Learning Machine Learning for the Future

Dr. Fayyaz ul Amir Afsar Minhas

PIEAS Biomedical Informatics Research Lab

Department of Computer and Information Sciences

Pakistan Institute of Engineering & Applied Sciences

PO Nilore, Islamabad, Pakistan

http://faculty.pieas.edu.pk/fayyaz/

Introduction

- Objective
  - What are some possibilities with Machine Learning?
- Download the presentation
  - http://goo.gl/CoMcW9

Ways of obtaining knowledge

- *Observation*
- *Experience*
- *Reason or Logic*
- Testimony
- Revelation

Man is essentially ignorant, and becomes learned through acquiring knowledge.
(Ibn Khaldun)

Intelligent Computers

- Computers are _____
  - Powerful
  - Great data storage and manipulation devices
  - Dumb!

- The science of making computers intelligent is called
  - **Artificial Intelligence**
  - Replicating ways of acquiring knowledge in the computer
- **Examples?**

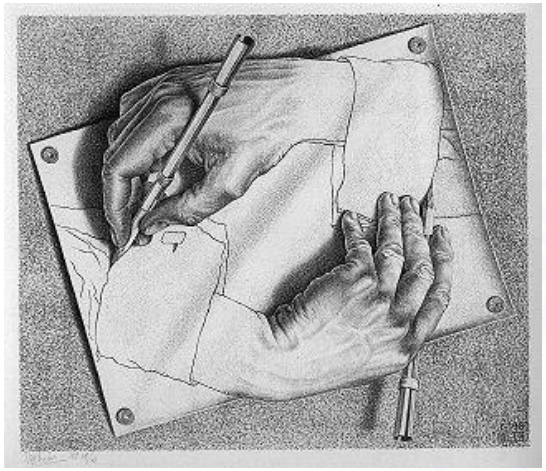
What is Machine Learning?

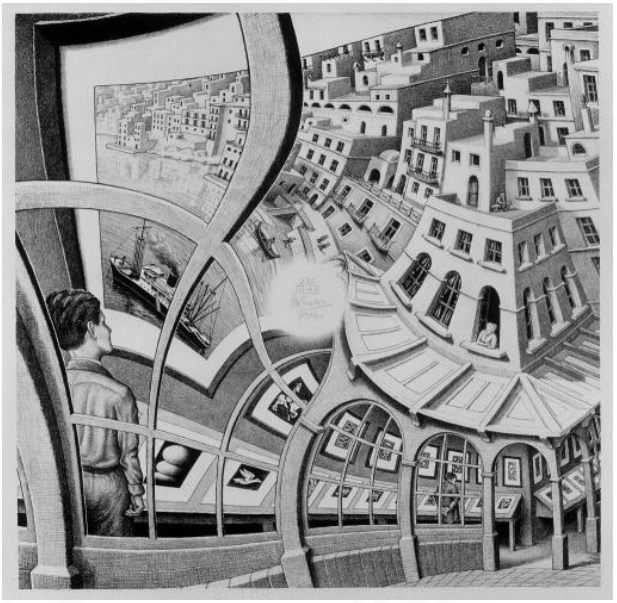
Apples

Oranges

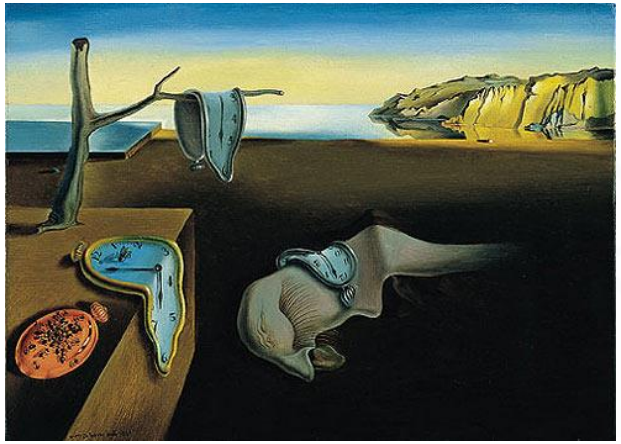
What is this?

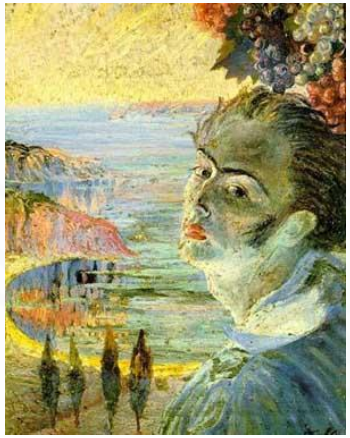
Paintings by two different painters





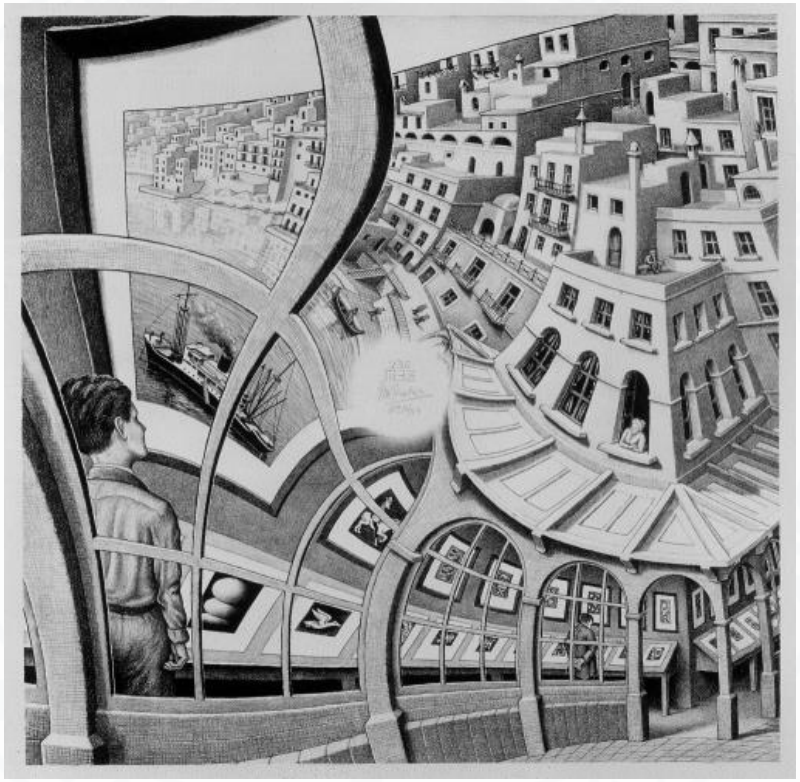
Escher



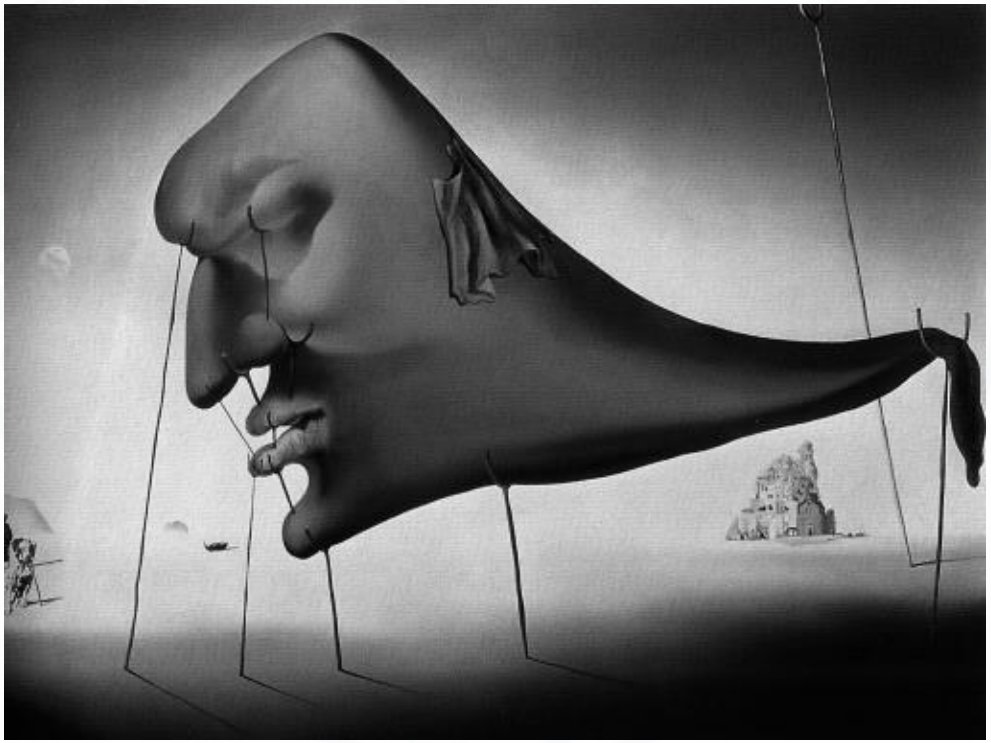


Dali

Whose painting is this?

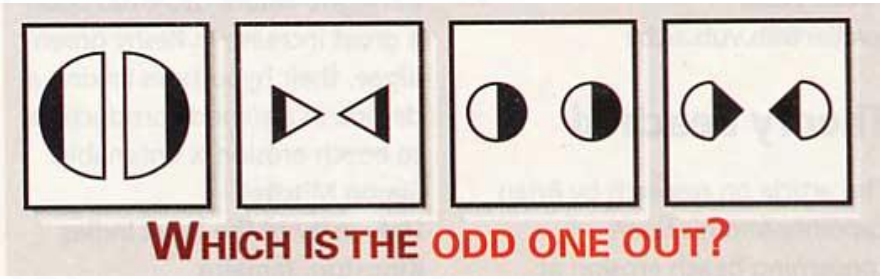


And this?



How many categories (clusters) are there?

Find the odd one out!



Predict the series

- 1,1,2,3,5,8,13,…

Question?

- Consider the vectors
  - $X_1=[1\ 2\ 1\ 4]^T$
  - $X_2=[2\ 4\ 2\ 4]^T$
  - $X_3=[0\ 0\ 0\ 4]^T$
  - $X_4=[3\ 6\ 3\ 4]^T$
  - $X_5=[4\ 8\ 4\ 4]^T$
- To store each vector, how many dimensions (or variables) do we need?

Learning to write

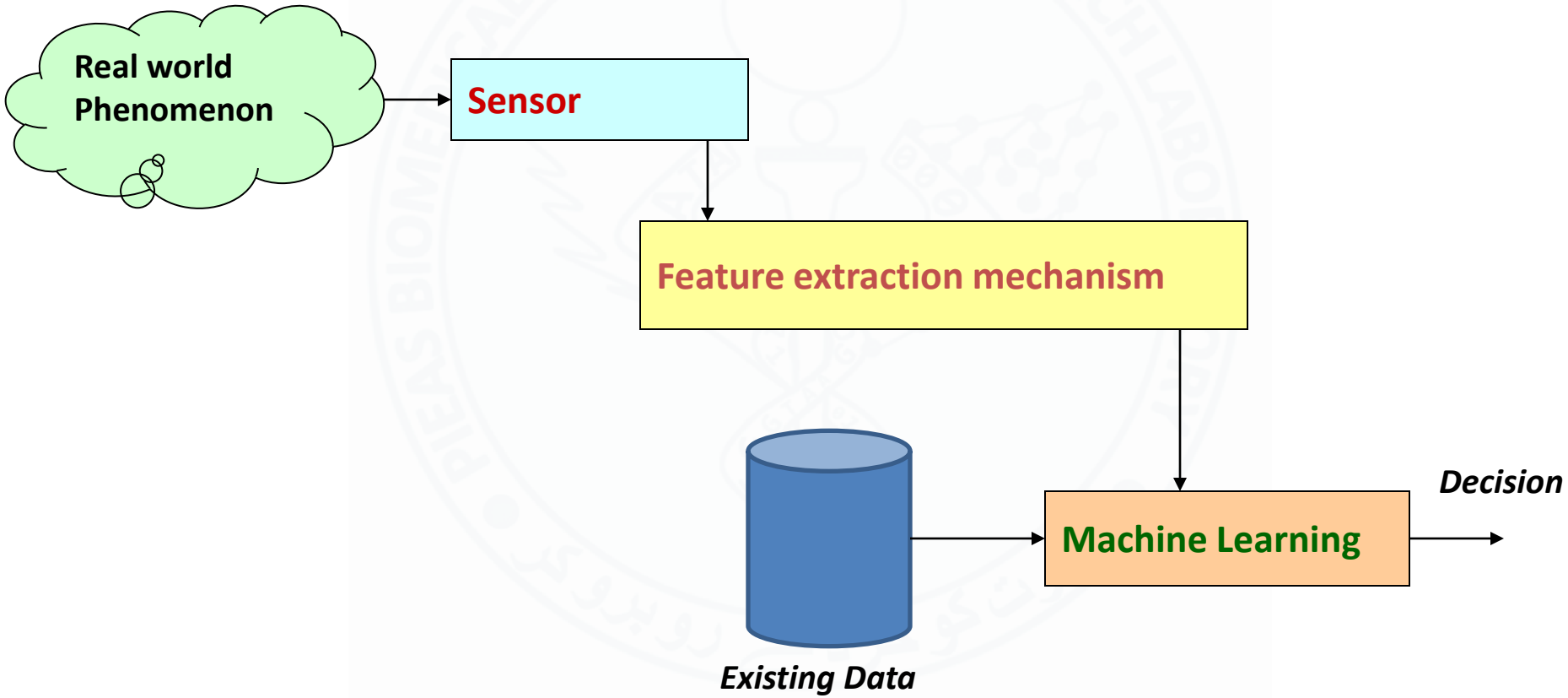
Questions

• How were you able to recognize that the object shown was indeed an apple?	Classification
• How were you able to discriminate between the paintings from two different painters?	Classification
• How were you able to find out the different types of apples in the picture?	Clustering
• How did you manage to find the next number in the series?	Regression
• How were you able to find which dimension was redundant?	Dimensionality Reduction
• How were you able to find the odd one out?	Anomaly Detection
• Learning to drive / write?	Reinforcement learning

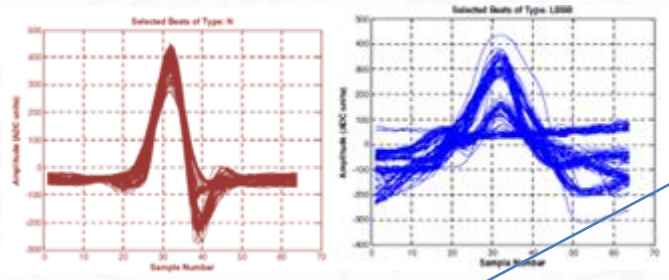
What is Machine Learning?

