Mathematics for AI

Lecture 1

Introduction and Logistics

Mathematics for Artificial Intelligence

K. N. Toosi
University of Technology

- Graduate Course
- 3 credits, 32 sessions
- Saturday, Wednesday, 15:30-17:30
- Instructor: Behrooz Nasihatkon
- Email: nasihatkon@kntu.ac.ir
- Office: Room 402

Exam Dates

- Midterm Exam: Thursday, 17 Aban, 9:00-12:00
- Final Exam: look at the schedule

K. N. Toosi
University of Technology

Ask Questions

K. N. Toosi
University of Technology

Special Needs

K. N. Toosi
University of Technology


Eating in class

How to get help

How to give feedback?

K. N. Toosi
University of Technology

Anonymous form: https://goo.gl/zPxBAS

Join the Telegram channel

K. N. Toosi
University of Technology

https://t.me/math4AI4031

My Telegram Channel

K. N. Toosi
University of Technology

t.me/behrooznasihatkon

دانشجویان ارشد ورودی ۱۴۰۳

K. N. Toosi
University of Technology

https://t.me/cemaster1403

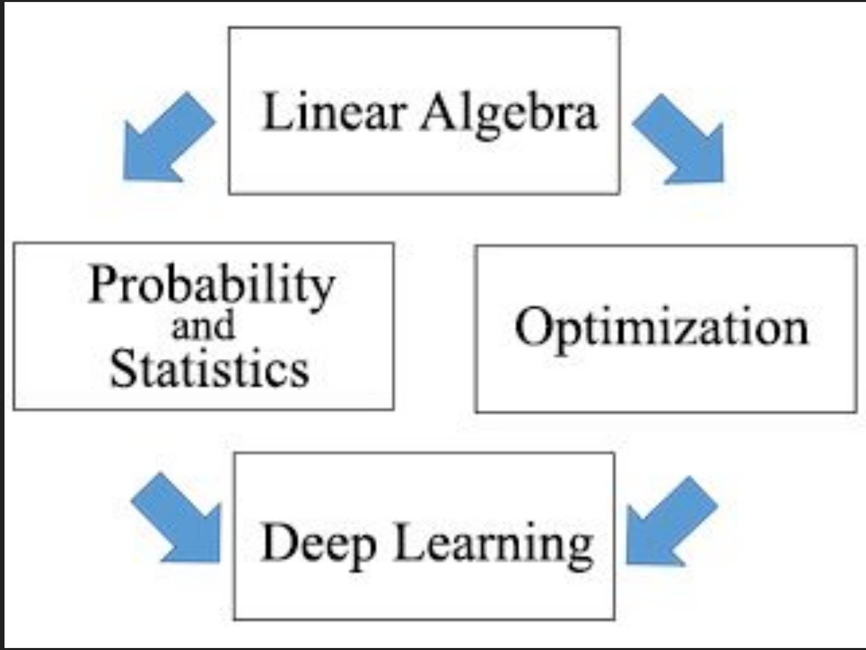
Why this course?

K. N. Toosi
University of Technology

-

Why this course?

Why this course?

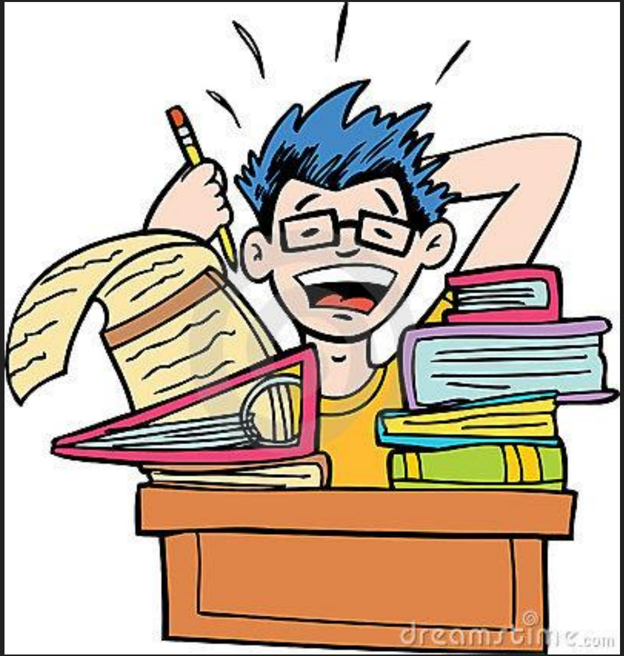


https://ocw.mit.edu/courses/18-065-matrix-methods-in-data-analysis-signal-processing-and-machine-learning-spring-2018/

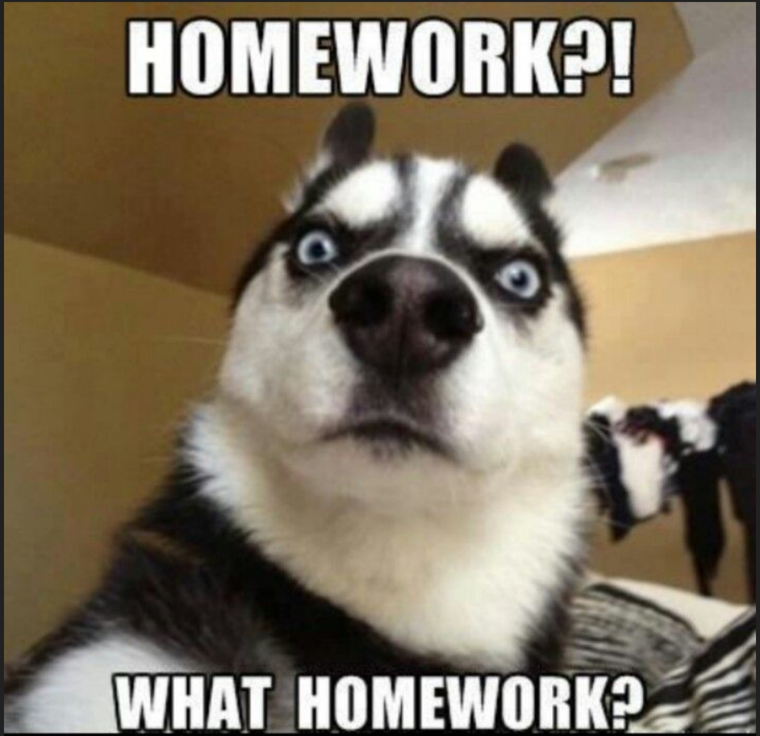
Evaluation

K. N. Toosi
University of Technology

- Lab Sessions (~2.5 points)
- Homework (~2.5 points)
- Projects (~3 points)
- Midterm Exam (~5 points)
- Final Exam (~7 points)



