Introduction to Machine Learning

C. V. Jawahar
IIIT Hyderabad

Plans for today (and lead to next lecture)

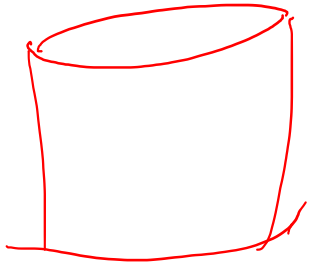
A. Introduction (the broad canvas)
B. Conceptual formulation for ML
C. Intro. to Supervised Learning algorithms
D. Worries During the Learning

A. Introduction

What is Modern “AI/ML” ?

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Data → **Computational Procedures** → **Intelligence**



Data Driven Intelligence

Input, Data, State → **Intelligent System** → **Decisions, Actions, Output**

Data Driven Problem Solving

Trivial problem, Simple, well-known model. (Y = X/3).
Can find the model with one example.

Complex problem. There could be more parameters (eg. Age, History)

Lot more data needed to solve. Complex biological process

X → [ ] → Y

Raju sells samosa. Some one observe the business data.
X = money paid by the customer
Y = no of samosa given in return
(X,Y) = (9,3), (24,8), (15,5)

How many samosas a new customer will get if she gives 30?

Dr. Sheela gives X cups of rice to different patients. They show Y increase in their glucose levels.

(X,Y) = (2,20), (1.5, 27), (1, 8)

For a new patient, if she gives 3 cups of rice, how much glucose will increase?

Knowledge

BI

Solvable?

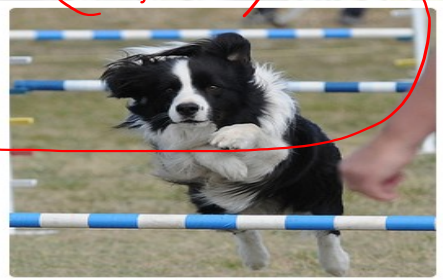
answer?

More Examples

(x_i, y_i) $y = f(x)$

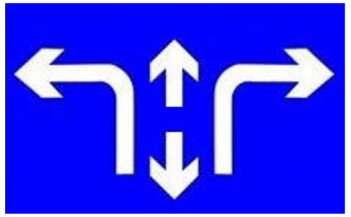
X

In this match , Rajasthan captain Ajinkya Rahane won the toss and decided to bowl first .



General Strategy: Given many examples of (X,Y), learn an automated solution to predict Y given a new X.

Y

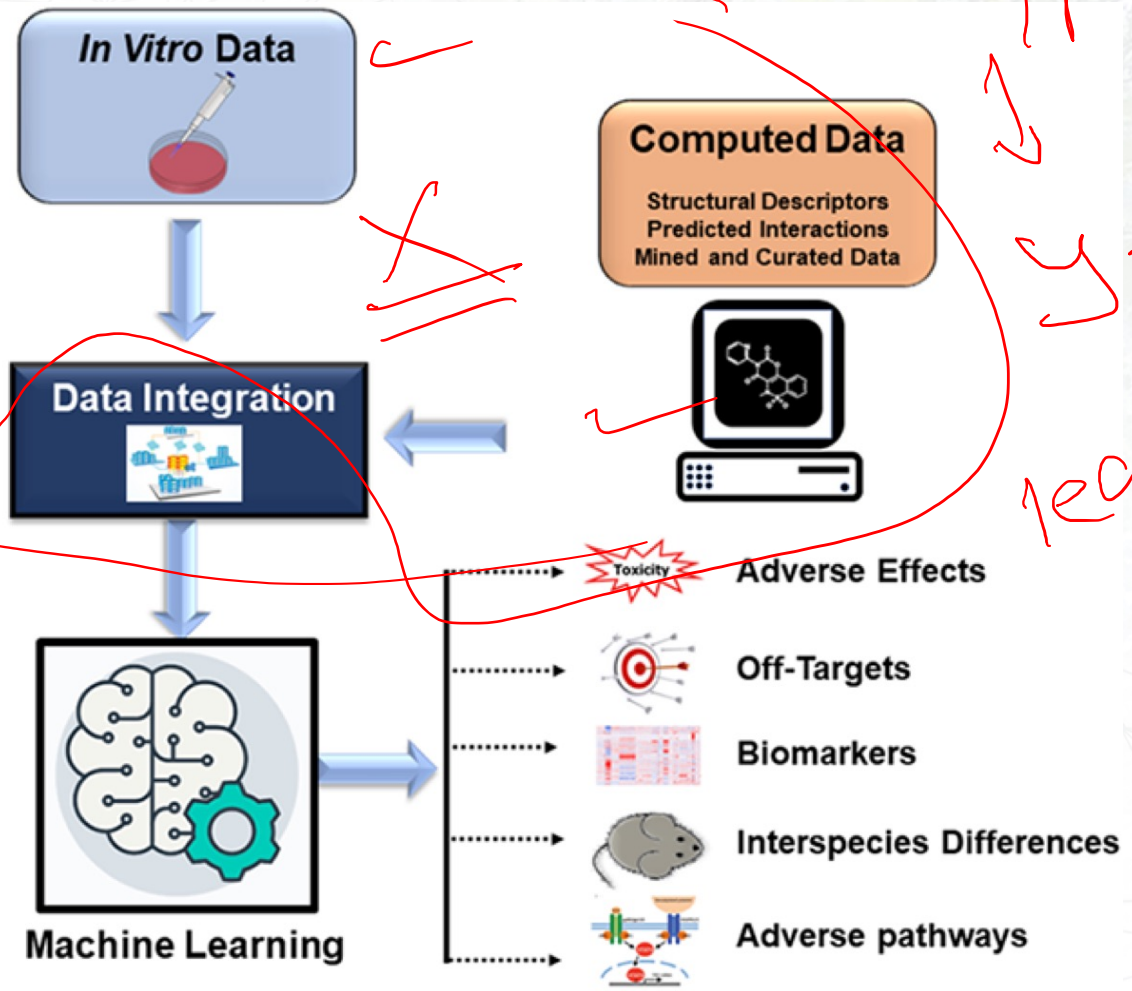


इस मैच में राजस्थान के कप्तान अजिंक्य रहाणे ने टॉस जीतकर पहले गेंदबाजी का फैसला किया ।

"Black and White Dog Jumps over Bars"

Drug Toxicity Prediction

Task: Classification
Predict Toxicity of drug molecule under development
Output: Type of toxicity

