TOCI : DEEP LEARNING
LECTURE 2
FUNDAMENTALS OF MACHINE LEARNING

MS. FARZEEN ASHFAQ

What is learning?

- H. Simon: Any process by which a system improves its performance
- M. Minsky: Learning is making useful changes in our minds
- R. Michalsky: Learning is constructing or modifying representations of what is being experienced
- L. Valiant: Learning is the process of knowledge acquisition in the absence of explicit programming

What is Machine Learning?

- How can we solve a specific problem?
  - As computer scientists we write a program that encodes a set of rules that are useful to solve the problem
  - In many cases is very difficult to specify those rules, e.g., given a picture determine whether there is a cat in the image

What is Machine Learning?

- How can we solve a specific problem?
  - As computer scientists we write a program that encodes a set of rules that are useful to solve the problem
  - In many cases is very difficult to specify those rules, e.g., given a picture determine whether there is a cat in the image
- Learning systems are not directly programmed to solve a problem, instead develop own program based on:
  - Examples of how they should behave
  - From trial-and-error experience trying to solve the problem
- Different than standard CS:
  - Want to implement unknown function, only have access e.g., to sample input-output pairs (training examples)
- Learning simply means incorporating information from the training examples into the system

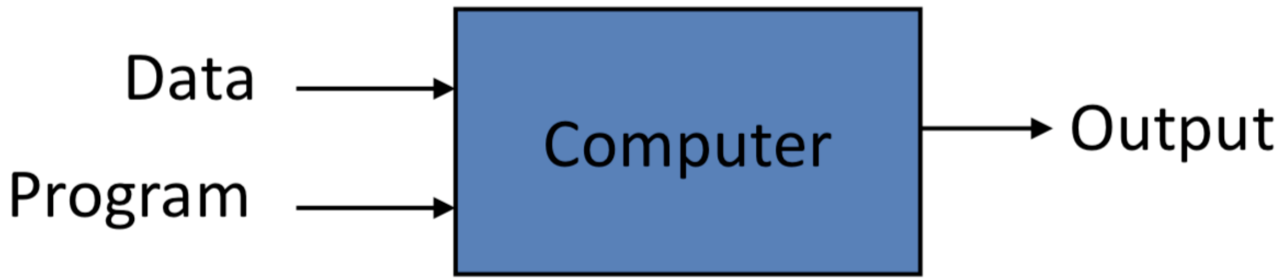
So What is ML……

Study of making machines learn a concept ***without*** having to explicitly program it.

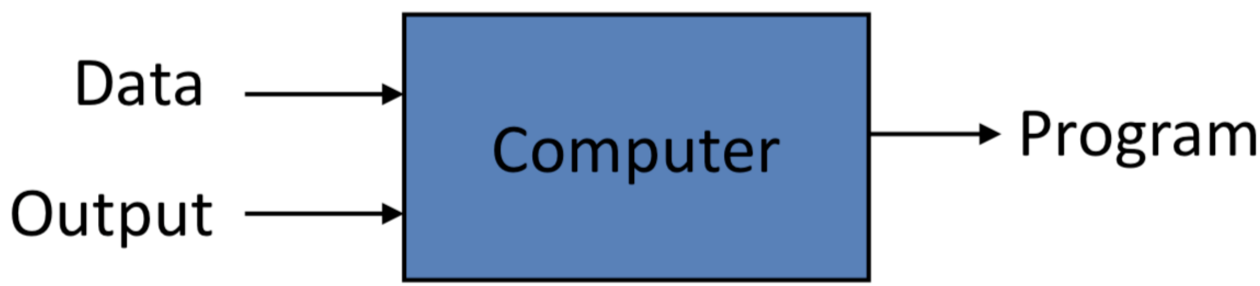
- Constructing algorithms that can:
    - learn from input data, and be able to make predictions.
    - find interesting patterns in data.
- Analyzing these algorithms to understand the limits of ‘learning’

Non-explicit programming

Traditional Programming



Machine Learning



Machine Learning models can learn by example

- Algorithms learn rules from <u>labelled examples</u>
- A set of labelled examples used for learning is called <u>training data.</u>
- The learned rules should also be able to <u>generalize</u> to correctly recognize or predict new examples not in the training set.

<u>Audio signal</u>

<u>Output text</u>

How do I get to Ann Arbor?

