

1

1

LMS Enrollment Code

- Enrollment Code :
 - 426708913

2

2

WhatsApp Group

- Follow this link to join WhatsApp group MachineLearning2025:
- <https://chat.whatsapp.com/GfU7KrUxXDI332yysFEp9Q>



3

Announcements

- Assignment 1 is online
 - Deadline: **24 Sep 2025 at 2359**



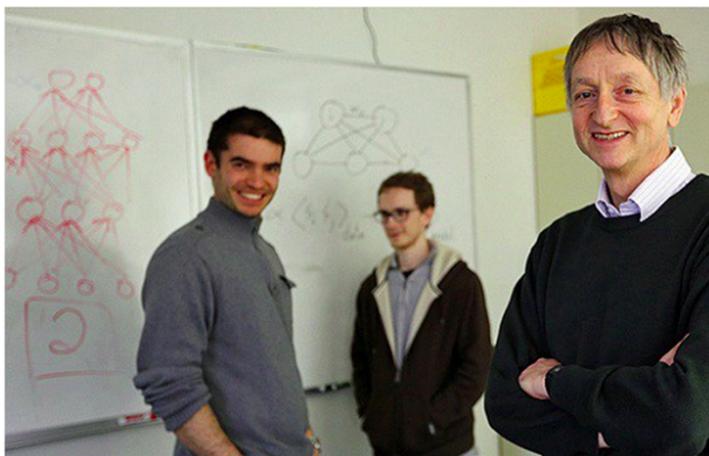
4

Outline

- Types of Machine Learning (ML)
- Example:
 - A ML Problem
- A Typical ML system design cycle/ Pipeline
- Different Types of Learning
- Reading Work !!!
- ChatGPT/Python Demo



Dedication



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.toronto.edu

Ilya Sutskever
University of Toronto
ilya@cs.toronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.toronto.ca

Abstract

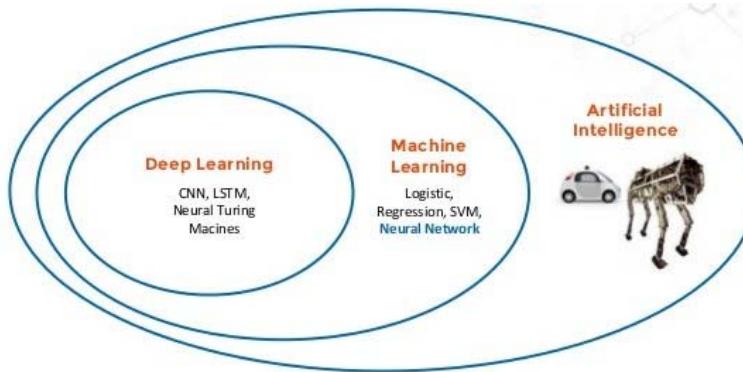
We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different categories. On our data, we achieve top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than previous work. The network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers. We then use three fully-connected layers to obtain 1000 outputs. To train faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers, we employed a rectified linear unit (ReLU) activation function called “deep” that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

1 Introduction

Current approaches to object recognition make essential use of machine learning methods. To improve their performance, we can collect larger datasets, learn more powerful models, and use better techniques for preventing overfitting. Until recently, datasets of labeled images were relatively small — on the order of tens of thousands of images (e.g., NORB [16], Caltech-101/256 [8, 9], and



AI vs ML vs DL



Definition of Machine Learning

- Machine Learning is the science (and art) of programming computers so they can *learn from data*.
- *[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.*

—Arthur Samuel, 1959

- *[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.*

—Arthur Samuel, 1959



Definition of Machine Learning

- A computer program is said to learn from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .

-- Tom Mitchell, 1997



What is Artificial Intelligence

- Field of study which studies how to create computing systems that are capable of intelligent behavior
- FUN Fact:
 - Fun Fact about AI is that after a problem is fully solved it's not called intelligent anymore... (ie: Make a computer play chess was the highest display of intelligence, now people don't consider that anymore)



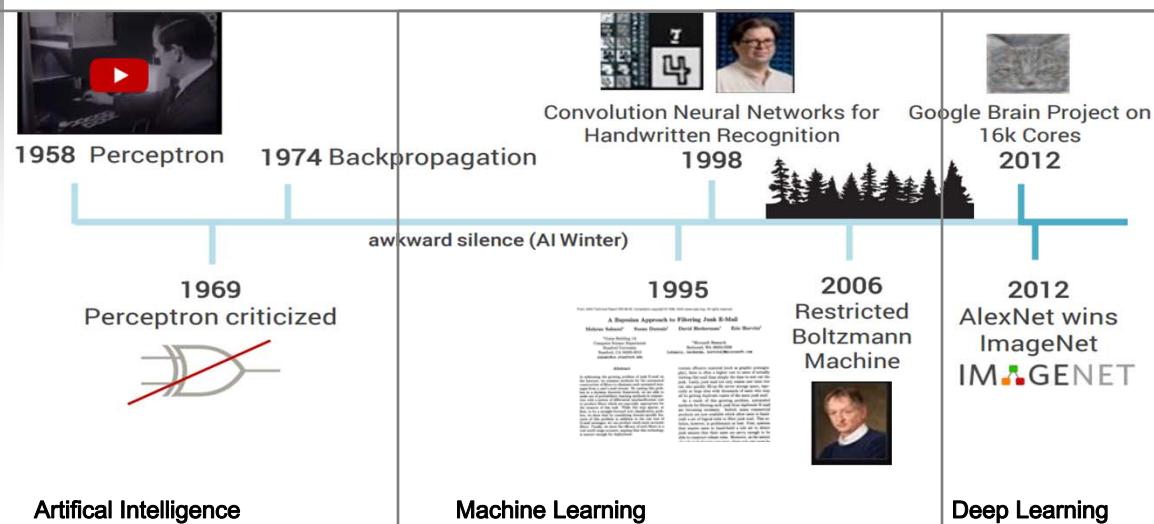
Machine Learning

- Machine learning (ML) is a **field of study** in artificial intelligence concerned with the development and study of **statistical algorithms** that can learn from data and generalize to unseen data and thus perform tasks without explicit instructions

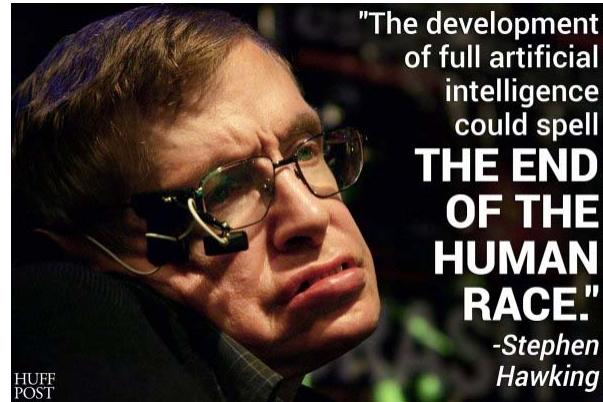
https://en.wikipedia.org/wiki/Machine_learning



History



Mix of hype/fear



Dr. Ali Hassan, DCSE, College of E&ME, NUST

13

13

sky news

17 Sep 21° 15° Watch Live

Home UK Politics World Israel-Hamas War US Climate Science & Tech Business Ents & Arts More

All-in-one Solution Built for Digital Goods Sales SIGN-UP NOW

Elon Musk tells Sky News AI is a 'risk' to humanity

The SpaceX, Tesla, and X owner is among more than 100 major figures at the UK's landmark safety summit on AI, alongside the likes of OpenAI's Sam Altman, Google DeepMind's Demis Hassabis, and US vice president Kamala Harris.

By Tom Acres, technology reporter

© Wednesday 1 November 2023 17:07, UK

Play Video – Elon Musk: 'AI is a risk'



Dr. Ali Hassan, DCSE, College of E&ME, NUST

14

14

The screenshot shows a news article from Business Insider. The title reads "'Godfather of AI' says he's 'glad' to be 77 because the tech probably won't take over the world in his lifetime". The author is Effie Webb. Below the title is a large photo of Geoffrey Hinton, an elderly man with white hair, wearing a dark suit and tie, gesturing with his hands while speaking. A caption below the photo states: "Geoffrey Hinton gave a 'sort of 10 to 20% chance' that AI systems could one day seize control. PONTUS LUNDH/TT NEWS AGENCY/AFP via Getty Images". To the right of the main content is a vertical advertisement for Mobileye, featuring the text "Mobileye Driving" and an image of a car's internal components.

15

Why Now

- The AI/ ML algorithms (Even the deep ones) are there for decades so why we have now this buzzword?
- Due to the advance of **computing power through** (**GPUs**, multi-core CPU systems, and **FPGAs**) and the availability of data (**Big data**) through internet.



16

Why Now

- Also the **amount of data** that need to be classified nowadays become too big to be handled manually,
- So big companies Google, Microsoft, Facebook, start to invest heavily on the **data**.
- **Data is the new gold rush**



Three Driving Factors...

Big Data Availability

facebook

350 millions images uploaded per day

Walmart

2.5 Petabytes of customer data hourly

YouTube

100 hours of video uploaded every minute

New ML Techniques

Deep Neural Networks

Compute Density

GPUs

ML systems extract value from Big Data



Batch Size	Training Time CPU	Training Time GPU	GPU Speed Up
64 images	64 s	7.5 s	8.5X
128 images	124 s	14.5 s	8.5X
256 images	257 s	28.5 s	9.0X

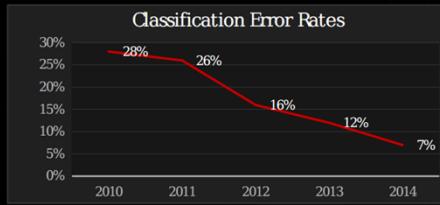
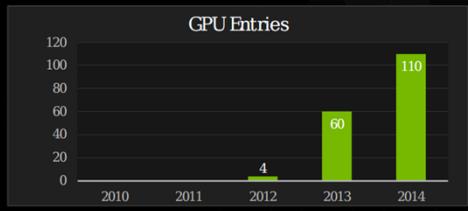


The new Hype

- The last years AI/ML models are **surpassing human intelligence**
 - Speech and natural language processing
 - Face Recognition
 - Image Classification, object detection
 - Self Driving Car
 - Playing complex games (Alpha Go)
 - Control strategies (Control engineering)
 - Live translation



Examples ...



Computing power Comparison

Type	Name	Flops	Cost
Mobile	Raspberry Pi 1 st Gen, 700 Mhz	0,04 Gflops	\$35
Mobile	Apple A8	1,4 Gflops	\$700 (in iPhone 6)
CPU	Intel Core i7-4930K (Ivy Bridge), 3.7 GHz	140 Gflops	\$700
CPU	Intel Core i7-5960X (Haswell), 3.0 GHz	350 Gflops	\$1300
GPU	NVidia GTX 980	4612 Gflops (single precision), 144 Gflops (double precision)	\$600 + cost of PC (~\$1000)
GPU	NVidia Tesla K80	8740 Gflops (single precision), 2910 Gflops (double precision)	\$4500 + cost of PC (~1500)



Deepmind hardware

- This picture is the hardware used to play against one of the best Go players in the world.



What is Machine Learning

- Machine learning is all about using your computers to "learn" how to deal with problems **without "explicit programming"**
- We need ML on cases that would be difficult to program by hand all possible variants of a classification/prediction problem
- **The basic Idea of ML is to make the computer learn something from the data**

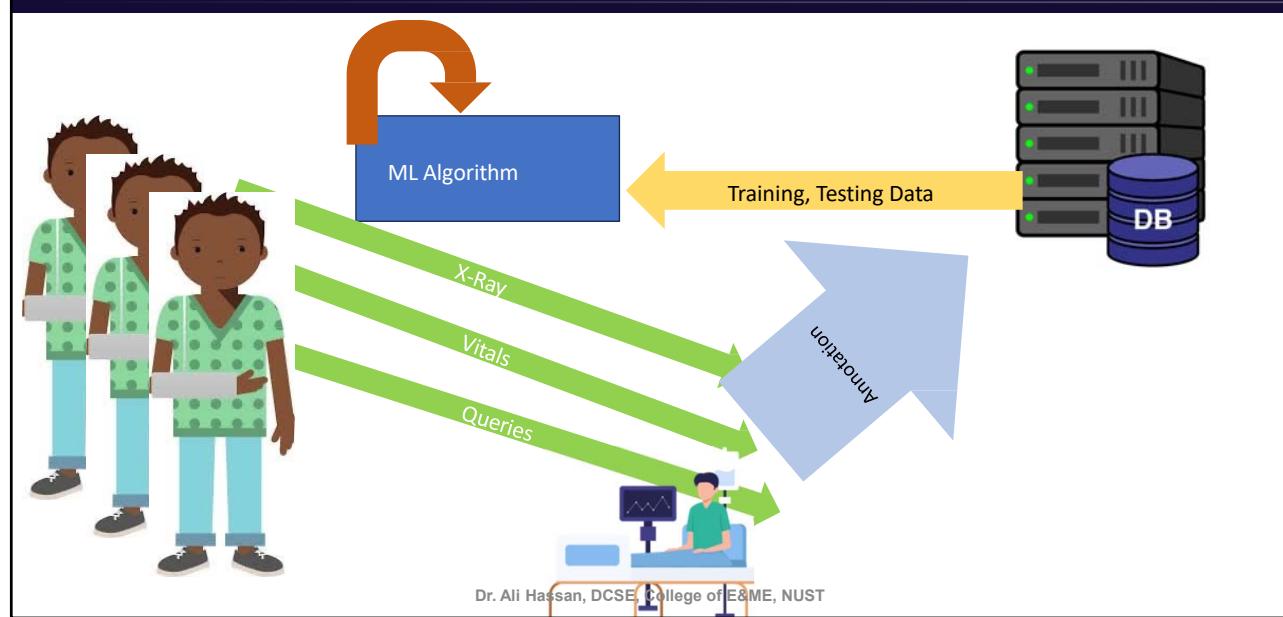


What is Machine Learning ...

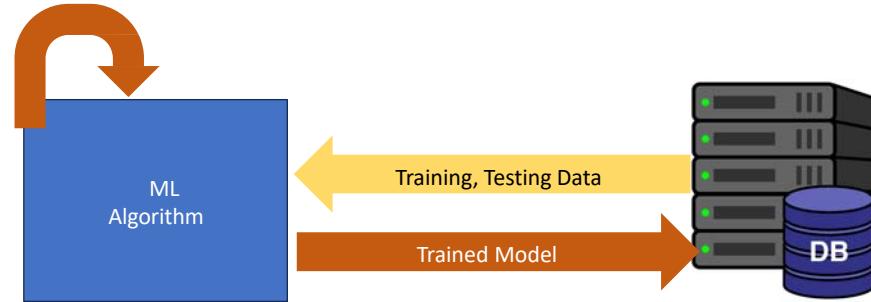
- ML is a way to make the computers create a program that gives some output with a known input and that latter give an intelligent output to a different but similar input.
 - Take some data,
 - Train a model on that data, and
 - Use the trained model to make predictions on new (UNSEEN) data.