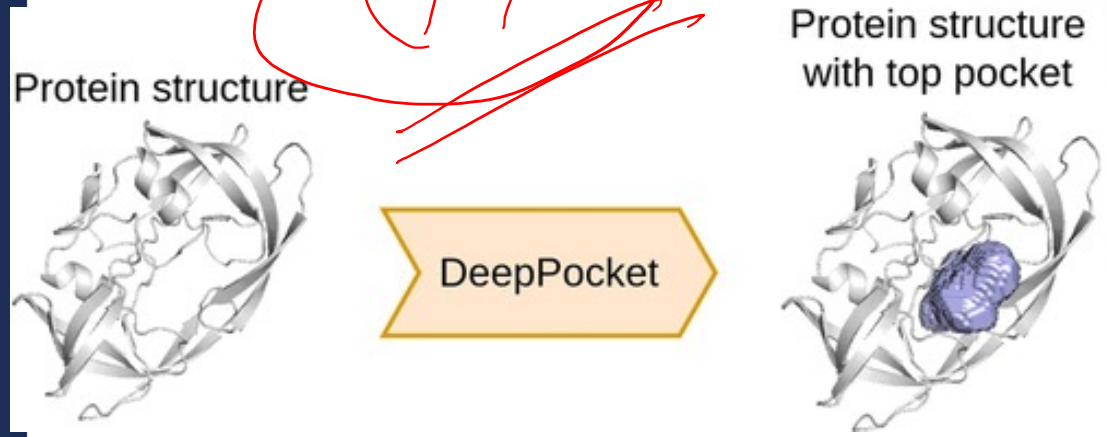


Chem. Res. Toxicol. 2020, 33, 20

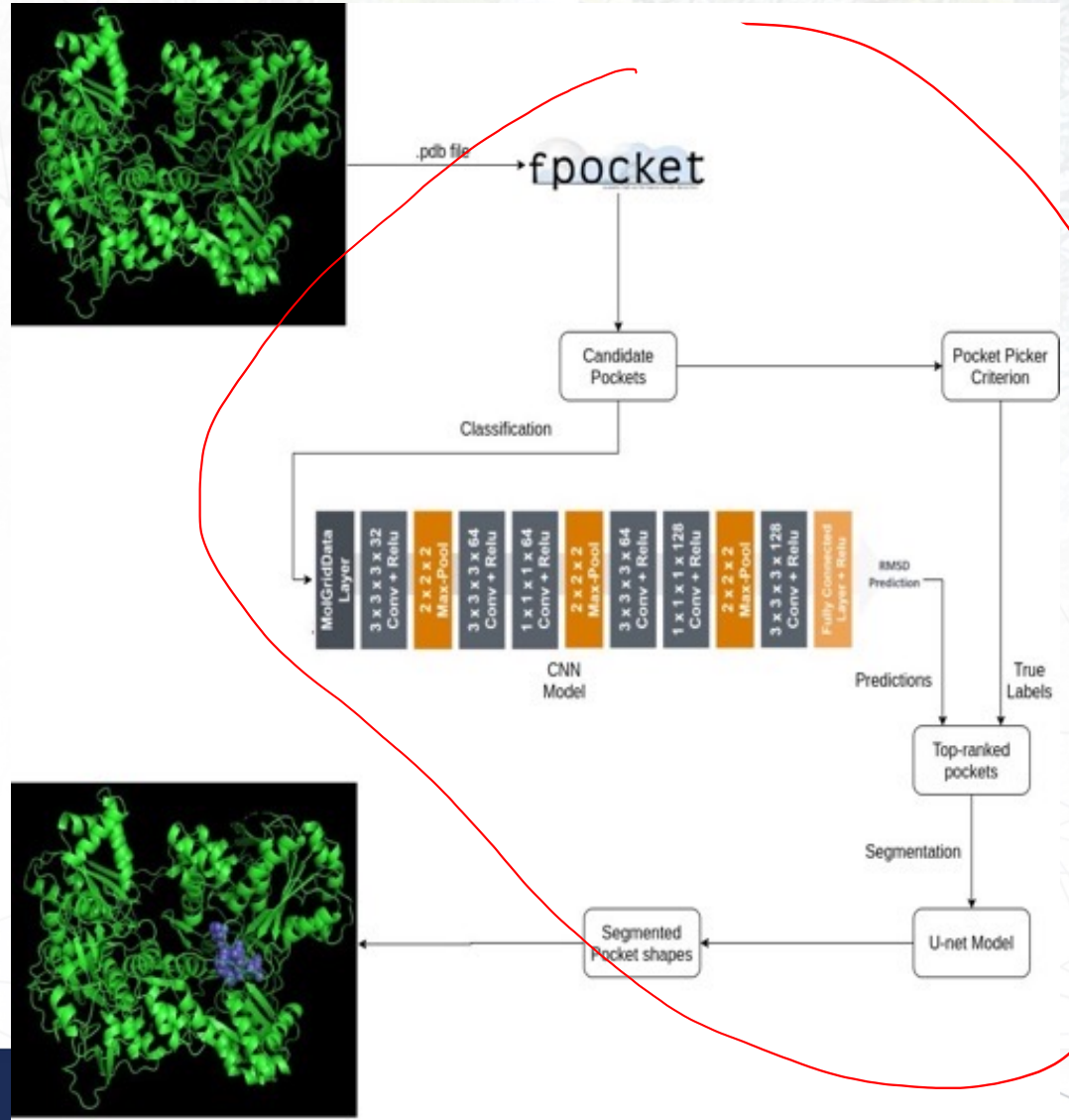
Drug Binding Pocket Identification in Proteins

Input: All pockets predicted using traditional (fpocket)
Output: Most probable binding pocket

(x, y)



J. Chem. Inf. Model. 2022 (doi:10.1021/acs.jcim.1c00799



ILL Posed Problems: Why do they work?

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Can Human(experts) do this?

How do they do?

X

Y

Title: Biscuits

Ingredients:
Flour, butter, sugar, egg, milk, salt.

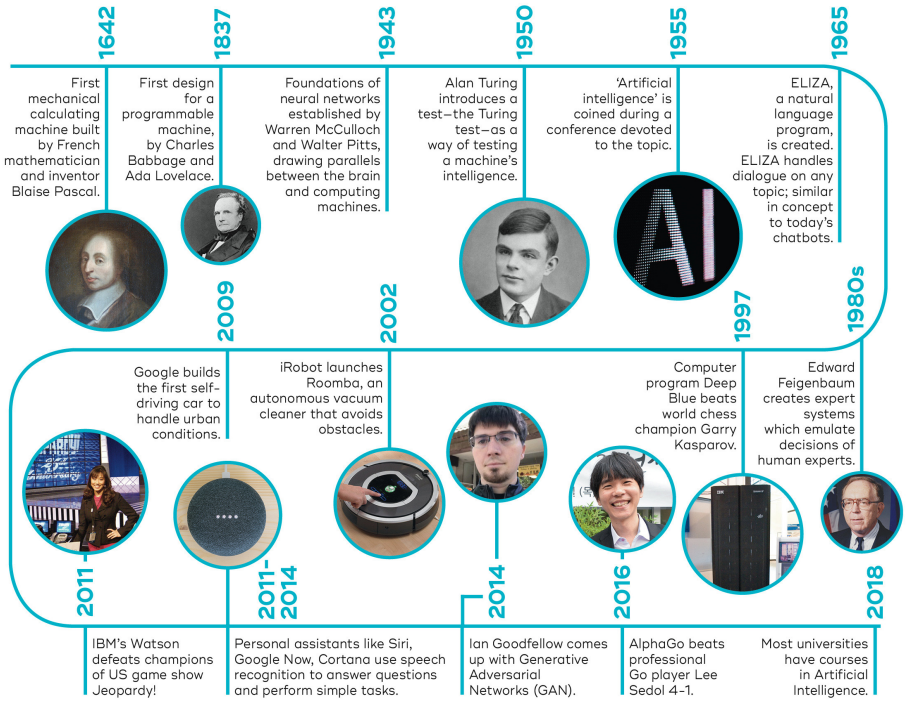
Instructions:
- Preheat oven to 450 degrees.
- Cream butter and sugar.
- Add egg and milk.
- Sift flour and salt together.
- Add to creamed mixture.
- Roll out on floured board to 1/4 inch thickness.
- Cut with biscuit cutter.
- Place on ungreased cookie sheet.
- Bake for 10 minutes.

