CSE 417A: Introduction to Machine Learning

Introductory Lecture
Prof. Sanmay Das

Plan for today

Welcome & introductions

What is this class about?

Class logistics

What is machine learning?

“Enabling computers to learn from data”

Supervised Learning: Generalizing from seen data to unseen data

Unsupervised Learning: Finding patterns in input data

Reinforcement Learning: Learning how to act, based on rewards for actions

Supervised Learning

Can you define a tree?

A brown trunk coming up from the ground, with branches extending out?

Are these trees?

Hard to define
“Know it when I see it”
I’ve **learned** it from data!

The general supervised learning problem

Unknown target function: $f : \mathcal{X} \to \mathcal{Y}$

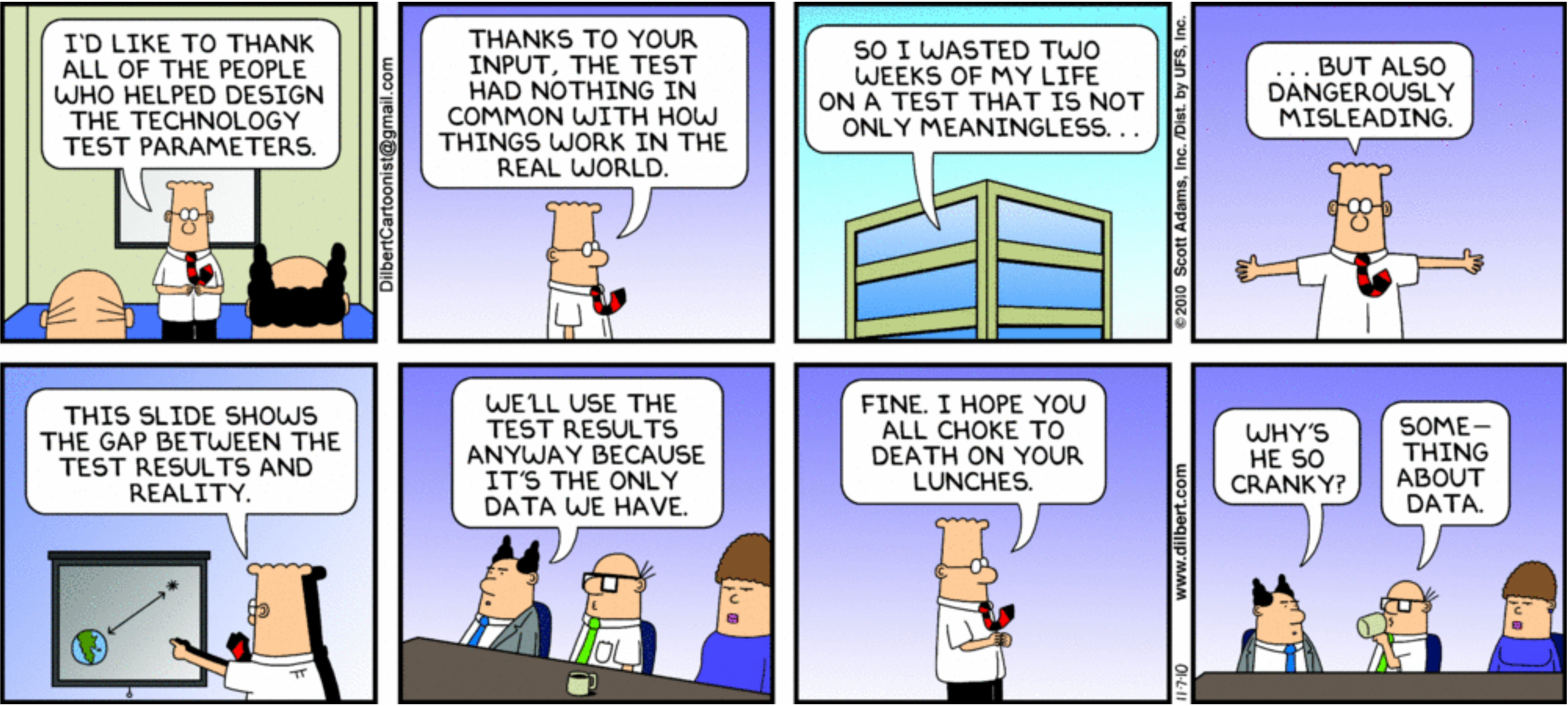
Training data: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)$

where $y_i = f(\mathbf{x}_i)$

Want to learn g "close to" f

Two central questions:

- How do we learn g?
- What can we say about how close g is to f?