Homework/Projects/Labs



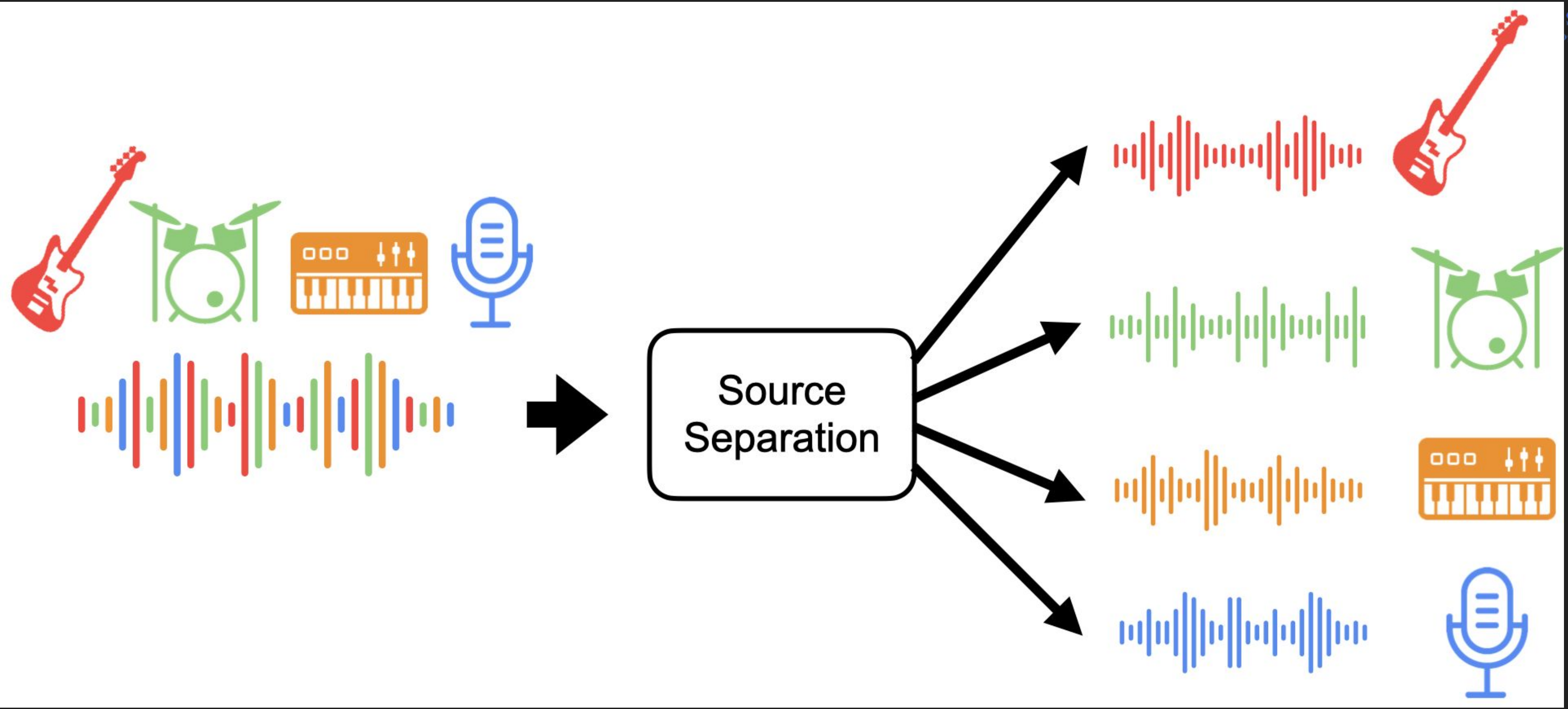
What is considered cheating?

K. N. Toosi
University of Technology

Programming Languages

- Matlab
- C++
- Python
- R
- Julia
- ...