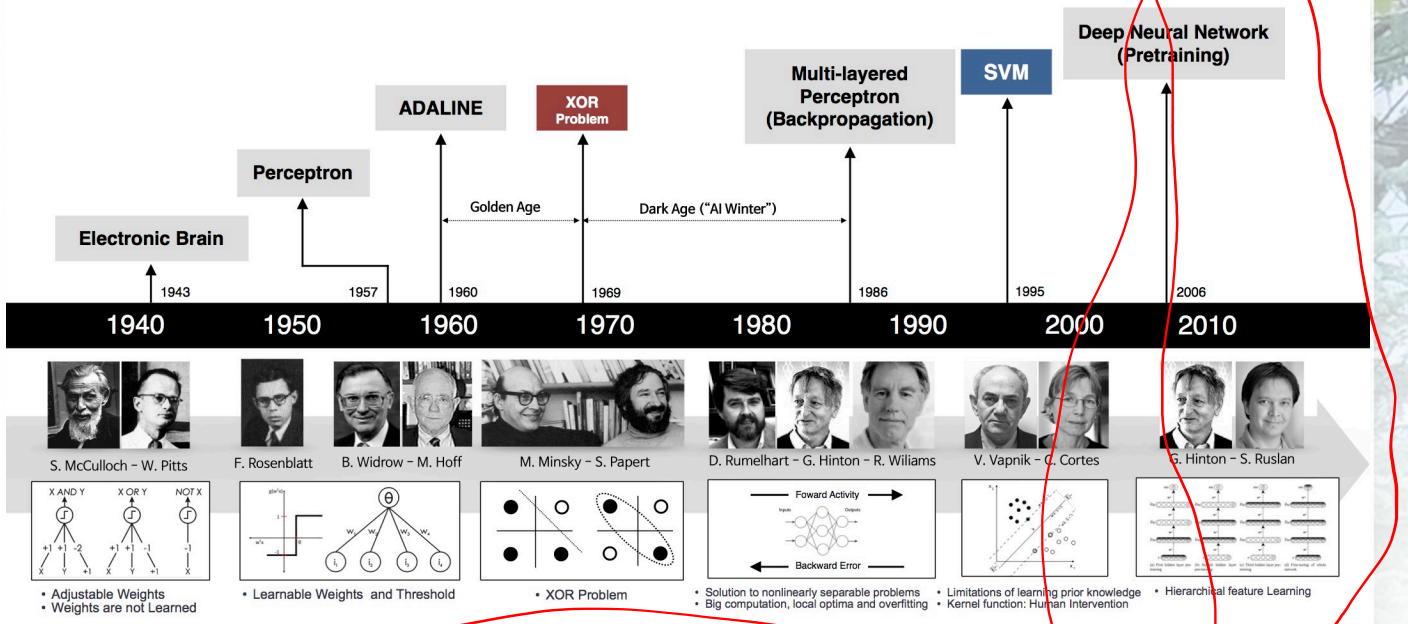
Extensive use of Prior Knowledge.

Composition of parts seen in the past.

Inverse Cooking (CVPR 2019)

Is AI, ML, DL really New?



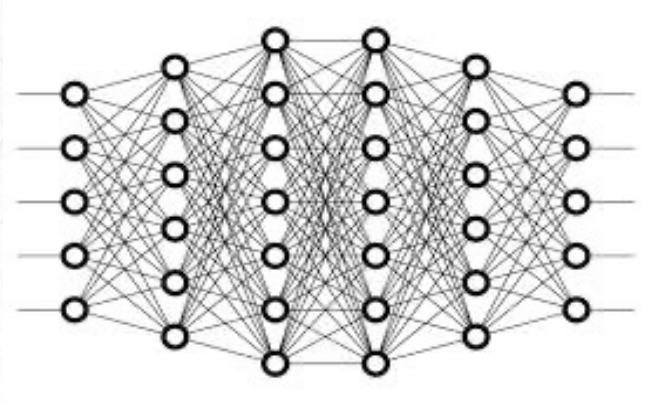


70 years

Why AI started to work now?

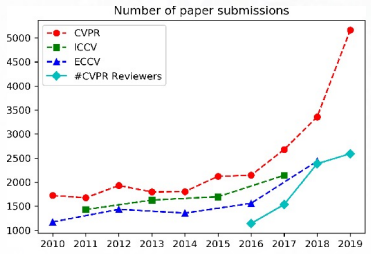
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

