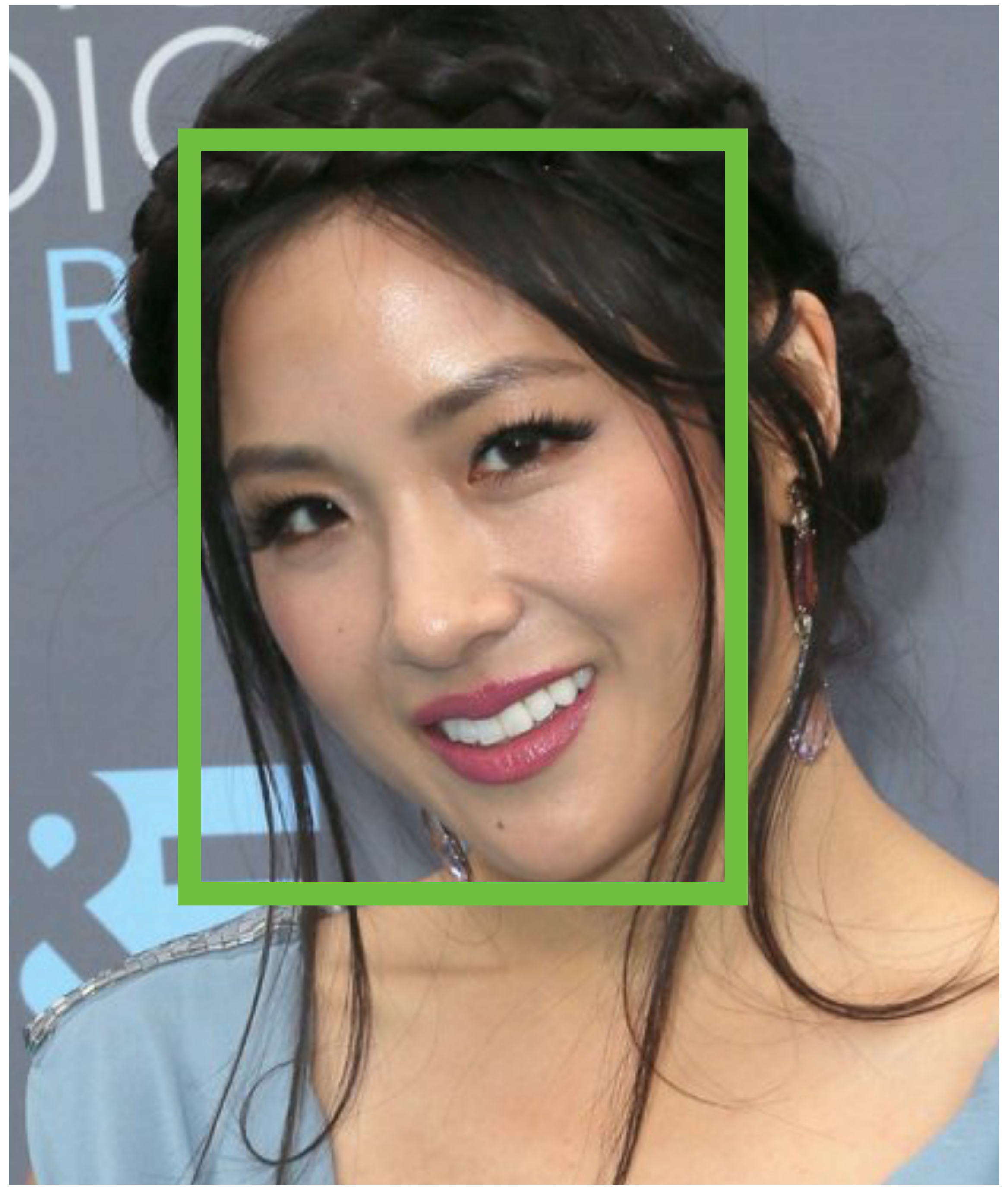
- Three machine learning stories/cautionary tales
- Deep learning definition
- Types of machine learning
- Best practices

Machine Learning Story 1
Face Detection & Recognition


SLO-MO
VIDEO
PHOTO
SQUARE
PANO


google.com
Google Photos
Download for
Go to Google Photos
7:20
Search
People
Places
Find photos by what's in them
Looking for that photo of your pup? Just tap "dog" or the place you took it to find it faster.

What Does a Human Face Look Like?



Apple II image from wikipedia.com.
Eyes added digitally.