Hello!

Please order me a pizza.

Machine Learning models learn from experience

- **Labeled examples**
  (Email spam detection)

- **User feedback**
  (Clicks on a search page)

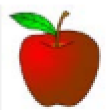
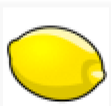
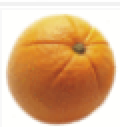
- **Surrounding environment**
  (self-driving cars)

Machine Learning brings together statistics, computer science, and more..

- **Statistical methods**
  - *Infer conclusions from data*
  - *Estimate reliability of predictions*
- **Computer science**
  - *Large-scale computing architectures*
  - *Algorithms for capturing, manipulating, indexing, combining, retrieving and performing predictions on data*
  - *Software pipelines that manage the complexity of multiple subtasks*
- **Economics, biology, psychology**
  - *How can an individual or system efficiently improve their performance in a given environment?*
  - *What is learning and how can it be optimized?*

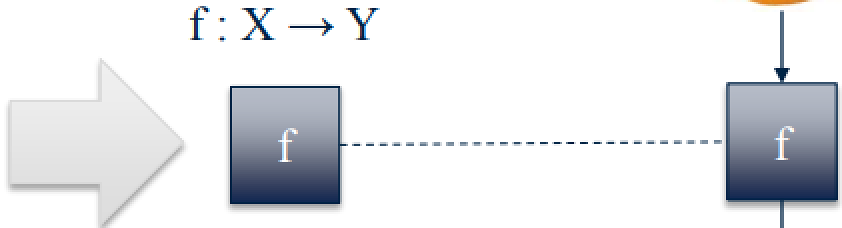
Supervised Learning (classification example)

Training set

X Sample		Y Target Value (Label)	
	x_1	Apple	y_1
	x_2	Lemon	y_2
	x_3	Apple	y_3
	x_4	Orange	y_4

Classifier

$f : X \rightarrow Y$



At training time, the classifier uses labelled examples to learn rules for recognizing each fruit type.

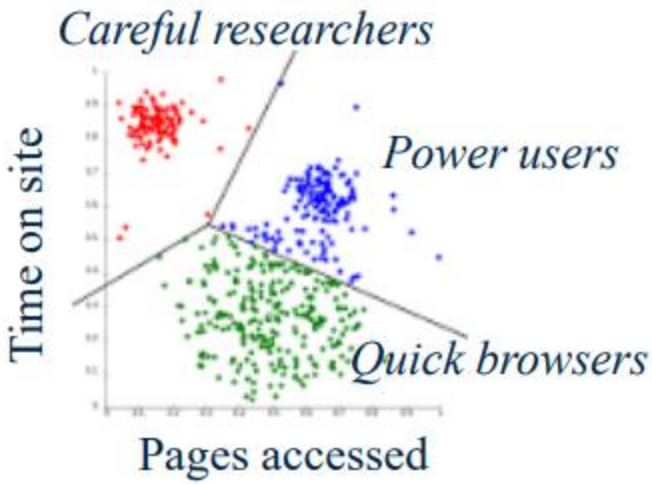
Future sample

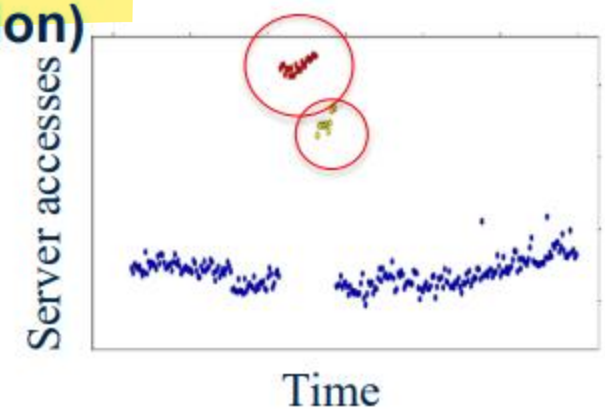
Label: Orange

After training, at prediction time, the trained model is used to predict the fruit type for new instances using the learned rules.

