Introduction to Machine Learning

CMSC 422

SOHEIL FEIZI
sfeizi@cs.umd.edu

What is Learning?

- Is memorization a form of “learning”?
- One definition: Learning is the process of obtaining *expertise* from *experience*
- Our goal: *learning* “Machine Learning”

What is this course about?

- Machine learning studies algorithms for learning to perform certain tasks
- By finding and exploiting patterns in data

Machine Learning

- Paradigm: “Programming by example”
  - Replace ``human writing code'' with ``human supplying data''
- Most central issue: generalization
  - How to abstract from ``training'' examples to ``test'' examples?

What can we do
with machine learning?

This text has been automatically translated from Arabic:

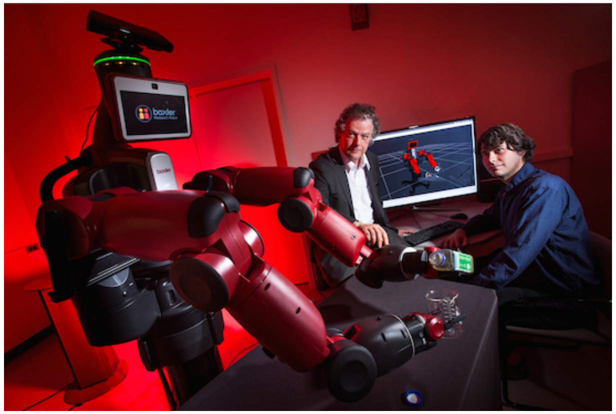
Moscow stressed tone against Iran on its nuclear program. He called Russian Foreign Minister Tehran to take concrete steps to restore confidence with the international community, to cooperate fully with the IAEA. Conversely Tehran expressed its willingness

Translate text

شددت موسكو لهجتها ضد إيران بشأن برنامجها النووي.
ودعا وزير الخارجية الروسي طهران إلى اتخاذ خطوات
ملموسة لاستعادة الثقة مع المجتمع الدولي والتعاون
الكامل مع الوكالة الذرية. بالمقابل أبدت طهران
استعدادها لاستئناف السماح بعمليات التفتيش
المفاجئة بشرط إسقاط مجلس الأمن ملفها النووي.