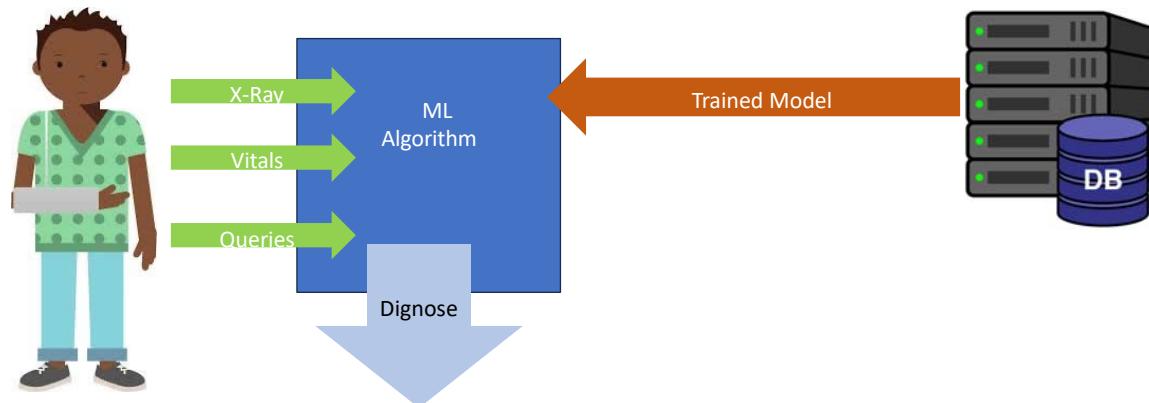
Machine Learning

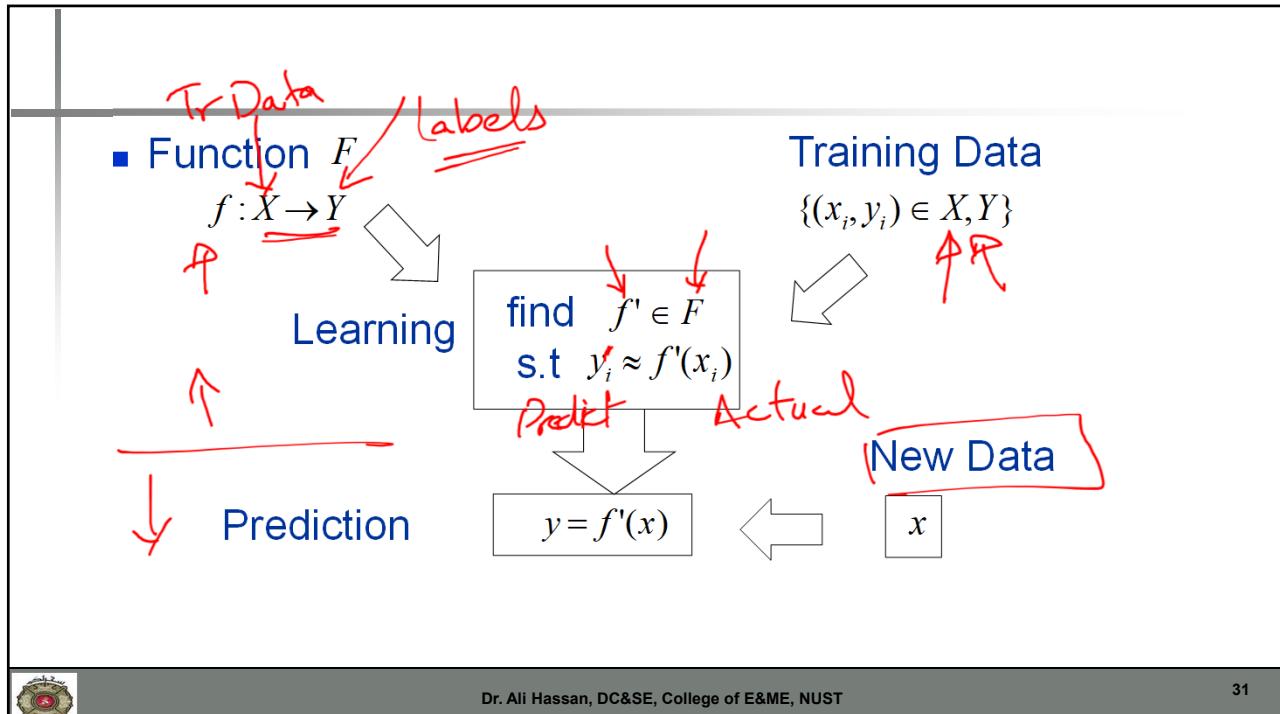


Machine Learning Training



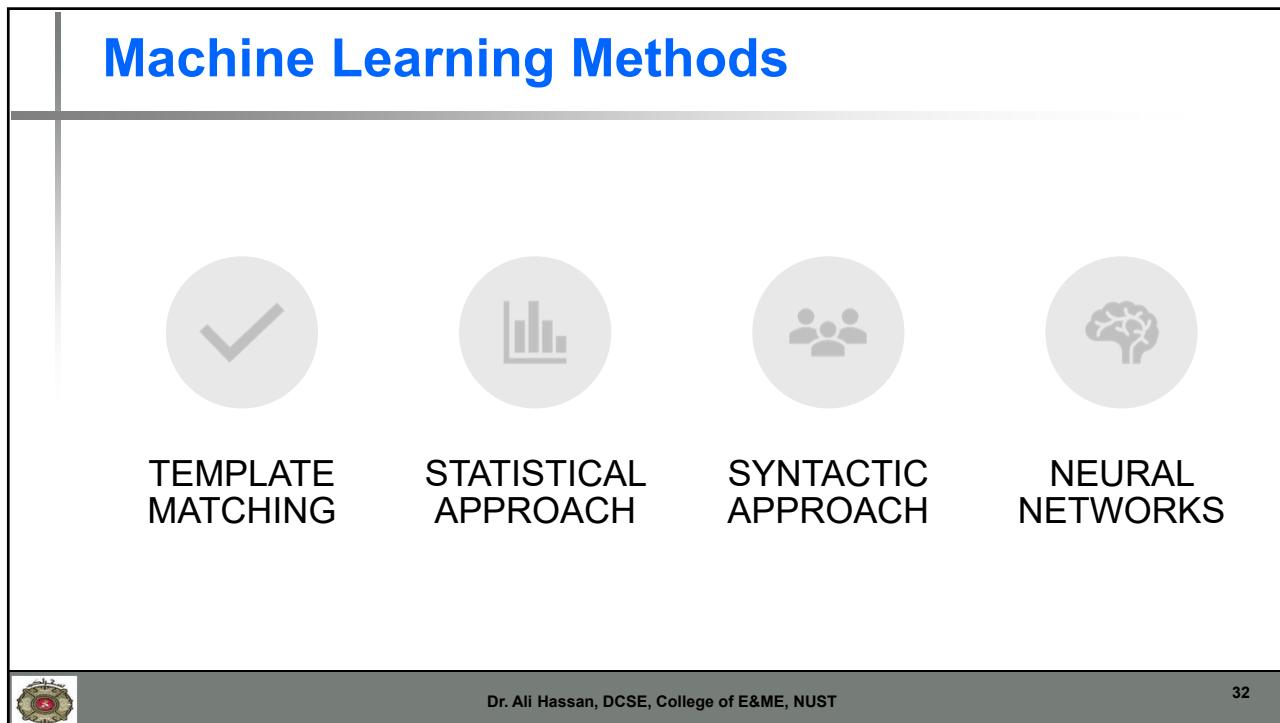
Machine Learning -- Testing





31

31



32

32

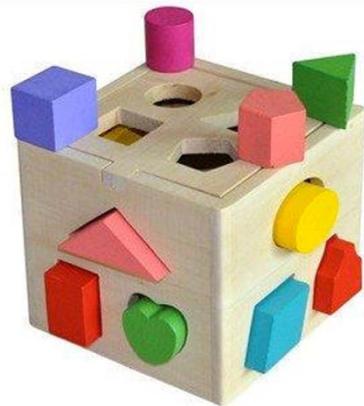
TEMPLATE MATCHING



Template Matching

- A template (typically a 2D shape) or a prototype of the pattern to be recognized is available.
- Compute the similarity between the template and the pattern to be matched.
- Take into account pose (rotation, translation) and scale changes.





A typical template matching problem



Issues of concern

- Take an example of an image of size 28x28 pixels.
- You have to use all $28 \times 28 = 784$ pixels
- Computational complexity
- **Rigidity** assumption (use deformable template models)
- Choice of template



STATISTICAL APPROACH



Statistical Approach

- Each pattern is represented in terms of d- features, and is viewed as a point in a d-dimensional space
- The goal is to choose those features that allow pattern vectors belonging to **different categories** to occupy **compact and disjoint regions** in a d-dimensional feature space.

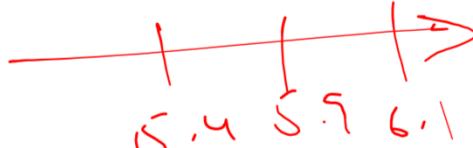


$$\tilde{x}^{(1)} = [\text{height}] \quad X = \begin{bmatrix} 5.4 \\ 5.9 \\ 6.1 \end{bmatrix}$$

$$\tilde{x}^{(2)} = 5.4$$

$$\tilde{x}^{(3)} = 5.9$$

$$\tilde{x}^{(4)} = 6.1$$

+ 

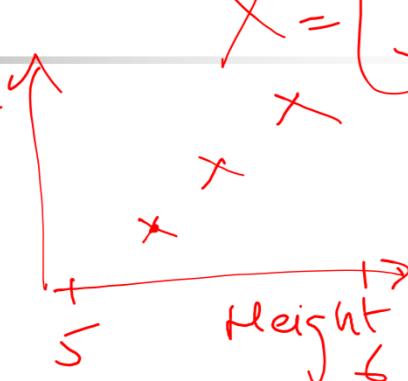


$$\tilde{x}^{(1)} = \begin{bmatrix} \text{Height} \\ \text{weight} \end{bmatrix}$$

$$\tilde{x}^{(2)} = \begin{bmatrix} 5.4 \\ 60 \end{bmatrix}$$

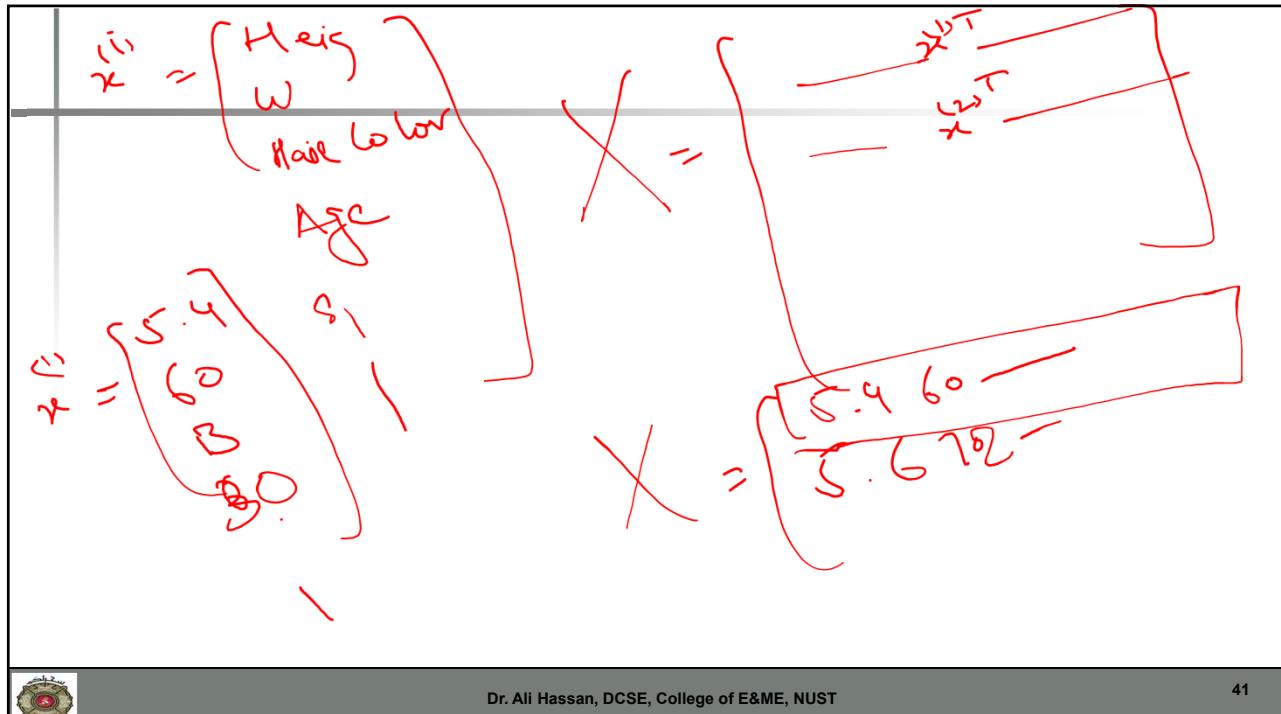
$$\tilde{x}^{(3)} = \begin{bmatrix} 5.9 \\ 72 \end{bmatrix}$$

$$\tilde{x}^{(4)} = \begin{bmatrix} 6.1 \\ 90 \end{bmatrix}$$

+ 

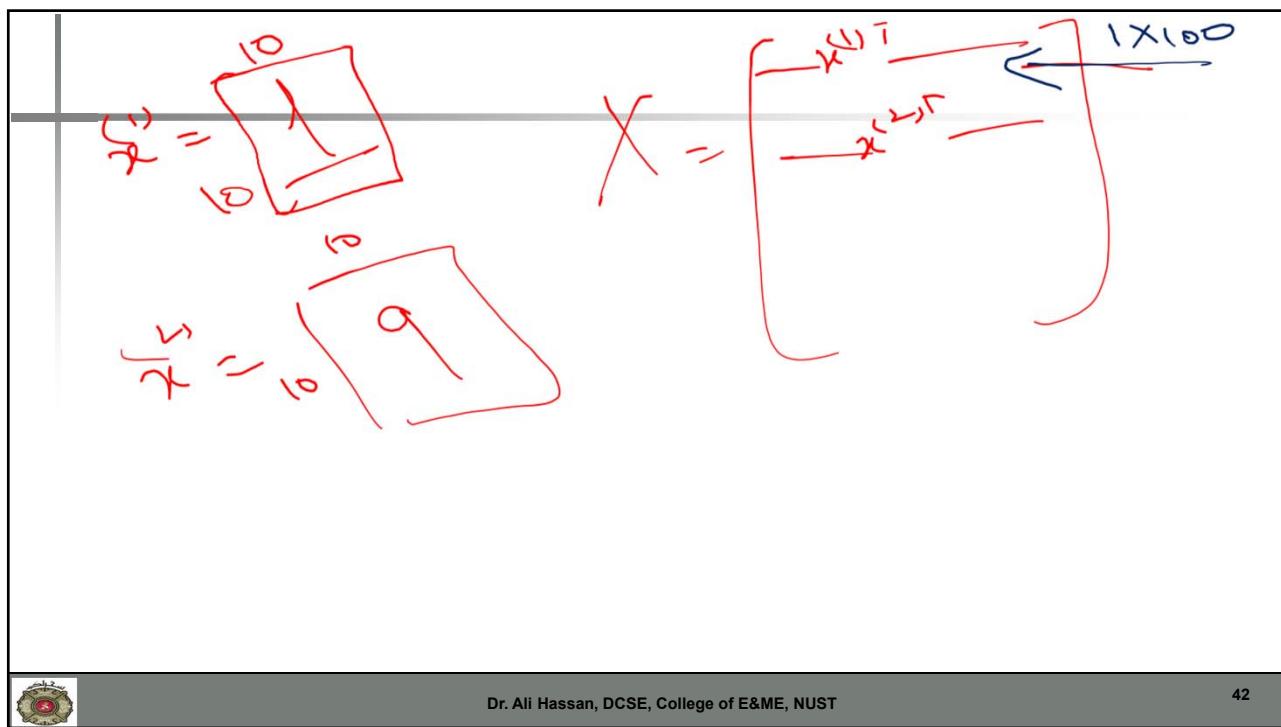
$$X = \begin{bmatrix} 5.4 & 60 \\ 5.9 & 72 \\ 6.1 & 90 \end{bmatrix}$$