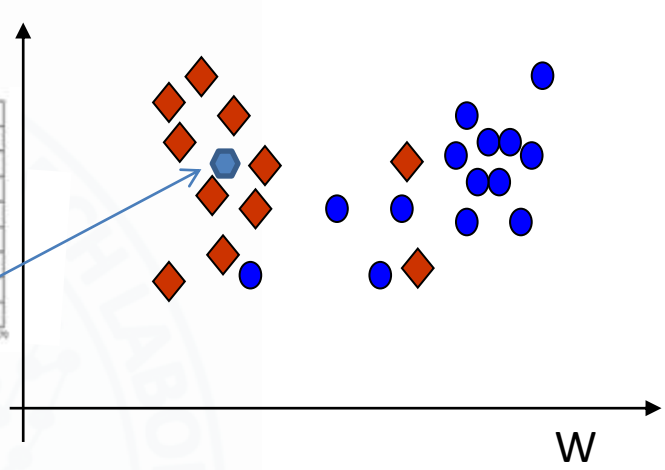
- Computers are _____.
  - **Dumb**
- Making a machine (computer) perform the same tasks which you have just done is called _____
  - **Artificial Intelligence**
- If you learn to do these tasks using existing data, then this is called _____
  - **Machine Learning**

How to use Data to Produce Knowledge?



Example

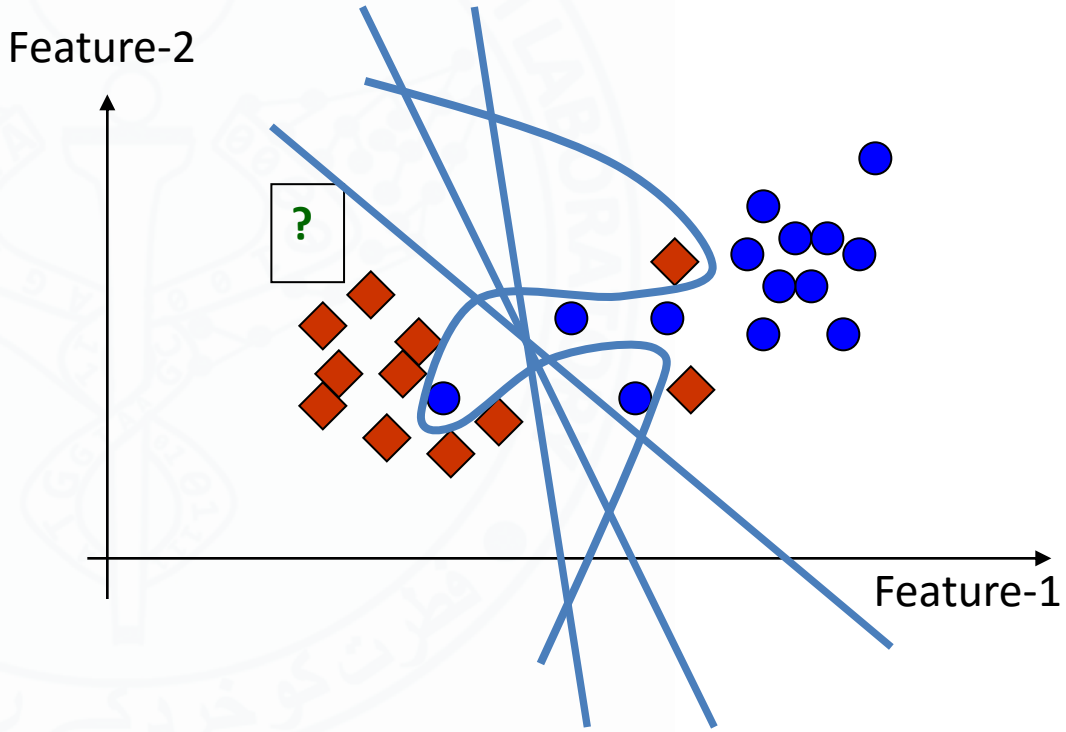




- Objective: Make good predictions not only on known data but previously unseen one
  - Generalization

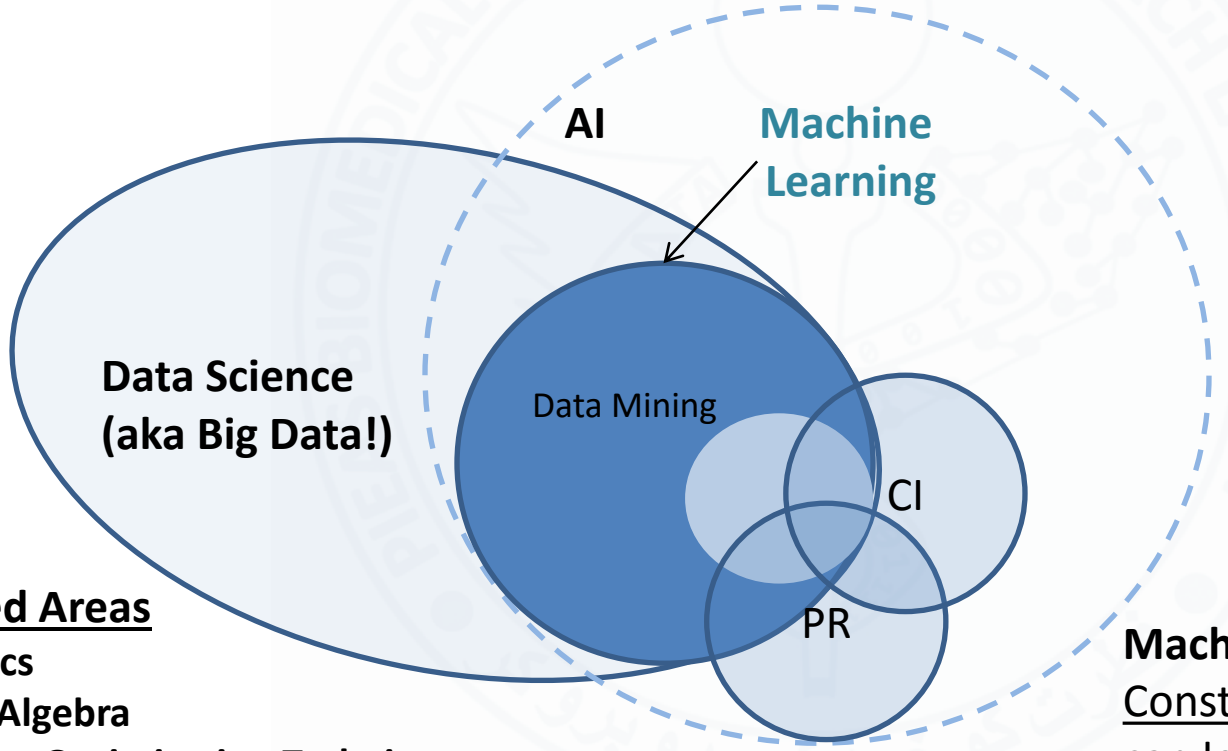
Classification Example

- Making a prediction rule
  - Nearest Neighbor
  - Linear Discriminant
  - Support Vector Machine
    - Margin
  - Non-Linear Boundaries



What is machine learning?

- Learning from observations, experience or Inductive Reasoning



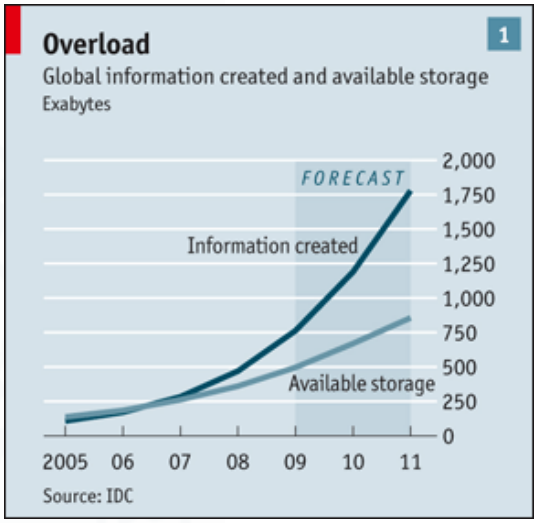
The Machine Learning Centric View of AI
(not to scale)

Related Areas
Statistics
Linear Algebra
Calculus, Optimization Techniques
High Performance Computing
Algorithms, Data structures and Programming
Information Retrieval, NLP, Computer Vision, Signal Analysis

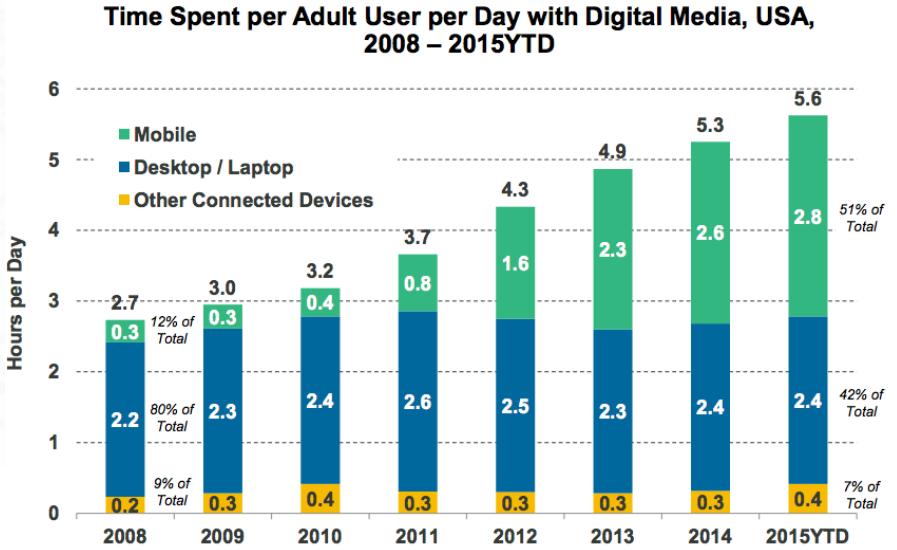
Machine Learning is the Construction of algorithms that can learn from data to "explain" the data and make predictions

When to Apply Machine Learning?

- Information Explosion in the Global Village
- Machine Learning is particularly suited for areas that can exploit "The Unreasonable Effectiveness of Data"
  - When making rules or theories about a phenomenon is hard or impossible?
  - Examples

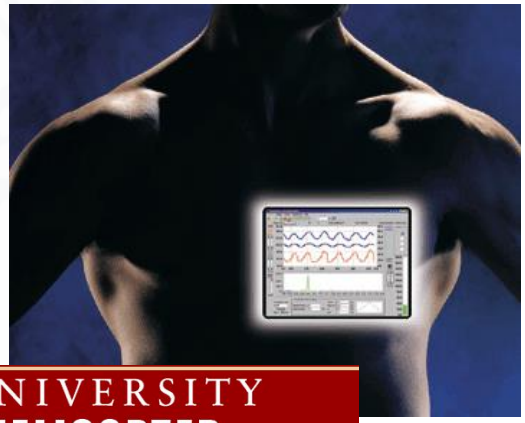


The Economist. 2010. "Data, Data Everywhere," February 25, 2010. http://www.economist.com/node/15557443.



Halevy, Alon, Peter Norvig, and Fernando Pereira. "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems*, 2009.