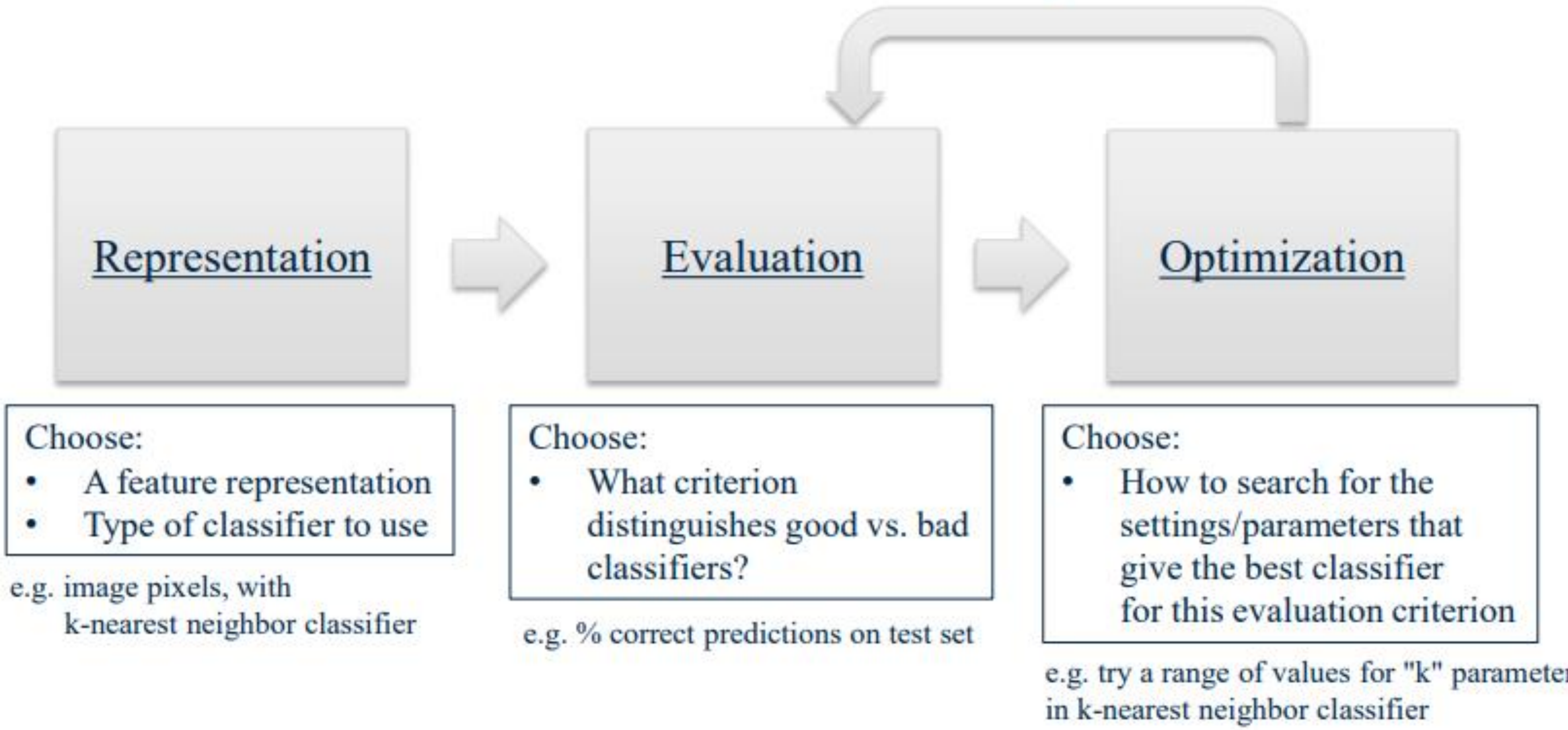
Unsupervised learning: finding useful structure or knowledge in data when no labels are available

- Finding clusters of similar users (clustering)
- Detecting abnormal server access patterns (unsupervised outlier detection)





A Basic Machine Learning Workflow



Feature Representations

Feature representation

Email

```
To:  Chris Brooks
From:  Daniel Romero
Subject:  Next course offering
Hi Daniel,
Could you please send the outline for the
next course offering?  Thanks! -- Chris
```

Feature	Count
to	1
chris	2
brooks	1
from	1
daniel	2
romero	1
the	2
...	