41

41



42

42

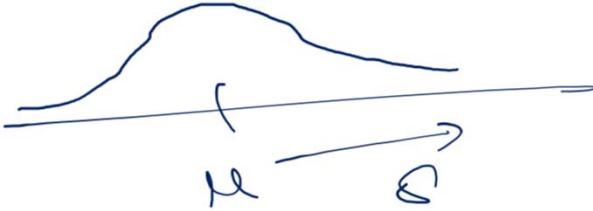
Statistical	Structural
Number of segments: 4 Number of horizontal segments : 2 Number of vertical segments: 2 Number of diagonal segments : 0	 <div style="display: flex; justify-content: space-around;"> horizontal vertical </div> <div style="display: flex; justify-content: space-around;"> vertical horizontal </div>
Number of segments: 3 Number of horizontal segments : 1 Number of vertical segments: 0 Number of diagonal segments : 2	 <div style="display: flex; justify-content: space-around;"> diagonal diagonal </div> <div style="display: flex; justify-content: space-around;"> horizontal </div>

 Dr. Ali Hassan, DCSE, College of E&ME, NUST 43

43

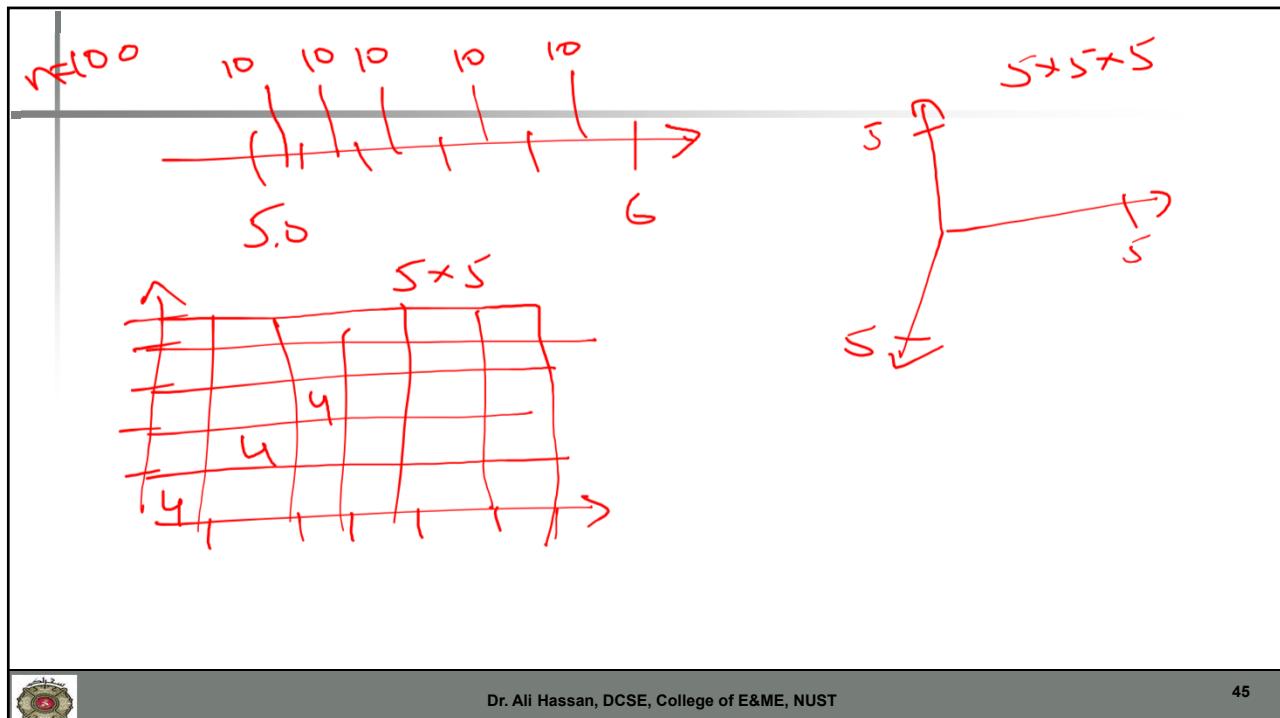
Issues of concern

- Usually $d \ll D$ (dimensions of image/signal)
- What should be the value of d ??
- We might encounter **curse of dimensionality**



 Dr. Ali Hassan, DCSE, College of E&ME, NUST 44

44



45

Dr. Ali Hassan, DCSE, College of E&ME, NUST

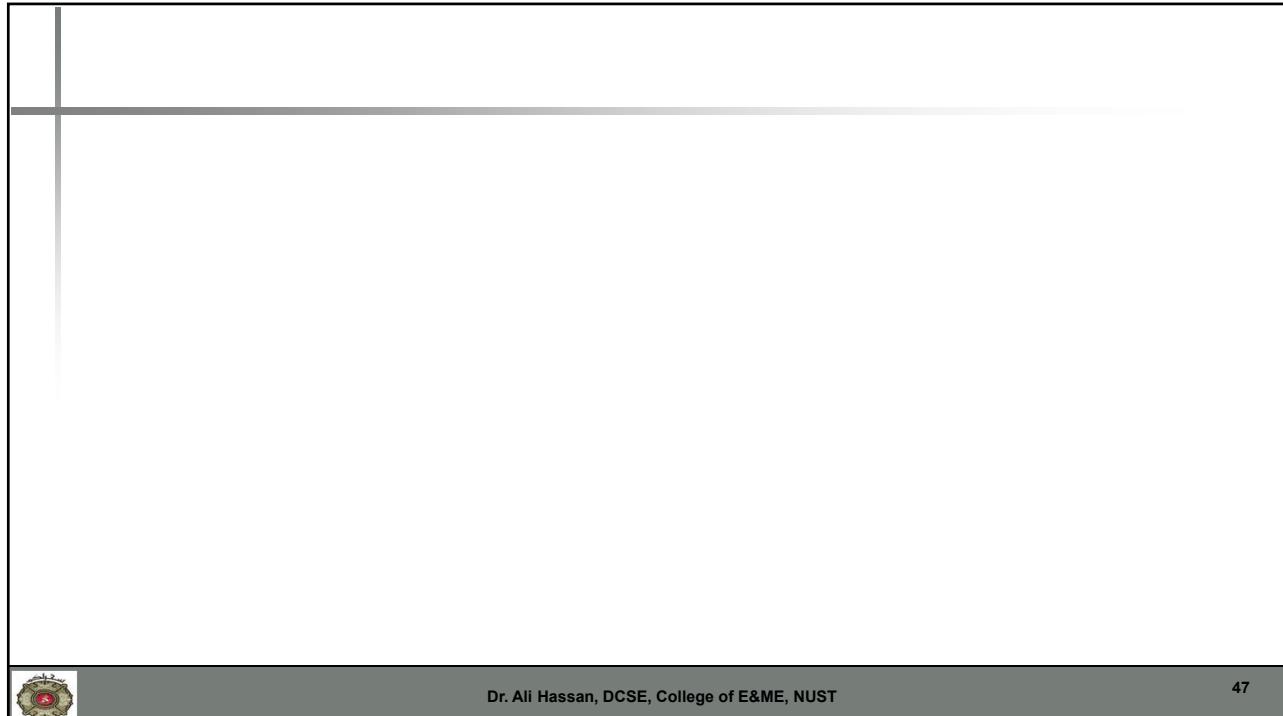
45



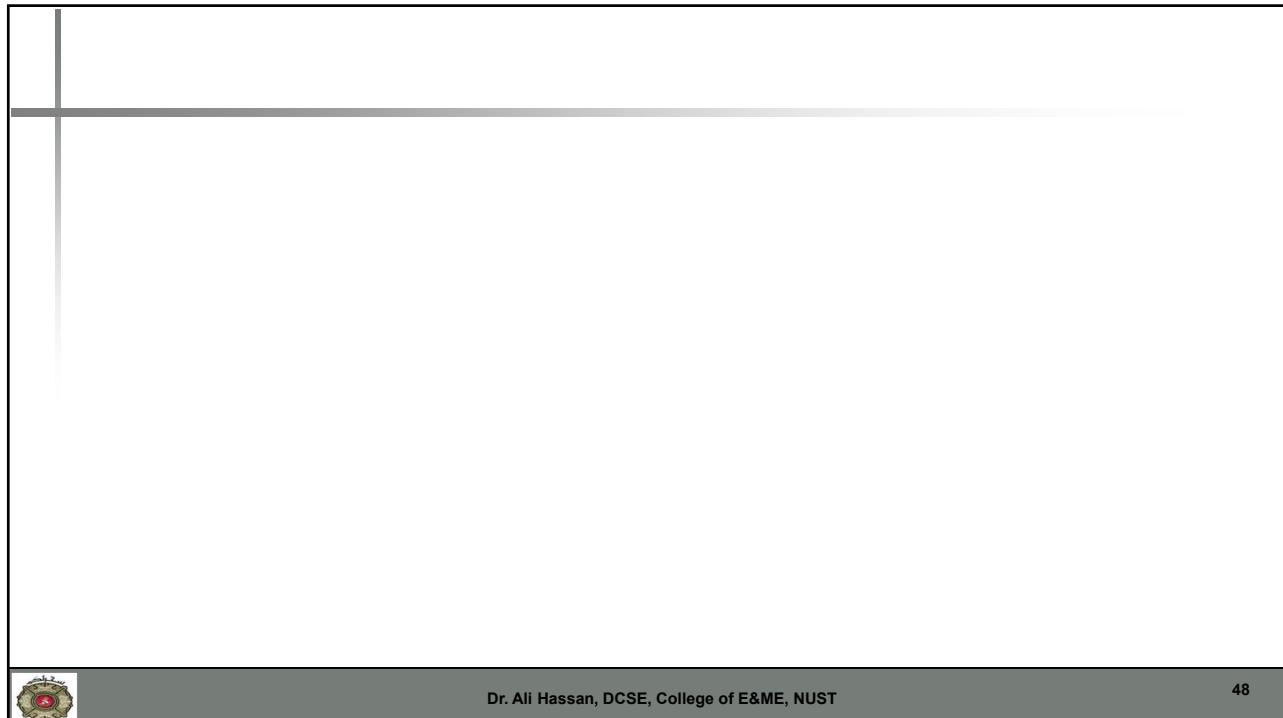
Dr. Ali Hassan, DCSE, College of E&ME, NUST

46

46



47

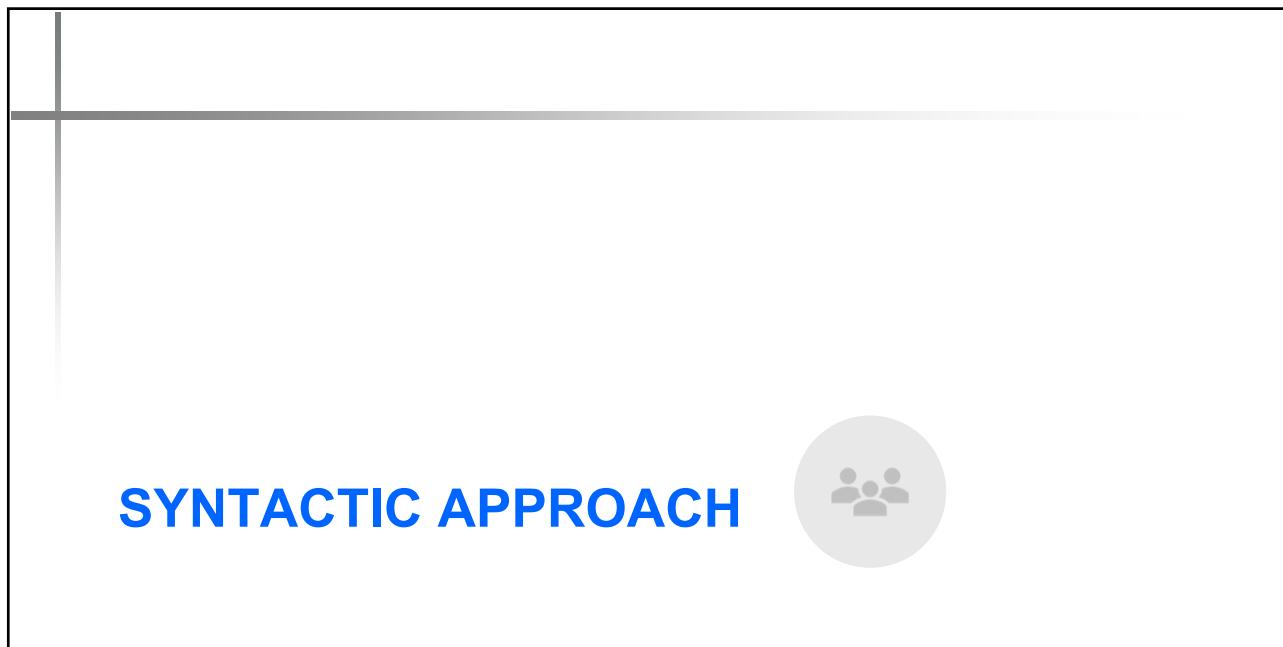


48



 Dr. Ali Hassan, DCSE, College of E&ME, NUST 49

49



 Dr. Ali Hassan, DCSE, College of E&ME, NUST 50

50

Syntactic Approach

- Use hierarchical structures to represent complex patterns.
- The simplest unit is called: *primitives*
- Complex pattern is represented in terms of the interrelationships (*grammars*) between the primitives.
- Learn Grammatical rules from Data
- Trend analysis using tweets
 - NLP, NLG



Issues of concern

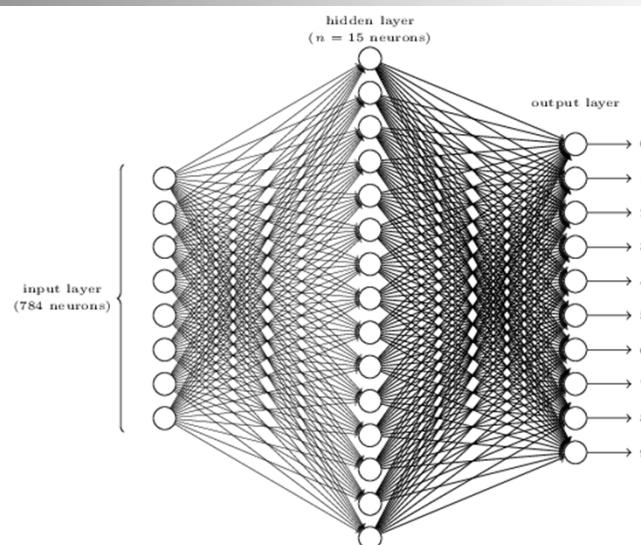
- However, it's usually difficult to segment noisy patterns and infer grammar from the training set.
- May yield a combinatorial explosions of possibilities to be investigated.