Applications



STANFORD UNIVERSITY
AUTONOMOUS HELICOPTER

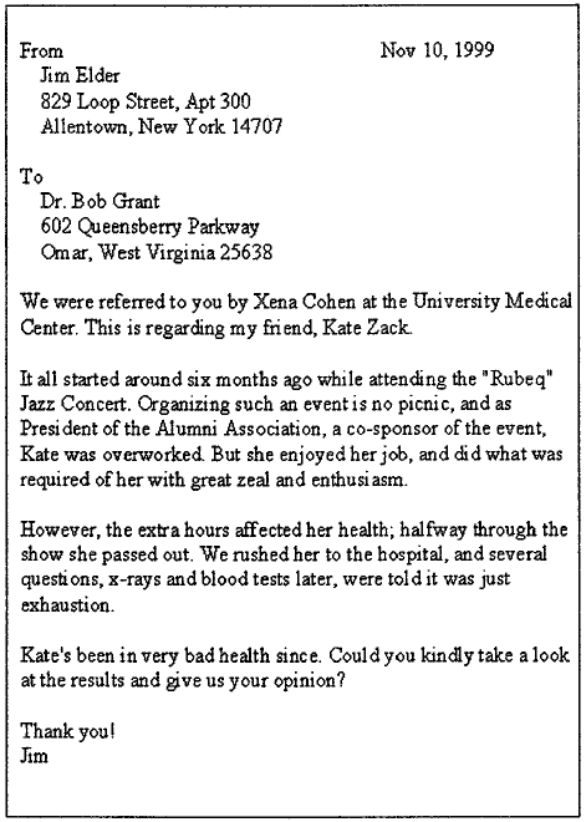
Overview

The goal of this project is to push the state-of-the-art in autonomous helicopter flight: extreme aerobatics under computer control.

Featured Videos

http://heli.stanford.edu/

Handwriting Recognition / OCR



From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

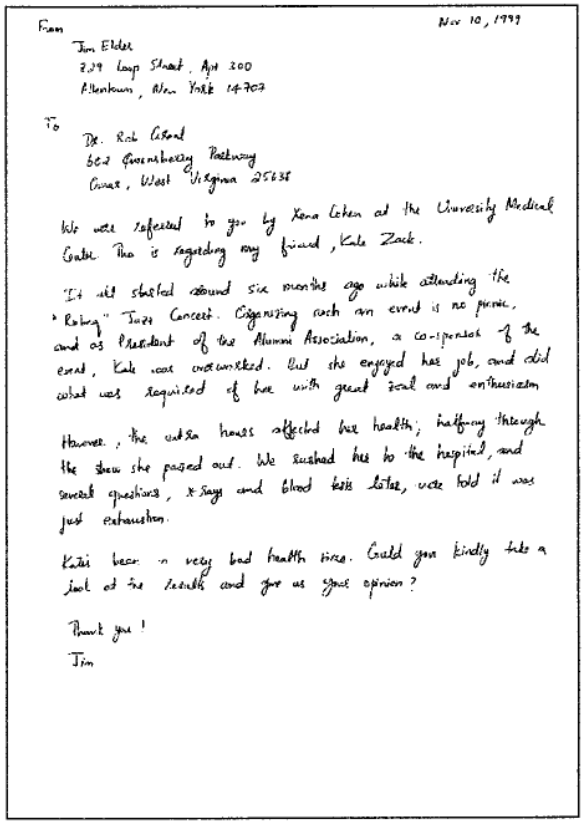
We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim



From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

Nov 10, 1999

To
Dr. Bob Grant
602 Queensberry Parkway
Omar, West Virginia 25638

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the "Rubeq" Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

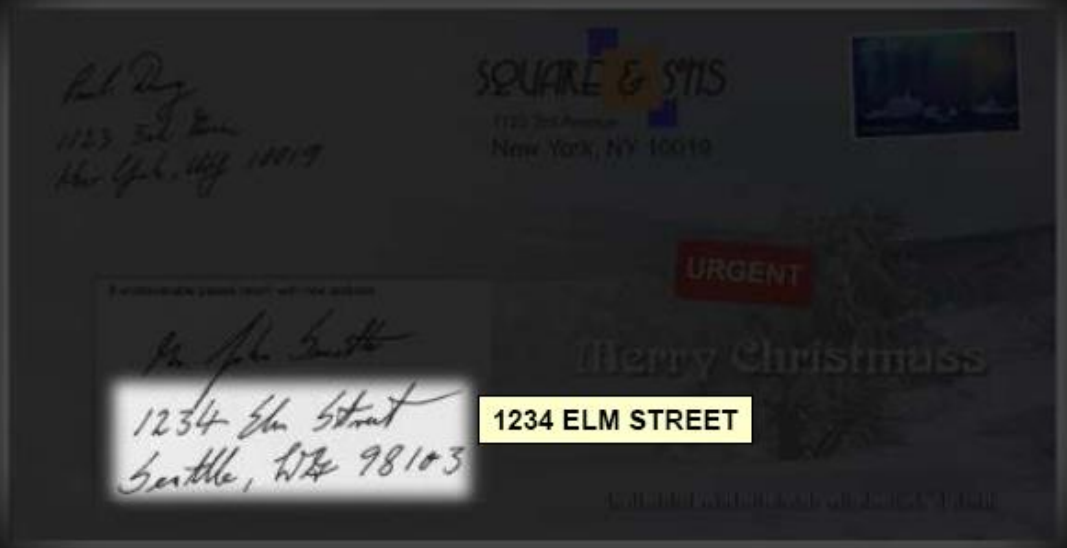
Kate's been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim

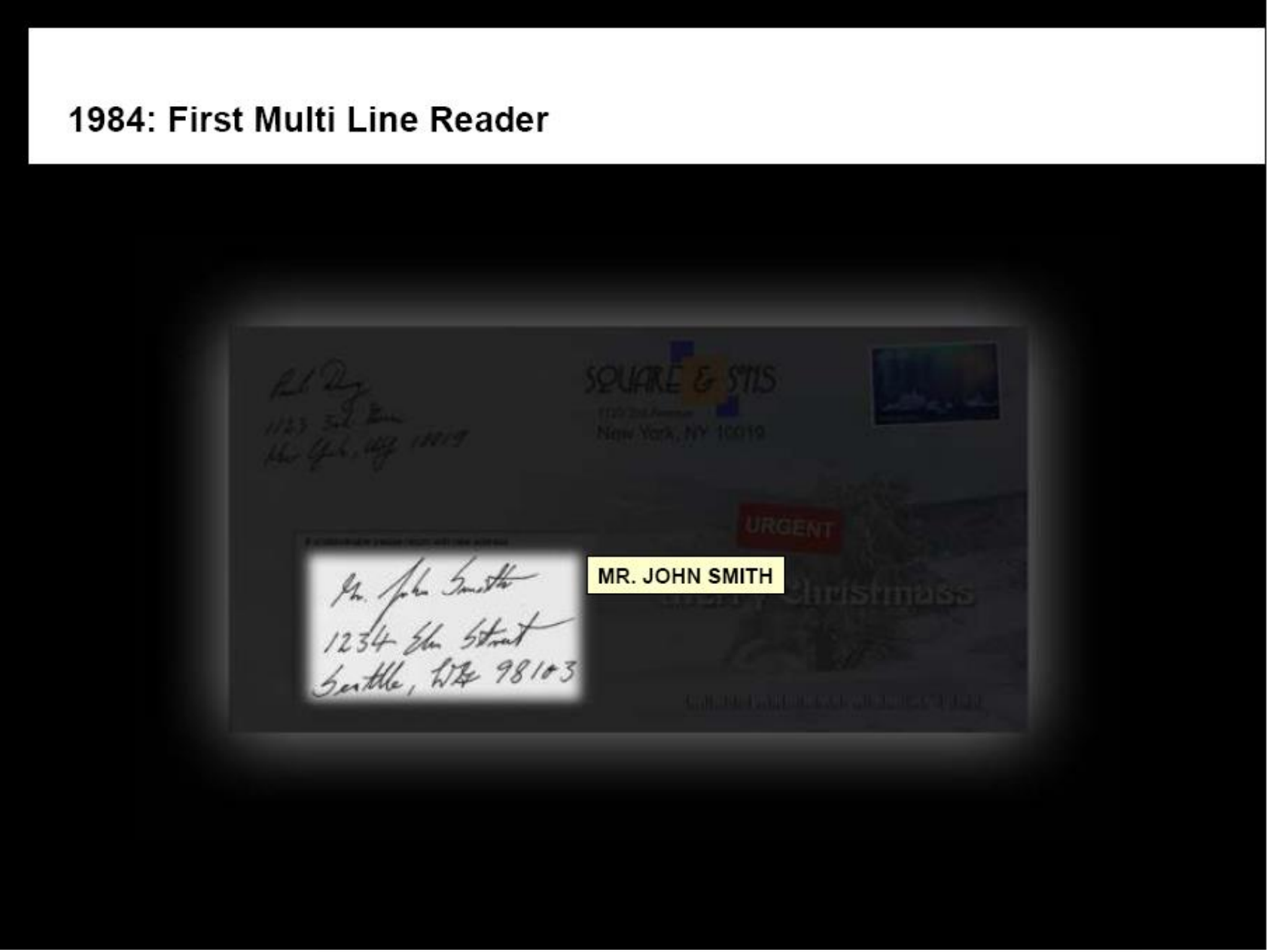
The Letter

1978: First Postal Code Reader Worldwide

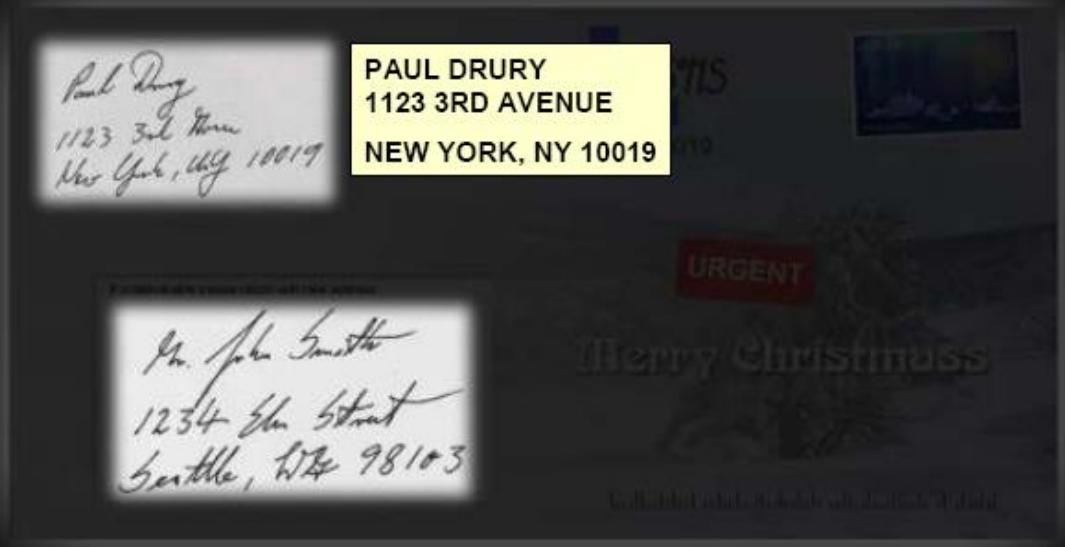
1982: First Address Reader Worldwide



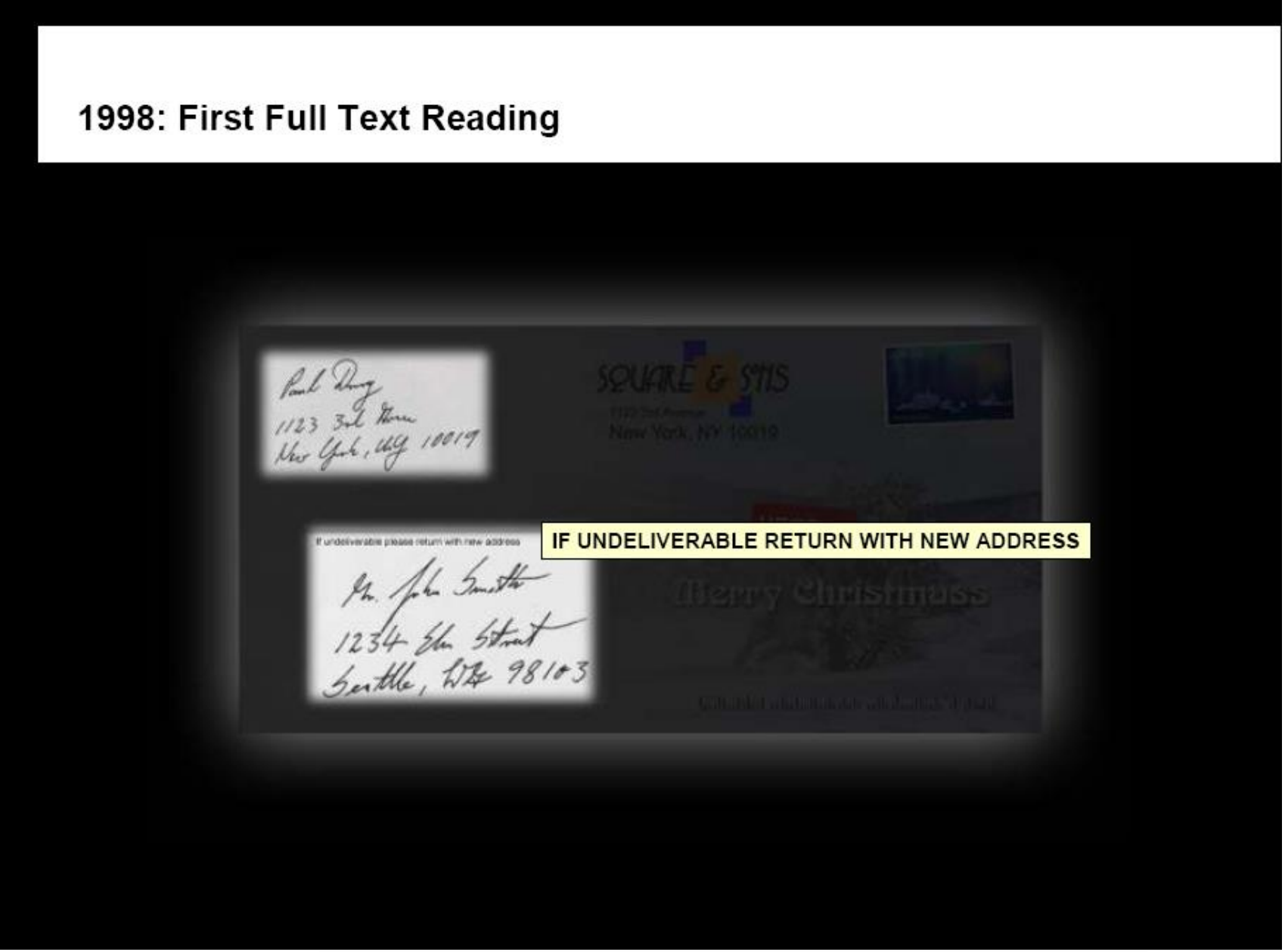
1984: First Multi Line Reader



1996: First Sender's Address Reader



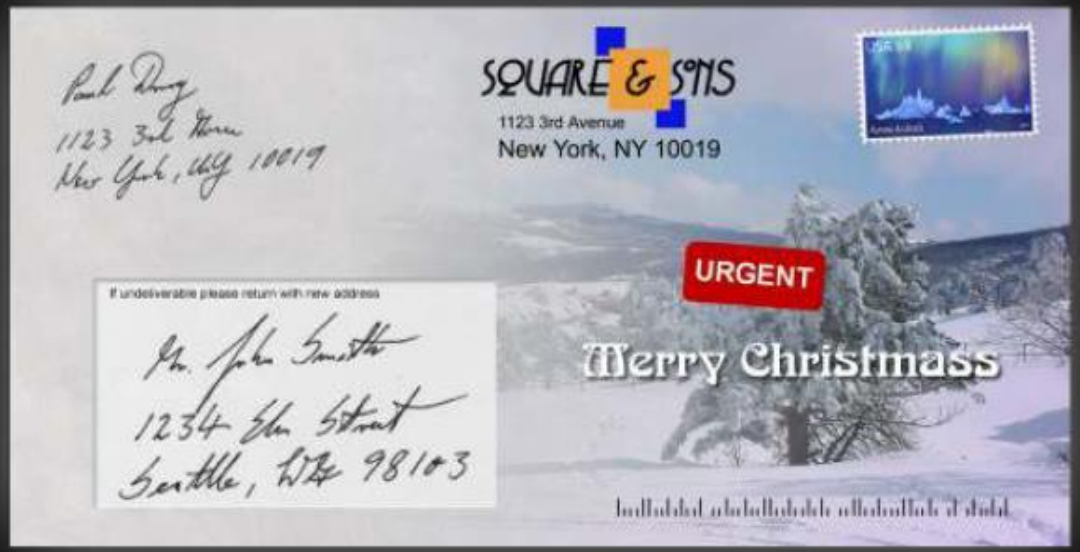
1998: First Full Text Reading



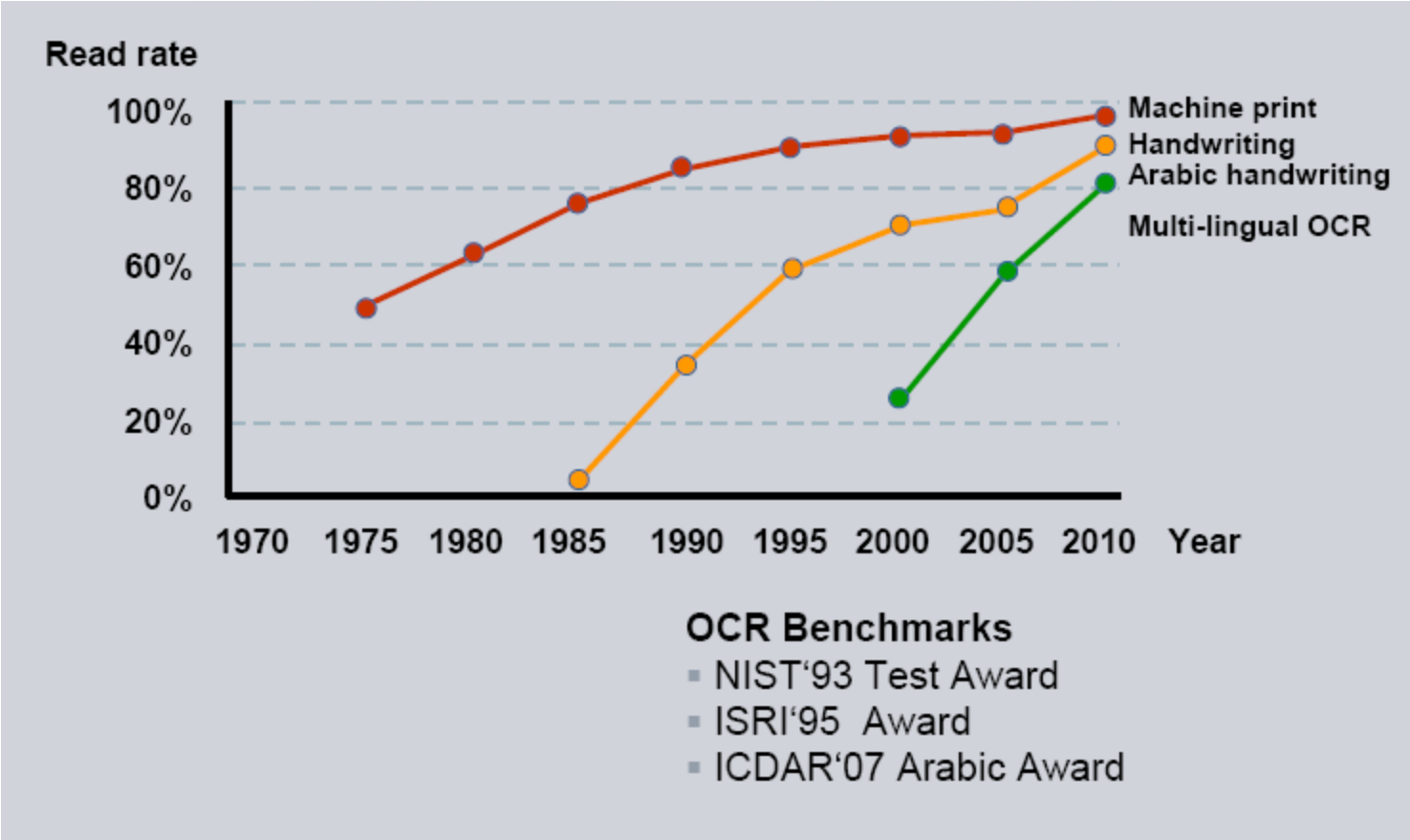
2000: First Graphics Recognition

2004: First Full Recognition

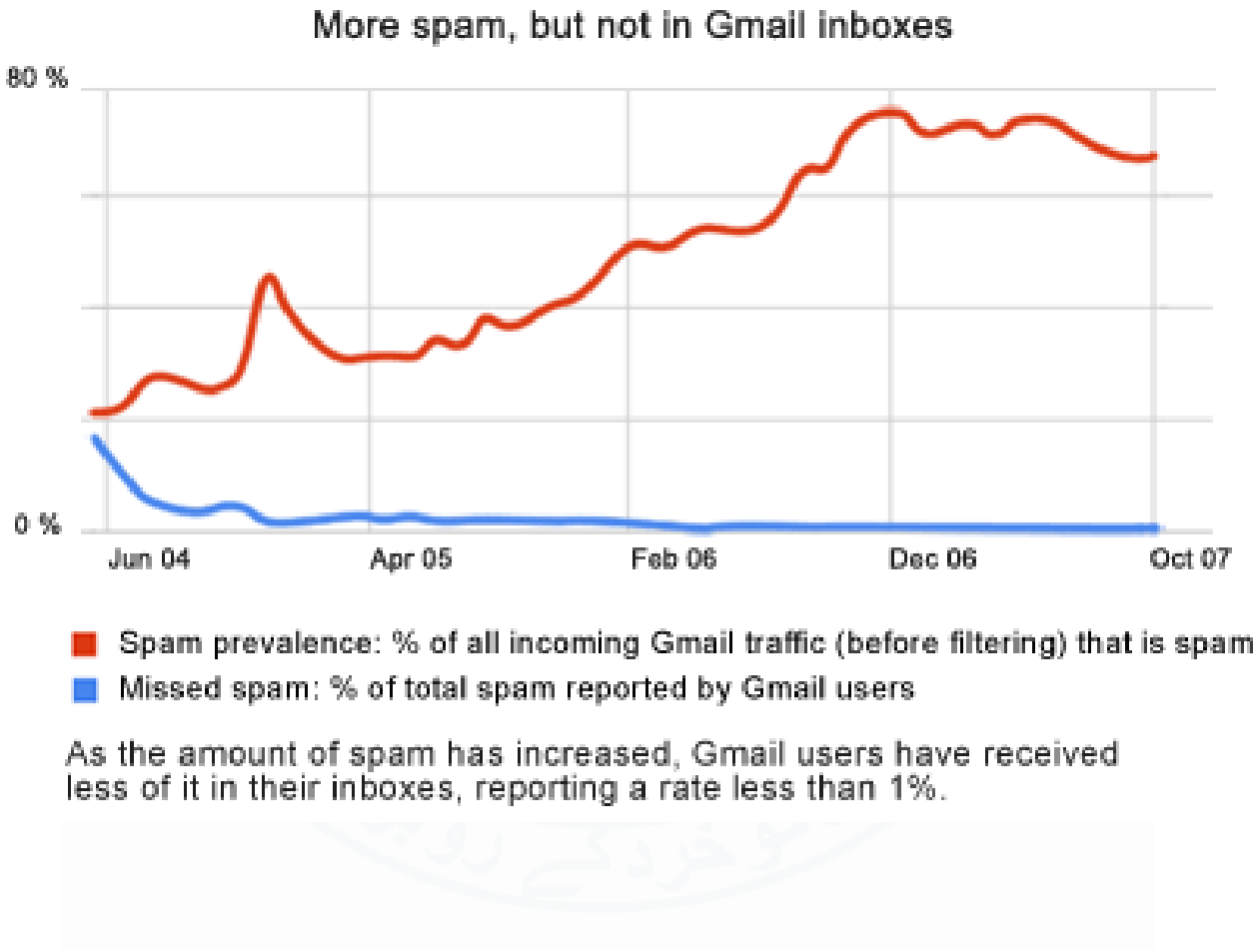
2008: Recognition on Both Sides of Envelope