Example: Source Separation



https://source-separation.github.io/tutorial/landing.html

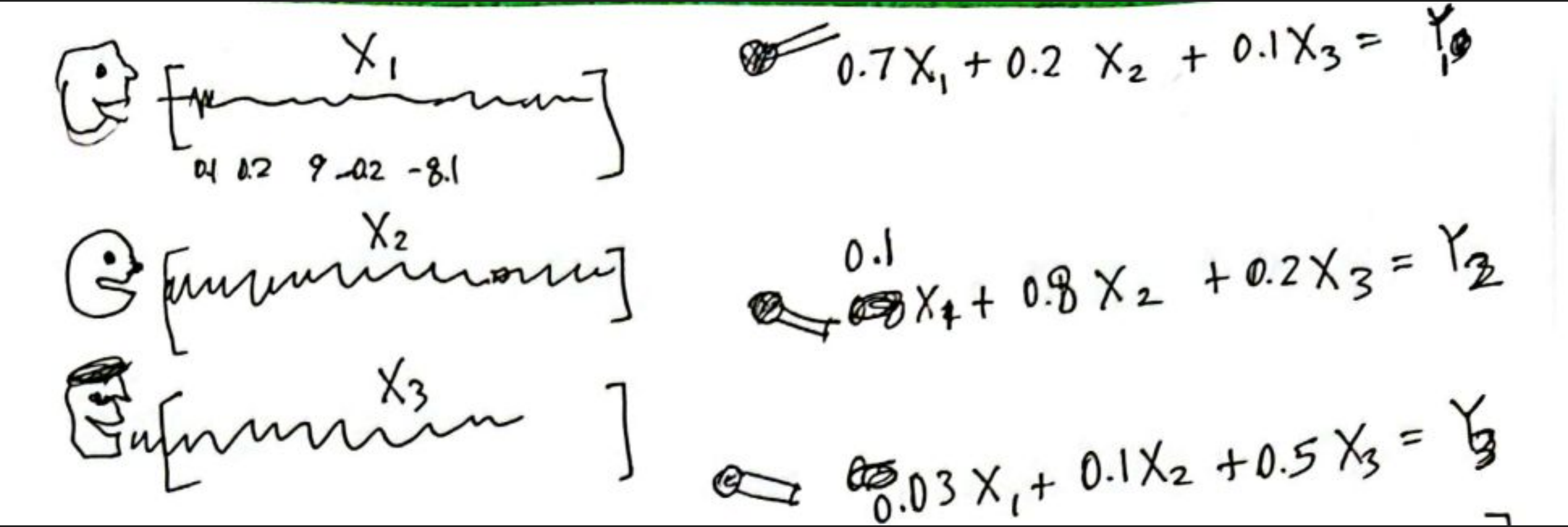
Source Separation

https://youtu.be/n7y2rLAnd5I

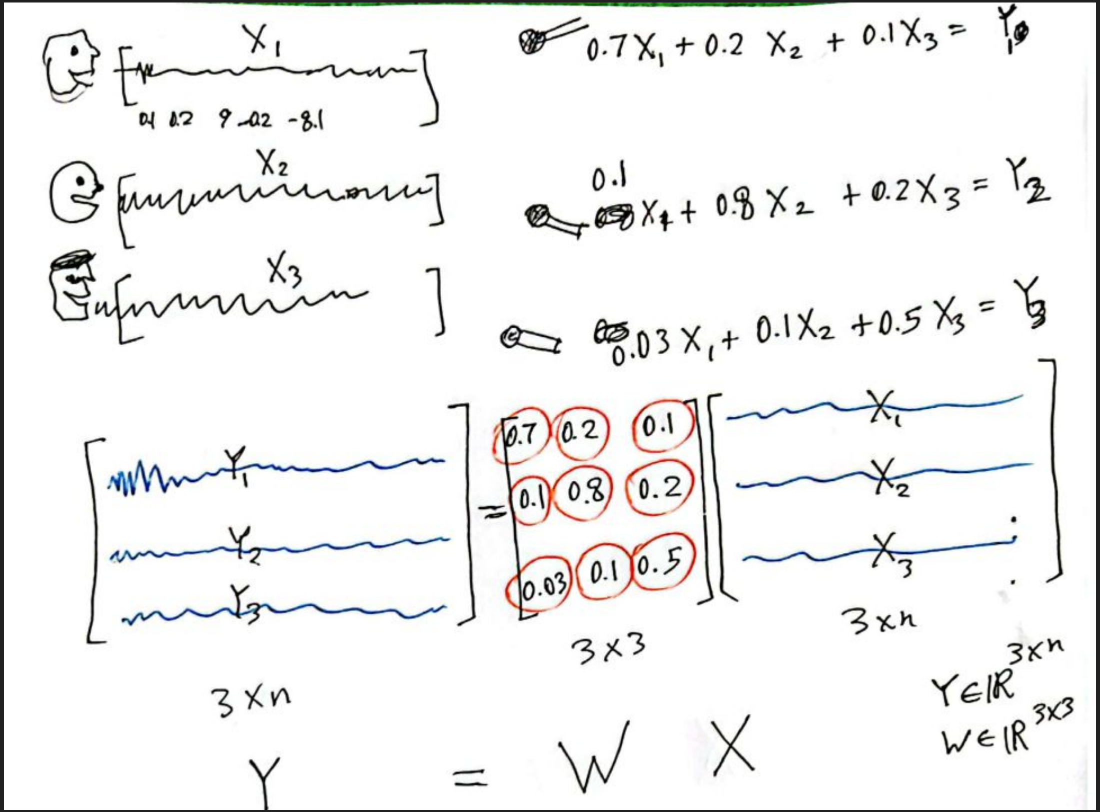
Source Separation

https://www.youtube.com/watch?v=tkkm6zVUDXo

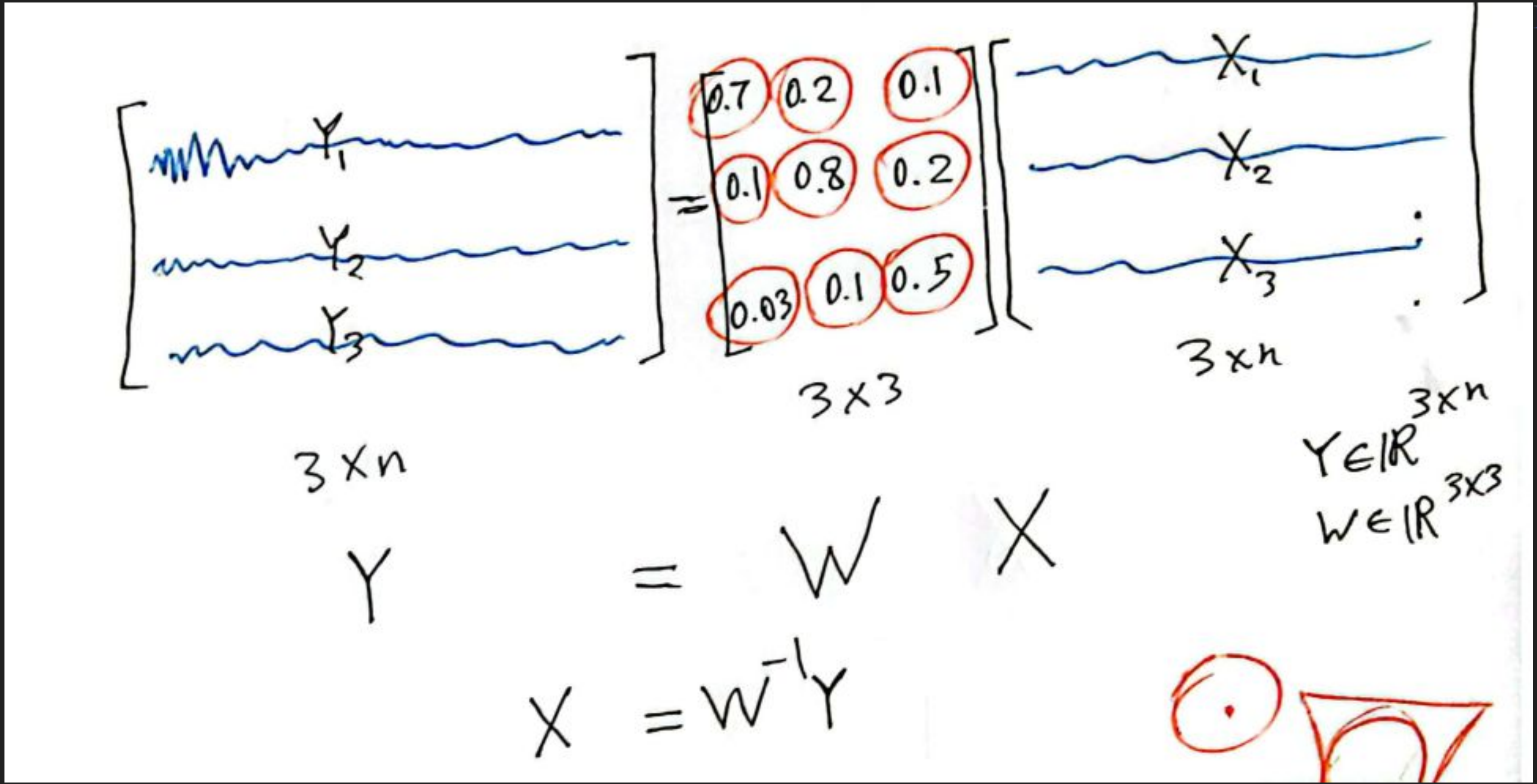
Example: Source Separation



Example: Source Separation



Example: Source Separation



K. N. Toosi
University of Technology

Example: Source Separation

K. N. Toosi
University of Technology

$$\underset{3\times n}{Y} = \underset{3\times 3}{W} \; \underset{3\times n}{X}$$

decompose Y into $W \in \mathbb{R}^{3\times 3}$ by $X \in \mathbb{R}^{3\times n}$

$$Y = I\,Y$$

$$Y = WX = \underbrace{WA}_{W'}\,\underbrace{A^{-1}X}_{X'} = W'X'$$

Example: Source Separation

- Linear Algebra:

  Y = A X   => X = A^{-1} Y
- Probability and Statistics:

  Mutual Information

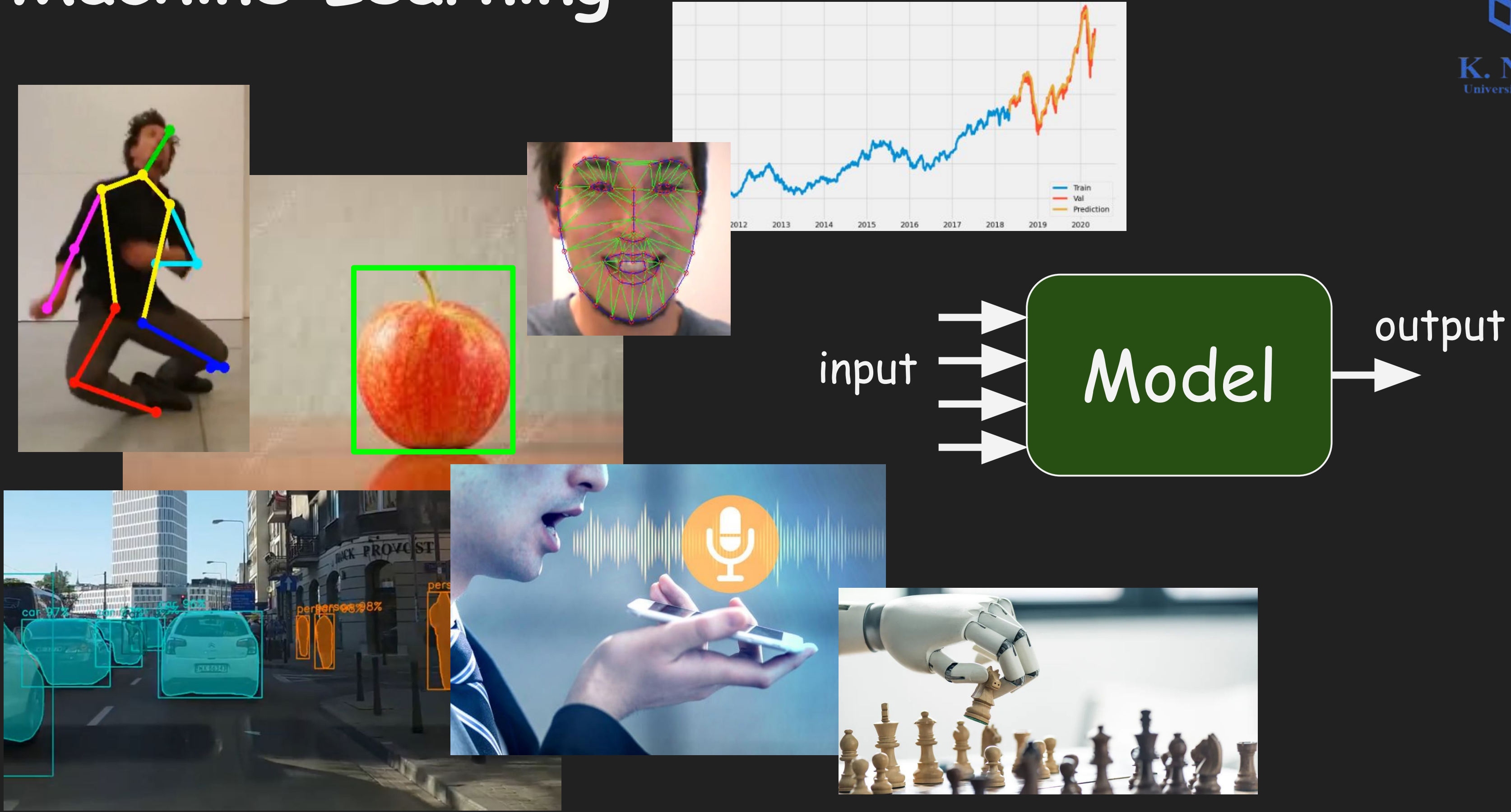
$$I(X_1;X_2;X_3) = \int_{x_1}\int_{x_2}\int_{x_3} p(x_1,x_2,x_3)\log\left(\frac{p(x_1,x_2,x_3)}{p(x_1)p(x_2)p(x_3)}\right)dx_1dx_2dx_3$$

- Optimization

$$\max_{A,X}\ I(X_1;X_2;X_3) \quad \text{subject to} \quad Y = AX$$

Machine Learning

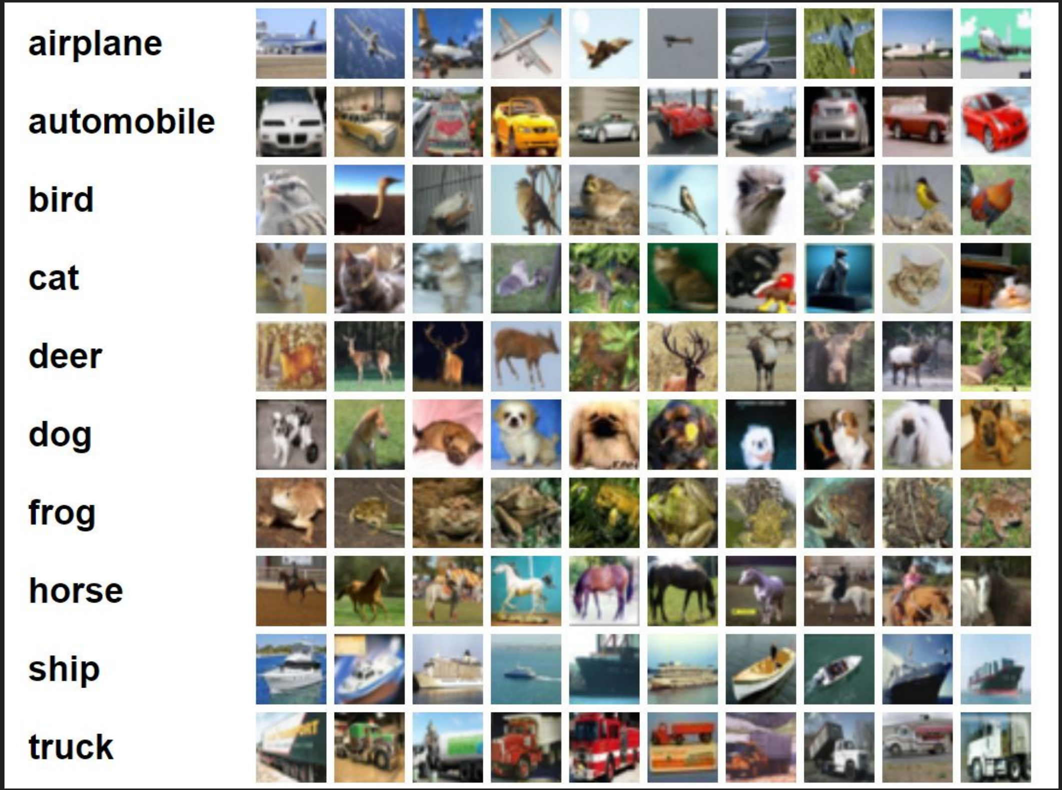
K. N. Toosi
University of Technology



Learning from data

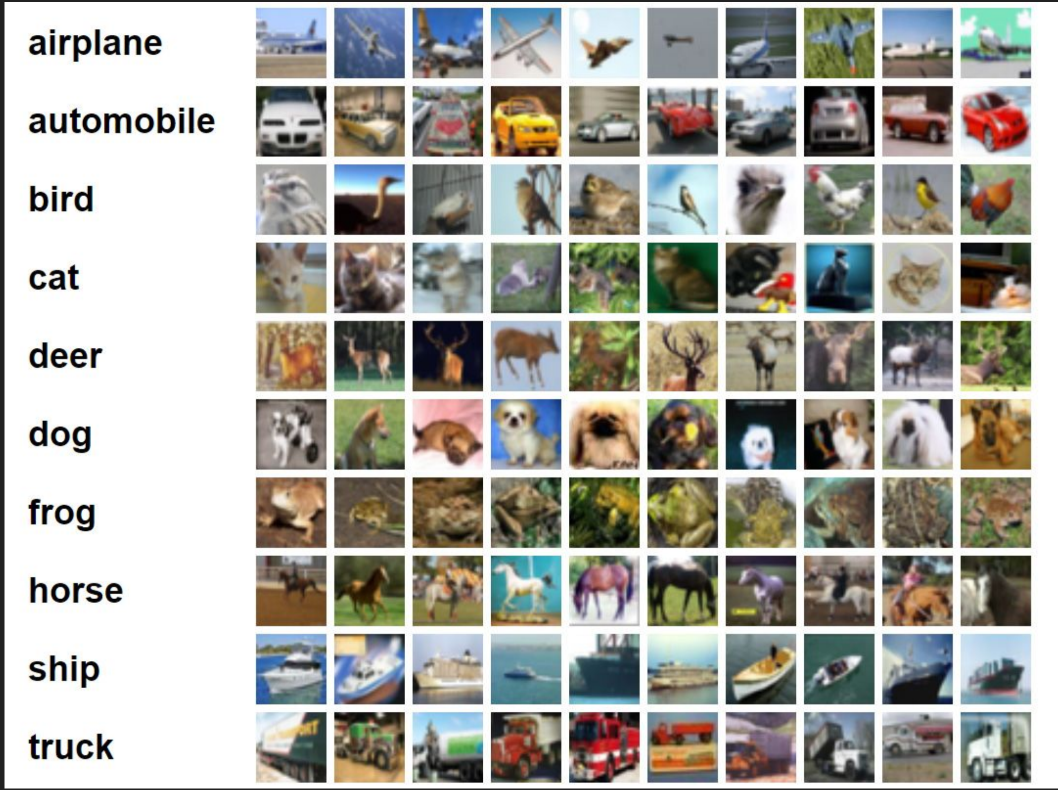
https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/