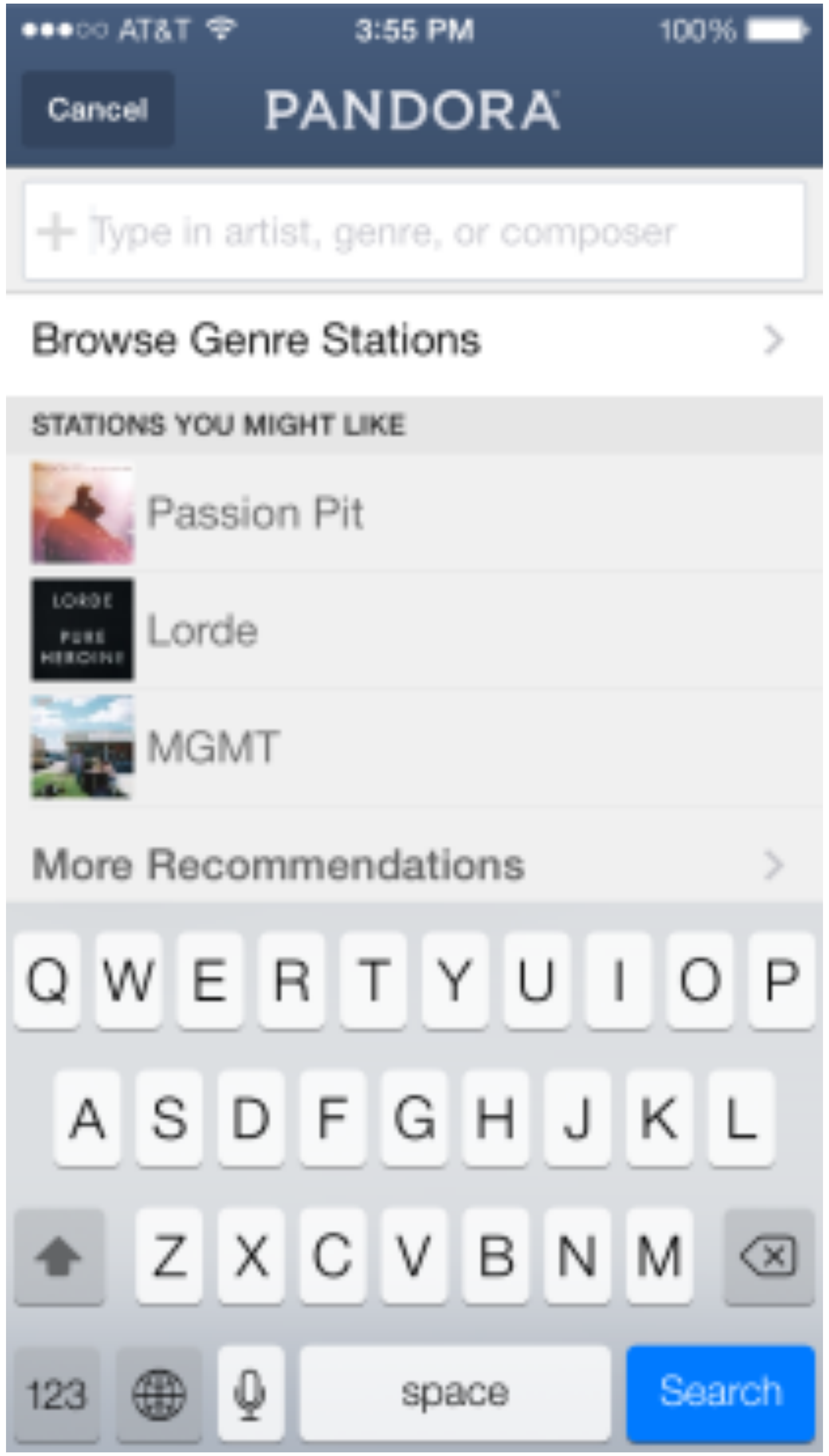
Apple II image from wikipedia.com.
Eyes added digitally.

```
if pixel153 > 128 & pixel154 > 128 &
   pixel155 > 128 & pixel156 < 64 &
   sqrt(pixel157) < 82 &
   log(pixel1132 * pixel1133) > 1  ….
      then image is a face*
```

* (not a real face recognition program)