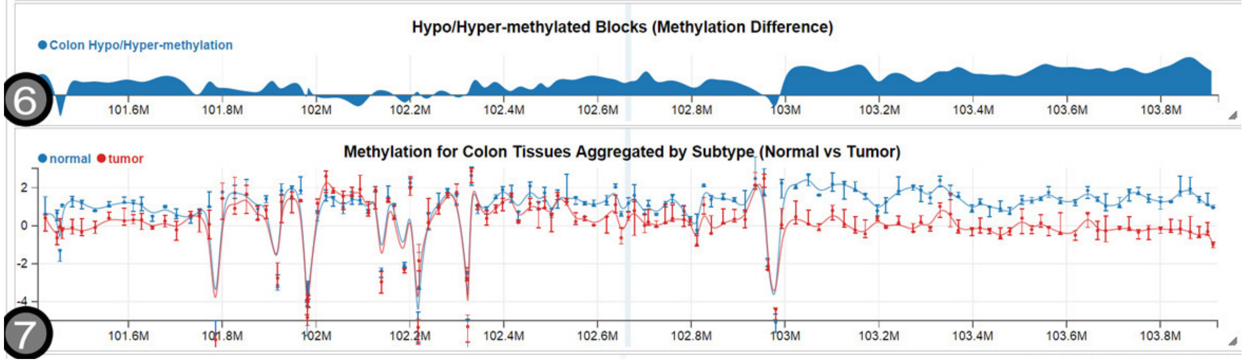
from Arabic to English BETA

Analyze text & speech



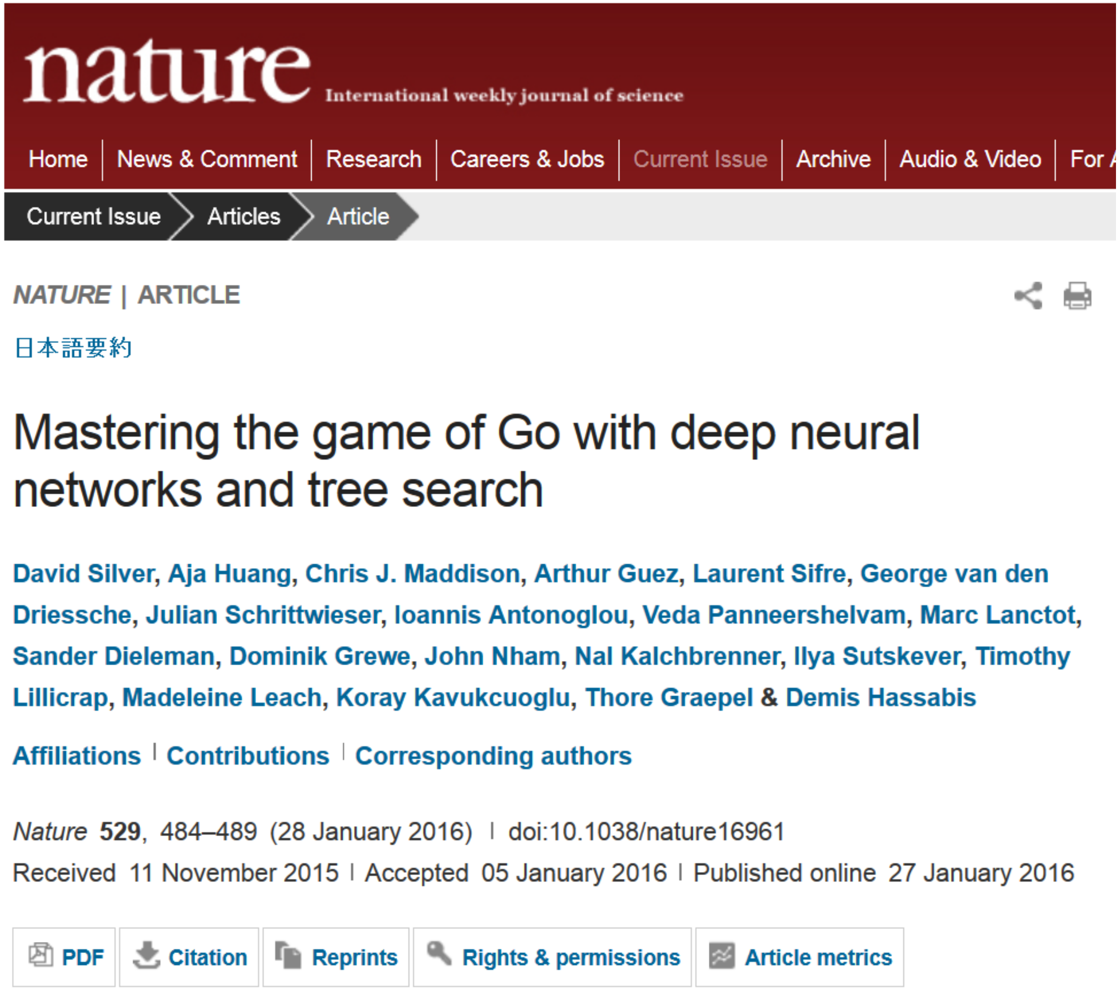
Teach robots how to cook from youtube videos

Recognize objects in images



Analyze genomics data

Sometimes machines even perform better than humans!



NATURE | ARTICLE

日本語要約

Mastering the game of Go with deep neural networks and tree search

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis

Affiliations | **Contributions** | **Corresponding authors**

Nature **529**, 484–489 (28 January 2016) | doi:10.1038/nature16961

Received 11 November 2015 | Accepted 05 January 2016 | Published online 27 January 2016

Question Answering system beats Jeopardy champion Ken Jennings at Quiz bowl!

Why Machine Learning?

- A growing and fast moving field with broad applicability: Finance, robotics, vision, machine translation, medicine, etc.

- Close connection between theory and practice

- Open field, lots of room for new work!

Course Goals

- By the end of the semester, you should be able to
  - Look at a problem
  - Identify if ML is an appropriate solution
  - If so, identify what types of algorithms might be applicable
  - Apply those algorithms

- This course is not
  - A survey of ML algorithms
  - A tutorial on ML toolkits such as Weka, TensorFlow, …

Prerequisites

- MATH461: Linear Algebra for Scientists and Engineers
- MATH240: Introduction to Linear Algebra
- CMSC351: Introduction to Algorithms
- CMSC330: Organization of Programming Languages
- CMSC320: Introduction to Data Science

What you can expect from the instructors

4 Teaching Assistants

- Samuel Dooley (PhD)
- Sahil Singla (PhD)
- Chris DeCarolis (UG)
- Rex Ledesma (UG)

We are here to help you learn by

- Introducing concepts from multiple perspectives
    - Theory and practice
    - Readings and class time
- Providing opportunities to practice, and feedback to help you stay on track
    - Homeworks
    - Programming assignments
    - Office hours

What I expect from you

- Work hard (this is a 3-credit class!)
  - Do a lot of math (calculus, linear algebra, probability)
  - Do a fair amount of programming

- Come to class prepared
  - Do the required readings!