Supervised Learning



http://seansoleyman.com/effect-of-dataset-size-on-image-classification-accuracy/

Supervised Learning

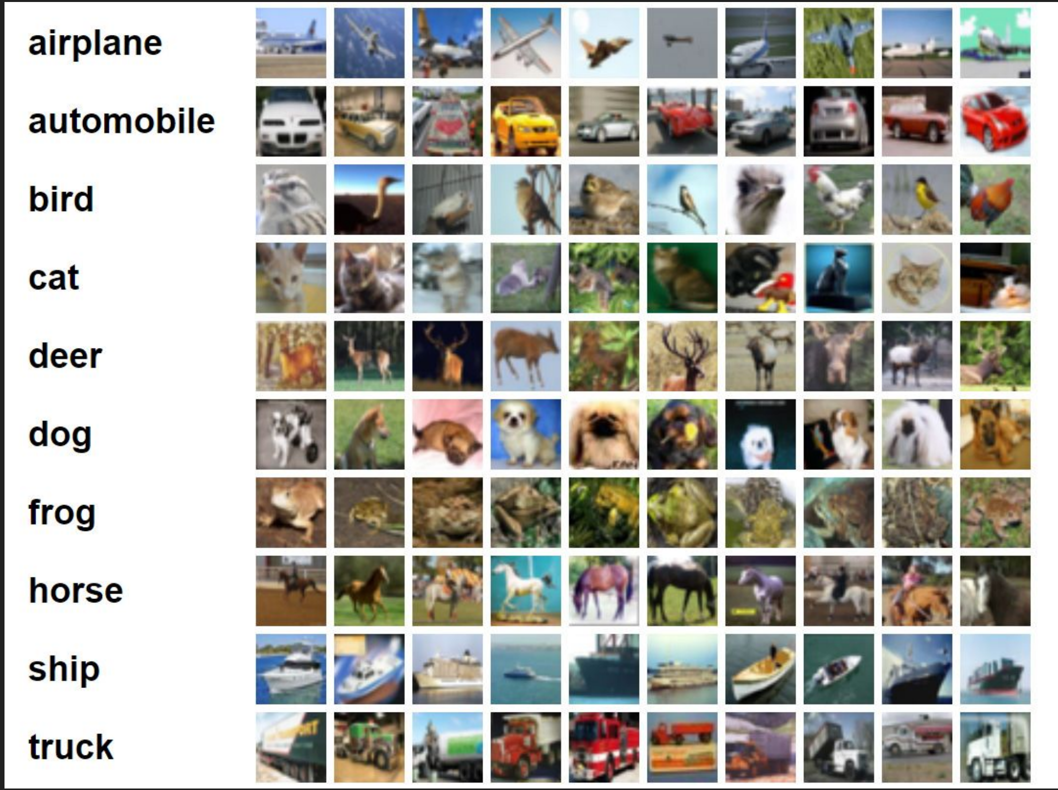


http://seansoleyman.com/effect-of-dataset-size-on-image-classification-accuracy/

Training data:

X_1, y_1
X_2, y_2
X_3, y_3
$\vdots$
X_n, y_n

Supervised Learning



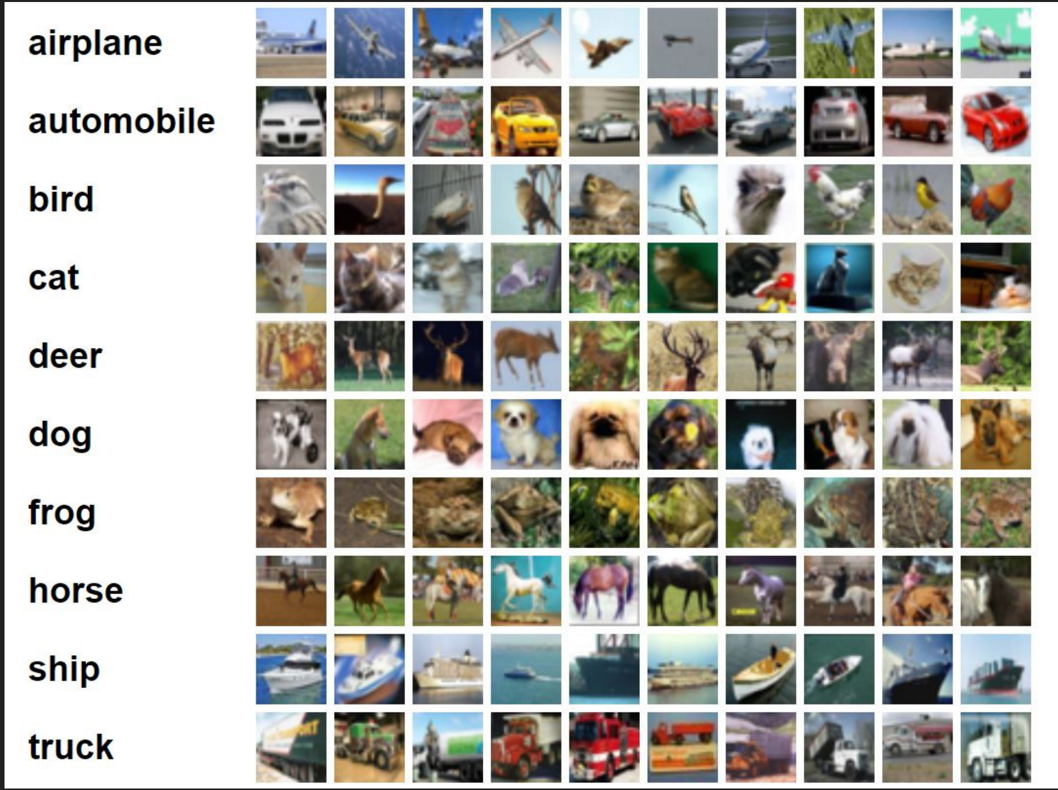
http://seansoleyman.com/effect-of-dataset-size-on-image-classification-accuracy/

Training data:

	Apple
	Apple
	Orange
⋮	
	Orange

K. N. Toosi
University of Technology

Supervised Learning



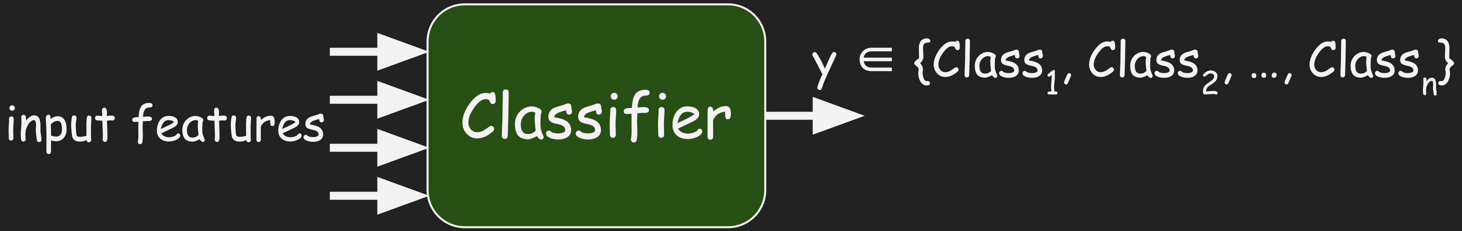
http://seansoleyman.com/effect-of-dataset-size-on-image-classification-accuracy/

Training data:

	0
	0
	1
⋮	
	1

Supervised Learning

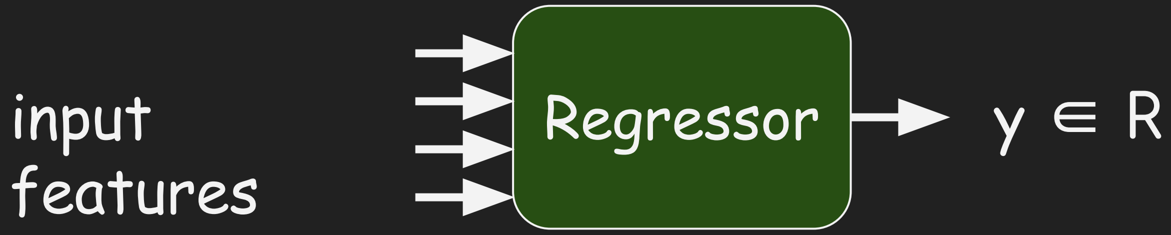
Classification



Classification

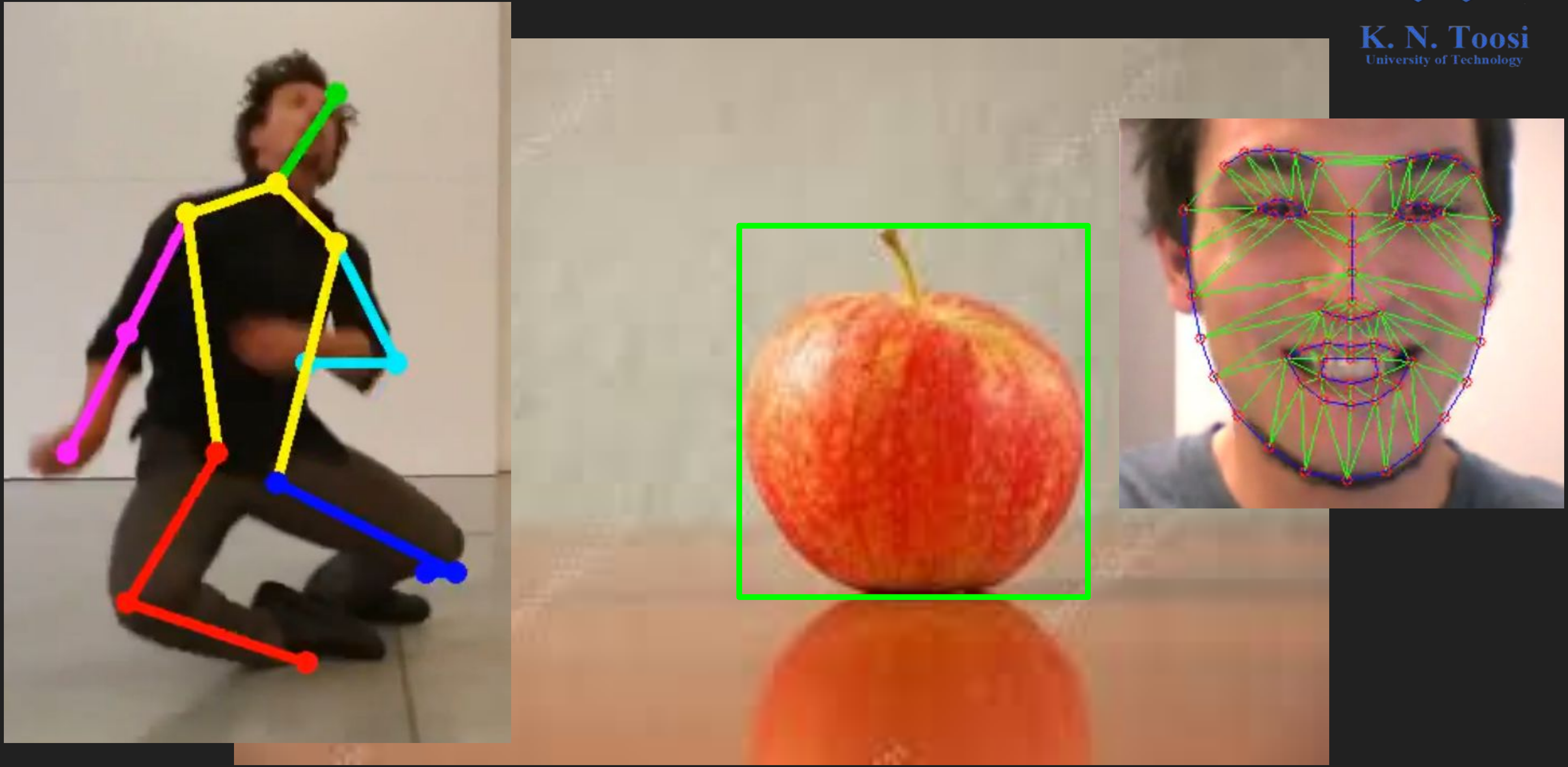
Classification

Regression



Regression

Regression



Learnable Models

Learnable Models: Example

Learnable Models: Example

Learnable Models: Input-output map

$$y = f(x)$$

$$f: R^m \rightarrow R^n$$

Learnable Models: Input-output map

K. N. Toosi
University of Technology

$$y = f(x,\theta)$$

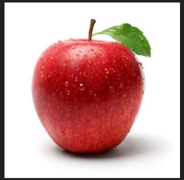
$$f: R^m \rightarrow R^n$$

Learnable Models: Input-output map

$$y = f(x, \theta) \qquad \theta \in R^k$$

$$f: R^m \times R^k \rightarrow R^n$$

Learnable Models: Example



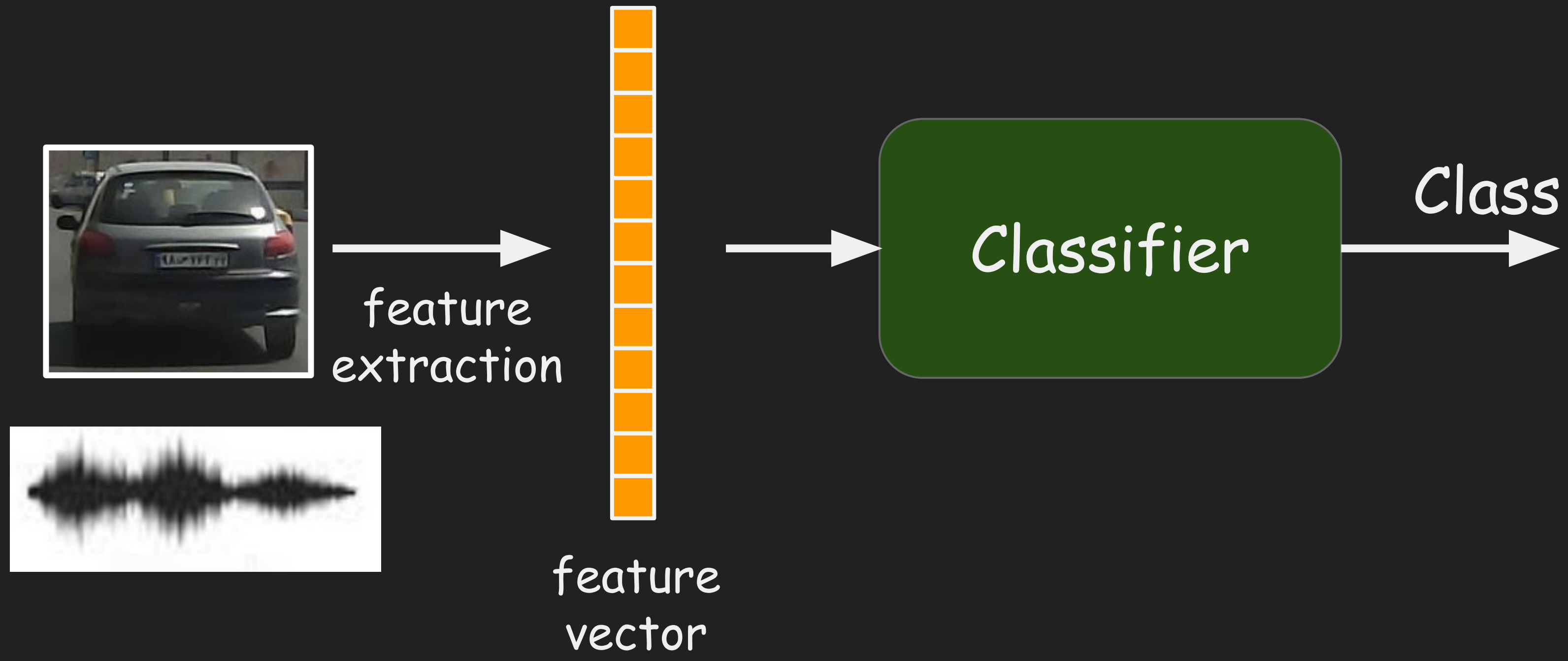
I

x = features(I) → f → y = 0

$$y = f(x, \theta)$$

$$f: R^m \times R^k \rightarrow R^n$$

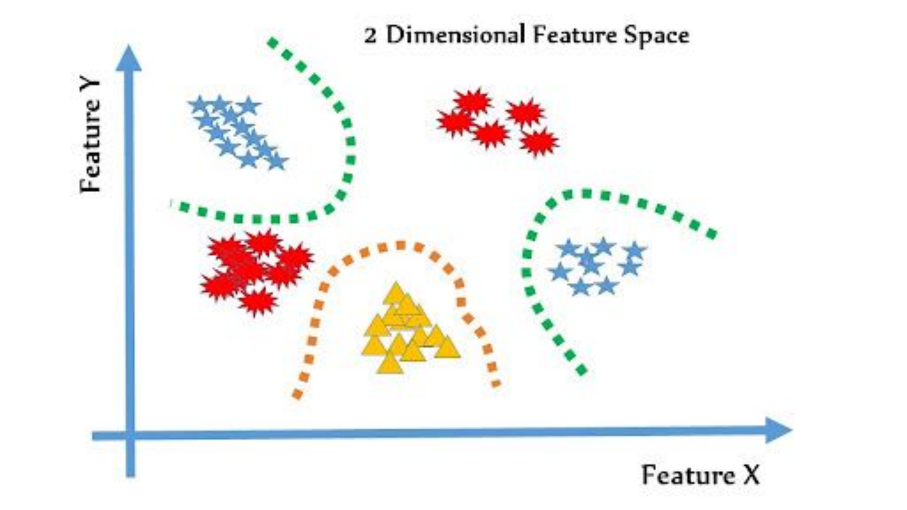
Features



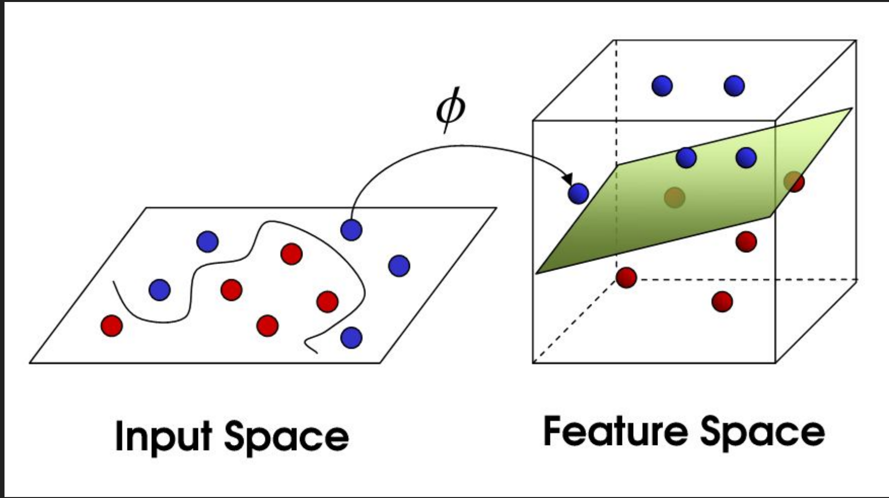
K. N. Toosi
University of Technology

Feature space

K. N. Toosi
University of Technology

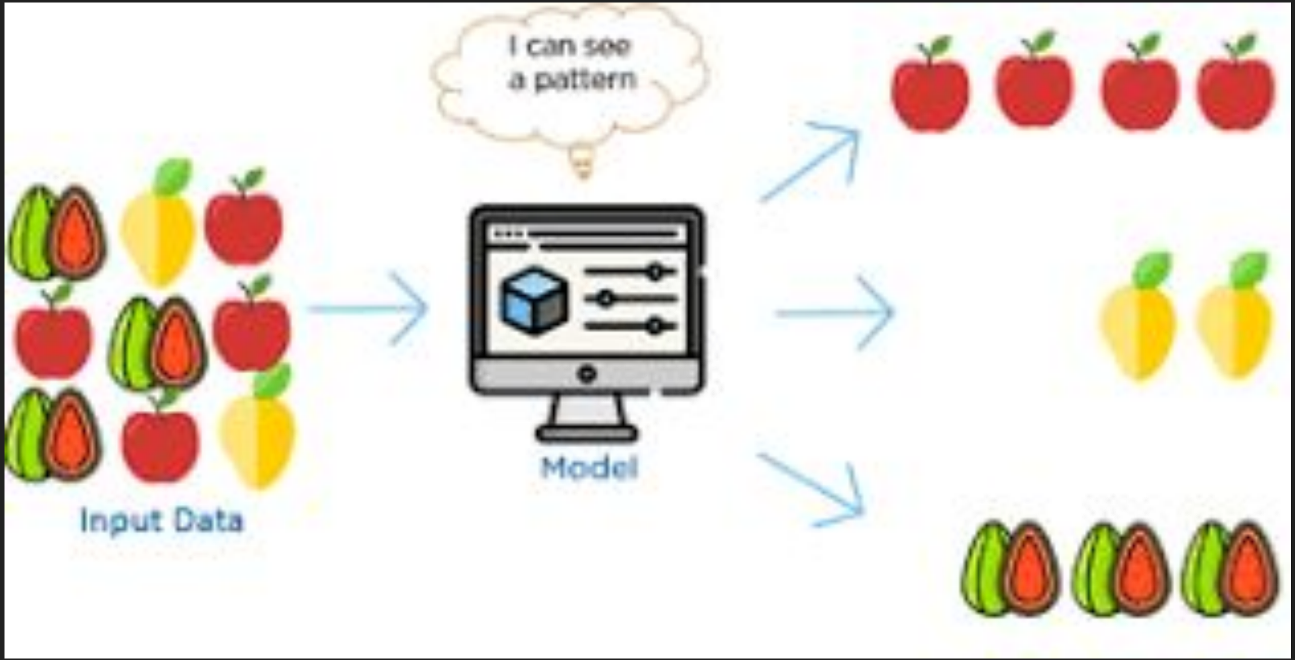


https://www.petersincak.com/news/why-i-do-not-believe-in-error-backpropagation/



https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

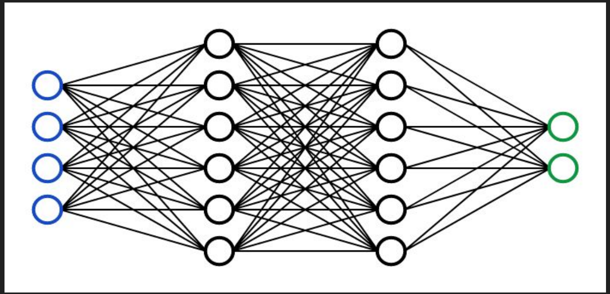
Unsupervised Learning

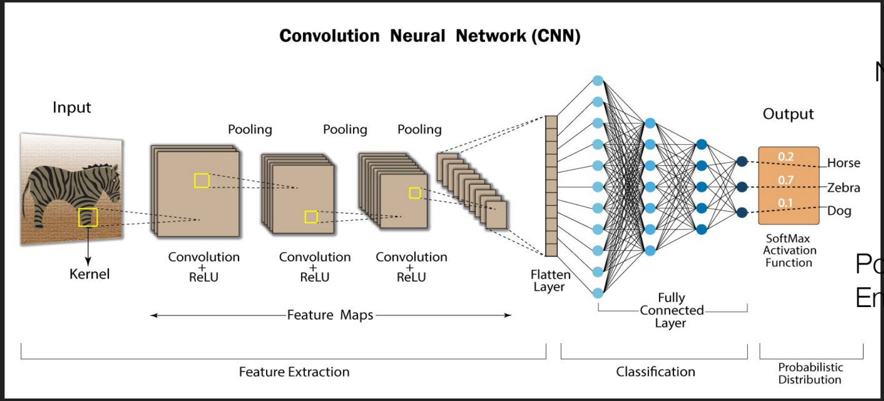


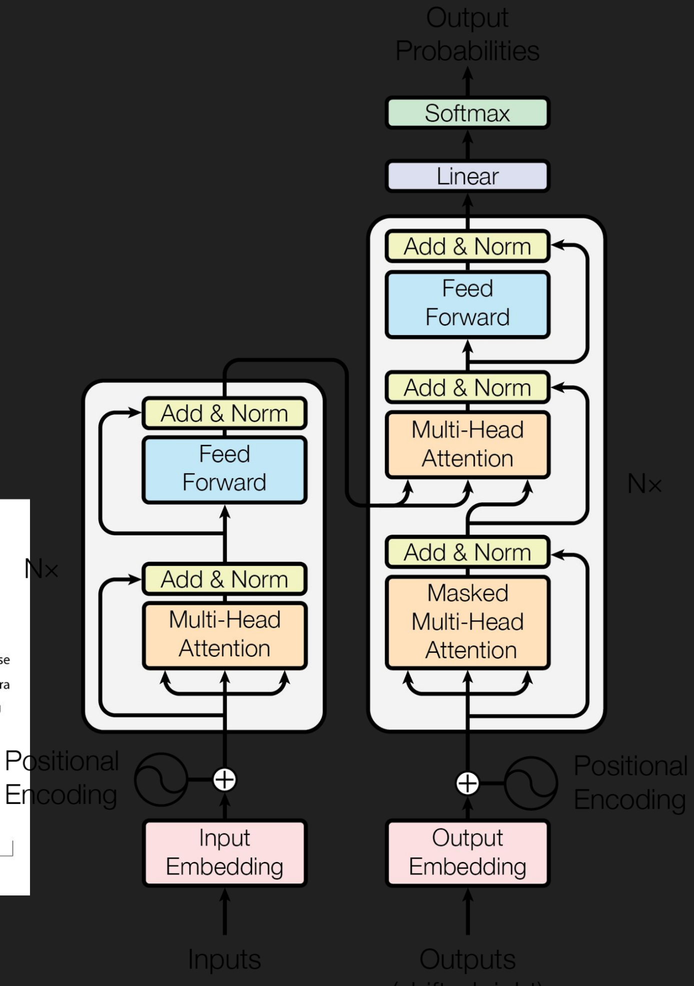
https://towardsdatascience.com/machine-learning-types-and-algorithms-d8b79545a6ec

Neural Networks

K. N. Toosi
University of Technology







Natural Language Processing

ChatGPT

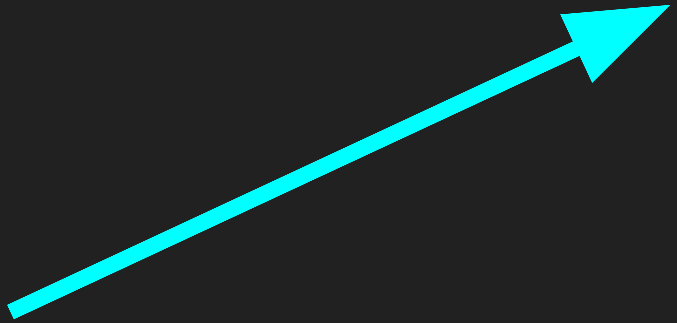
What are the applications of mathematics in AI?