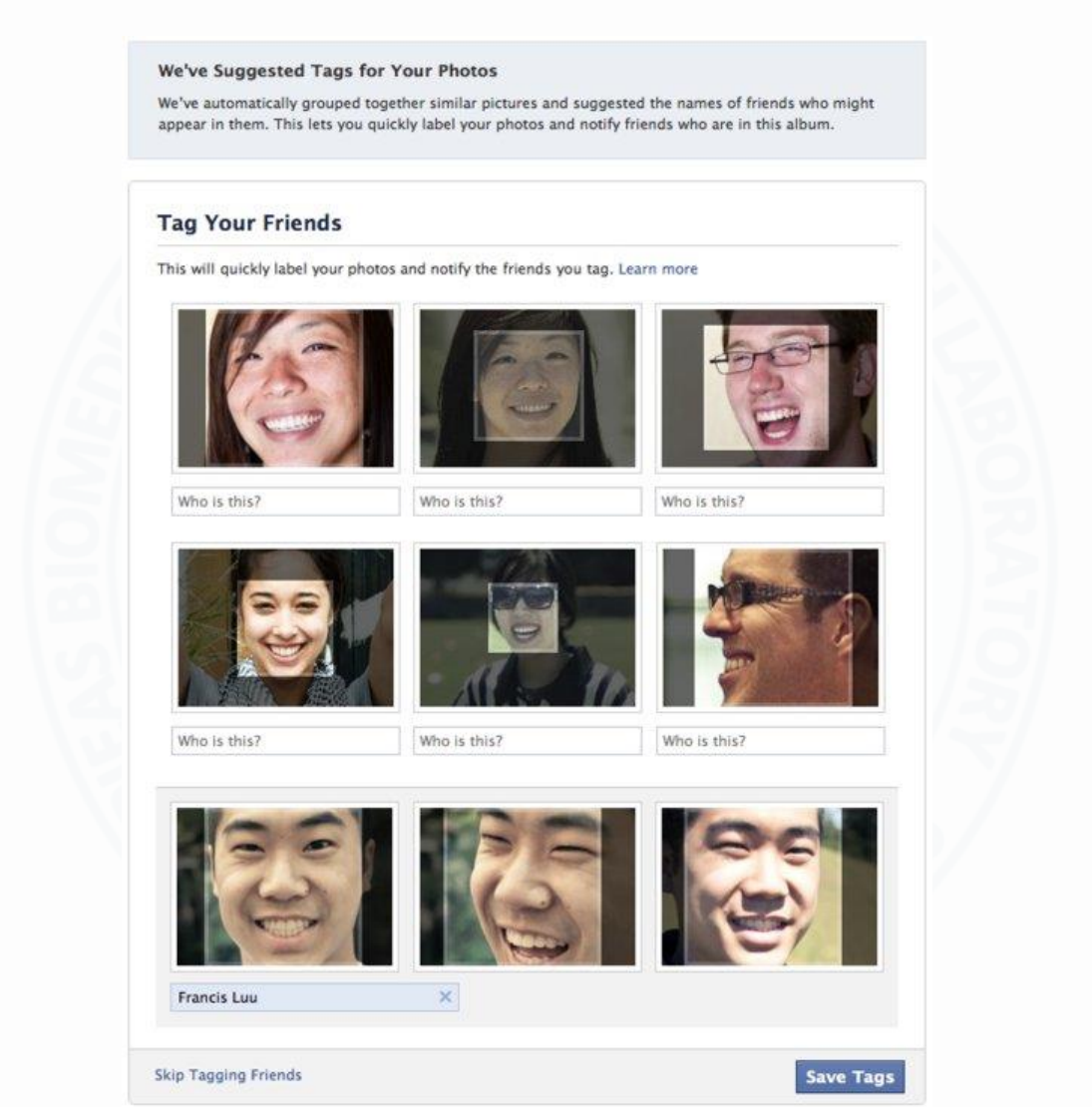
OCR Accuracy



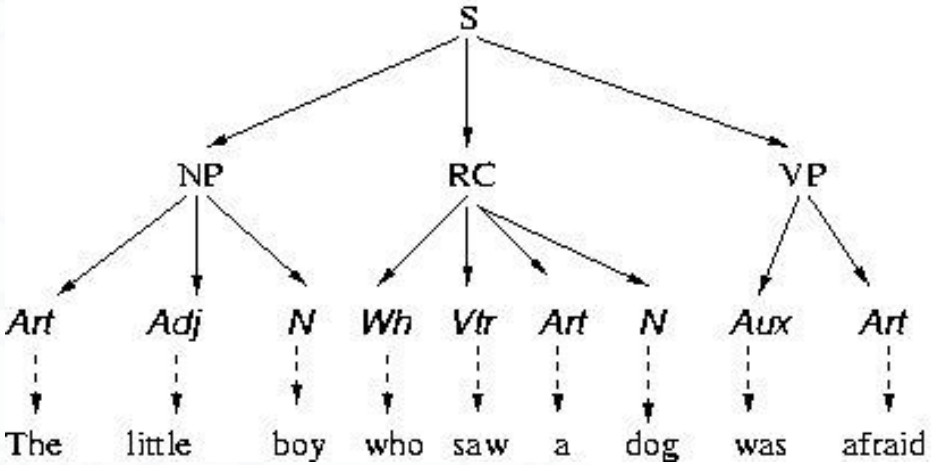
Gmail: ML in NLP

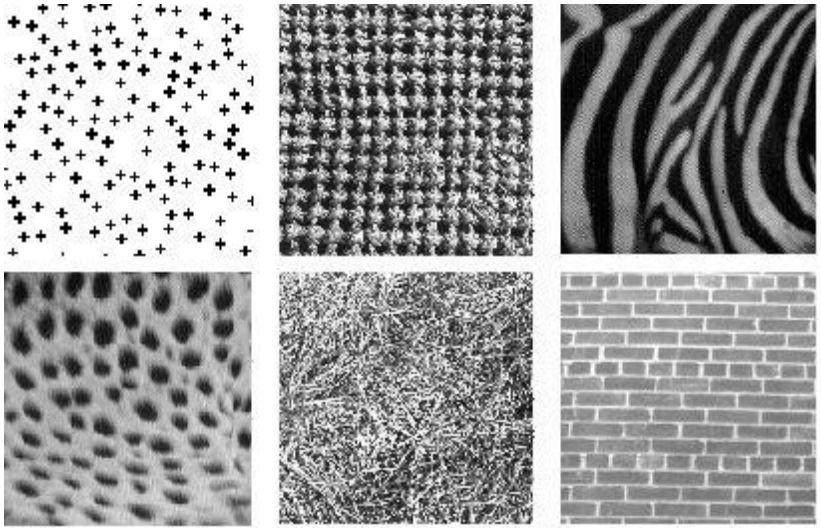


Facebook Friends Tagging



Applications of PR

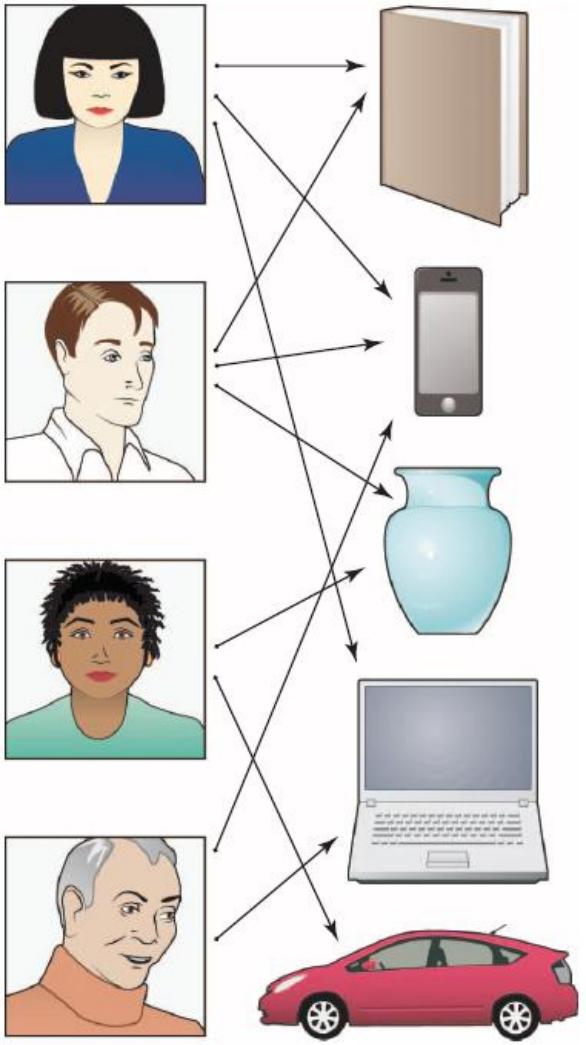




Recommender Systems

NETFLIX

- Recommend movies based on user preferences, interests and likes
- Similar ideas for facebook…
  - Find friends that share your interests

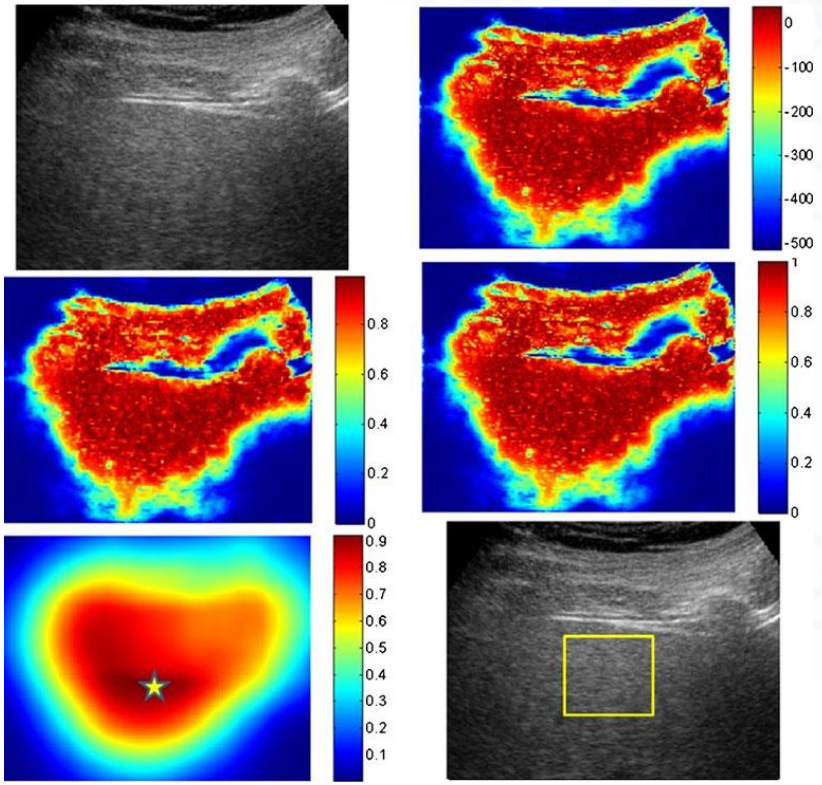


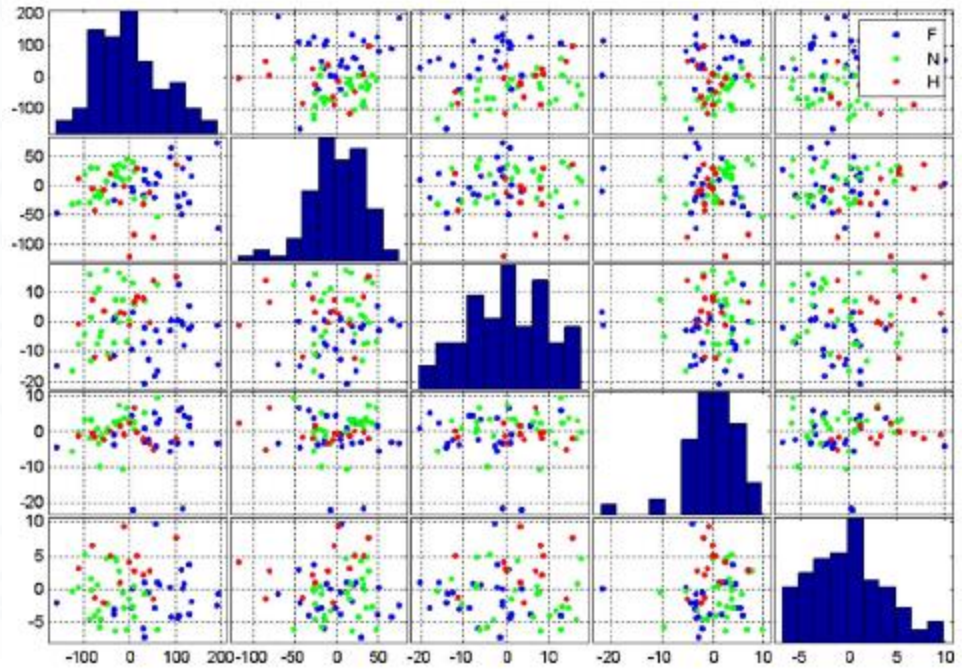
PIEAS Bio-Medical Informatics Lab

- Objective
  - Development of Intelligent Computational Solutions to problems in Biology and Medicine

BMI Lab Projects

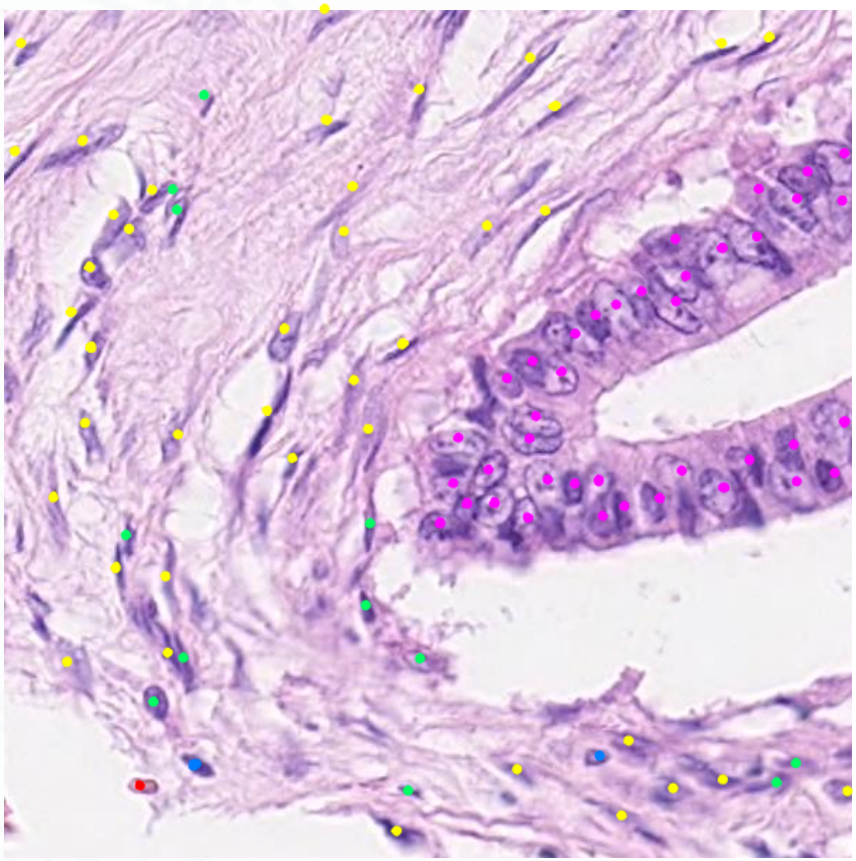
- Predicting Liver Disorders
  - Given: Liver ultrasound Images
  - Output: Diagnose surface & textural irregularities