A list of words with their frequency counts

Picture

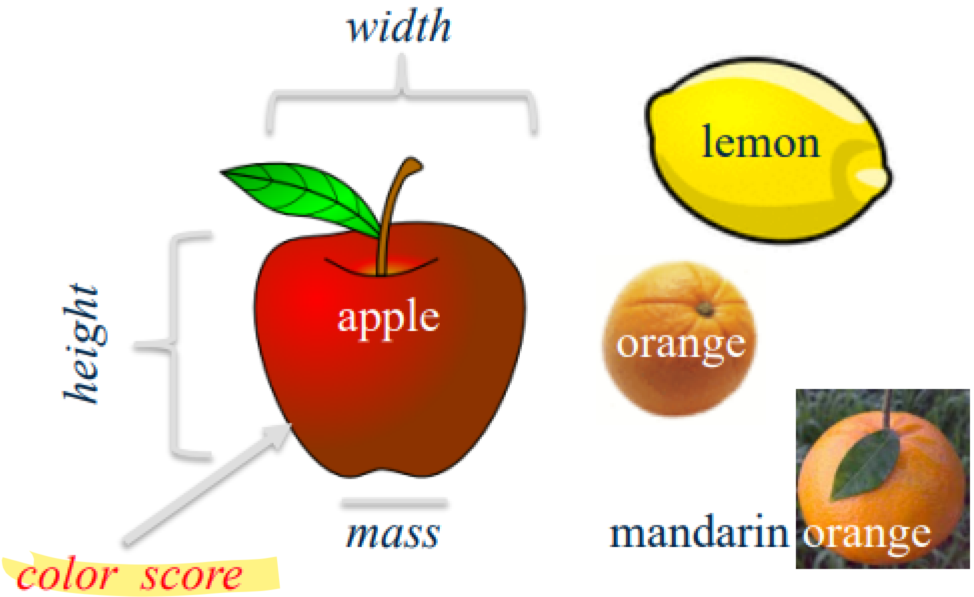
A matrix of color values (pixels)

Sea Creatures

Feature	Value
DorsalFin	Yes
MainColor	Orange
Stripes	Yes
StripeColor1	White
StripeColor2	Black
Length	4.3 cm