Highlights from course logistics

Grading

- *Homework and Programming assignments (30%)*
- *Final project (15%)*
- *Midterm exam (25%)*, in class on **March 7**
- *Final exam (30%)*, cumulative, in class.

- HW01 is due next Thursday 11:59 pm
- No late homeworks
- Read syllabus here

Where to…

- find the schedule and slides: [Course webpage]
- find the readings: A Course in Machine Learning
- view and submit assignments: Canvas
- check your grades: Canvas
- ask and answer questions, participate in discussions and surveys, contact the instructors, and everything else:
  - Piazza Please use piazza instead of email
  - Office hours

Data

Data comes in different formats:

$$\left\{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\right\}$$

→ Training set

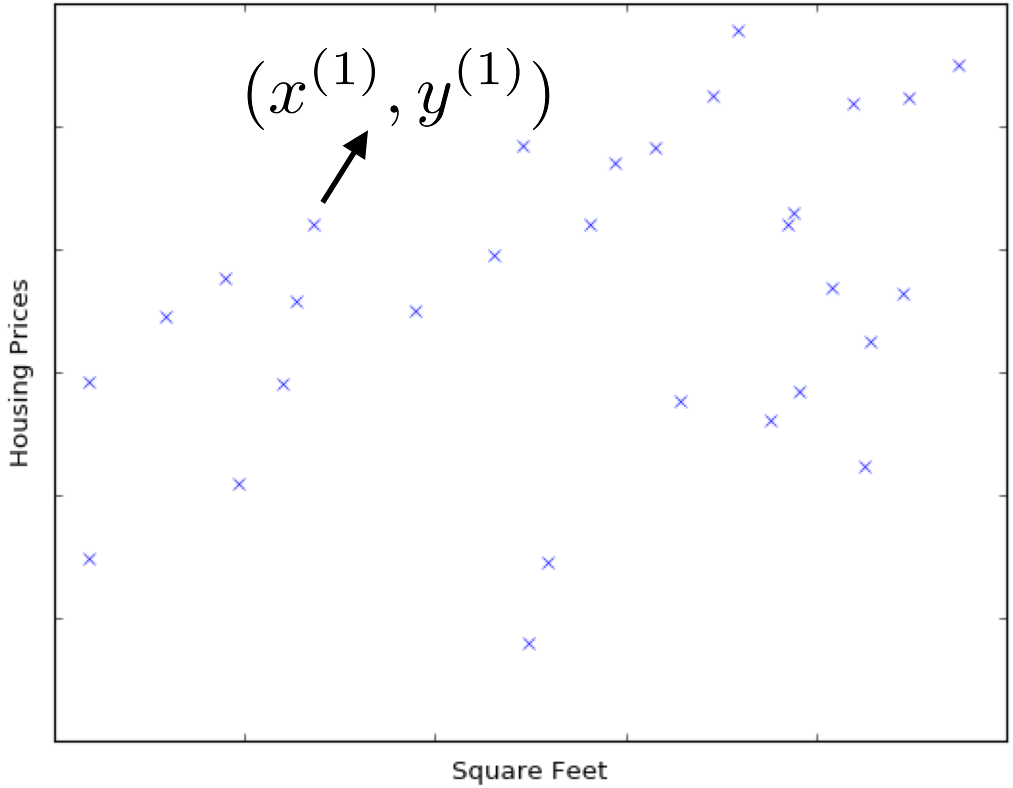
features

target/label variable

Goal: predict the label/target using features

→ Supervised Learning

Example

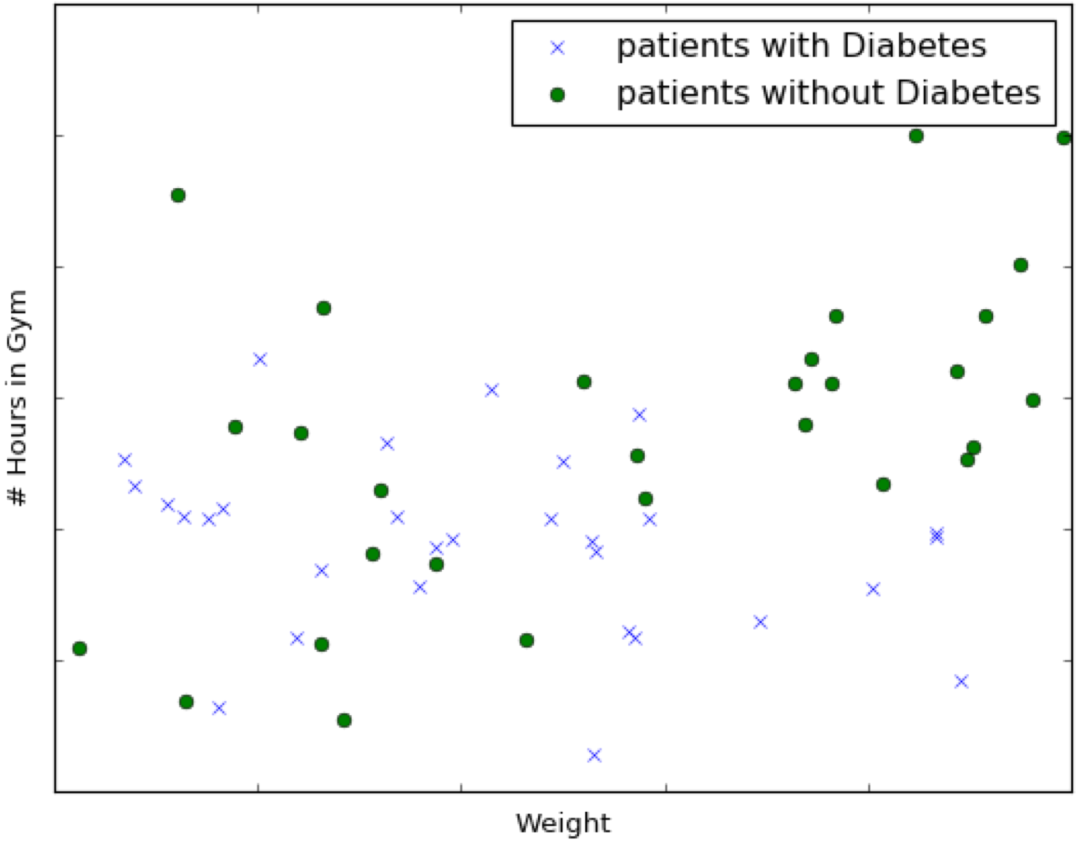


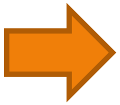
Regression
Problem

Q1. What is the dimension of the feature var? one $x^{(i)} \in \mathbb{R}$

Q2. What is the dimension of the target var? one $y^{(i)} \in \mathbb{R}$

Example





Classification

Problem

Q1. What is the dimension of the feature var? two $x^{(i)} \in \mathbb{R}^2$

Q2. What is the dimension of the target var? one $y^{(i)} \in \{0, 1\}$

Data

Data comes in different formats:

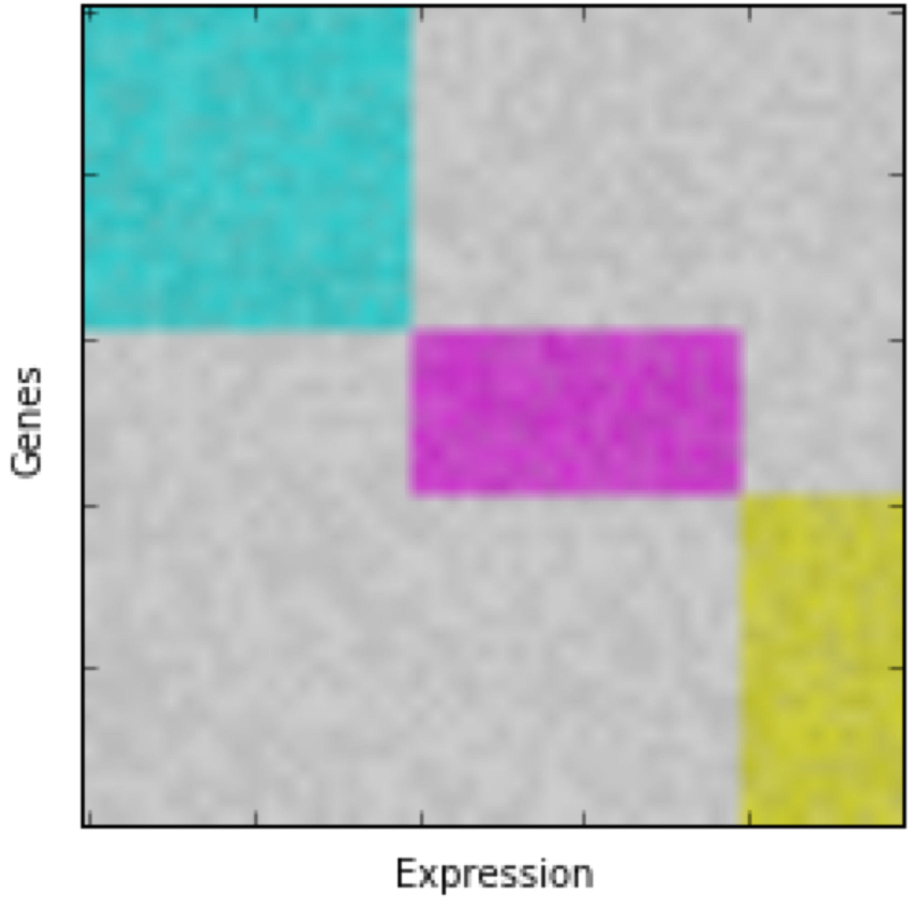
$$\left\{x^{(1)}, x^{(1)}, \ldots, x^{(N)}\right\}$$ ➡ Training set

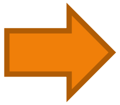
features

Goal: find "interesting" patterns in data

➡ Unsupervised Learning

Example





Clustering
Problem

Who are these people?

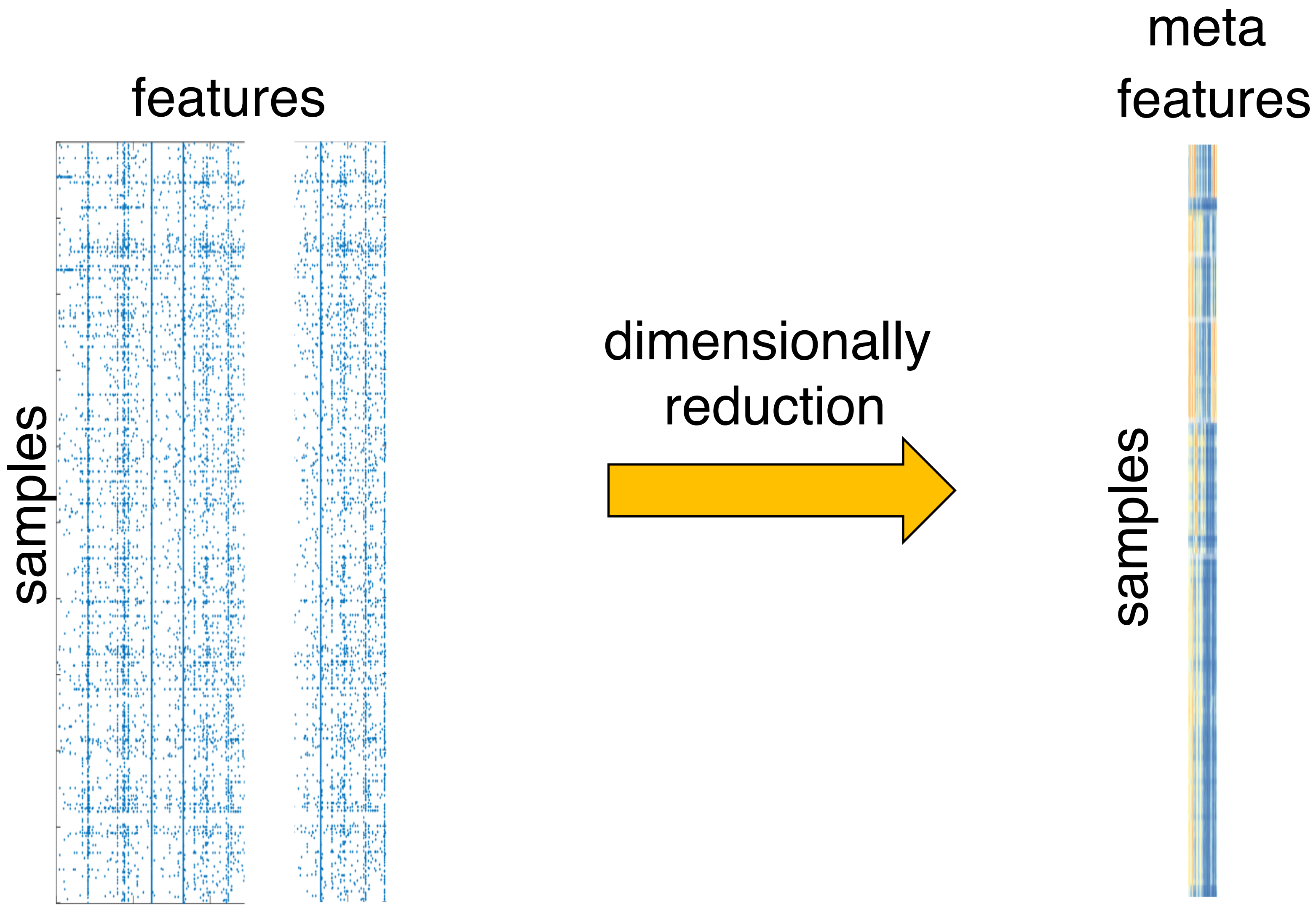
Generative Models (GANs)