NN BASED APPROACHES



Neural Networks

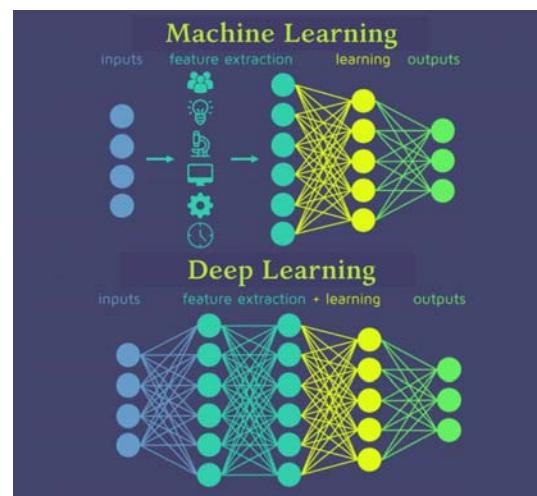


Neural Networks

- Massively parallel computing unit consisting of an extremely large number of simple processors with many interconnections.
- Can learn **complex non-linear** input-output relationships.
- Feed-forward networks such as multilayer perceptron and Radial Basis Function network are useful for pattern classification.



Machine vs Deep Learning



Spam Filter

MACHINE LEARNING EXAMPLE # 1

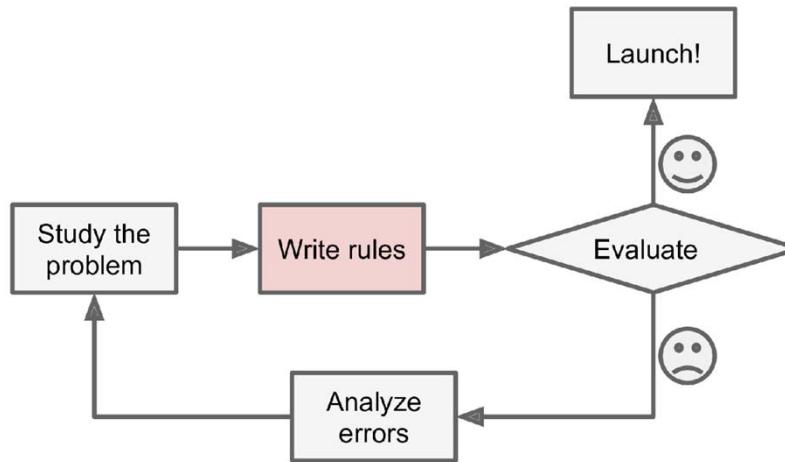


What is spam filter

- Your spam filter is a Machine Learning program that, given examples of spam emails (e.g., flagged by users) and examples of regular (nonspam, also called “ham”) emails, can learn to flag spam
- How will you go about it?



Traditional Approach



Rules

- Consider what spam typically looks like. (such as “4U,” “credit card,” “free,” and “amazing”)
- You would write a detection algorithm for each of the patterns that you noticed, and your program would flag emails as spam
- You would test your program and repeat steps 1 and 2 until it was good enough to launch.
- Any Issues ??

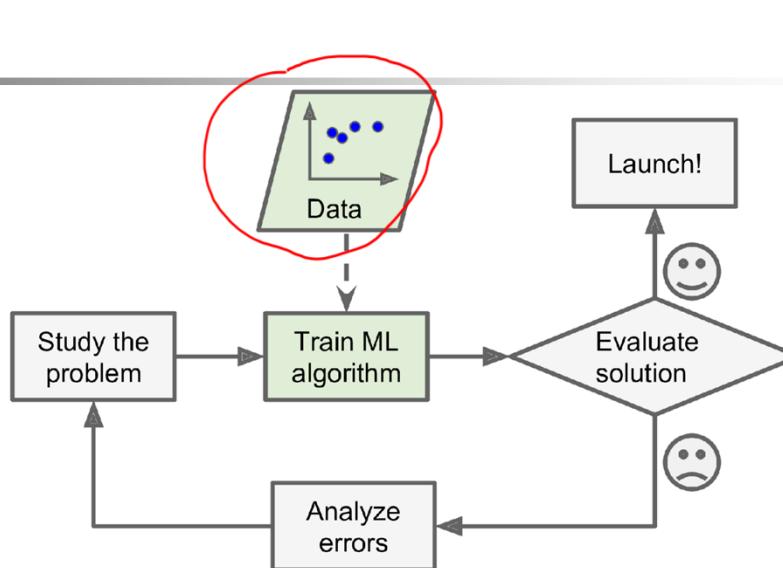


Issue

- What if spammers notice that all their emails containing “4U” are blocked? They might start writing “For U” instead

Solution

- Learn from Data
- A spam filter based on ML techniques automatically notices that “For U” has become unusually frequent in spam flagged by users, and it starts flagging them



Separating Salmon from Sea Bass

MACHINE LEARNING EXAMPLE # 2

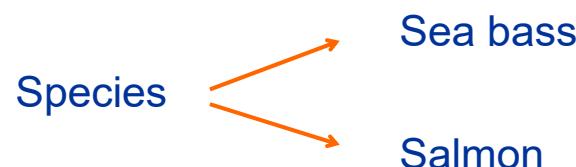


63

63

An Example

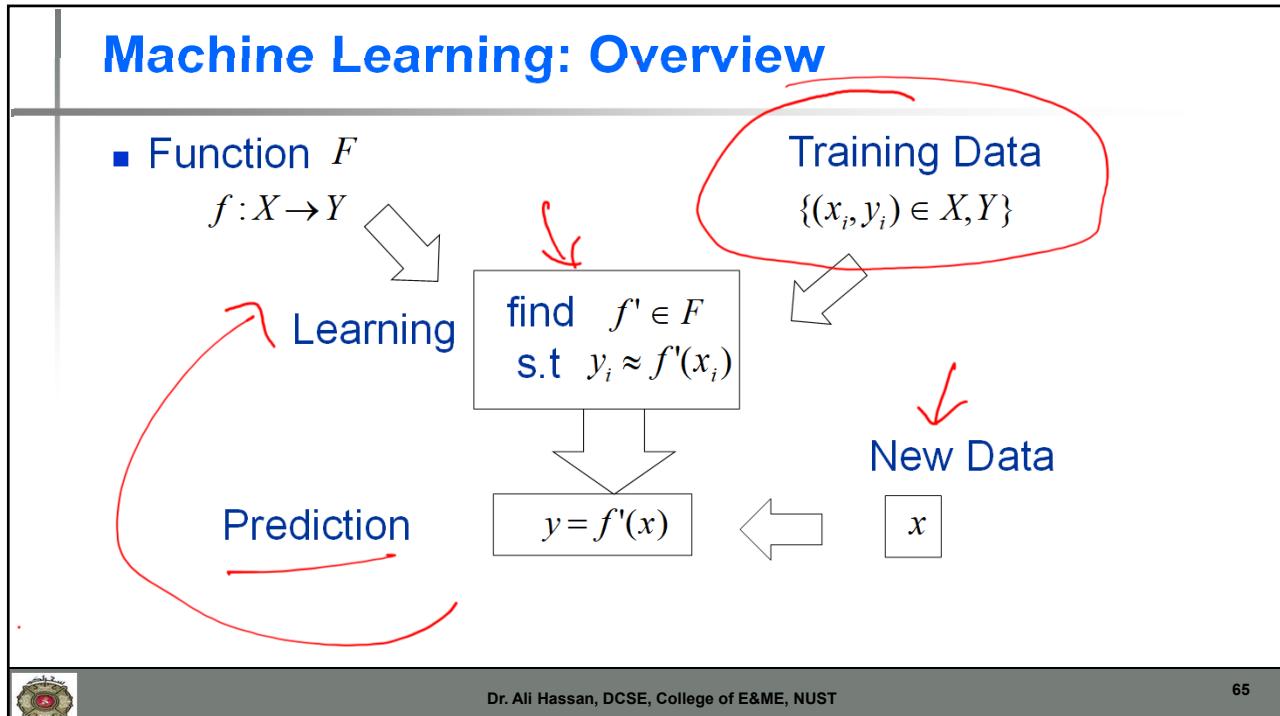
- “Sorting incoming Fish on a conveyor according to species using optical sensing”



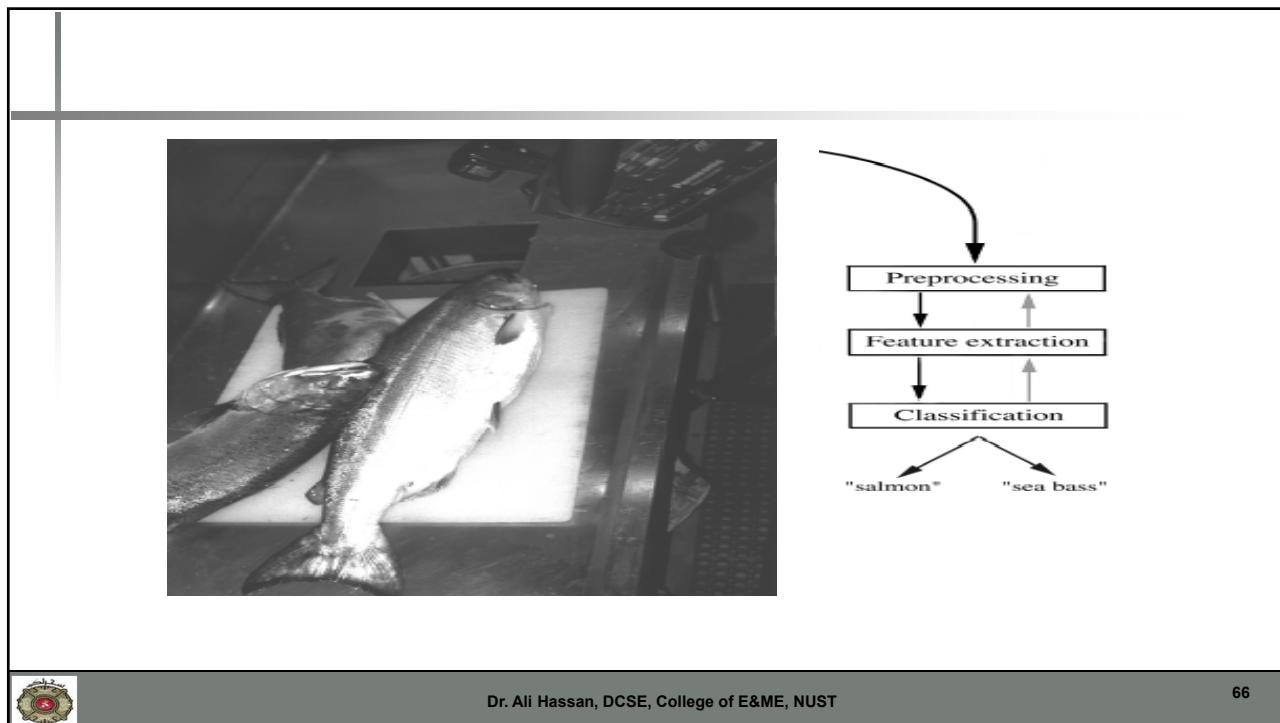
Dr. Ali Hassan, DCSE, College of E&ME, NUST

64

64



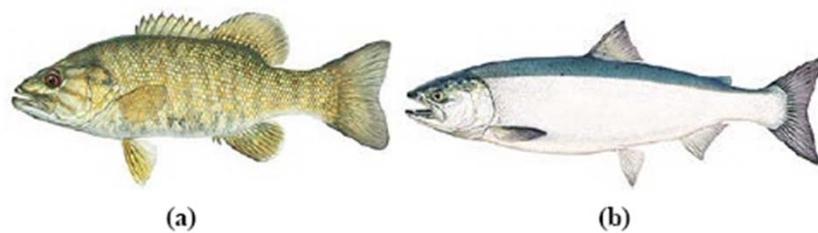
65



66

Pre-Processing

- Use a **segmentation** operation to isolate fishes from one another and from the background
- Information from a single fish is sent to a **feature extractor** whose purpose is to reduce the data by measuring certain features
- The features are passed to a **classifier** or **decision maker**



The objects to be classified; (a). Sea bass,
and (b). salmon.



Problem Analysis

- Set up a camera and take some sample images to extract features
 - Length
 - Lightness
 - Width
 - Number and shape of fins
 - Position of the mouth, etc...
- This is the set of all suggested features to explore for use in our classifier!



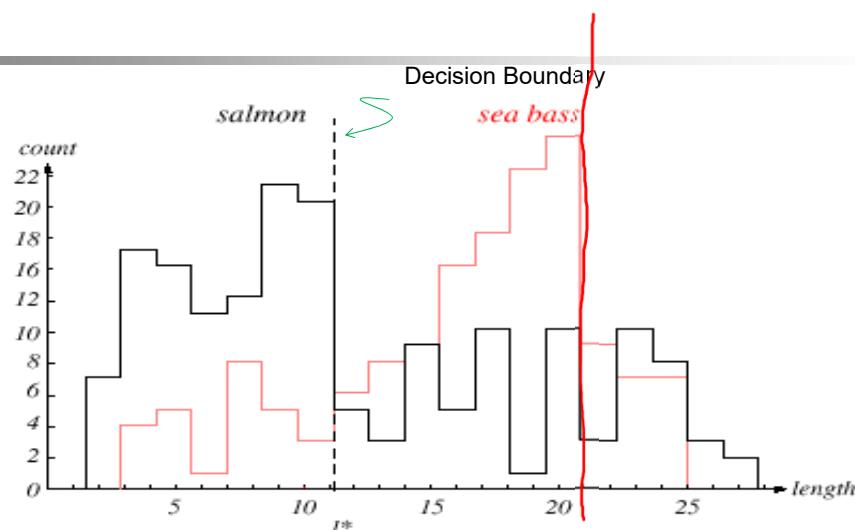
System Design Process

- Domain knowledge:
 - a sea bass is generally longer than a salmon
- Selected Feature
 - Select the **length** of the fish as a possible feature for discrimination
 $x = [x_1]$, where x_1 is length
- Model:
 - Sea bass have some typical length, and this is greater than the length of a salmon



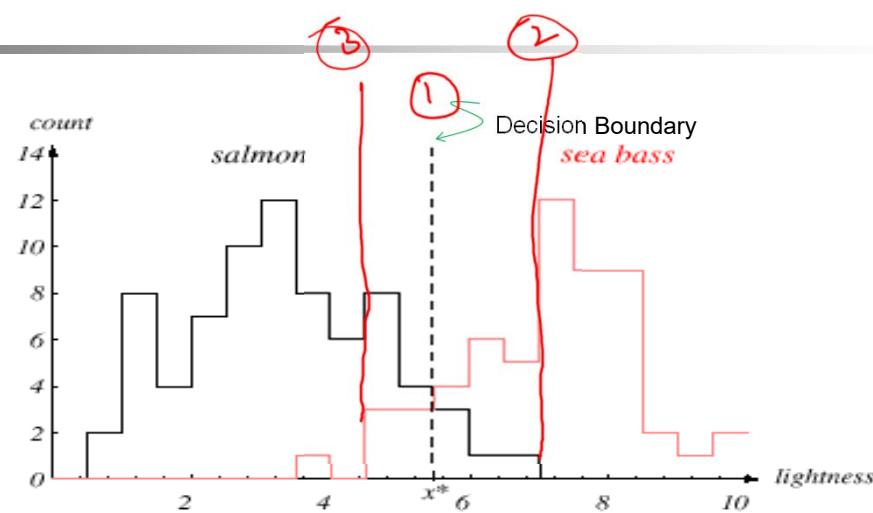
System Design Process

- Classification Rule
 - $f' = \begin{cases} \text{If } Length \geq l^* \text{ then sea bass} \\ \text{otherwise salmon} \end{cases}$
- How to choose l^* ???
- Use *Training Data*



Analysis

- The **length** is a poor feature alone!
- New Feature:
 - Select the **lightness of fish** as a possible feature



Adding More Features

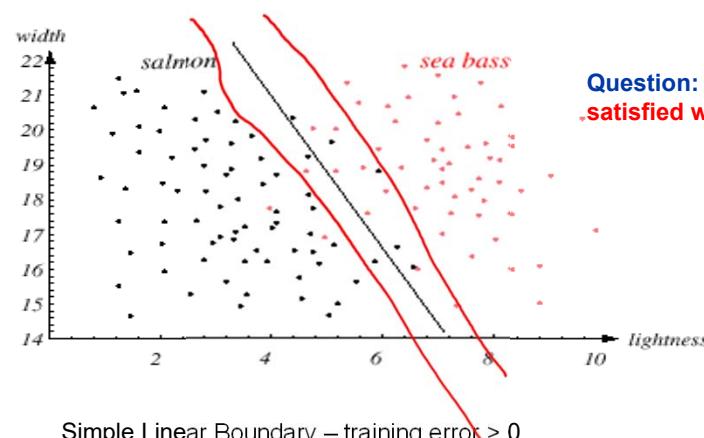
- Adopt the lightness and add the width of the fish

Fish $\rightarrow \mathbf{x}^T = [x_1, x_2]$

Lightness Width



Decision rule: Use the linear boundary



Question: Should we be satisfied with this result?