I'D LIKE TO THANK ALL OF THE PEOPLE WHO HELPED DESIGN THE TECHNOLOGY TEST PARAMETERS.
DilbertCartoonist@gmail.com
THANKS TO YOUR INPUT, THE TEST HAD NOTHING IN COMMON WITH HOW THINGS WORK IN THE REAL WORLD.
SO I WASTED TWO WEEKS OF MY LIFE ON A TEST THAT IS NOT ONLY MEANINGLESS. . .
© 2010 Scott Adams, Inc. /Dist. by UFS, Inc.
. . . BUT ALSO DANGEROUSLY MISLEADING.
THIS SLIDE SHOWS THE GAP BETWEEN THE TEST RESULTS AND REALITY.
WE'LL USE THE TEST RESULTS ANYWAY BECAUSE IT'S THE ONLY DATA WE HAVE.
FINE. I HOPE YOU ALL CHOKE TO DEATH ON YOUR LUNCHES.
www.dilbert.com
11-7-10
WHY'S HE SO CRANKY?
SOME-THING ABOUT DATA.

© Scott Adams: http://dilbert.com/strips/comic/2010-11-07/

A concrete example: Credit approval

You apply for a credit card

Bank decides whether to approve or deny

What is the form of each *(x, y)* example?

What are we trying to learn, and from what?

- We have past data on customers, and want to learn the "ideal credit approval function"
- *x*'s can consist of demographic, income, personal data
- *y*'s are some relevant outcome

How can I learn *f*?

- Pick a *hypothesis set* $\mathcal{H} = \{h_1, h_2, \ldots,\}$
- Use a *learning algorithm* to select a hypothesis g from $\mathcal{H}$ based on the training data

The choice of $\mathcal{H}$ and the learning algorithm are deeply tied to each other