Generating realistic but fake samples

CelebA dataset

Karras et al. 2017

Dimensionality Reduction



features

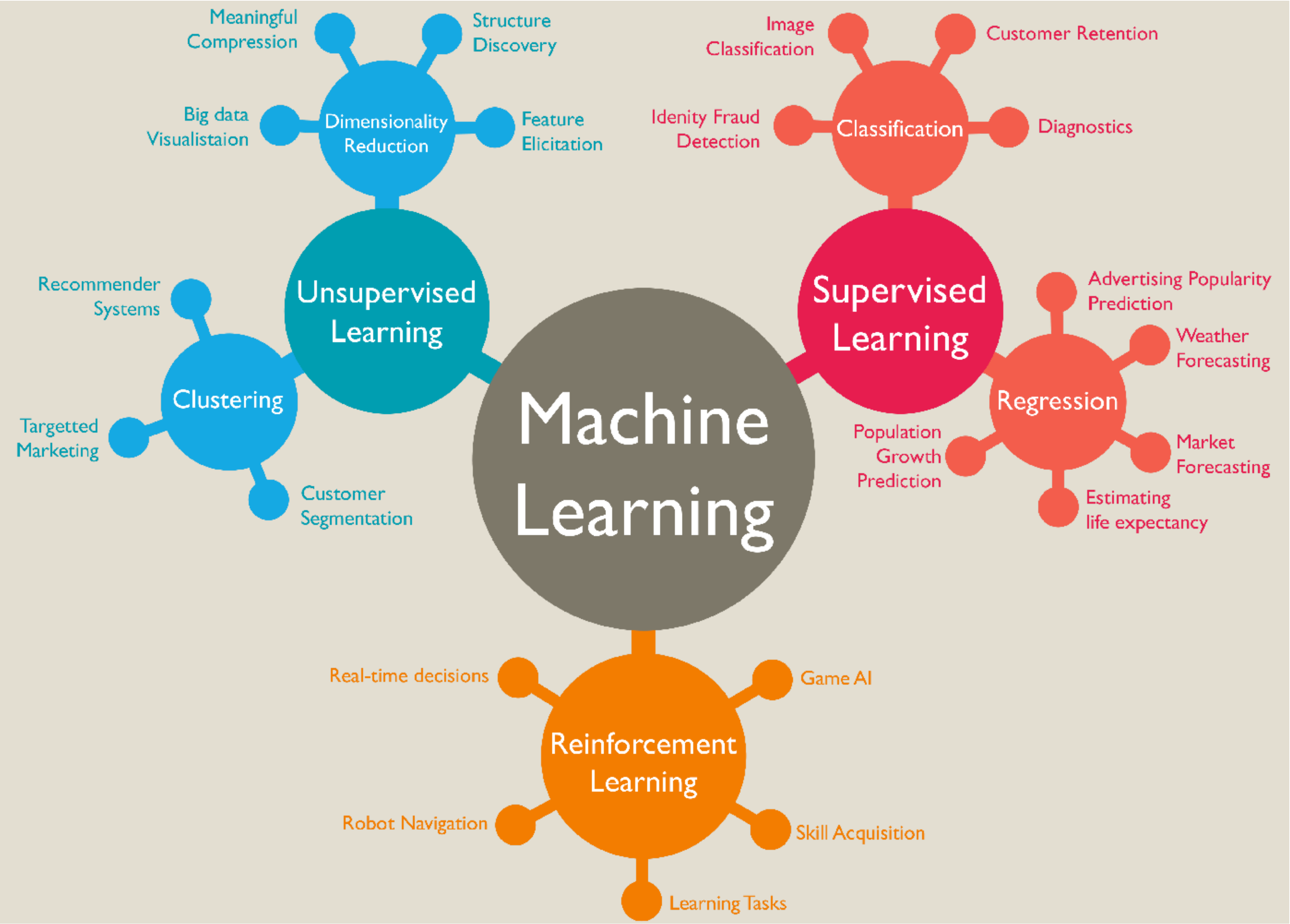
samples

dimensionally
reduction

meta
features

samples

http://www.7wdata.be/big-data/10-companies-using-machine-learning-in-cool-ways/



A Closer Look:
the classification problem

What does it mean to “learn by example”?

- Classification tasks

- Inductive bias

- Formalizing learning

Classification tasks

- How would you write a program to distinguish a picture of a cat from a picture of a dog?