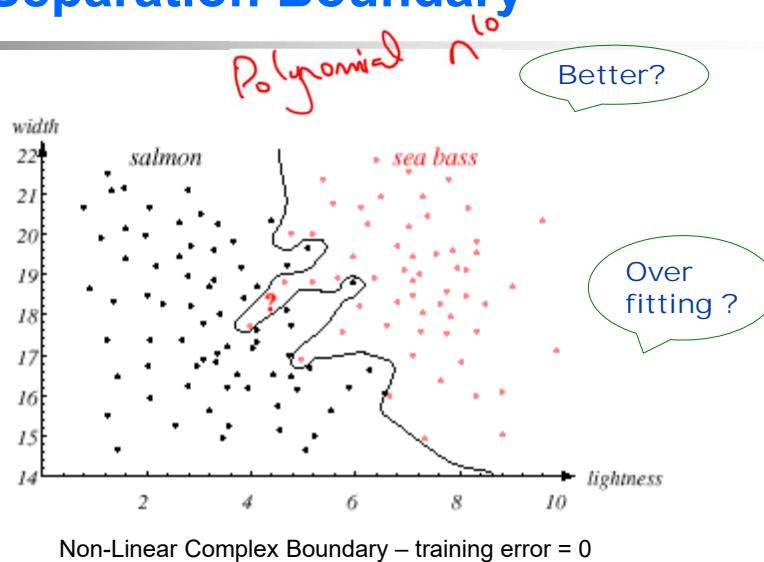
Simple Linear Boundary – training error > 0



- We might add other features that are not correlated with the ones we already have. A precaution should be taken not to reduce the performance by adding such “noisy features”
- Ideally, the best decision boundary should be the one which provides an optimal performance
- Use a **more complex model**



Optimal Separation Boundary



- However, our satisfaction is premature because the central aim of designing a classifier is to correctly classify novel input



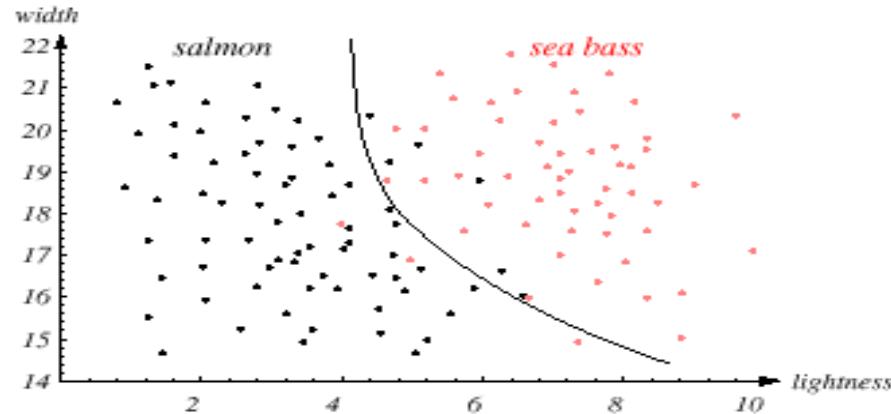
Issue of generalization!



Generalization

- The classifier should capture the underlying characteristics of the categories
- The classifier should NOT be tuned to the specific (accidental) characteristics of the training data
- A good classifier should be able to generalize, i.e. perform well on unseen data
- Training data in practice contain some noise





Simpler non-Linear Boundary – training error > 0



Decision Costs

- A classifier, intuitively, is designed to minimize classification error, the total number of instances (fish) classified incorrectly.
 - Is this the best objective (cost) function to minimize?
 - What kinds of error can be made? Are they all equally bad? What is the real cost of making an error?
 - Sea bass misclassified as salmon: Pleasant surprise for the consumer, tastier fish/ manufacturer lose money for selling expensive fish for the cost of inexpensive fish
 - Salmon misclassified as sea bass: Customer upset, paid too much for inferior fish

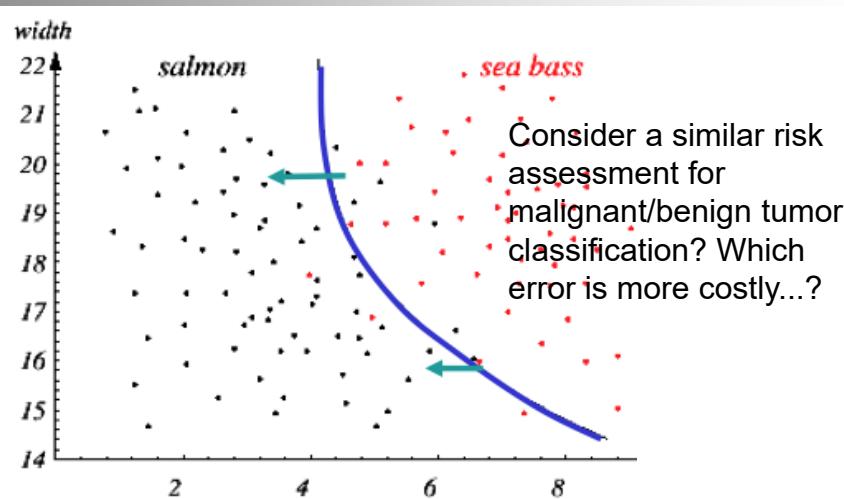


Decision Costs ...

- Threshold decision boundary and cost relationship
 - Move our decision boundary toward smaller values of lightness in order to minimize the cost (reduce the number of sea bass that are classified salmon!)



Task of decision theory



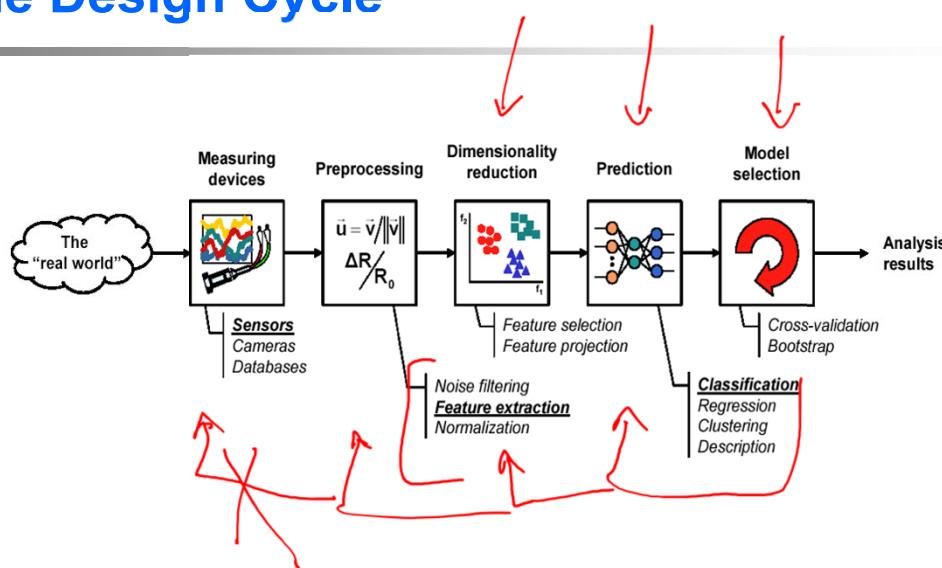
MACHINE LEARNING DESIGN CYCLE



85

85

The Design Cycle



Dr. Ali Hassan, DCSE, College of E&ME, NUST

86

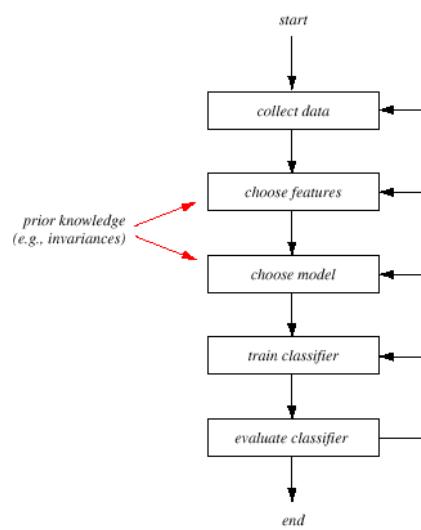
86

The Design Cycle

- Data collection
- Feature Choice
- Model Choice
- Training
- Evaluation
- Computational Complexity



*figure taken from *Pattern Classification*, Duda Hart



- Data Collection
 - How do we know when we have collected an adequately large and representative set of examples for training and testing the system?
- Pre-Processing
 - Filtering
 - Normalisation
 - How many examples of each class should be obtained



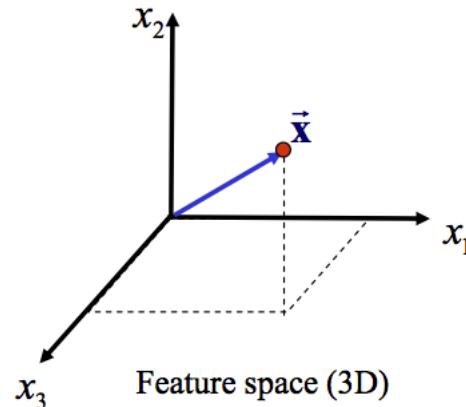
Features

- **Features:** a set of variables believed to carry discriminating and characterizing information about the objects under consideration
- **Feature vector:** A collection of d features, ordered in some meaningful way into a d -dimensional column vector, that represents the signature of the object to be identified.
- **Feature space:** The d -dimensional space in which the feature vectors lie. A d -dimensional vector in a d -dimensional space constitutes a point in that space.



Features

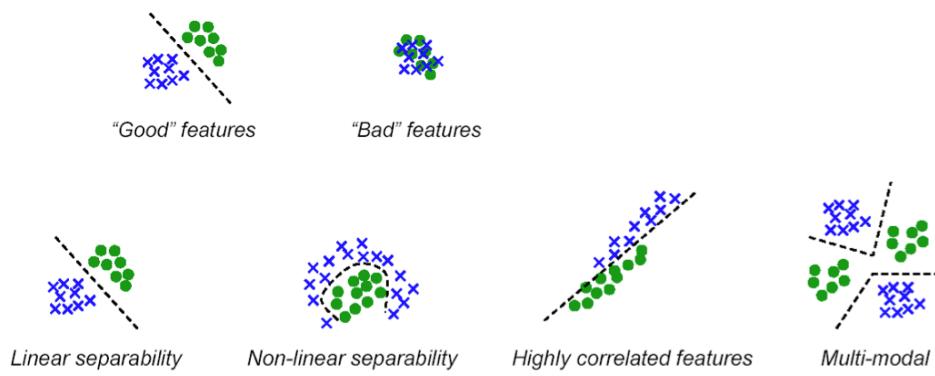
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{matrix} \text{feature 1} \\ \text{feature 2} \\ \vdots \\ \text{feature } d \end{matrix}$$



Features

- Feature Choice
 - Good Features
 - Ideally, for a given group of patterns coming from the same class, feature values should all be similar
 - For patterns coming from different classes, the feature values should be different.
 - Bad Features
 - irrelevant, noisy, outlier?
- Bottom Line
 - In any PR problem, you get what you put in:
 - **garbage in – garbage out (GIGO)!**





■ Model Choice

- What type of *classifier* shall we use? How shall we select its parameters? Is there best classifier...?
- How do we train...? How do we adjust the parameters of the model (*classifier*) we picked so that the model fits the data?