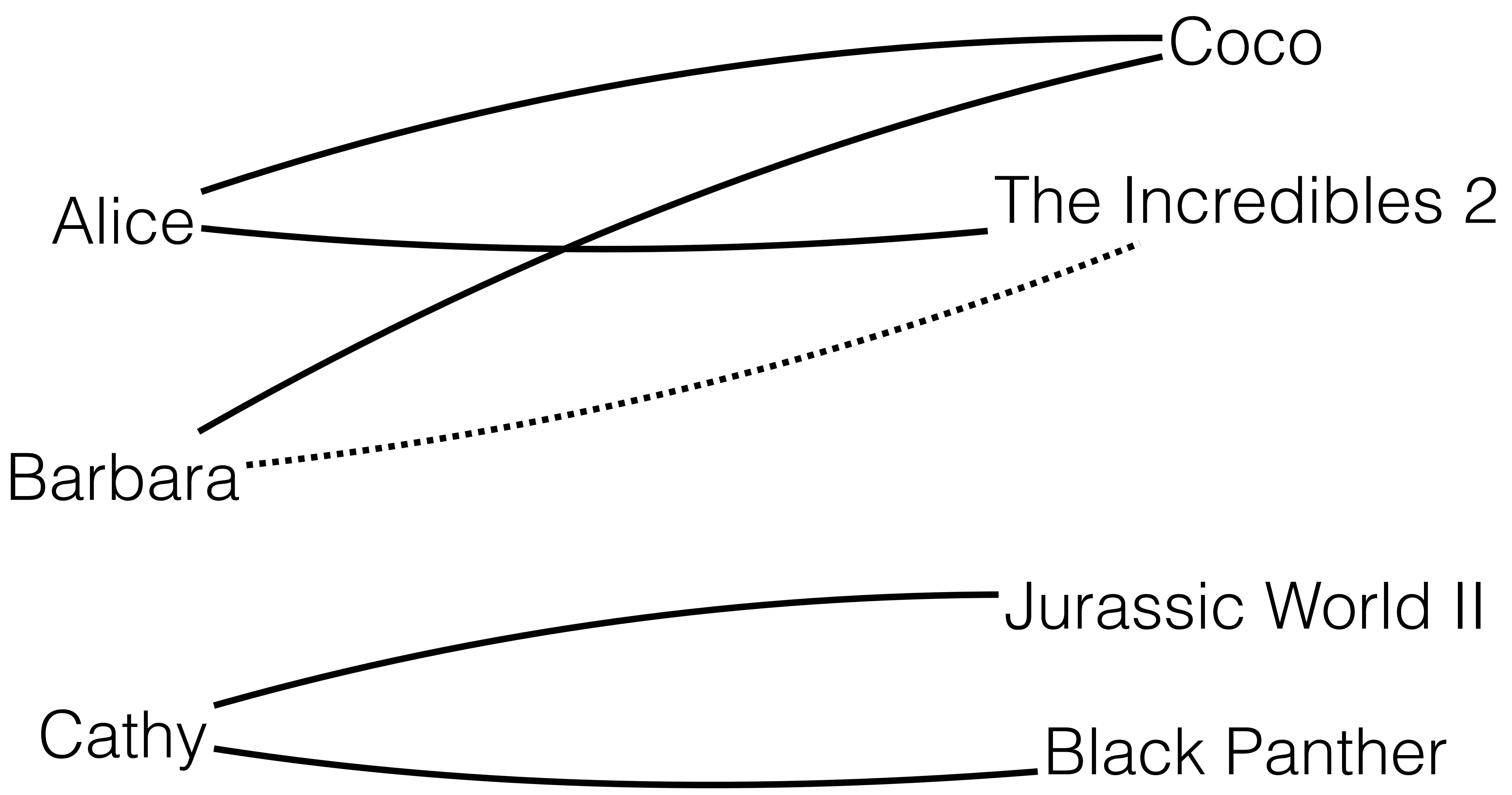
Apple II image from wikipedia.com.
Eyes added digitally.

Machine Learning Story 2
Recommender Systems


AT&T
3:55 PM
100%
Cancel
PANDORA
Type in artist, genre, or composer
Browse Genre Stations
STATIONS YOU MIGHT LIKE
Passion Pit
Lorde
MGMT
More Recommendations
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M
123
space
Search


People You May Know
see all
Jim M
Add as Friend
Erin Elizabeth K
Add as Friend
Josh S
Add as Friend


Recommended for You
se recommendations are based on items you own and more.
All | New Releases | Coming Soon
Cybertext: Perspectives on Ergodic Literature
by Espen J. Aarseth (Aug 6, 1997)
Average Customer Review: (3)
In Stock
List Price: $22.95
Price: $19.55
29 used & new from $10.82
Add to cart
Rate it
Recommended because you added Hamlet on the Holodeck to your Shopping Cart and more (Fix this)
Narrative as Virtual Reality: Immersion and Interactivity in Lit
Media (Parallax: Re-visions of Culture and Society)