- Provide examples pictures of cats and pictures of dogs and let a **classifier** learn to distinguish the two.

Classification tasks

- How would you write a program to distinguish a sentence is grammatical or not?

- Provide examples of grammatical and ungrammatical sentences and let a **classifier** learn to distinguish the two.

Classification tasks

- How would you write a program to distinguish cancerous cells from normal cells?

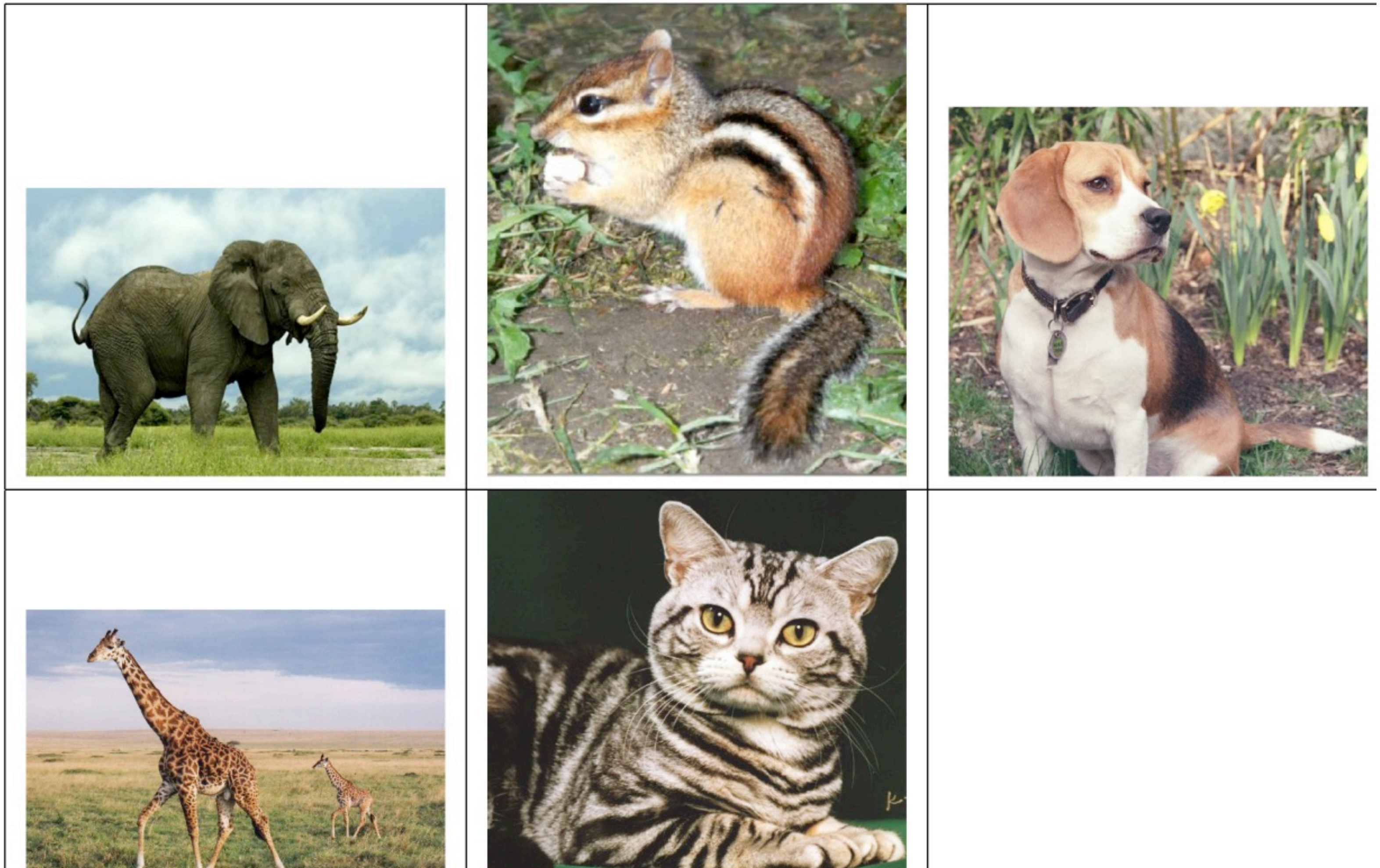
- Provide examples of cancerous and normal cells and let a **classifier** learn to distinguish the two.