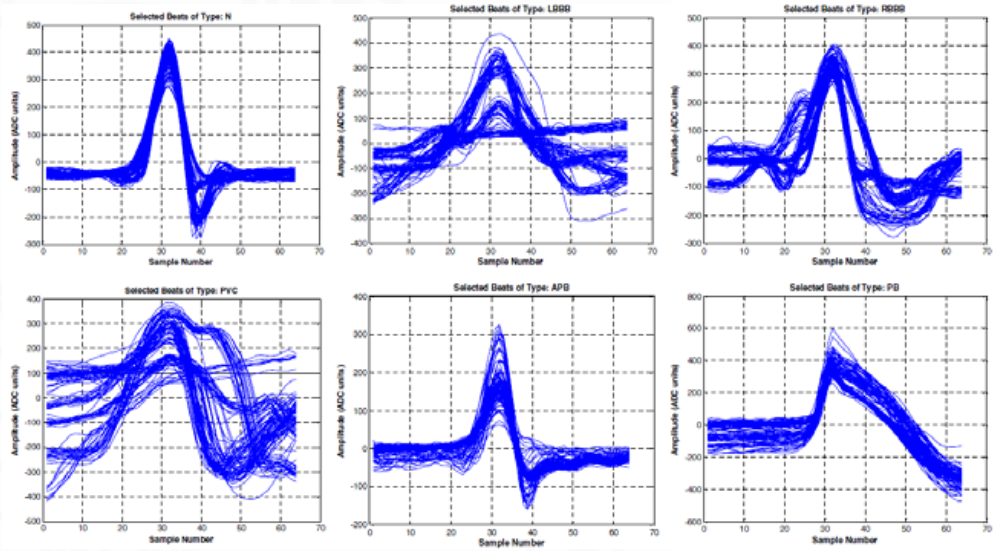
BMI Lab Projects

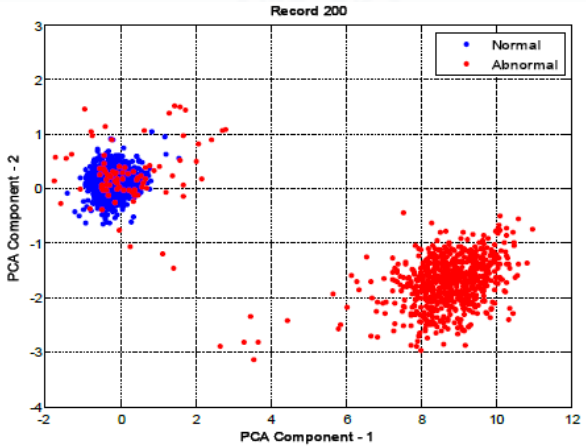
- Detecting cells
  - Input: Histopathology Images
  - Output: Identifying location and types of cells

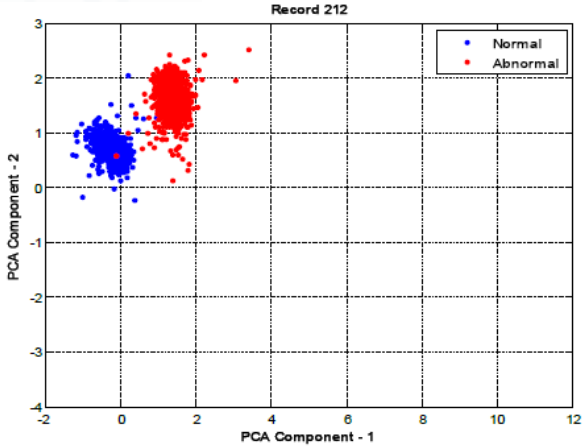


BMI Lab Projects

- Predicting ECG Abnormalities
  - Input: ECG Recording

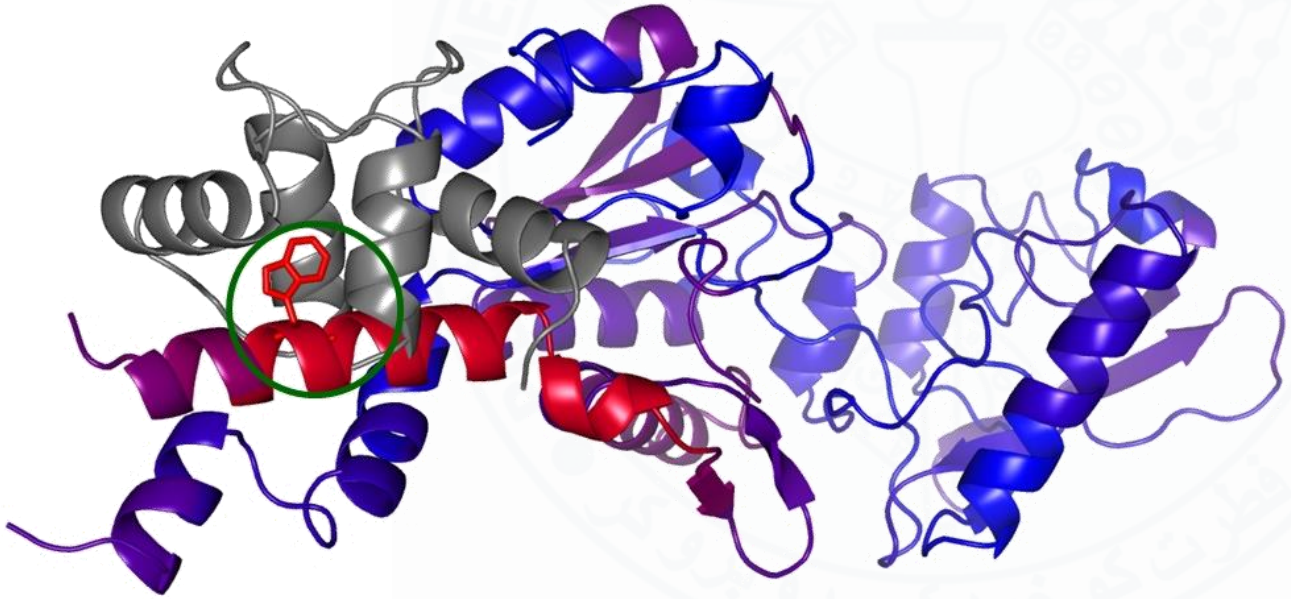


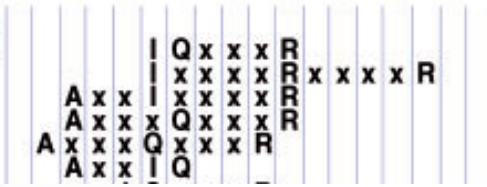




BMI Lab Projects

- Predicting Protein Binding Sites
  - Input: Protein Sequences or 3D structures
  - Output: Identifying interfaces