Coco
Alice
The Incredibles 2
Barbara
Jurassic World II
Cathy
Black Panther

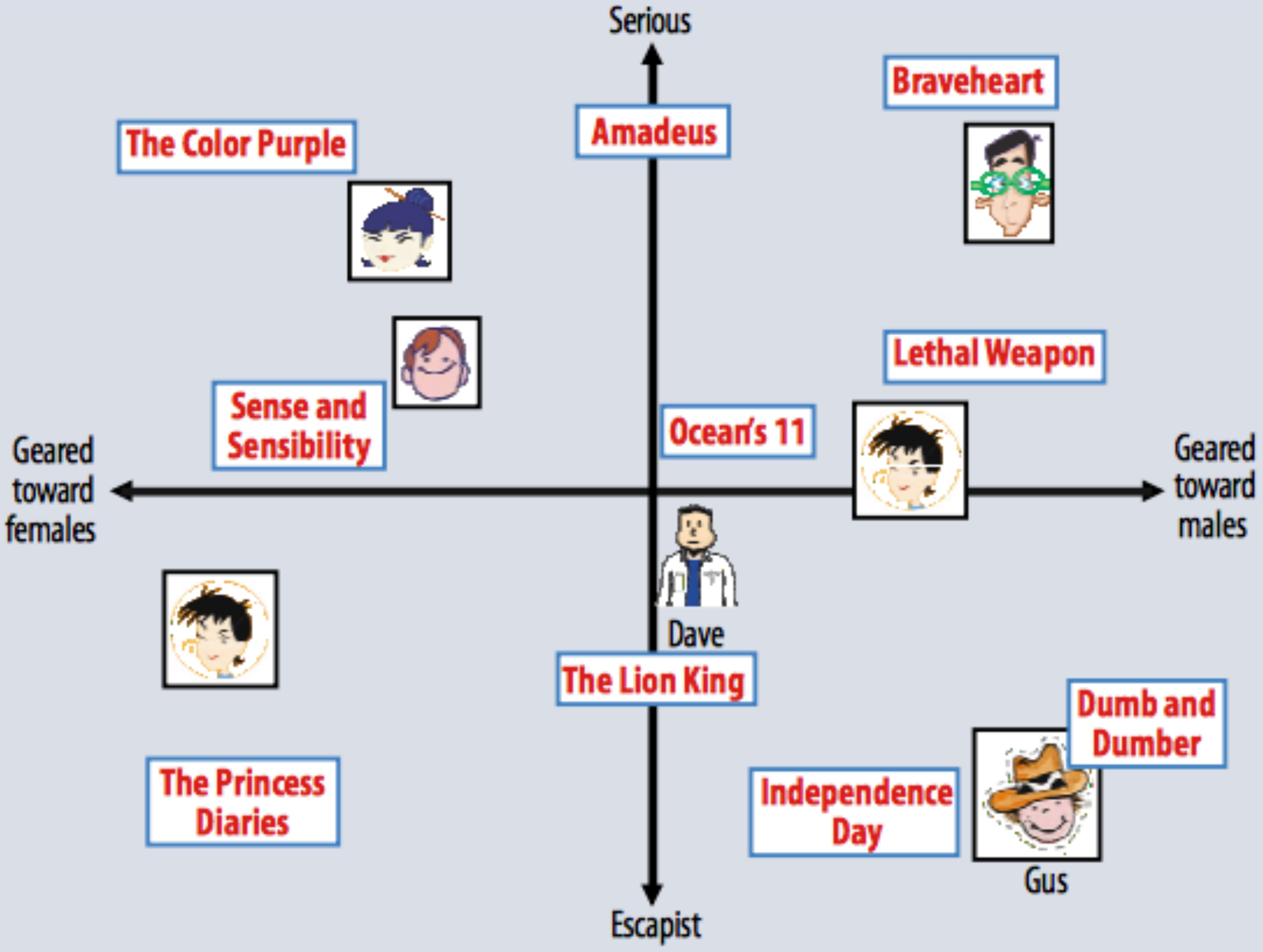


Figure from Koren, Bell, Volinksy, IEEE Computer, 2009

Applications of Recommendation

- Movies
- Books
- Music
- Medicine
- Education
- Jobs

Applications of Recommendation

- Movies
- Books
- Music
- Medicine
- Education
- Jobs

Machine Learning Story 3
Housing Markets

ASA Excellence in Statistical Reporting Award

The formula that killed Wall Street

Wall Street in the mid-1980s turned to the quants – brainy financial engineers – to invent new ways to boost profits. They and their managers, though laziness and greed, built a huge financial bubble on foundations that they did not understand. It was a recipe for disaster. The journalist **Felix Salmon** won the American Statistical Association's Excellence in Statistical Reporting Award for 2010. We reprint his article, first published as the cover story of *Wired* magazine, because it brilliantly conveys complex statistical concepts

A formula in statistics, misunderstood and misused, has devastated the global economy

In the years before 2008, it was hardly unth
that a math wizard like David X. Li might so
earn a Nobel Prize. After all, financial econo
even Wall Street quants – have received the
in economics before, and Li's work on measuri
has had more impact, more quickly, than p
Nobel Prize-winning contributions to the field
though, as dazed bankers, politicians, regulato
investors survey the wreckage of the biggest fi
meltdown since the Great Depression, Li is p
thankful he still has a job in finance at all. N
his achievement should be dismissed. He too
toriously tough nut – determining correlation,
seemingly disparate events are related – and

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

The formula that killed so many pension plans: David X. Li's Gaussian copula, as first published in 2000. Investors exploited it as a quick – and fatally flawed – way to assess risk.