Let’s try it out…

- Your task: learn a classifier to distinguish class A from class B from examples

- Examples of class A:

- Examples of class B



Let’s try it out…

✓ learn a classifier from examples

- Now: predict class on new examples using what you’ve learned

Key ingredients needed for learning

- Training vs. test examples
  - Memorizing the training examples is not enough!
  - Need to generalize to make good predictions on test examples
- Inductive bias
  - Many classifier hypotheses are plausible
  - Need assumptions about the nature of the relation between examples and classes

Machine Learning as Function Approximation

Problem setting

- Set of possible instances X
- Unknown target function $f: X \rightarrow Y$
- Set of function hypotheses $H = \{h \mid h: X \rightarrow Y\}$

Input

- Training examples $\{(x^{(1)}, y^{(1)}), \dots (x^{(N)}, y^{(N)})\}$ of unknown target function f

Output

- Hypothesis $h \in H$ that best approximates target function f

Formalizing induction: Loss Function

$l(y, f(x))$ where y is the truth and $f(x)$ is the system's prediction

e.g. $l(y, f(x)) = \begin{cases} 0 & if\ y = f(x) \\ 1 & otherwise \end{cases}$

Captures our notion of what is important to learn

Formalizing induction:
Data generating distribution

- Where does the data come from?
  - Data generating distribution
    - A probability distribution D over (x, y) pairs
  - We don’t know what D is!
    - We only get a sample from it: our training data

Formalizing induction: Expected loss

- f should make good predictions
  - as measured by loss l
  - on **future** examples that are also drawn from D

- Formally
  - ε , the expected loss of f over D should be small

$$\varepsilon \triangleq \mathbb{E}_{(x,y)\sim D}\{l(y,f(x))\} = \sum_{(x,y)} D(x,y)l(y,f(x))$$

Formalizing induction: Training error

- We can't compute expected loss because we don't know what D is

- We only have a sample of D
  - training examples $\{(x^{(1)}, y^{(1)}), \ldots (x^{(N)}, y^{(N)})\}$

- All we can compute is the training error

$$\hat{\varepsilon} \triangleq \sum_{n=1}^{N} \frac{1}{N} l(y^{(n)}, f(x^{(n)}))$$

Formalizing Induction

- Given
  - a loss function l
  - a sample from some unknown data distribution D

- Our task is to compute a function f that has low expected error over D with respect to l.

$$\mathbb{E}_{(x,y)\sim D}\{l(y, f(x))\} = \sum_{(x,y)} D(x, y) l(y, f(x))$$

Recap: introducing machine learning

What does “learning by example” mean?

- Classification tasks

- Learning requires examples + inductive bias
- Generalization vs. memorization

- Formalizing the learning problem
  - Function approximation
  - Learning as minimizing expected loss

Your tasks before next class

- Check out course webpage, Canvas, Piazza
- Start reading the reviews on probability and linear algebra (posted on course webpage)
- Get started on HW01