High predicted binding propensity

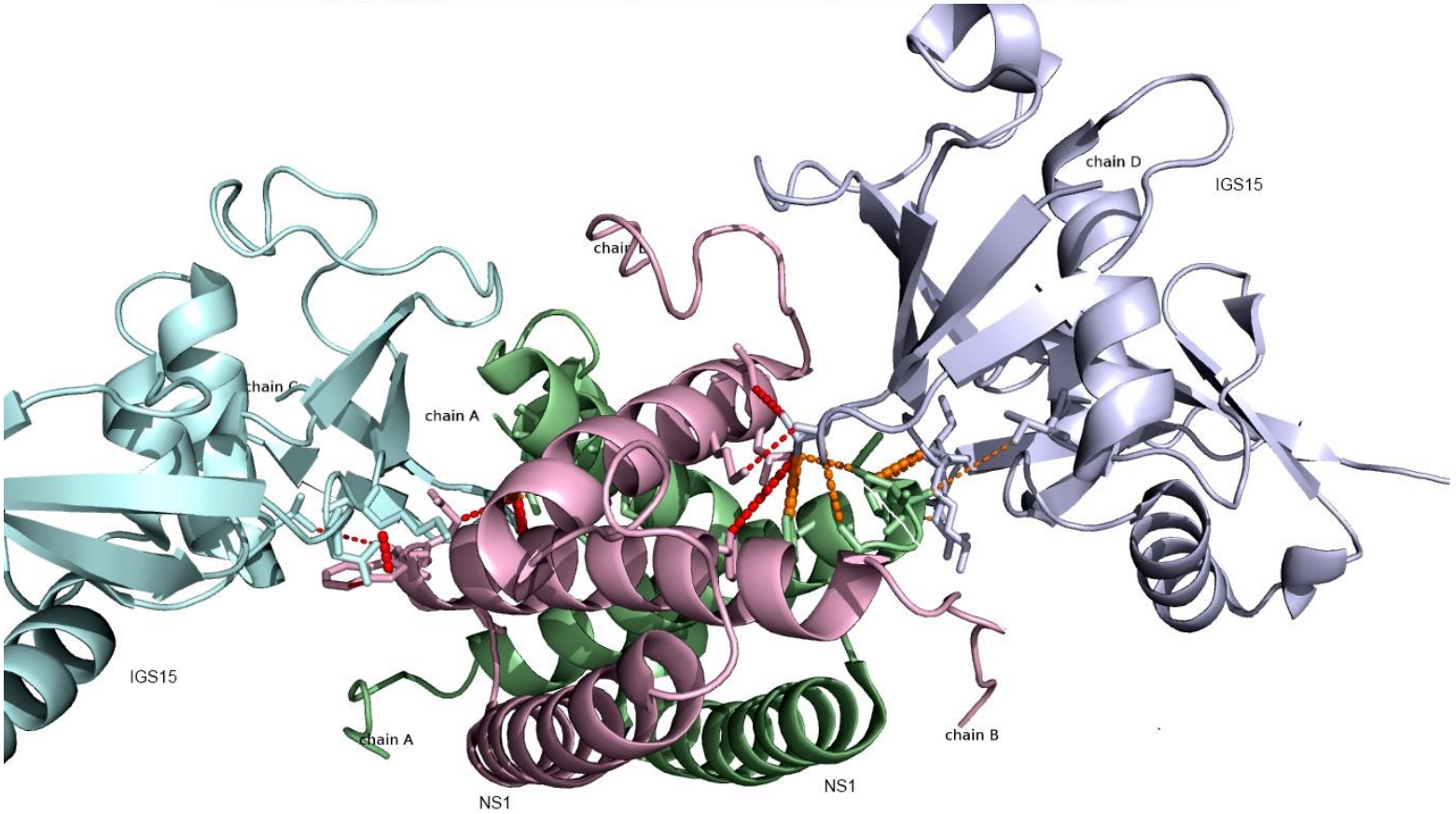
Low predicted binding propensity

Calmodulin

PDB: 1YRT

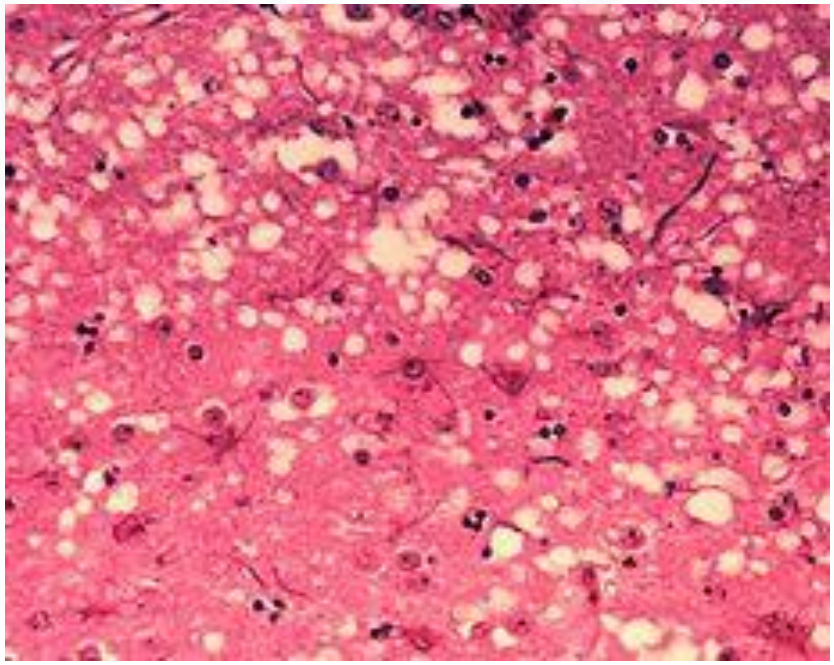
BMI Lab Projects

- Identifying Molecular Causes of Disease



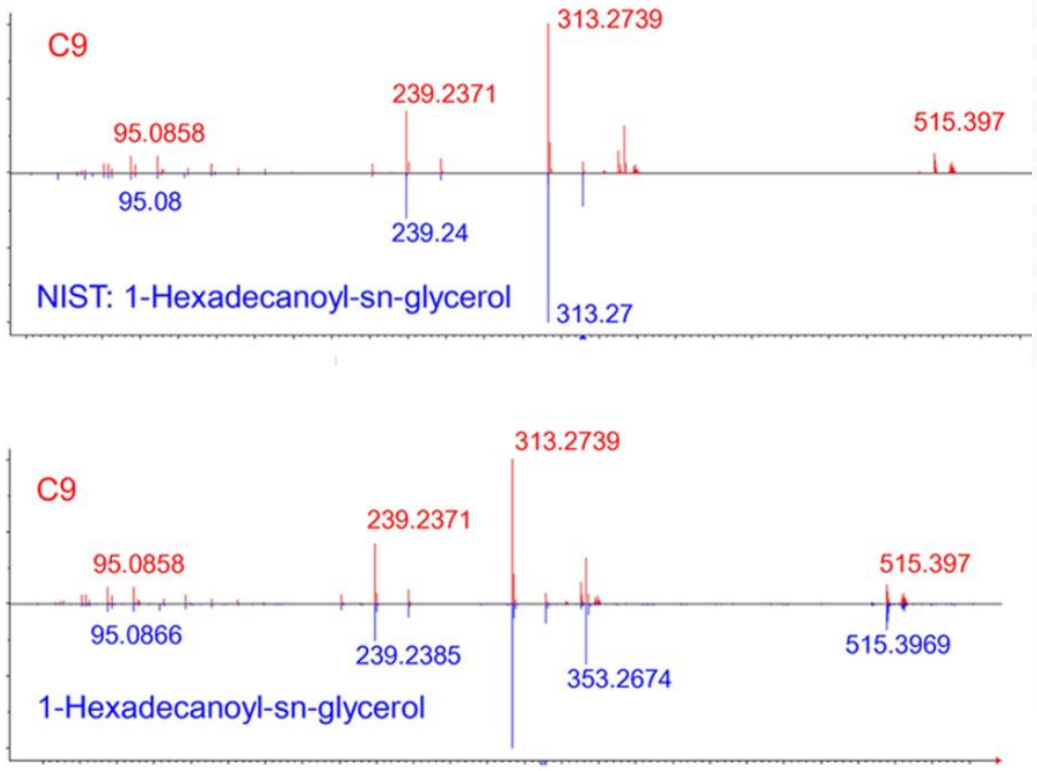
BMI Lab Projects

- Predicting Prion Proteins
  - Input: Protein Sequences
  - Output: Whether this protein can form prions



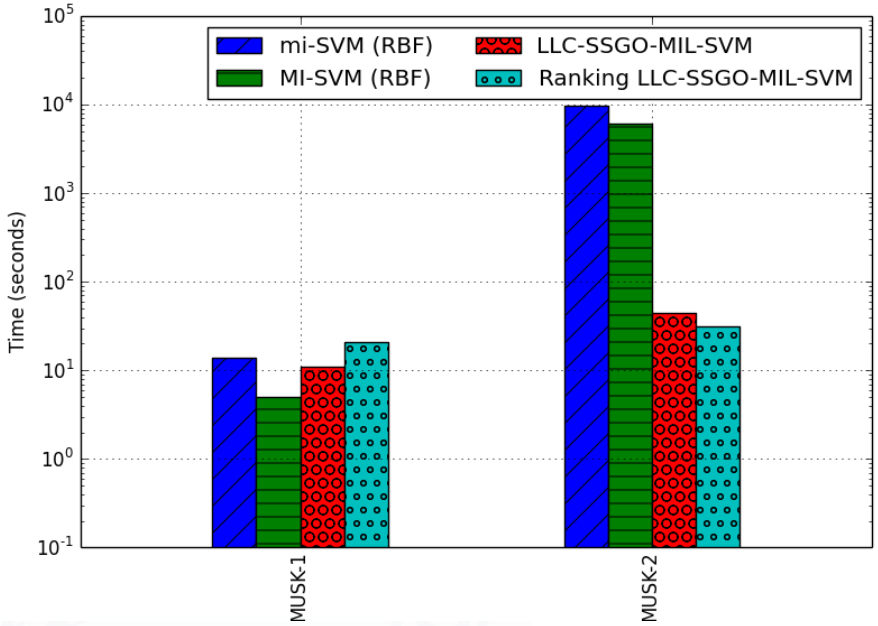
BMI Lab Projects

- Predicting Chemical Compounds in Mass-Spectrometry Data
  - RAMClust

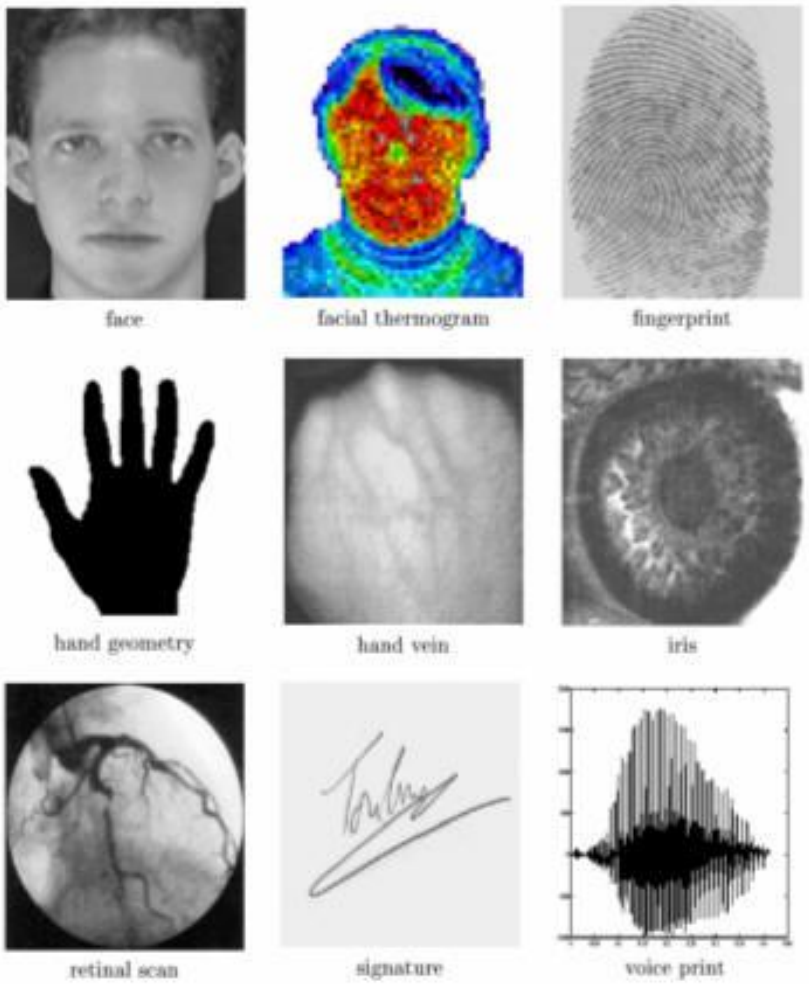


BMI Lab Projects

- Development of open-source machine learning tools and packages
  - PyLemmings: **Py**thon Based **L**arg**e** **M**argin **M**ultiple **I**nstance **L**earning **S**ystem
  - CAFÉ-Map: **C**ontext **A**ware **Fe**ature **Map**ping



BMI Lab Projects: Biometrics



Modern Issues in Machine Learning

- Extracting Features
  - Feature Engineering Takes a Long Time and Effort
  - Deep Learning
  - Graphical Processing Units

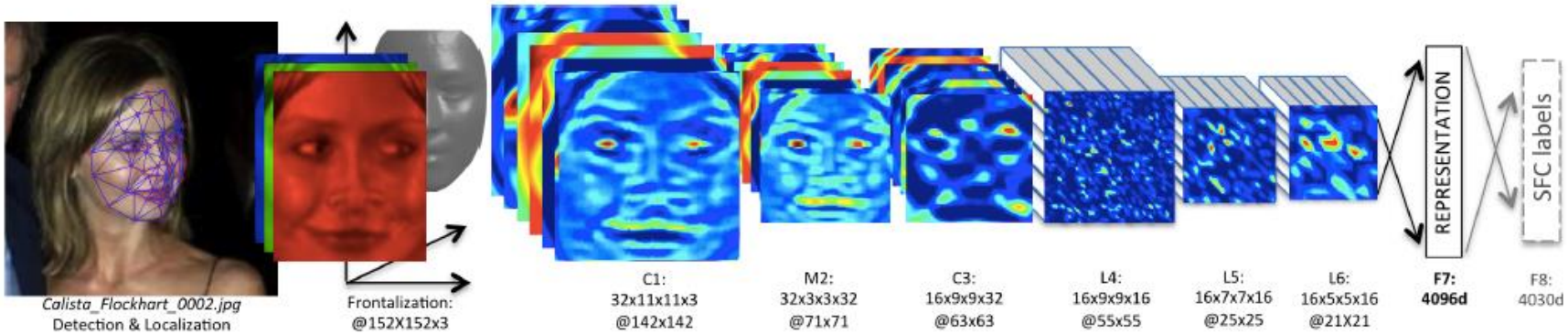
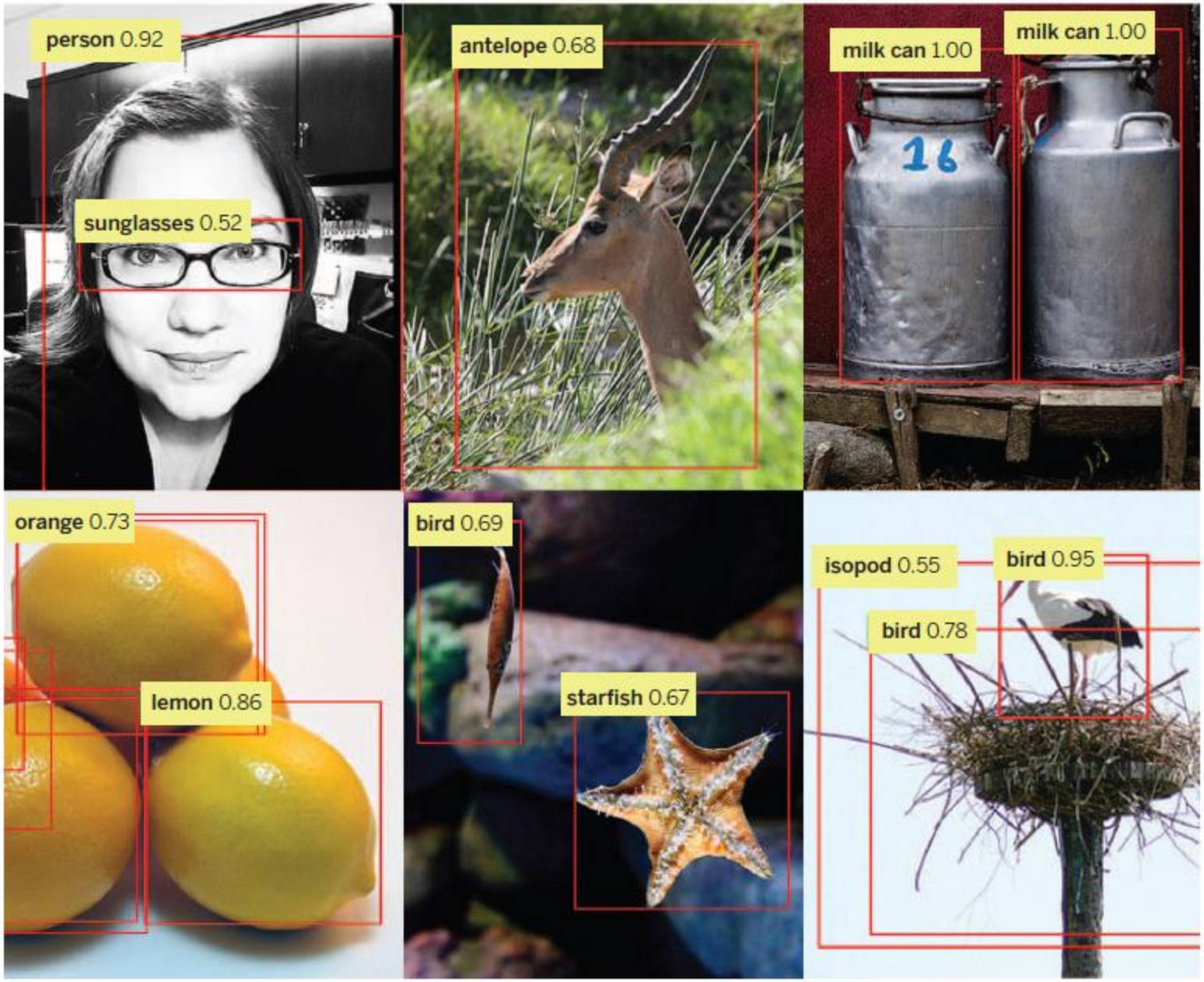