Probability

Specifically, this is a joint default probability – the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times

The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality

A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula

This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions

The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

Gamma

The all-powerful correlation parameter, which reduces correlation to a single constant – something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

Machine Learning Stories

- Face recognition
- Recommender systems
- Finance

What is deep learning?

raw image input
↓
image preprocessing
↓
edge detection
↓
object detection
↓
object identification

raw image input
↓
learnable component (neural network)
↓
another neural network
↓
another neural network
↓
object identification

Deep Learning

- Using machine learning to simultaneously train every part of the process from **raw input** to **raw output**
  - Considered “deep” when compared to “shallow” approach of training/designing each component on its own

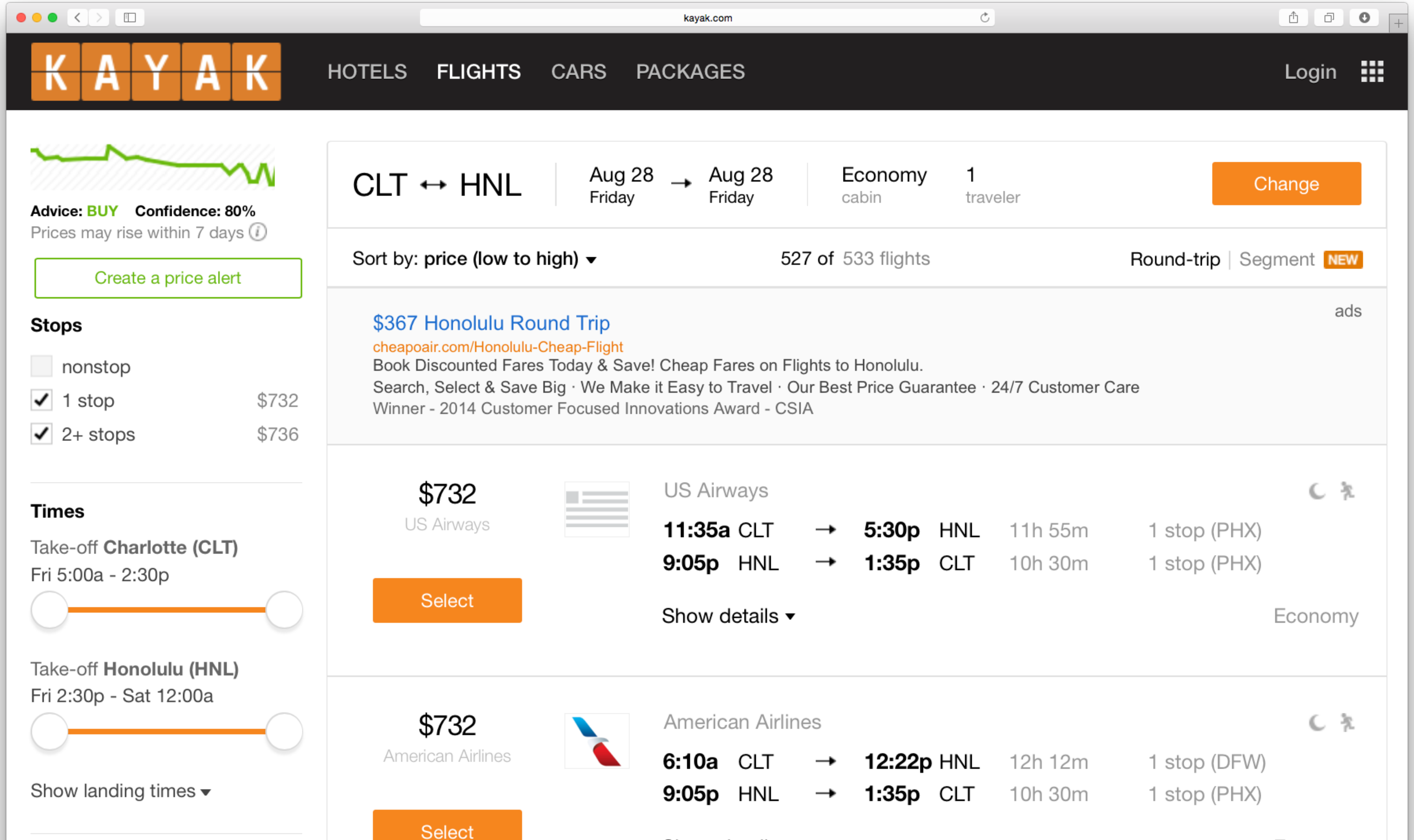
Types of Machine Learning

- Types of learning settings
  - Supervised learning
  - Unsupervised learning
- Types of learning algorithms
  - Batch learning
  - Online learning