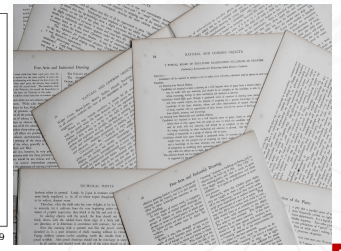
Data; Internet; Connectivity

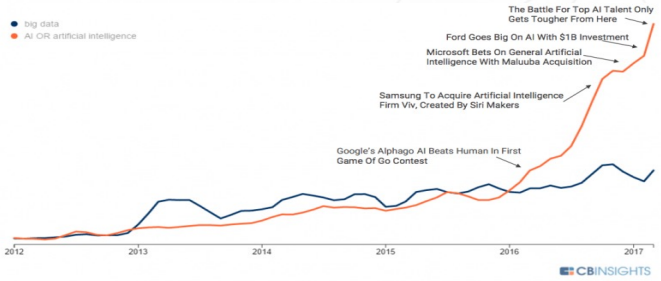


Algorithms, Deep Learning

Compute, Cloud, APIs, Libraries





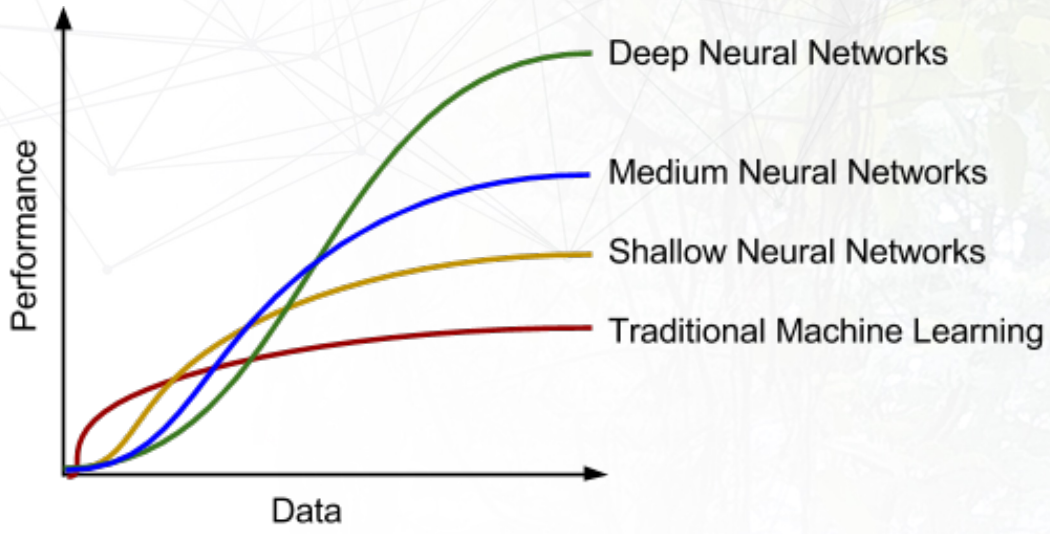


More People, Papers, Results, Funding, People. Positive Feedback.

Error Rates and Trends

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Low entropy bounded

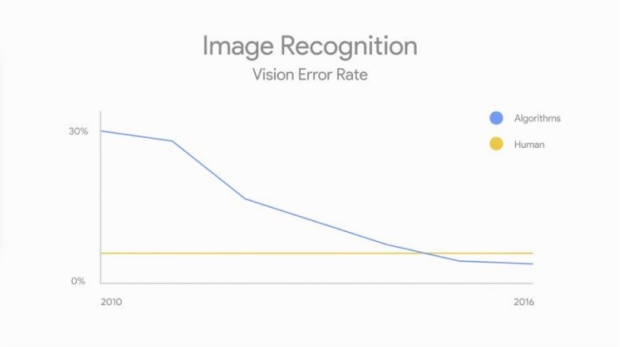


"Task"

"Meboic"

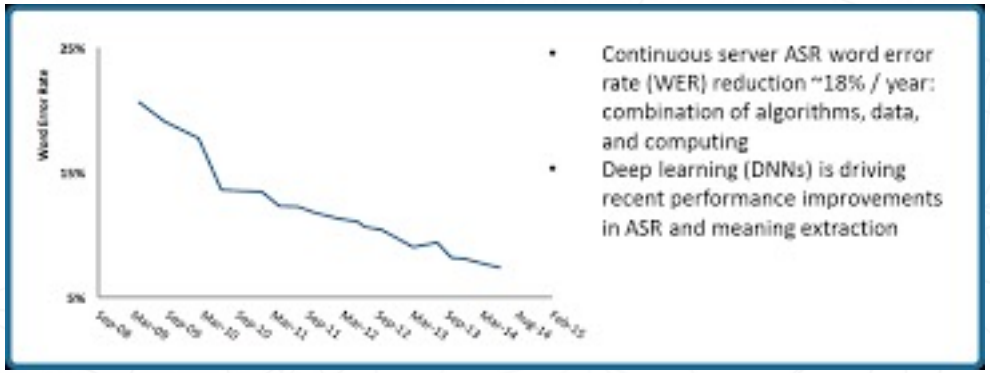
open

AI Versus Human Image Recognition Error Rates



Source: Unpacking the AI-Productivity Paradox,"
By Erik Brynjolfsson et al., *MIT Sloan Management Review*, January 2018
sloanreview.mit.edu

Error rates are better than human for many tasks



- Continuous server ASR word error rate (WER) reduction ~18% / year: combination of algorithms, data, and computing
- Deep learning (DNNs) is driving recent performance improvements in ASR and meaning extraction

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Domains, Applications and Problems

- Flood Prediction & Mitigation
- Safety on Roads
- Technology for Specially, Abled
- Fraud Detection in Banks
- Low Cost Medical Screening
- Crowd Management (Public Safety)
- Disease Prediction & Prevention
- Education & Skilling in all Indian Language
- Search Engine for Legal Judgment
- Air Quality Measurement

Algorithms and Areas

- Computer Vision
- Speech Recognition
- Language Processing
- Robotics
- Classification
- Clustering
- Regression
- Sequential Data Analysis
- Other Machine Learning

Data and Technology

- Data Annotation
- Data Platforms
- General (people) Internet Including Social Media
- Cyber Physical System (Sensor Networks)
- Remote Sensing
- Internet of Things
- Mobile Devices
- E – Governance & Workflow Systems

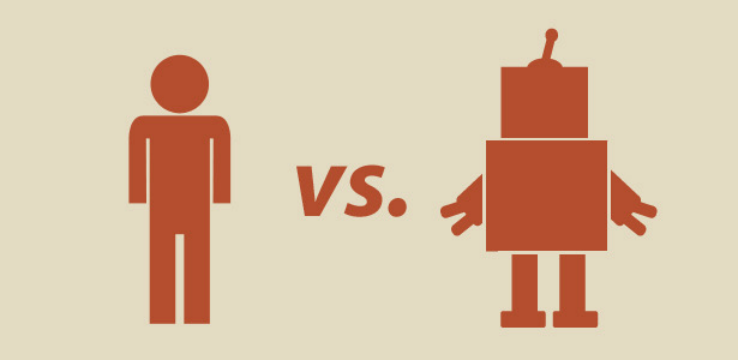
Worries

"People who use white bed sheets have higher probability of death"

Accuracy at the cost of human interpretability.

A new study finds a potential risk with self-driving cars: failure to detect dark-skinned pedestrians

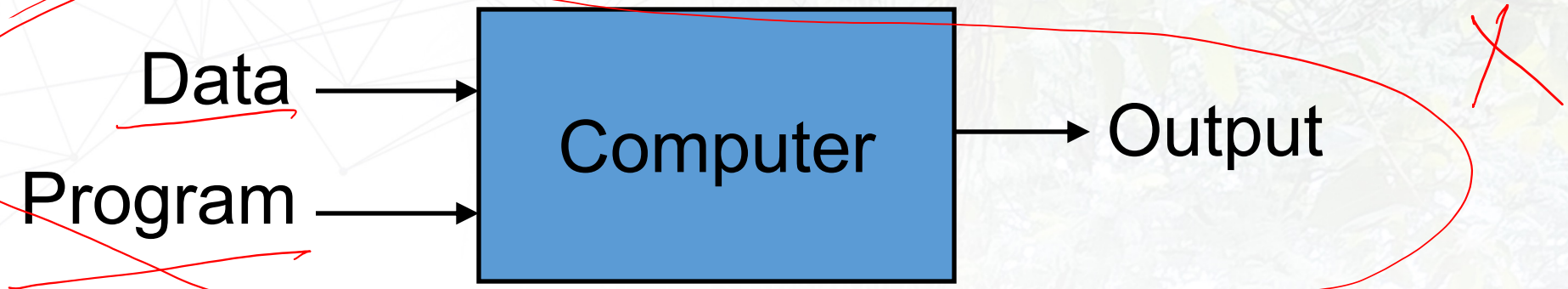
The findings speak to a bigger problem in the development of automated systems: algorithmic bias.