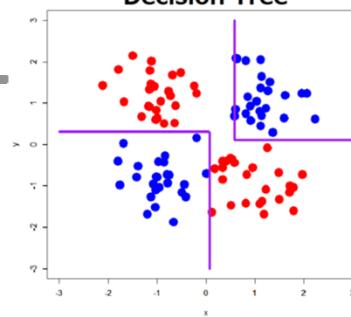


■ Training

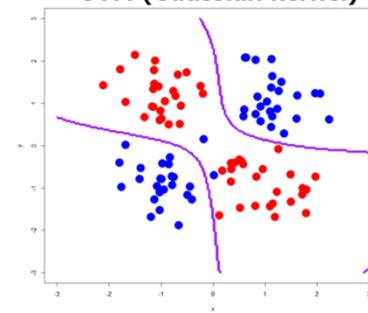
- Use data to determine the *classifier*. Many different procedures for training classifiers and choosing models



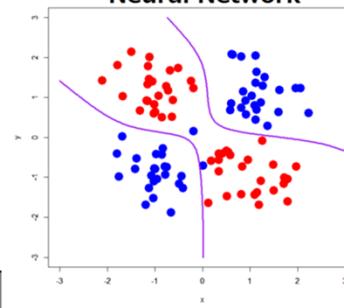
Decision Tree



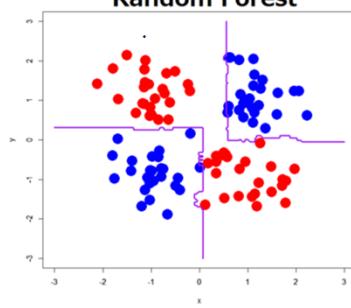
SVM (Gaussian kernel)

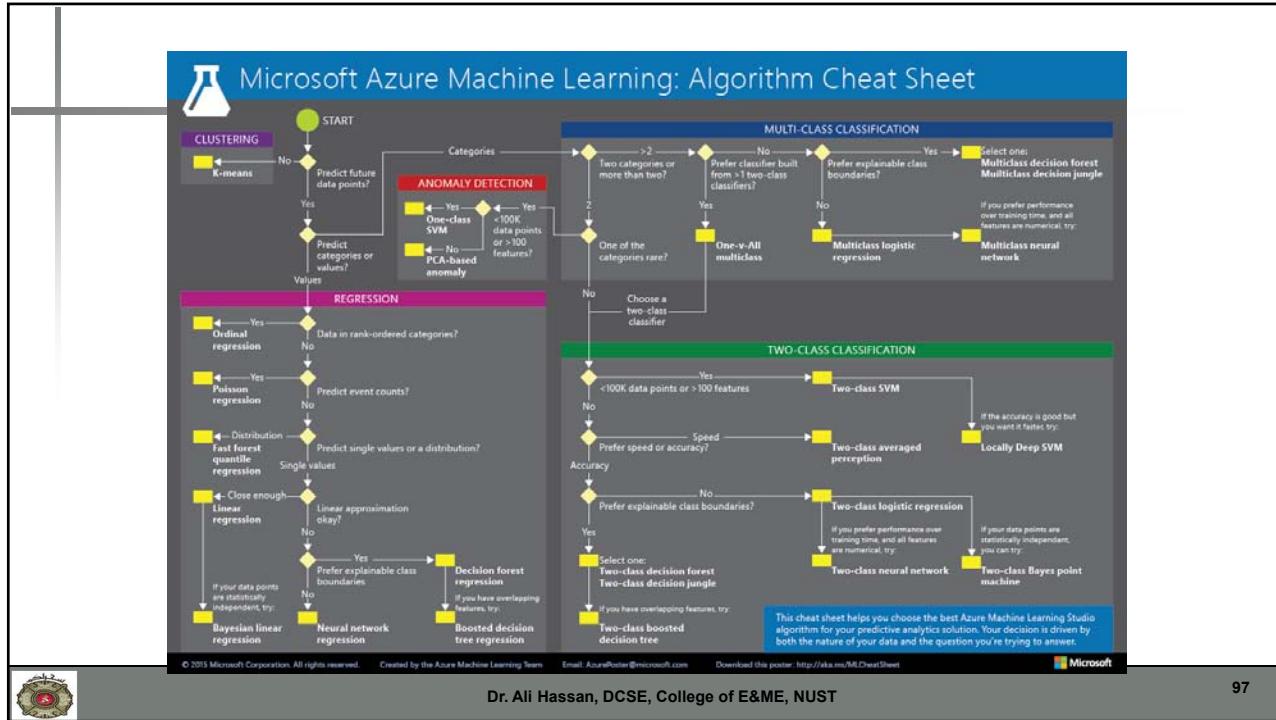


Neural Network



Random Forest





97

97

■ Evaluation

- Measure the error rate (or performance) and switch from one set of features to another one



98

48

■ Computational Complexity

- What is the trade-off between computational ease and performance?
- (How an algorithm scales as a function of the number of features, patterns or categories?)

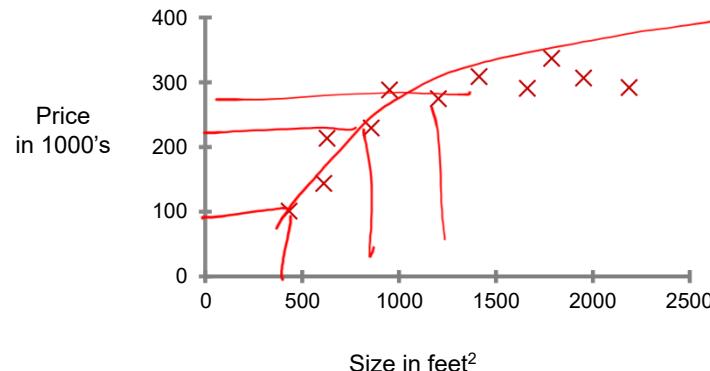


A set of handwritten red arrows and symbols on the left side of the slide, possibly indicating a flow or relationship between concepts.

SUPERVISED LEARNING



Housing price prediction.

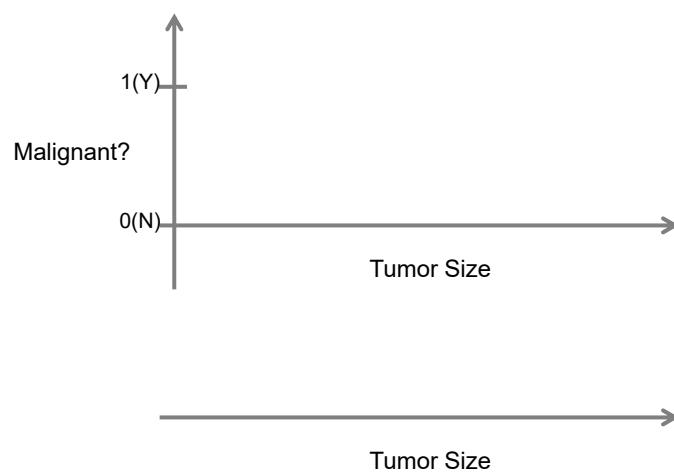


Supervised Learning
"right answers" given

Regression: Predict
continuous valued output
(price)

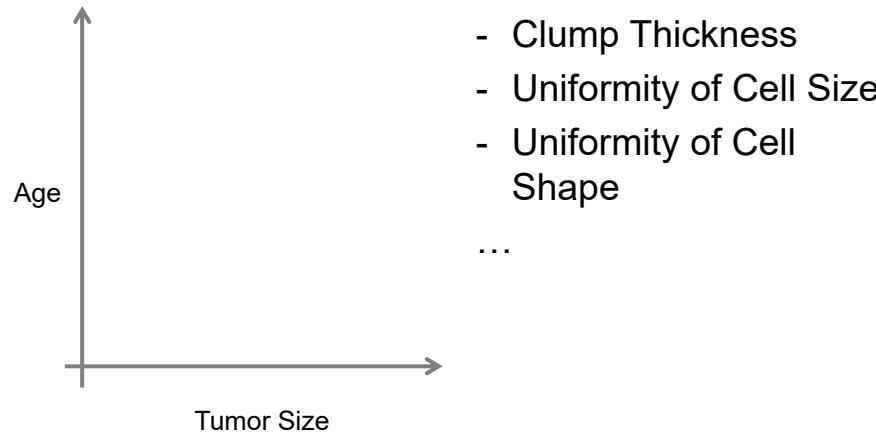


Cancer Problem (malignant, benign)



Classification
Discrete
valued output
(0 or 1)



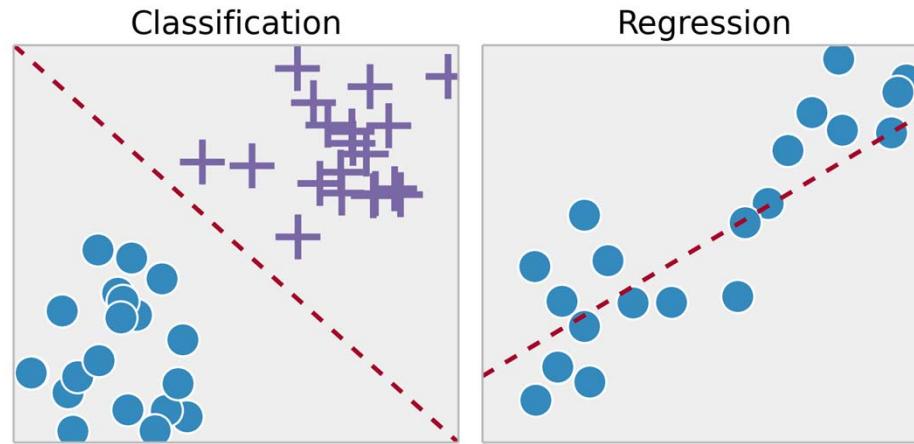


Examples of Supervised Learn.

- **Image Classification:**
 - You train with images/labels. Then in the future you give a new image expecting that the computer will recognise the new object (Classification)
- **Market Prediction:**
 - You train the computer with historical market data and ask the computer to predict the new price in the future (Regression)



Examples of Supervised Learn.



Learning from Data

- Convert real world data into features
- We are learning from data
- Different from Traditional AI—**rule based**
- Data is typically high dimensional (i.e. many features/attributes)



IRIS Dataset

- Historic Data collected by R. A. Fisher
- Determining iris species
- A multi class classification Problem



Setosa



Versicolor



Virginica



IRIS Attributes

- Attributes
 - sepal length in cm
 - sepal width in cm
 - petal length in cm
 - petal width in cm

	Min	Max	Mean	SD	Class Correlation
sepal length:	4.3	7.9	5.84	0.83	0.7826
sepal width:	2.0	4.4	3.05	0.43	-0.4194
petal length:	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width:	0.1	2.5	1.20	0.76	0.9565 (high!)



Data

- 50 measurements from each species

5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
5.5,2.3,4.0,1.3,Iris-versicolor
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica

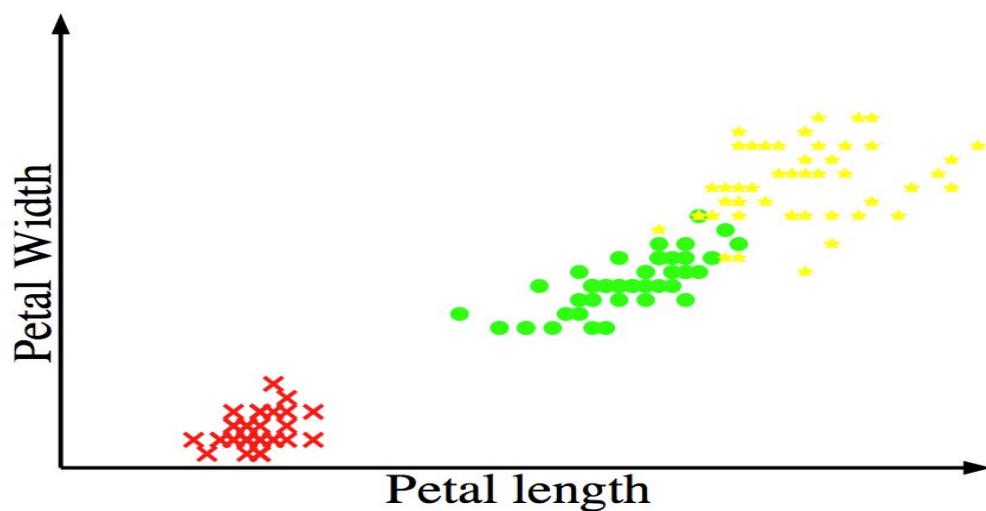
...

Dr. Ali Hassan, DCSE, College of E&ME, NUST

109

109

What Data Looks Like



Dr. Ali Hassan, DCSE, College of E&ME, NUST

110

110

HANDLING NON-NUMERIC DATA



111

111

Non-Numeric Data

- Inferring Salary from attributes
 - Age
 - Workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
 - Education: Bachelors, Some-college, HS-grad, Prof-school, Masters, Doctorate
 - marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse



Dr. Ali Hassan, DCSE, College of E&ME, NUST

112

112

Data

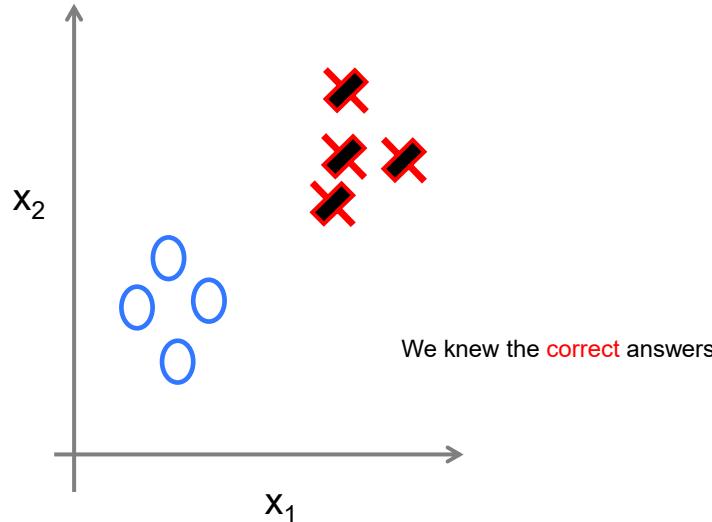
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, US, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, US, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, US, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, US, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 45, US, >50K
31, Private, 45781, Masters, 14, Never-married, Prof-specialty, Not-in-family, White, Female, 14084, 0, 50, US, >50K
19, ?, 37332, HS-grad, 9, Never-married, ?, Own-child, White, Female, 1055, 0, 12, US, <=50K



UNSUPERVISED LEARNING

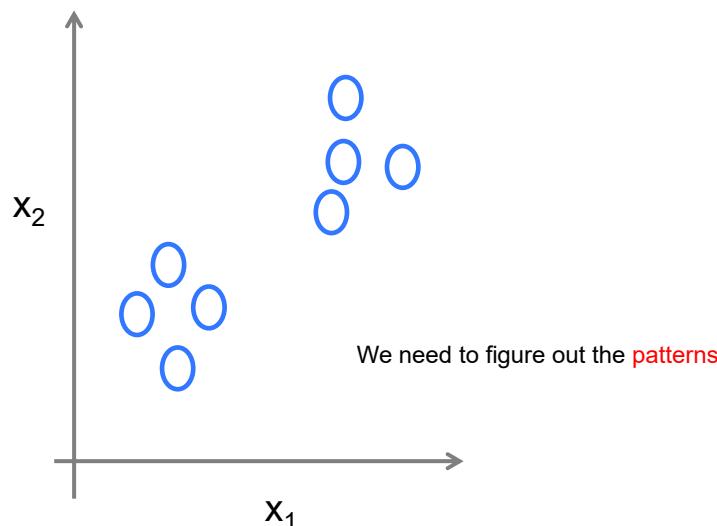


Supervised Learning



115

Unsupervised Learning



116

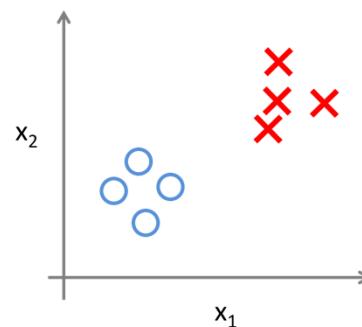
Types of ML

- Supervised Learning:
 - You give to the computer some pairs of inputs/outputs, so in the future when new inputs are presented you have an intelligent output.
- Unsupervised Learning:
 - You let the computer learn from the data itself without showing what is the expected output.

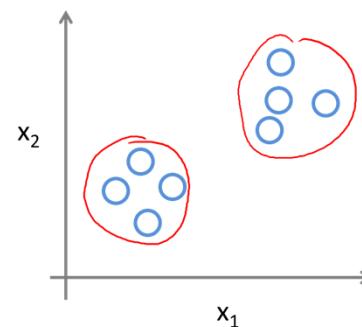


Types of ML

Supervised Learning



Unsupervised Learning



An Application of Unsupervised Learning

The screenshot shows the Google News homepage. On the left, there's a sidebar with 'Top stories' for COVID-19, including 'Imran Khan says cypher at centre of latest audio should now leak' and 'Audio leaks: Govt in the dark about Dark Web claims'. The main area displays 'Headlines' with a box for 'COVID-19 news: See the latest coverage of the coronavirus'. To the right, there's a 'Your local weather' section showing a sunny icon, 26°C, and a five-day forecast from Today to Sunday. Below the weather is a 'In the news' section with various topics like Apple, Elon Musk, Russia, iPhone, Ukraine, Mohammed bin Salman Al Saud, Suryakumar Yadav, Crown Prince of Saudi Arabia, Tropical Cyclone, and Pakistan.