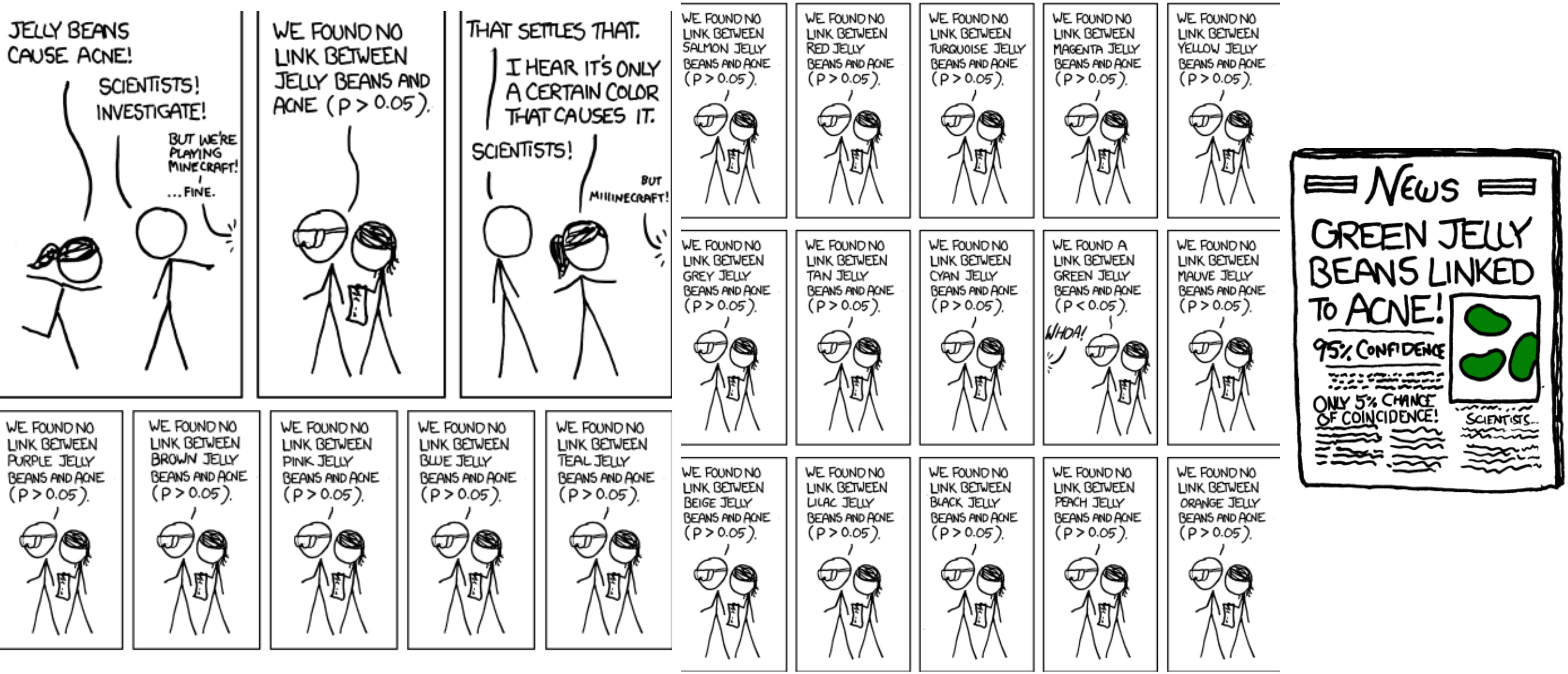
A linear hypothesis space

Suppose we have data on annual income (x_1), debt (x_2), average income in ZIP code (x_3). A possible hypothesis space can be expressed as (w_1, w_2, w_3, t) where the credit approval function is:

$$\text{Approve if } \sum_{i=1}^{3} w_i x_i \geq t$$

$$\text{Deny if } \sum_{i=1}^{3} w_i x_i < t$$

Note that the hypothesis space is infinite! How can we learn, and what can we say about what we've learned? That's what this class is all about!


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

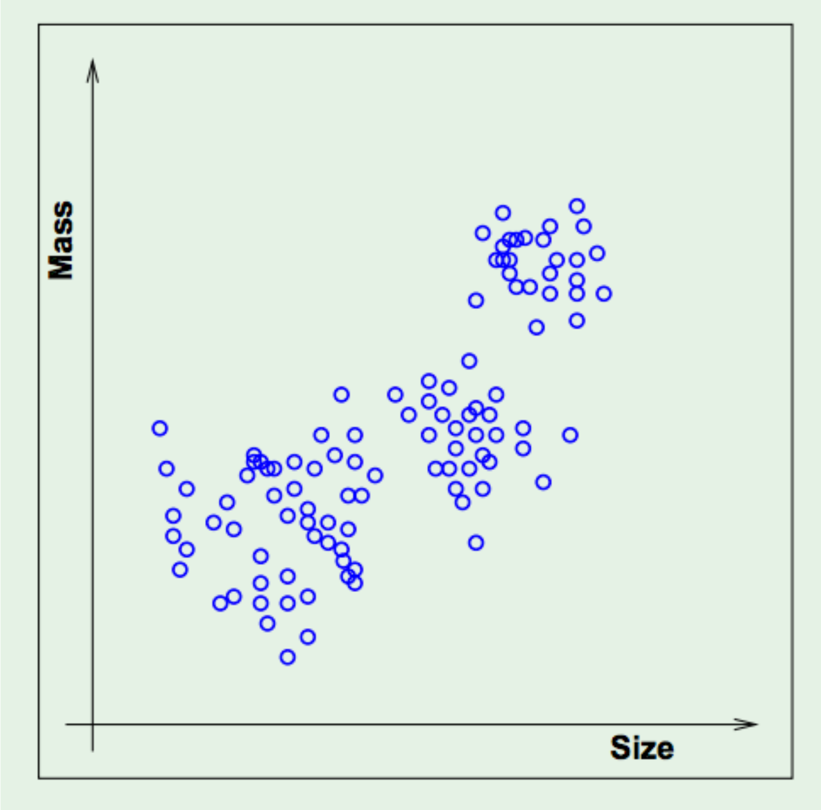
From xkcd, by Randall Munroe: http://xkcd.com/882/

Unsupervised Learning

Suppose you only have the feature vectors (*x*) and no labels.