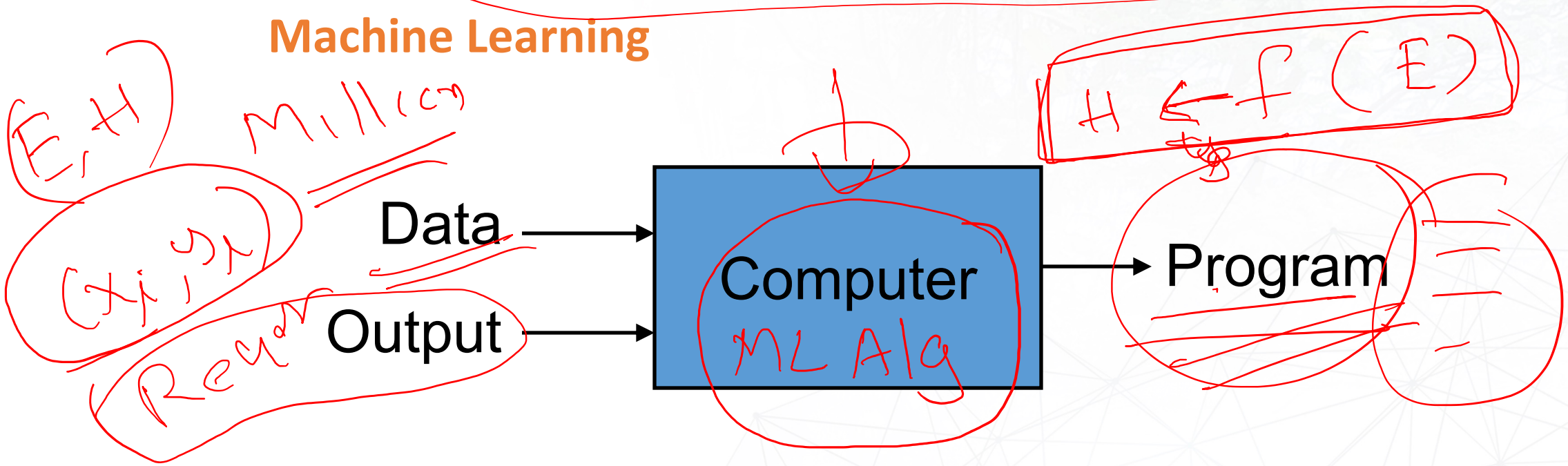
Humans does not require millions of examples to learn.

B. Conceptual Formulation

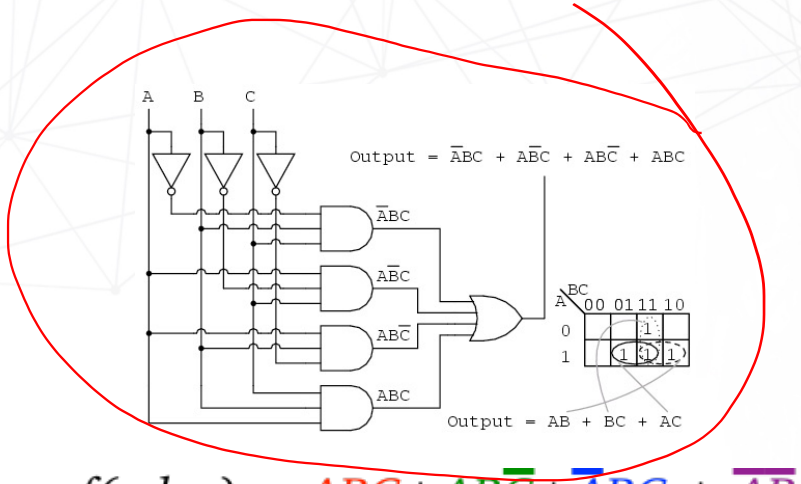
Traditional Programming

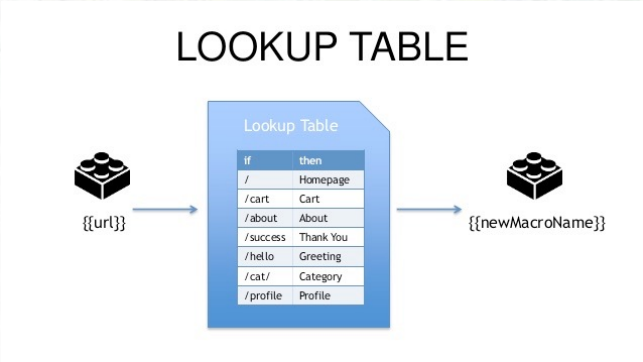


Machine Learning

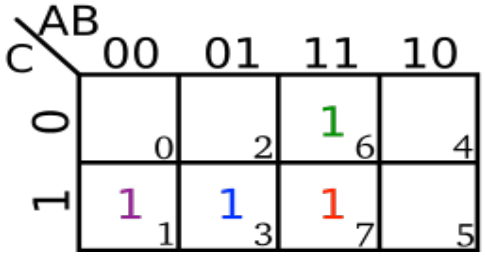


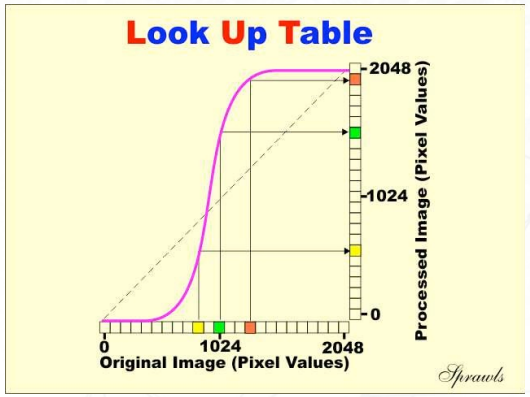
You already know about f() that fits samples.





$$f(a,b,c) = \underset{111}{ABC} + \underset{110}{AB\overline{C}} + \underset{011}{\overline{A}BC} + \underset{001}{\overline{A}\overline{B}C}$$