Dr. Ali Hassan, DCSE, College of E&ME, NUST

119

119

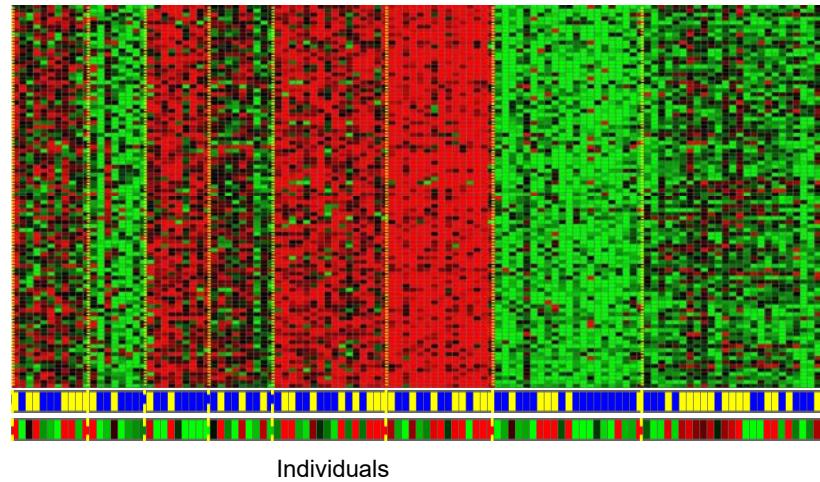
The screenshot shows three news websites displayed in separate browser tabs. The top tab is 'THE NEWS' dated Wednesday September 28, 2022, with a headline 'Audio leaks: Govt in the dark about Dark Web claims'. The middle tab is 'DAWN' dated September 27, 2022, with a headline 'The bugged office'. The bottom tab is 'THE TRIBUNE' dated Tuesday September 27, 2022, with a headline 'FBR asked to extend tax date'. All three sites feature similar layouts with navigation menus, search bars, and sections for various news categories.

Dr. Ali Hassan, DCSE, College of E&ME, NUST

120

120

Genes

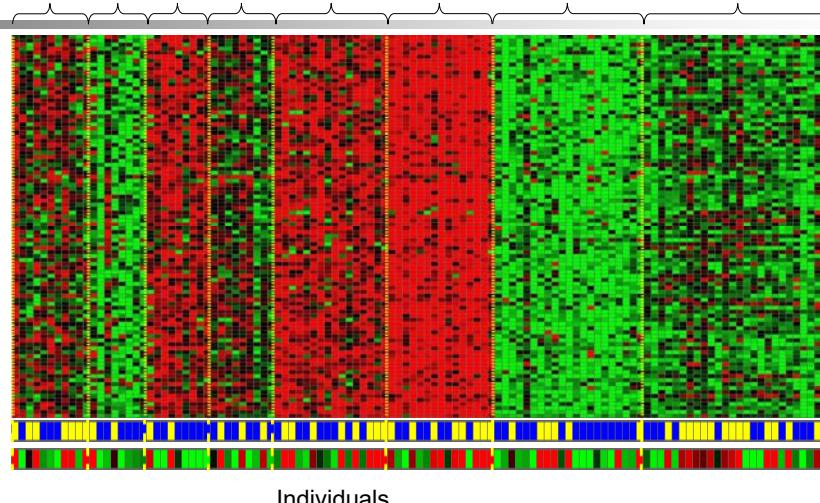


[Source: Daphne Koller]



121

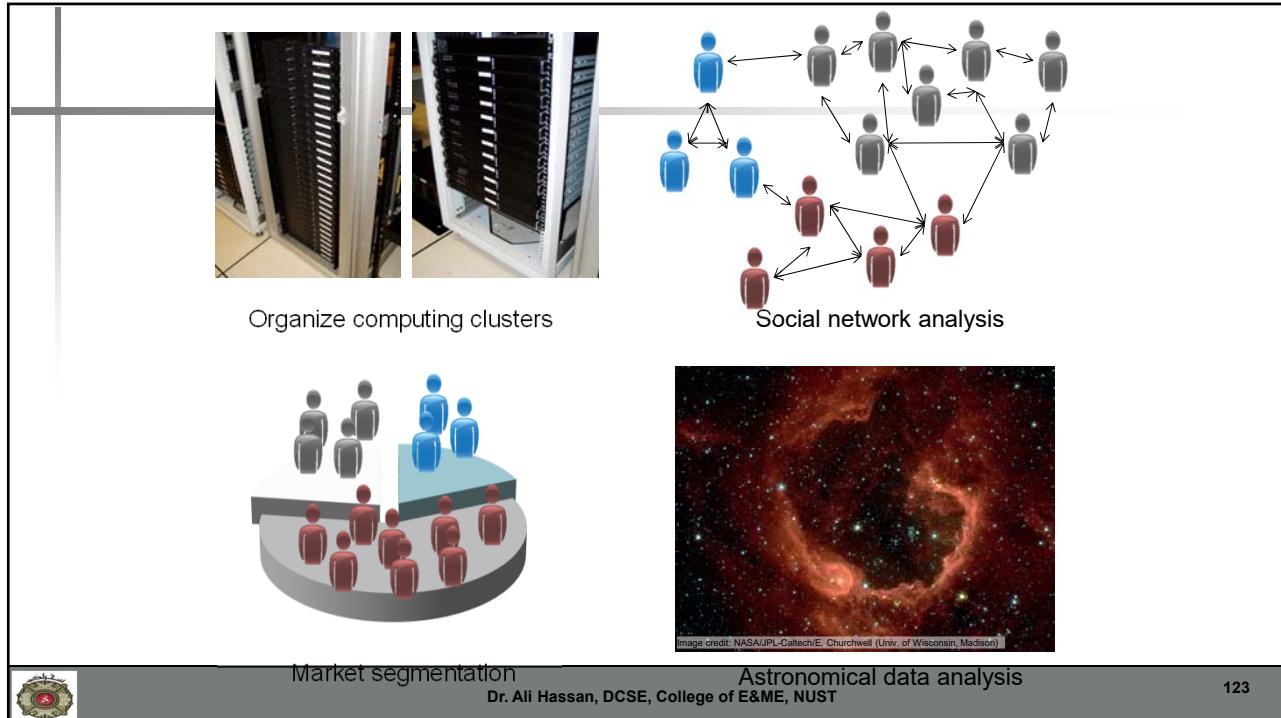
Genes



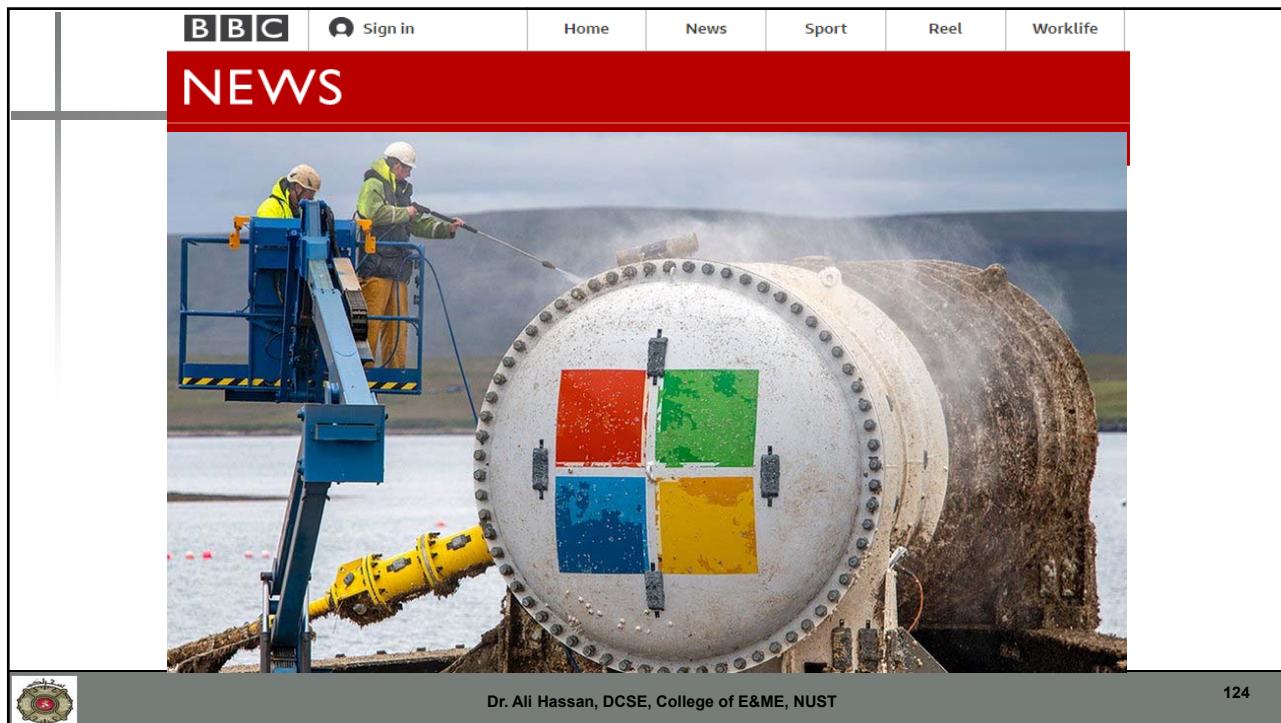
[Source: Daphne Koller]



122

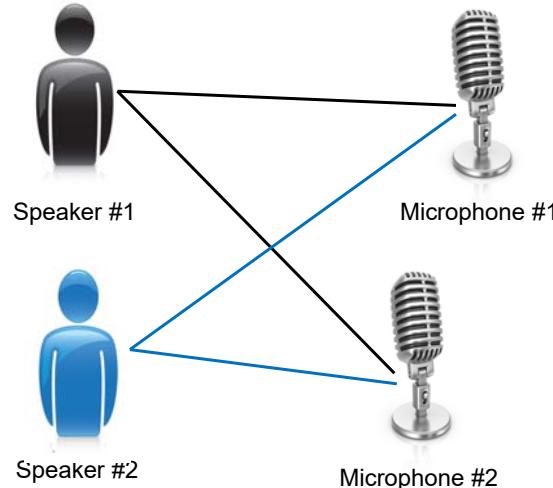


123



124

Cocktail party problem



Microphone #1: Output #1:
Microphone #2: Output #2:

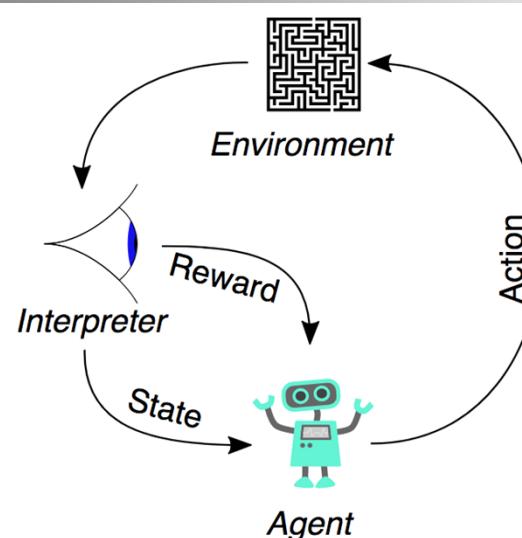


Cocktail party problem algorithm

```
[W,s,v] = svd((repmat(sum(x.*x,1),size(x,1),1).*x)*x');
```



Reinforced Learning

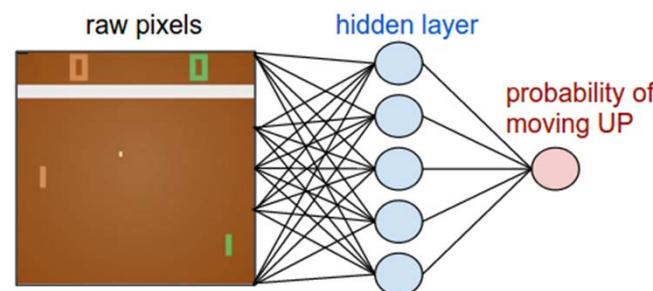
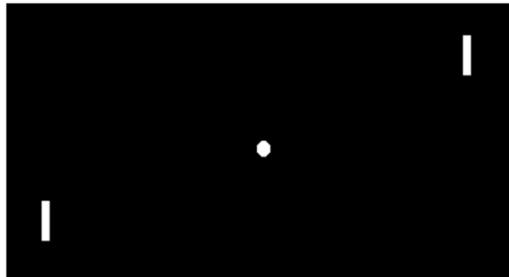
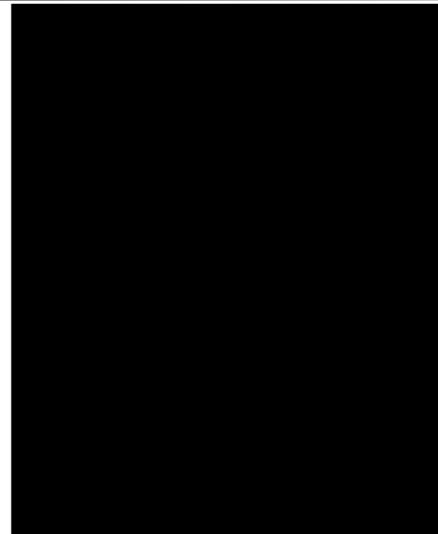


The typical framing of a Reinforcement Learning (RL) scenario: an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent.



Examples of RL

- How to do this using supervised machine learning?
- How to do this using RL?



Our policy network is a 2-layer fully-connected net



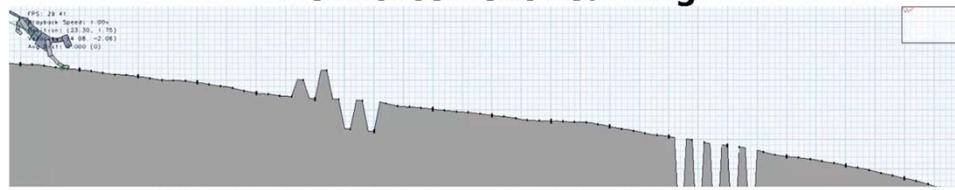


Dr. Ali Hassan, DCSE, College of E&ME, NUST

131

131

Terrain-Adaptive Locomotion Skills using Deep Reinforcement Learning



Xue Bin Peng, Glen Berseth, Michiel van de Panne
University of British Columbia
Supplementary Video

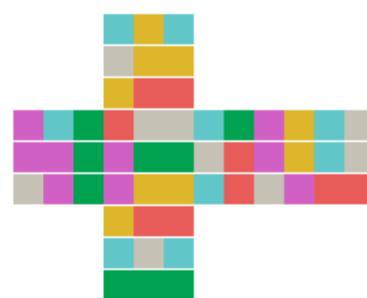
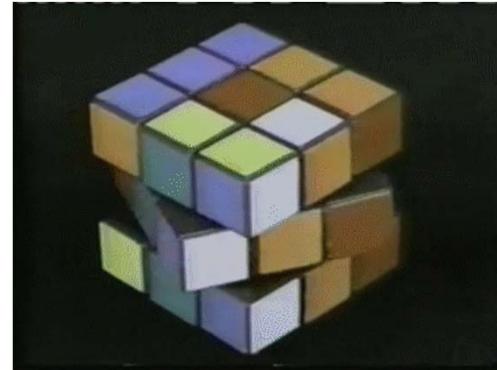


Dr. Ali Hassan, DCSE, College of E&ME, NUST

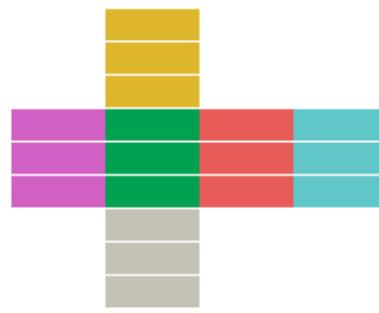
132

132

Learning To Solve a Rubik's Cube From Scratch using RL



Flattened Cube



Flattened solved Cube



Solving the Rubik's Cube Without Human Knowledge

Stephen McAleer*
 Department of Statistics
 University of California, Irvine
 smcaleer@uci.edu

Forest Agostinelli*
 Department of Computer Science
 University of California, Irvine
 fagostin@uci.edu

Alexander Shmakov*
 Department of Computer Science
 University of California, Irvine
 ashmakov@uci.edu

Pierre Baldi
 Department of Computer Science
 University of California, Irvine
 pfbaldi@ics.uci.edu

Abstract

A generally intelligent agent must be able to teach itself how to solve problems in complex domains with minimal human supervision. Recently, deep reinforcement learning algorithms combined with self-play have achieved superhuman proficiency in Go, Chess, and Shogi without human data or domain knowledge. In these environments, a reward is always received at the end of the game; however, for many combinatorial optimization environments, rewards are sparse and episodes are not guaranteed to terminate. We introduce Autodidactic Iteration: a novel reinforcement learning algorithm that is able to teach itself how to solve the Rubik's Cube with no human assistance. Our algorithm is able to solve 100% of randomly scrambled cubes while achieving a median solve length of 30 moves — less than or equal to solvers that employ human domain knowledge.