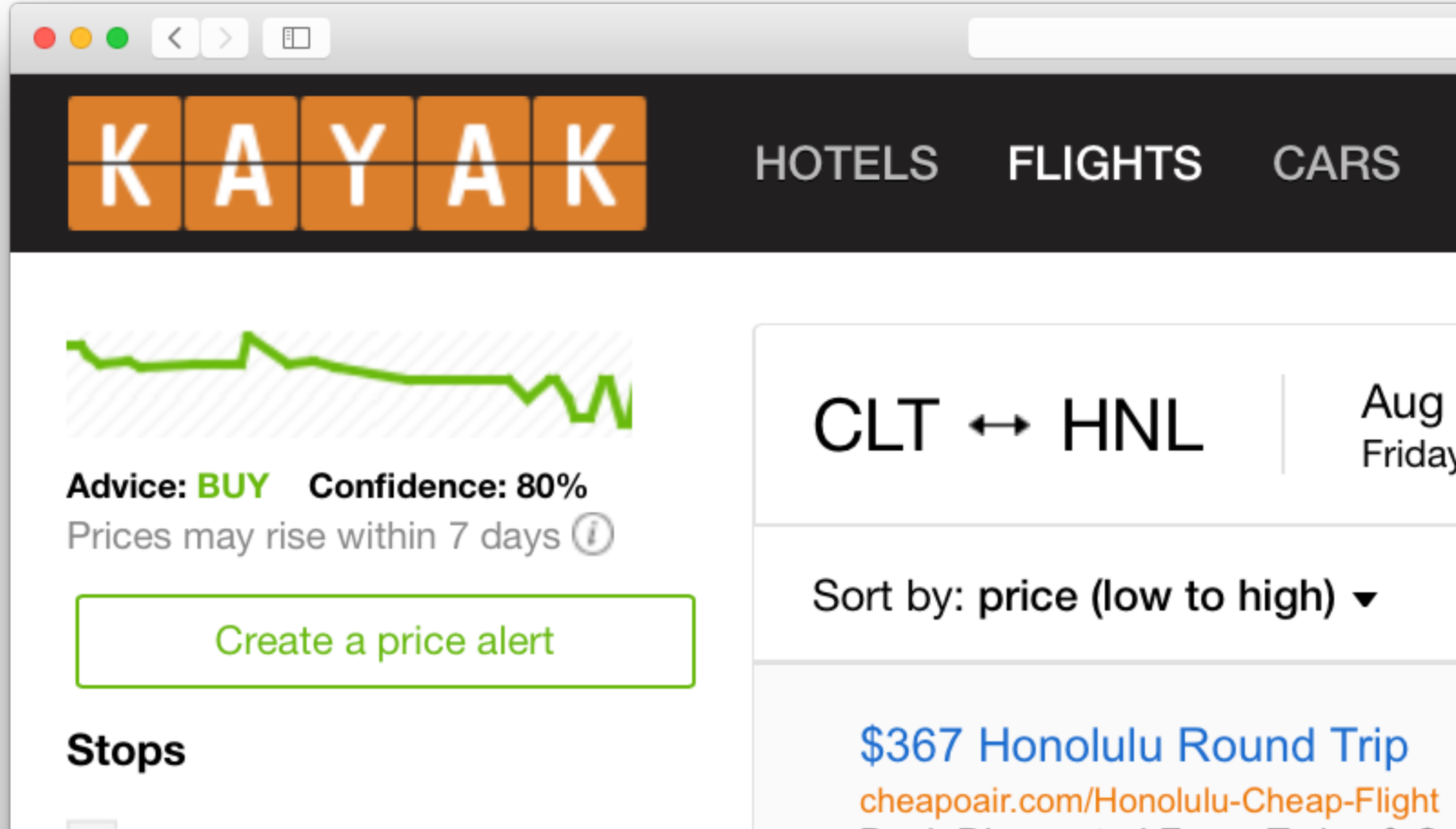
Example: Digit Classification

http://ufldl.stanford.edu/housenumbers/

Example: Airline Price Prediction



Example: Airline Price Prediction



Batch Supervised Learning

- Draw data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ from distribution $\mathbb{D}$
- Algorithm A learns hypothesis $h \in H$ from set H of possible hypotheses $A(D) = h$
- We measure the quality of h as the expected **loss**: $\underset{(x,y) \in \mathbb{D}}{E} [\ell(y, h(x))]$
  - This quantity is known as the **risk**
  - E.g., loss could be the Hamming loss $\ell_{\text{Hamming}}(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{otherwise} \end{cases}$

classification

Online Supervised Learning

- In step ***t***, draw data point ***x*** from distribution $\mathbb{D}$
- Current hypothesis ***h*** guesses the label of ***x***
- Get true label from oracle ***O***
- Pay penalty if ***h*(*x*)** is wrong (or earn reward if correct)
- Learning algorithm updates to new hypothesis based on this experience
  - Does not store history