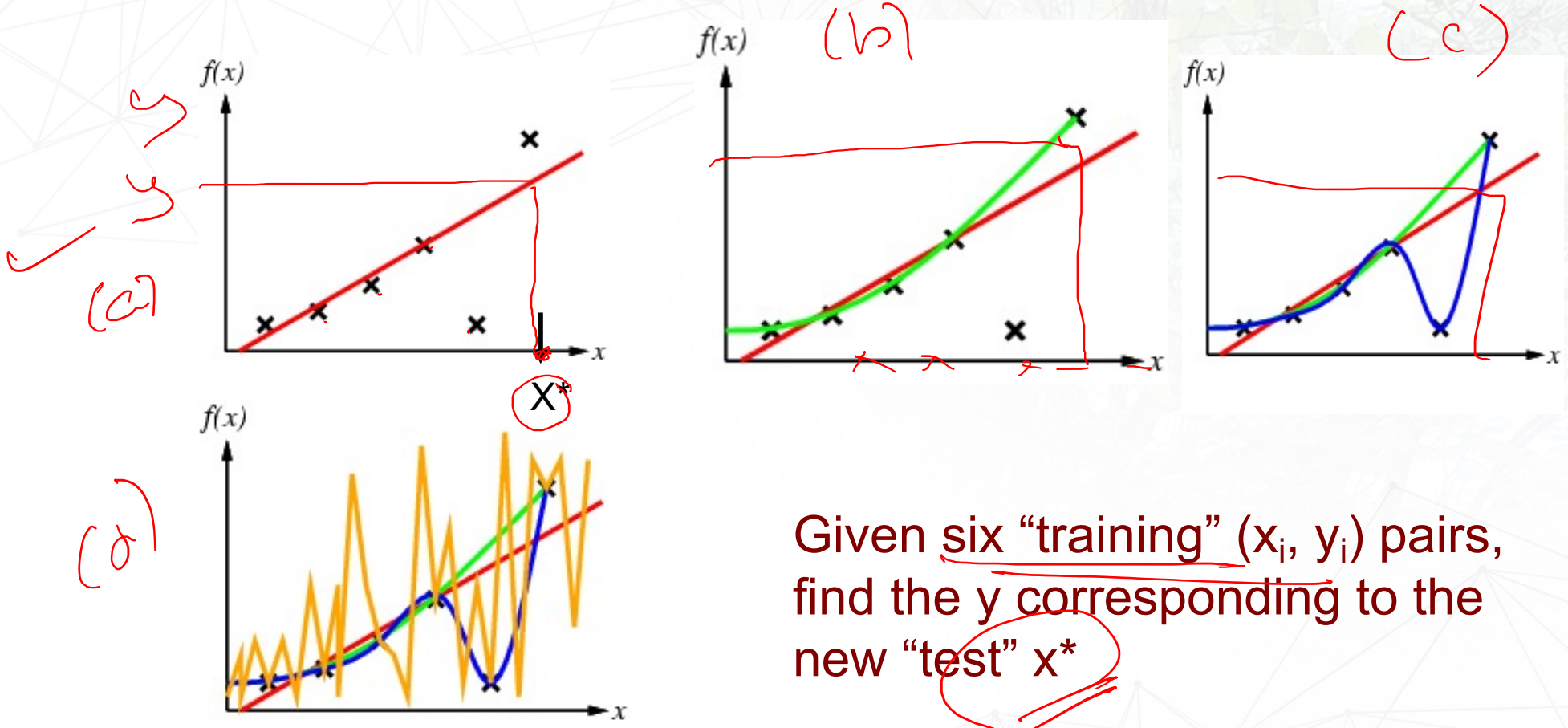
Is K-Map Learning?

Is LUT Learning?

Generalization

Learning is concerned with accurate prediction of future data, *not* accurate prediction of training or available data.

A Simple Fitting/Predicting Problem



Given six "training" (x_i, y_i) pairs, find the y corresponding to the new "test" x*

Which curve is the best?

Blank Slide

Occam's Razor

Select the simplest hypothesis (solution) that suits the data.

Min error of fit + λ degree or complexity

Regularization

Eg. Minimize Sum of "fit error" and "degree of the polynomial"

The machine learning framework

$$y = f(\mathbf{x})$$

output prediction function feature or representation

Rep
feat
embedd

(x_i, y_i)

$\sum_{i=1}^{N} (y_i - f(x_i))$

Compute
Inf e

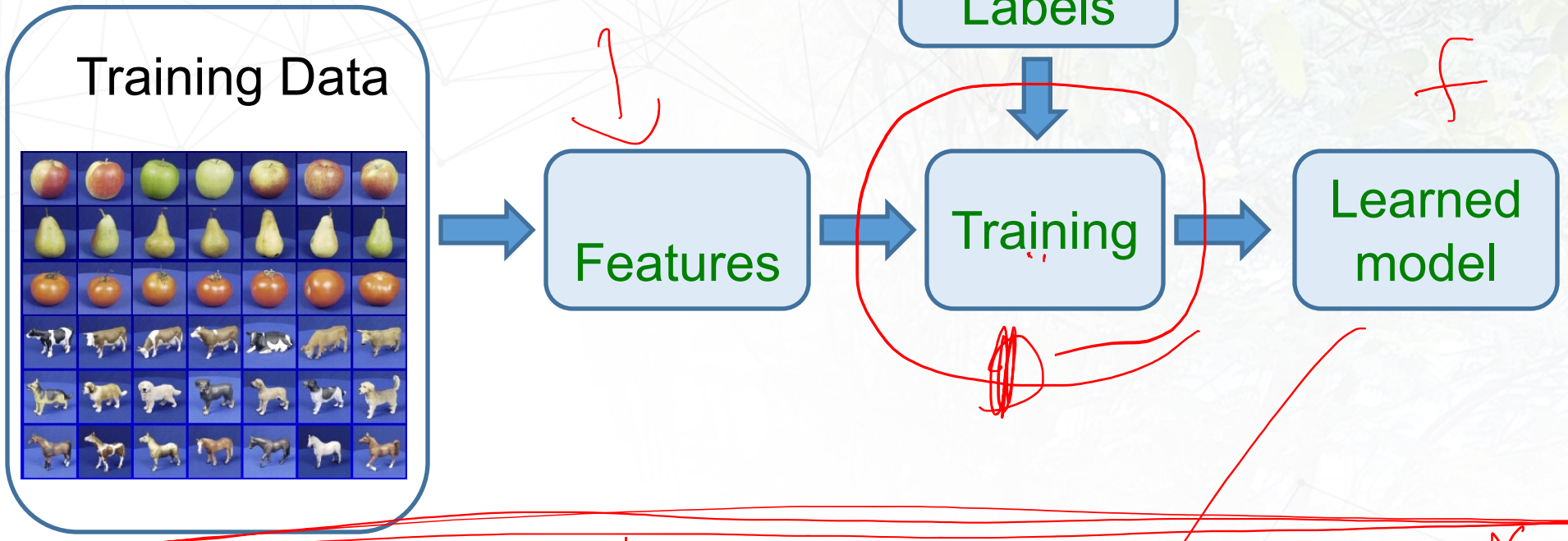
- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error.
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Infer
replc