5.07470v1 [cs.AI] 18 May 2018

Dr. Ali Hassan, DCSE, College of E&ME, NUST

135

LINEAR ALGEBRA



136

136

Scalar Vector Matrix

Scalar	Vector	Matrix
24	$\begin{bmatrix} 2 & -8 & 7 \end{bmatrix}$ <small>row or column</small> $\begin{bmatrix} -6 \\ -4 \\ 27 \end{bmatrix}$	$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}$ <small>row(s) × column(s)</small>

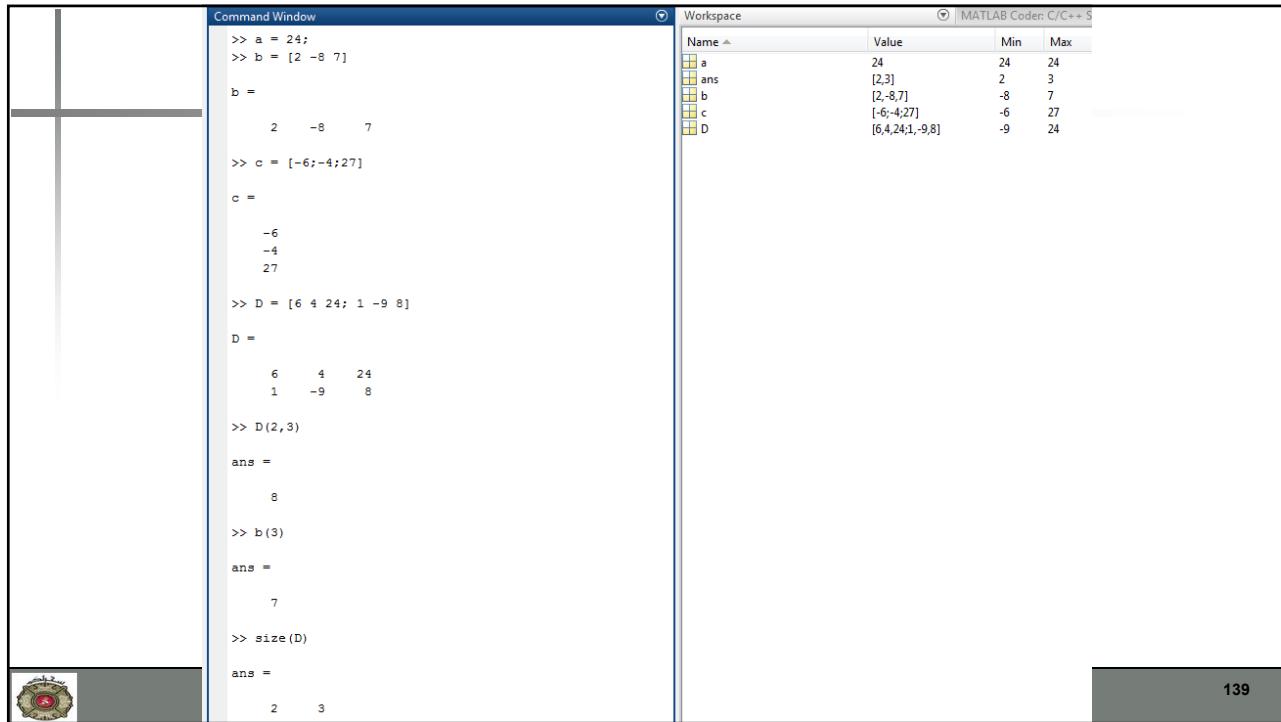


137

Dimensions	Example	Terminology
1	$\begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$	Vector
2	$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$	Matrix
3	$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$	3D Array (3 rd order Tensor)
N	$\begin{bmatrix} \dots \\ \dots \end{bmatrix}$	ND Array



138



The screenshot shows the MATLAB interface. The Command Window pane displays the following MATLAB session:

```

>> a = 24;
>> b = [2 -8 7]

b =
    2     -8      7

>> c = [-6;-4;27]

c =
    -6
    -4
    27

>> D = [6 4 24; 1 -9 8]

D =
    6     4     24
    1    -9      8

>> D(2,3)
ans =
    8

>> b(3)
ans =
    7

>> size(D)
ans =
    2     3

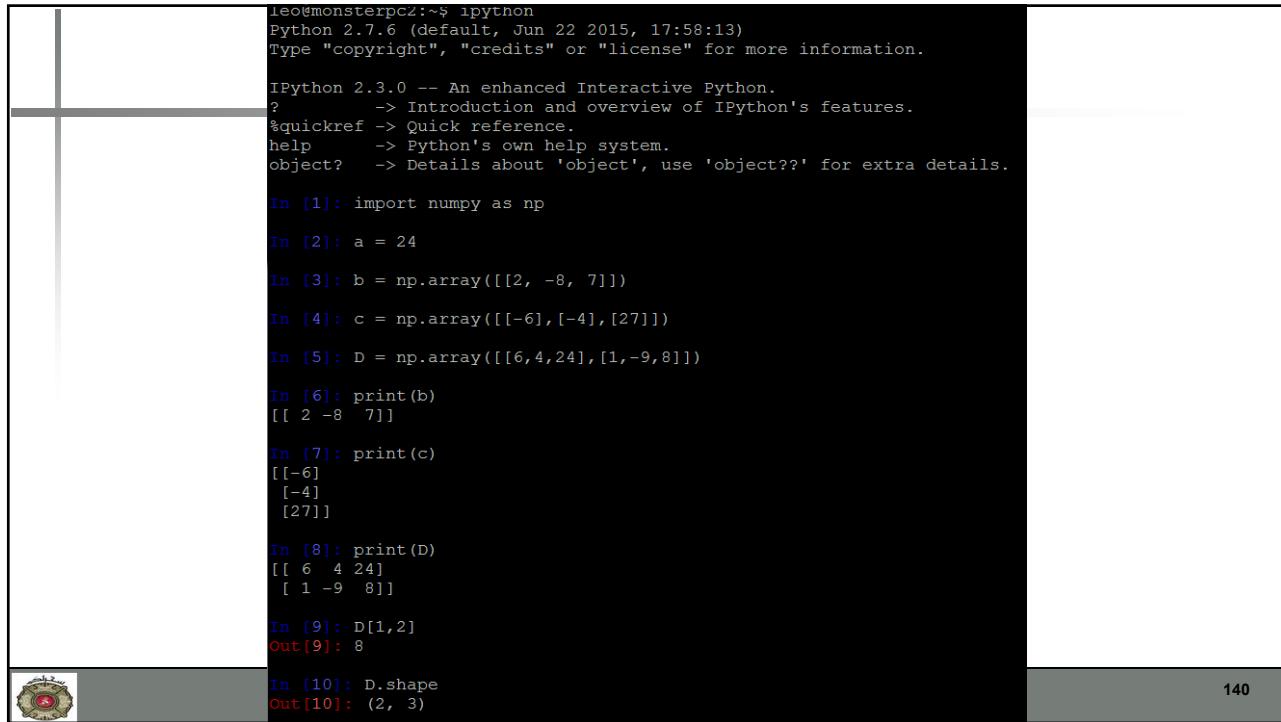
```

The Workspace pane shows the variables and their values:

Name	Value	Min	Max
a	24	24	24
ans	[2,3]	2	3
b	[2,-8,7]	-8	7
c	[-6;-4;27]	-6	27
D	[6,4,24;1,-9,8]	-9	24

139

139



The screenshot shows an IPython terminal session. The session starts with the Python 2.7.6 prompt and then enters an IPython 2.3.0 enhanced interactive session. The user imports numpy and performs operations similar to the MATLAB session above.

```

leo@monsterc2:~$ ipython
Python 2.7.6 (default, Jun 22 2015, 17:58:13)
Type "copyright", "credits" or "license" for more information.

IPython 2.3.0 -- An enhanced Interactive Python.
?           --> Introduction and overview of IPython's features.
%quickref --> Quick reference.
help        --> Python's own help system.
object?     --> Details about 'object', use 'object??' for extra details.

In [1]: import numpy as np

In [2]: a = 24

In [3]: b = np.array([[2, -8, 7]])

In [4]: c = np.array([-6, -4, 27])

In [5]: D = np.array([6,4,24], [1,-9,8])

In [6]: print(b)
[[ 2 -8  7]]

In [7]: print(c)
[-6
 -4
 27]

In [8]: print(D)
[[ 6  4 24]
 [ 1 -9  8]]

In [9]: D[1,2]
Out[9]: 8

In [10]: D.shape
Out[10]: (2, 3)

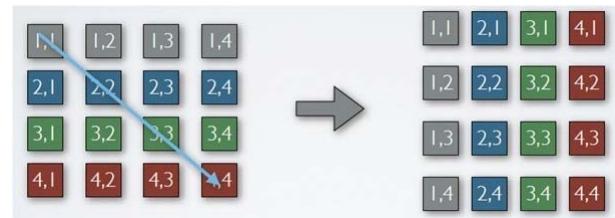
```

140

140

Matrix Operations

■ Transpose

**A**

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

- $[1 \ 2]^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
- $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$



Addition/Subtraction

$$\begin{bmatrix} 3 & 8 \\ 4 & 6 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} 7 & 8 \\ 5 & -3 \end{bmatrix}$$

3+4=7

$$\begin{bmatrix} 3 & 8 \\ 4 & 6 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} -1 & 8 \\ 3 & 15 \end{bmatrix}$$

3-4=-1



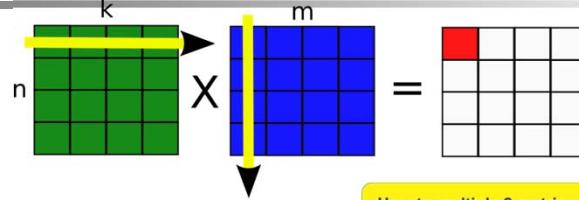
Multiply by scalar

$$2 \times \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 2 & -18 \end{bmatrix}$$

2x4=8



Matrix Multiplication



How to multiply 2 matrices?

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad 1 \times 1 + 2 \times 2 = 1 + 4 = 5$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad 1 \times 2 + 2 \times 4 = 2 + 8 = 10$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad 2 \times 1 + 4 \times 2 = 2 + 8 = 10$$

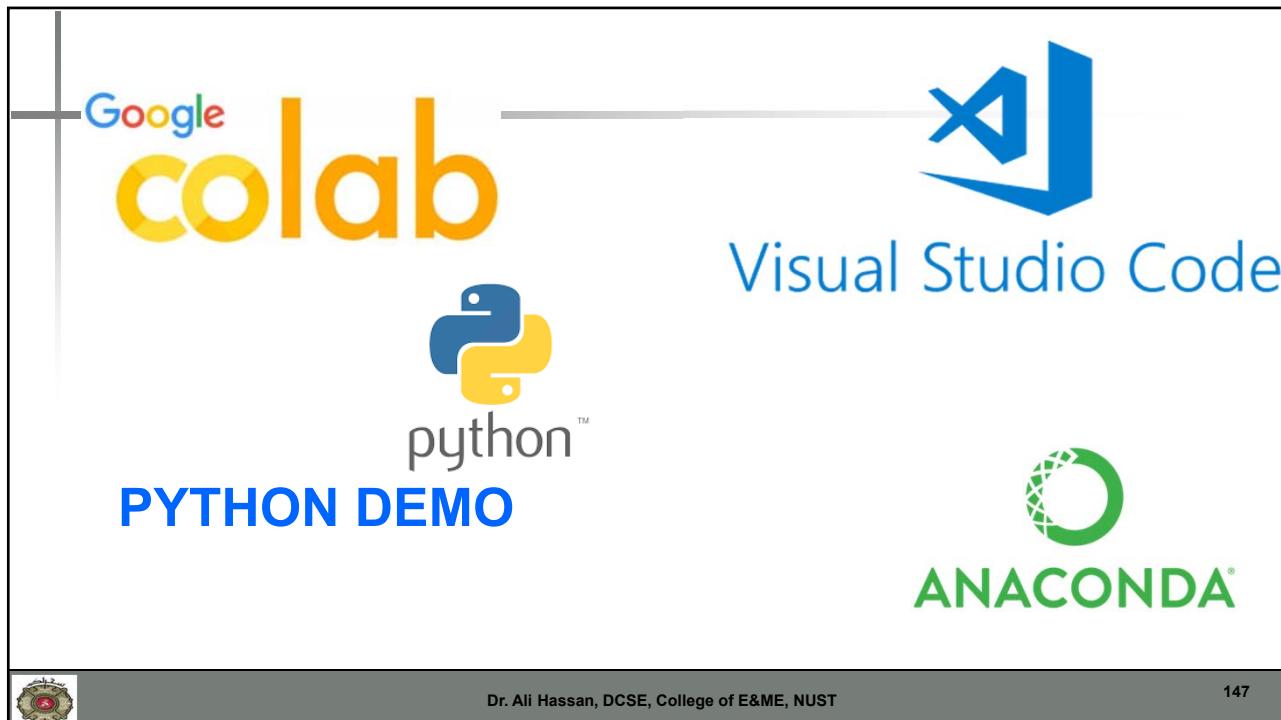
& so on



Types of Matrix

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$





147

Reading ...

- Pattern Classification, Duda, Hart, Stork
 - Chapter 1: Introduction
- Deep Reinforcement Learning: Pong from Pixels
 - (<http://karpathy.github.io/2016/05/31/rl/>)
- Solving the Rubik's Cube Without Human Knowledge
 - <https://arxiv.org/pdf/1805.07470.pdf>

Dr. Ali Hassan, DCSE, College of E&ME, NUST

148

Next Class

- Linear Regression – Uni-variate