Figure 2. **Outline of the *DeepFace* architecture.** A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate outputs for each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

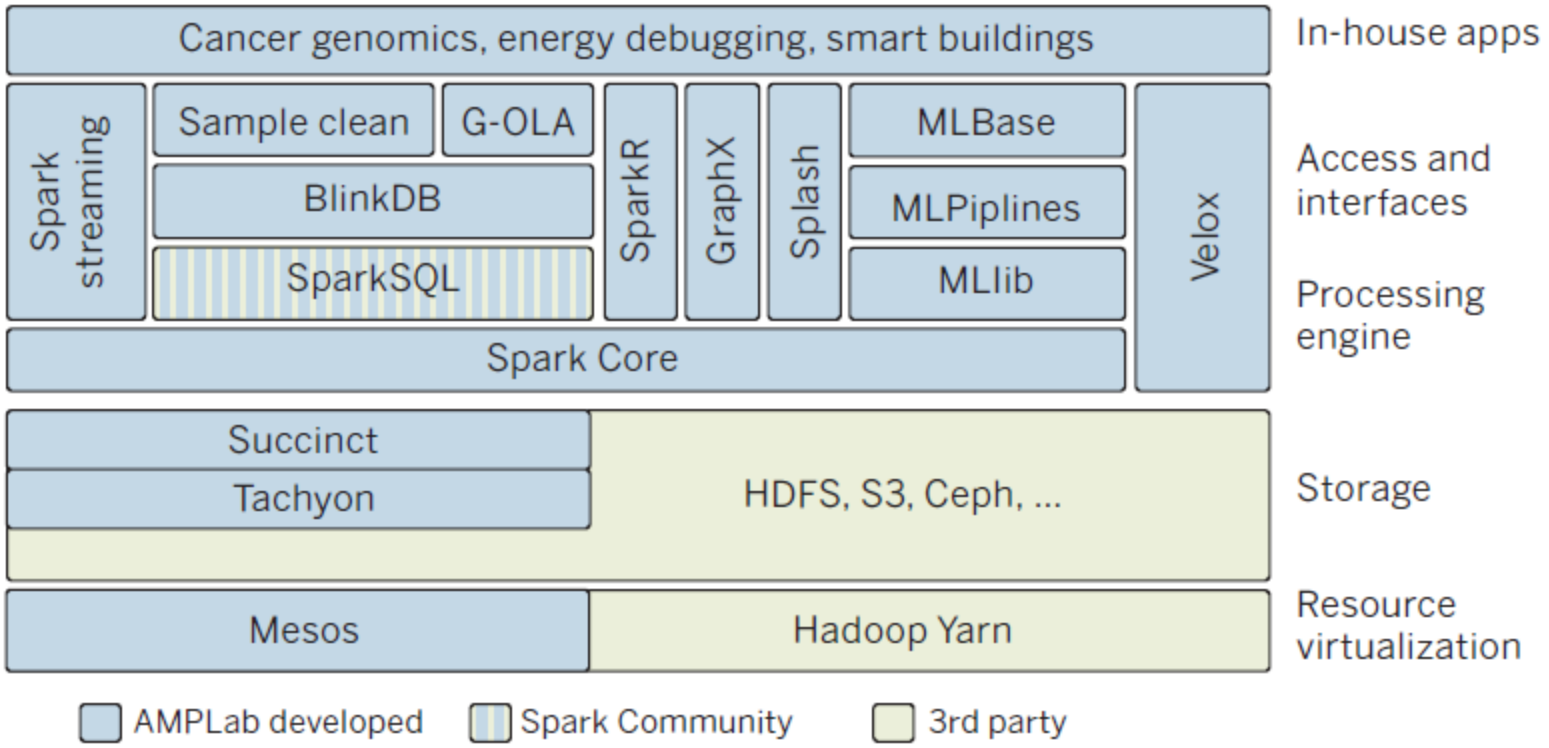

person 0.92
sunglasses 0.52
antelope 0.68
milk can 1.00
milk can 1.00
orange 0.73
lemon 0.86
bird 0.69
starfish 0.67
isopod 0.55
bird 0.95
bird 0.78

Modern Issues in Machine Learning

- Labeling Data
  - Getting labeled data is hard
    - Easier to obtain a large amount of unlabeled or partially labeled data
- Develop machine learning models that can learn from unlabeled or ambiguously labeled data
  - Multiple Instance Learning
  - Active Learning
  - Semi-Supervised Learning
  - Self-Taught Learning

Modern Issues in Machine Learning

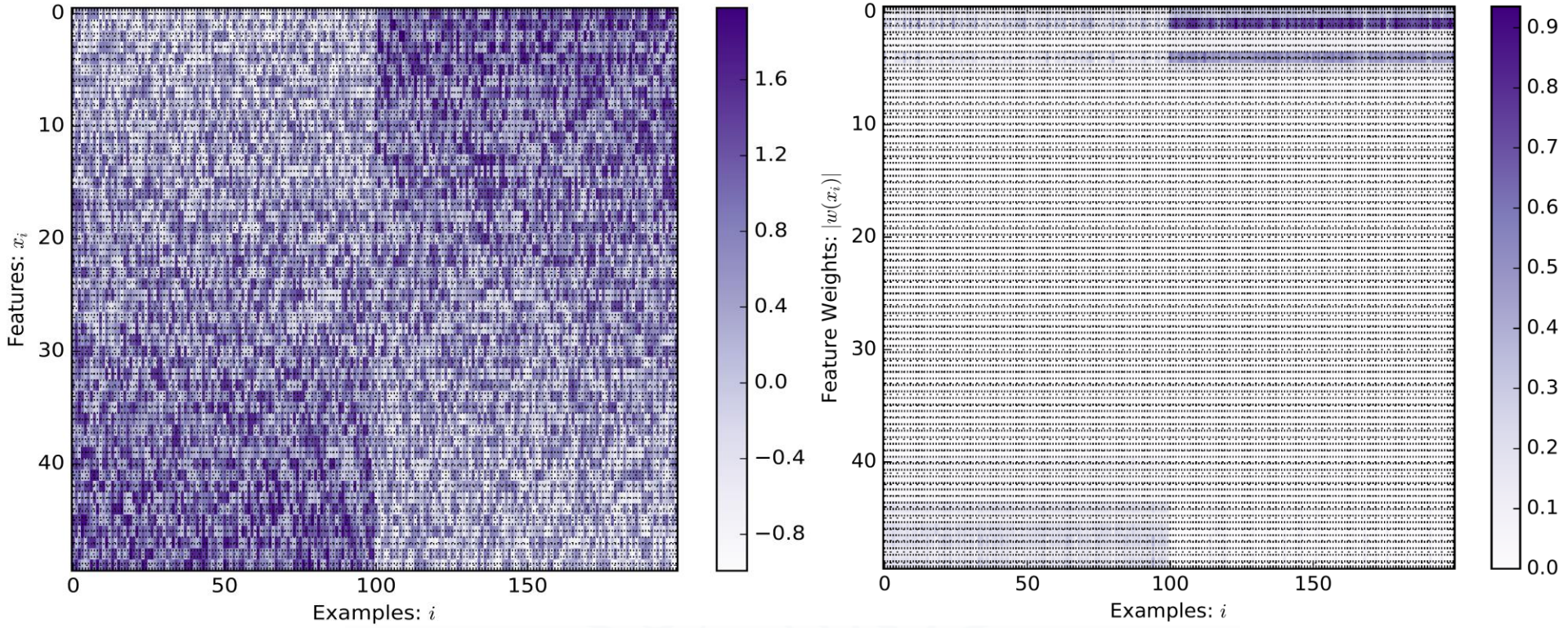
- Scalability and “Big Data”
  - GPUs
  - Cluster and Cloud Computing
  - Machine Learning as a service



Modern Issues in Machine Learning

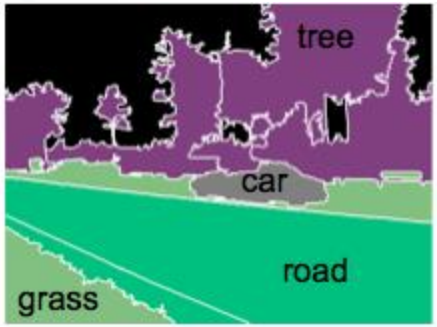
- Tall Data
  - Large number of dimensions
    - Many dimensions are unrelated
  - Small number of examples
    - Curse of Dimensionality
- Application areas
  - Bioinformatics

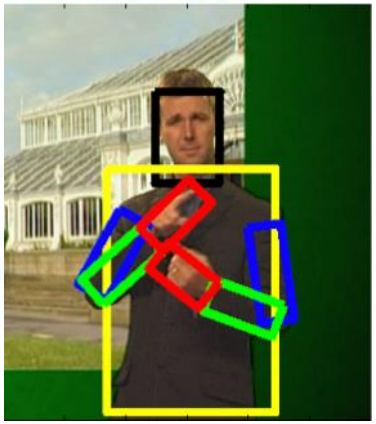
Example



Modern Issues in Machine Learning

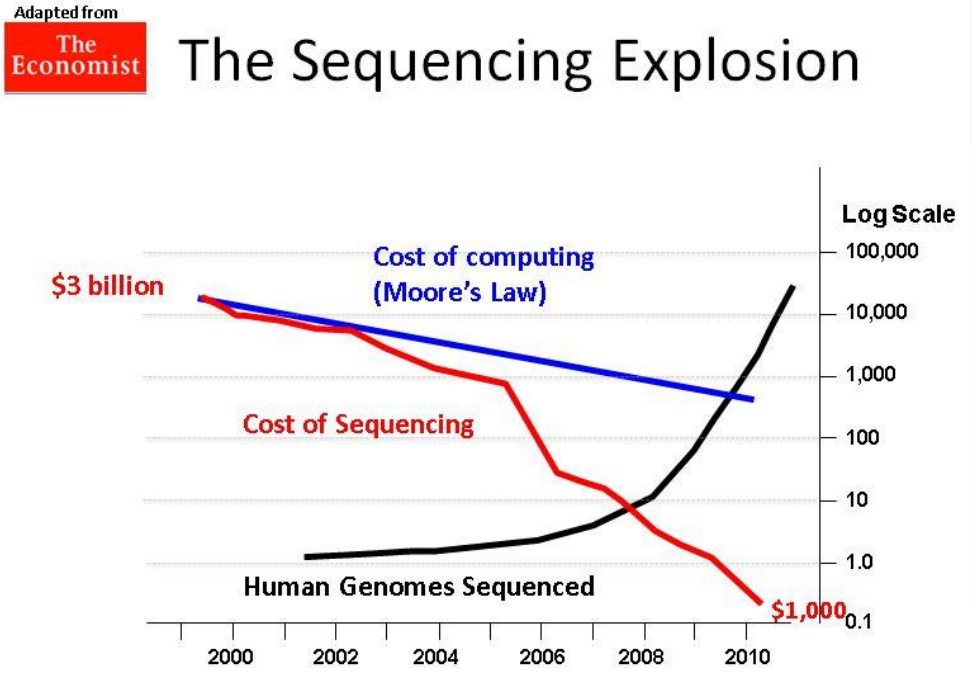
- Structured Outputs
  - Required output is not a simple decision
    - $f: X \rightarrow y$
  - Rather a complex data object
    - $f: X \rightarrow Y$
- Unstructured data
  - For example webpages or documents





Modern Application Areas: Bioinformatics

- The cost of DNA sequencing has come down
- Large amounts of data
- Few people to fill the gap
- Impactful Applications



How do I get started?

- Learn to program
  - Python
  - PIEAS Offering courses on Python Programming
- Take online courses or attend University ones
  - Coursera
- University Courses

How to Join the Lab?

- Passion for Biological and Medical Informatics
- Good Programming Skills
- Good Mathematics

- Need to know
  - Cross-Disciplinary Area
  - Application Oriented

References

- Publications
  - http://faculty.pieas.edu.pk/fayyaz/pubs.html
  - http://faculty.pieas.edu.pk/fayyaz/bmi.html
- Interesting Machine Learning Papers
  - Jordan, M. I., and T. M. Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” *Science* 349 (6245): 255–60. doi:10.1126/science.aaa8415.
  - Domingos, Pedro. 2012. “A Few Useful Things to Know About Machine Learning.” *Commun. ACM* 55 (10): 78–87. doi:10.1145/2347736.2347755.
  - Wagstaff, Kiri. 2012. “Machine Learning That Matters.” *arXiv:1206.4656 [cs, Stat]*, June. http://arxiv.org/abs/1206.4656.

We want to make a machine that will be proud of us.

- Danny Hillis