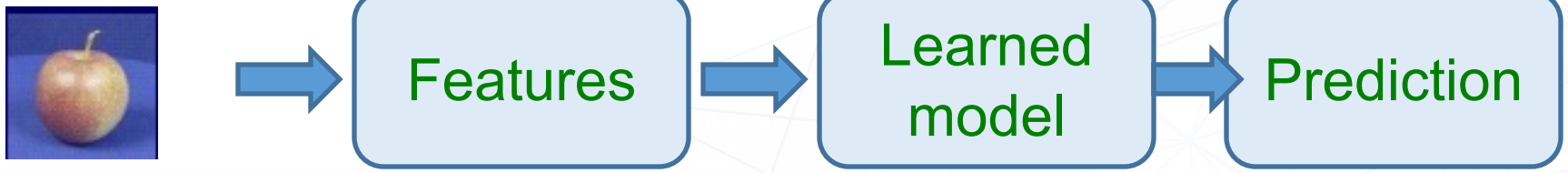
Slide credit: L. Lazebnik

Steps

Training



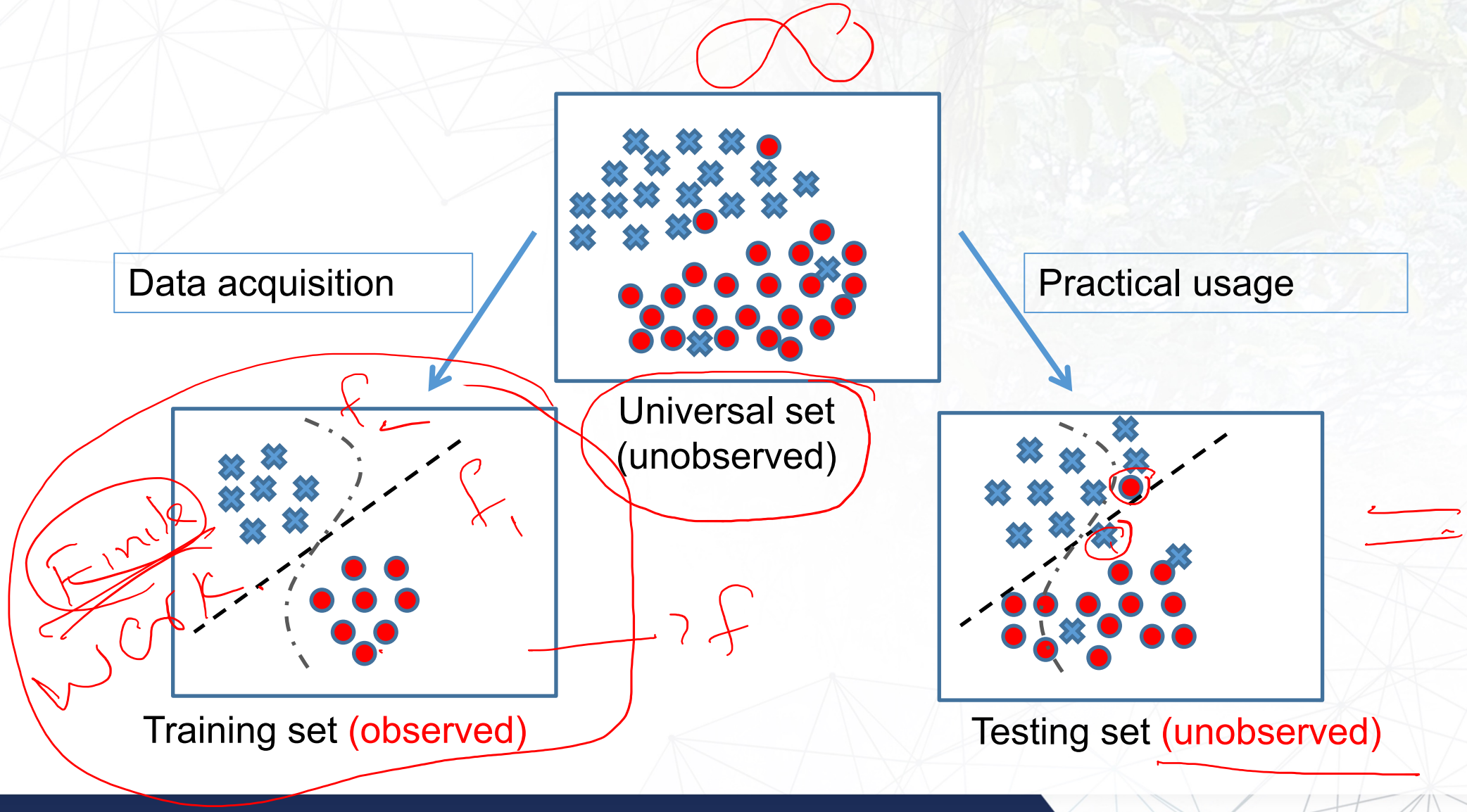
Testing



Test sample

Slide credit: D. Hoiem and L. Lazebnik

Training and Testing

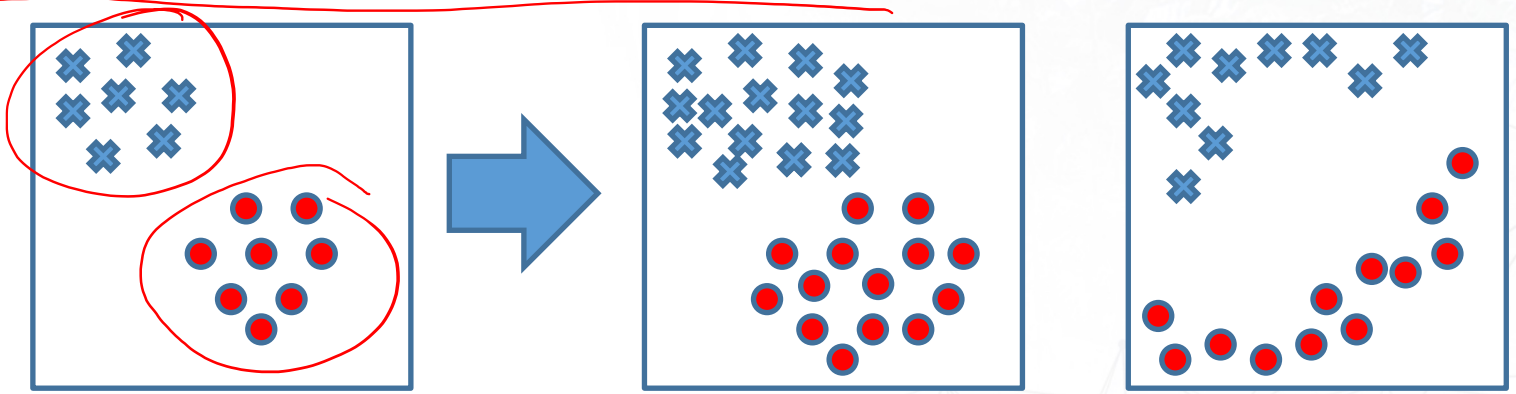


Blank Slide

Training and Testing

- Training is the process of making the system able to learn.