Still want to describe the data in some useful way

Example from the book:



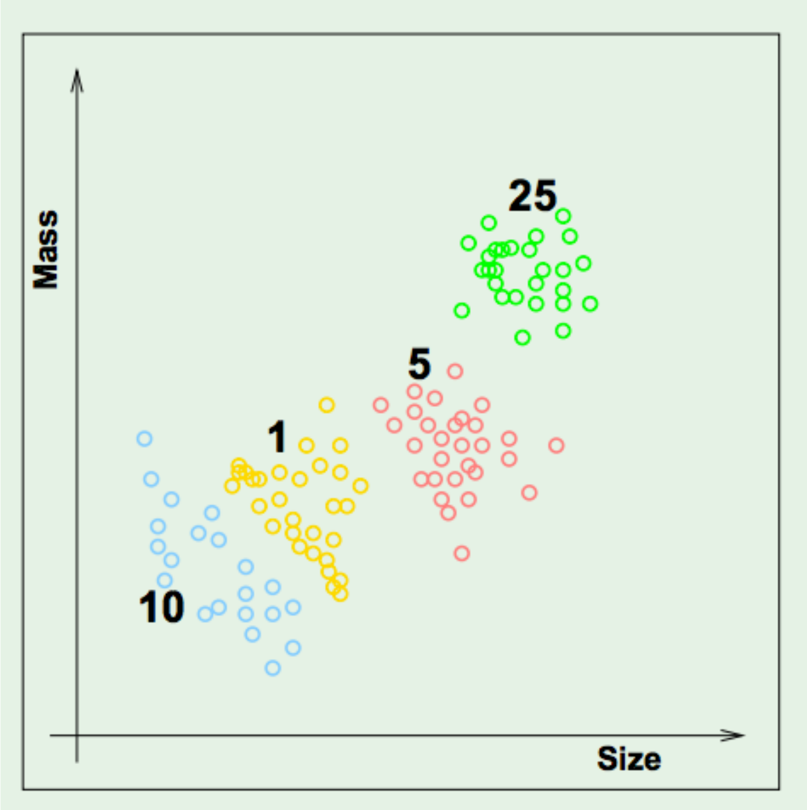
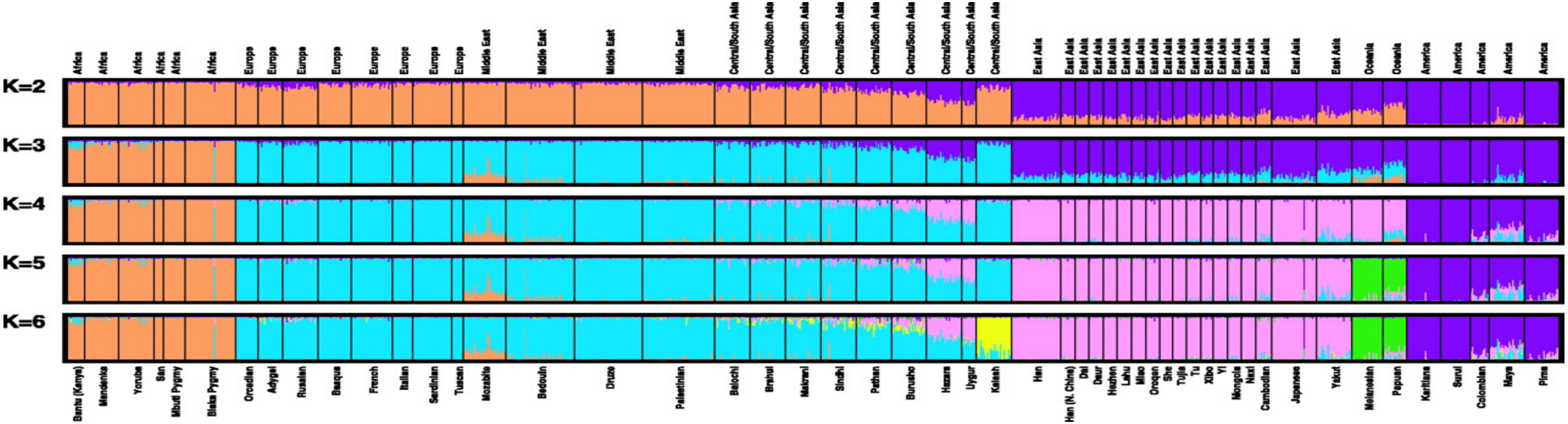


Figure 1 Estimated population structure.



N A Rosenberg et al. Science 2002;298:2381-2385

Science
AAAS

Published by AAAS

Reinforcement Learning

Agent interacts with the world by taking *actions*

Feedback is in the form of *rewards* (or *costs*)

Agent must learn a *policy*, which maps from the state of the world to an action

Major issues:

- Delayed reward / credit assignment
- Exploration / exploitation

Standard framework: *Markov Decision Processes*

Examples

Robot soccer

Learning to walk

Syllabus and course logistics

Role of lectures

Programming and math

Let’s look over the syllabus and discuss