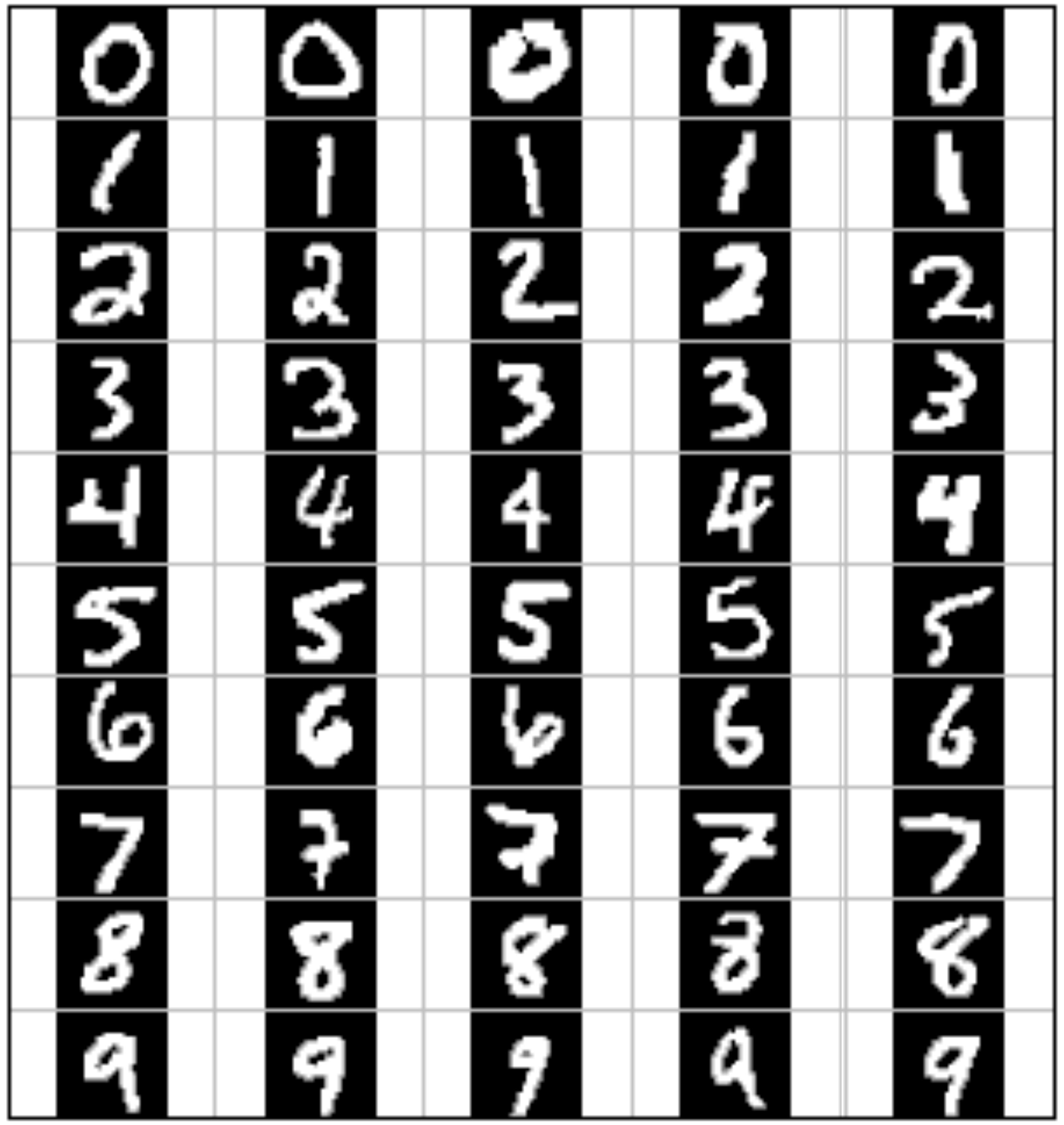
Learning Settings

- Supervised or unsupervised (or semi-supervised, weakly supervised, transductive…)
- Online or batch (or reinforcement…)
- Classification, regression
  - (or structured output, clustering, dimensionality reduction…)