- Assumptions:
  - Training set and testing set come from the same distribution
  - Need to make some assumptions or bias



Two popular problem

- Classification
  - Binary, Multi-Class, 1 Vs Rest,
- Regression
  - Real number, Real vector

$\{0, 1, 2\}$

$\mathbb{R}$

"Normal" Machine Learning:

$$f : \mathcal{X} \to \mathbb{R}.$$

- inputs $\mathcal{X}$ can be any kind of objects
- output y is a real number
  - classification, regression, density estimation, ...

Structured Output Learning:

$$f : \mathcal{X} \to \mathcal{Y}.$$

Struct
arrs
graph

- inputs $\mathcal{X}$ can be any kind of objects
- outputs $y \in \mathcal{Y}$ are complex (structured) objects
  - images, text, audio, folds of a protein

Blank Slide

What is structured output prediction?

Ad hoc definition: predicting *structured* outputs from input data
(in contrast to predicting just a single number, like in classification or regression)

- Natural Language Processing:
  - Automatic Translation (output: sentences)
  - Sentence Parsing (output: parse trees)
- Bioinformatics:
  - Secondary Structure Prediction (output: bipartite graphs)
  - Enzyme Function Prediction (output: path in a tree)
- Speech Processing:
  - Automatic Transcription (output: sentences)
  - Text-to-Speech (output: audio signal)
- Robotics:
  - Planning (output: sequence of actions)

(x, y)

Space of Algorithms

- **Supervised learning** ($\{x_n \in R^d, y_n \in R\}_{n=1}^N$)
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised learning** ($\{x_n \in R^d\}_{n=1}^N$)
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-supervised learning**
  - Partly labelled; Partly unlabelled
  - Label (extra information) as pairwise constraints
- **Reinforcement learning**
  - Decision making (robot, chess machine)

"Self-Supervis"

Problem Statement: Refined

Given the training data

$$\mathcal{X}_{Tr} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$$

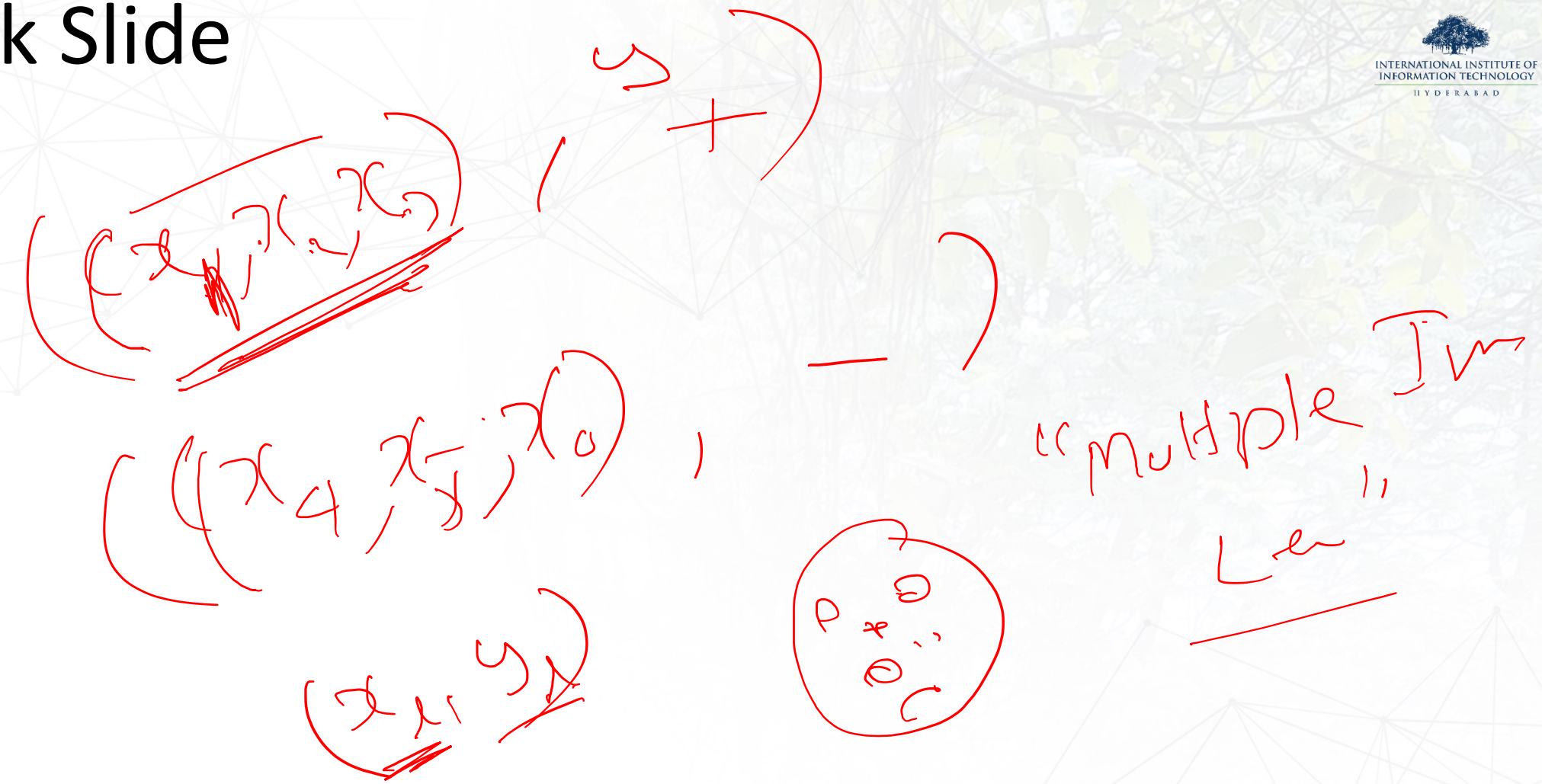
Do well on the test data

$$\mathcal{X}_{Te} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$$

with no possibility of seeing

or processing the test data

Blank Slide



What makes supervised learning difficult?

- Type of X and Y
- Nature of Function f()
- Amount of Data (more the better)
- Noise in the data/measurement X
- Noise in the tag/prediction Y
- Weaker label (a superset of Y)
- Un-Distilled X (X has unwanted stuff)
- More than one Y for an X
- Label at the level of "bags"

- Partly unlabeled data
- Transferring knowledge across tasks
- Incorporating human prior
- Challenge in optimization (eg. Non convexity)

Many of these are well studied sub-areas of ML

Thanks!!

Questions/Comments?