A set of attribute values

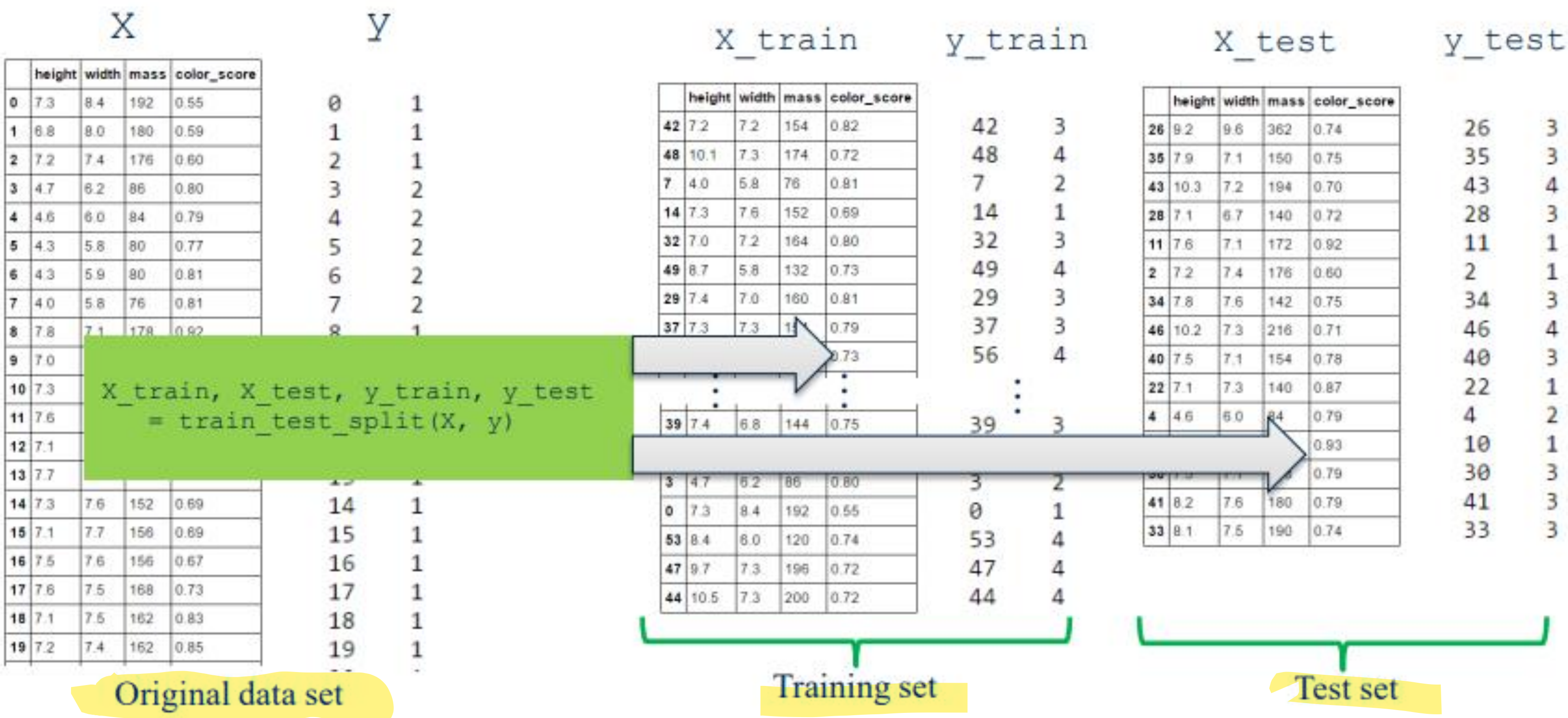
The Fruit Dataset



	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
0	1	apple	granny_smith	192	8.4	7.3	0.55
1	1	apple	granny_smith	180	8.0	6.8	0.59
2	1	apple	granny_smith	176	7.4	7.2	0.60
3	2	mandarin	mandarin	86	6.2	4.7	0.80
4	2	mandarin	mandarin	84	6.0	4.6	0.79
5	2	mandarin	mandarin	80	5.8	4.3	0.77
6	2	mandarin	mandarin	80	5.9	4.3	0.81
7	2	mandarin	mandarin	76	5.8	4.0	0.81
8	1	apple	braeburn	178	7.1	7.8	0.92
9	1	apple	braeburn	172	7.4	7.0	0.89
10	1	apple	braeburn	166	6.9	7.3	0.93
11	1	apple	braeburn	172	7.1	7.6	0.92
12	1	apple	braeburn	154	7.0	7.1	0.88
13	1	apple	golden_delicious	164	7.3	7.7	0.70
14	1	apple	golden_delicious	152	7.6	7.3	0.69
15	1	apple	golden_delicious	156	7.7	7.1	0.69
16	1	apple	golden_delicious	156	7.6	7.5	0.67

`fruit_data_with_colors.txt`

Creating Training and Testing Sets



Some reasons why looking at the data initially is important

Examples of incorrect or missing feature values

- Inspecting feature values may help identify what cleaning or preprocessing still needs to be done once you can see the range or distribution of values that is typical for each attribute.
- You might notice missing or noisy data, or inconsistencies such as the wrong data type being used for a column, incorrect units of measurements for a particular column, or that there aren't enough examples of a particular class.
- You may realize that your problem is actually solvable without machine learning.

	fruit_label	fruit_name	fruit_subtype	mass	width	height	color_score
0	1	apple	granny_smith	192	8.4	7.3	0.55
1	1	apple	granny_smith	180	8.0	6.8	0.59
2	1	apple	granny_smith	176	7.4	7.2	192
3	2	mandarin	mandarin	86	6.2	4.7	0.80
4	2	mandarin	mandarin	84	6.0	4.6	0.79
5	2	mandarin	apple	80	5.8	4.3	0.77
6	2	mandarin	mandarin	80	5.9	4.3	0.81
7	2	mandarin	mandarin	76	5.8	4.0	0.81
8	1	apple	braeburn	178	7.1	7.8	0.92
9	1	apple	braeburn		7.4	7.0	0.89
10	1	apple	braeburn		6.9	7.3	0.93
11	1	apple	braeburn		7.1	7.6	0.92
12	1	apple	braeburn		7.0	7.1	0.88
13	1	apple	golden_delicious	[illegible]	7.3	7.7	0.70
14	1	apple	golden_delicious	152	7.6	7.3	0.69