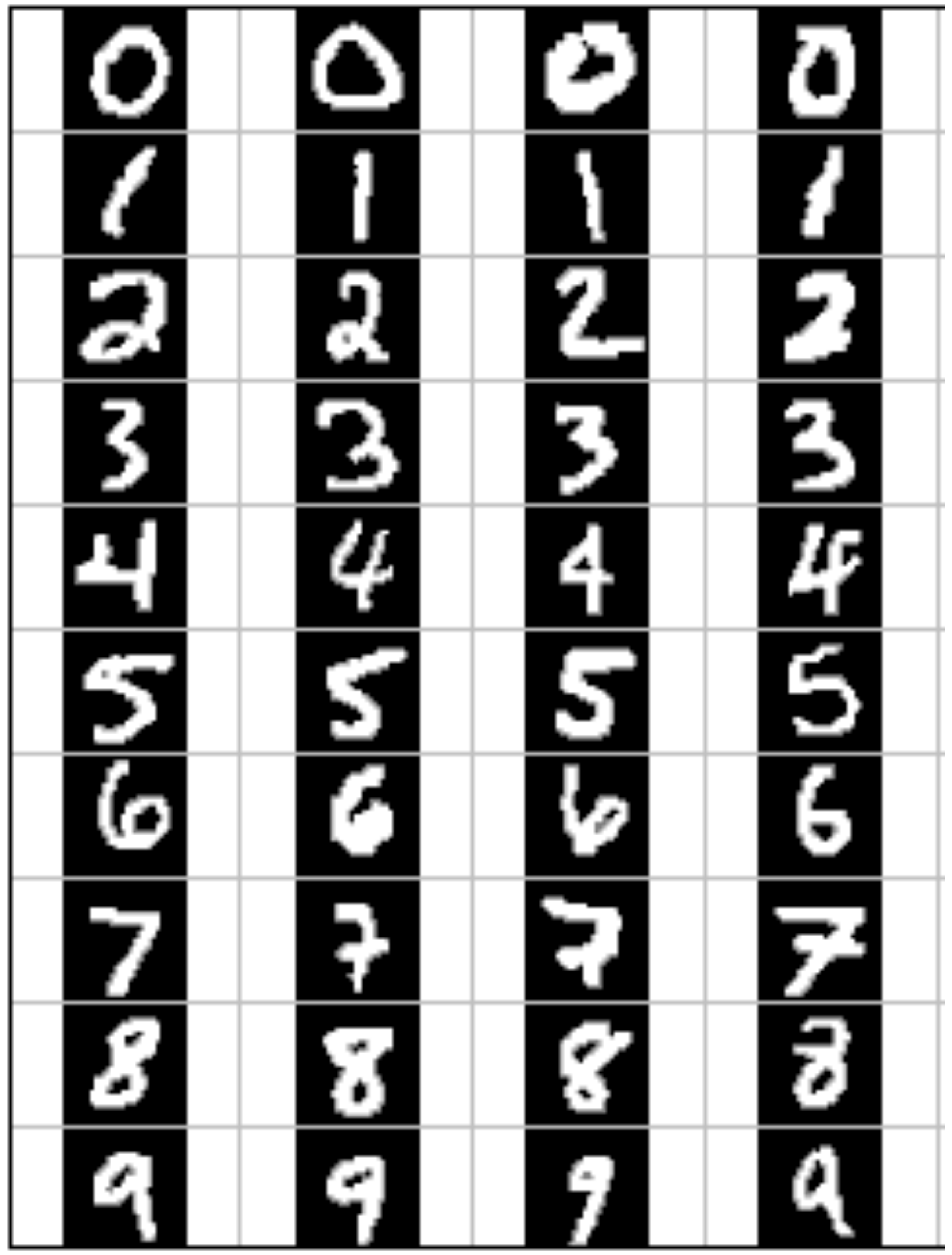
Best Practices

- Try range of models with different **capacity**
- Split data into training, validation, and testing sets
- Measure performance on evaluation set to tune parameters
- Measure performance on testing set as final check

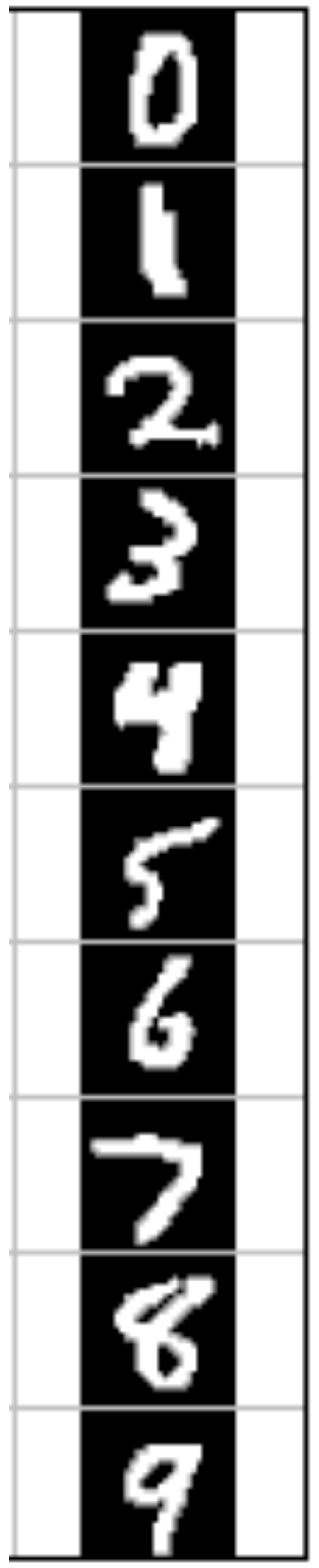
Held-out Validation



Held-out Validation



	Accuracy on training data	Accuracy on validation data
Simple	0.91	0.83
Medium	0.95	0.88
Complex	0.99	0.79
Super Complex	1.0	0.54



training data

validation data

Summary

- Three machine learning stories
  - One cautionary tale
- Deep learning definition
- Types of machine learning
- Best practices