Mathematics forms the foundation of many key concepts in AI, providing the theoretical frameworks, algorithms, and models that drive AI systems. Here are the main areas of mathematics and their applications in AI:

1. Linear Algebra

- **Applications**: Linear algebra is essential for representing and manipulating data in AI models. Vectors, matrices, and tensors are widely used to represent inputs, outputs, and transformations in algorithms like neural networks.
- **Specific Use Cases**:

Vectors



Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i

Cost function

Learning from data

K. N. Toosi University of Technology

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

data output

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

model output given x_i

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

distance

Learning from data

K. N. Toosi
University of Technology

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n} \| f(\theta, x_i) - y_i \|^2$$

distance

Learning from data

K. N. Toosi
University of Technology

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i \,)$$

choose θ such that $C(\theta)$ is small

Learning from data

- Training data $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
  - choose θ such that $f(\theta, x_i)$ is close to y_i
  - cost function:

$$C(\theta) = \Sigma_{i=1..n}\, d(\, f(\theta, x_i), y_i\,)$$

$$\theta^* = \text{argmin}_\theta\, C(\theta)$$

Learning from data

What if we need to model to also give us a measure of uncertainty?

Learning from data

How to choose the cost function?

$$C(\theta) = \Sigma_{i=1..n} \| f(\theta, x_i) - y_i \|^2$$

why not $C(\theta) = \Sigma_{i=1..n} \| f(\theta, x_i) - y_i \|$?

Learning from data

Sometimes we need a **probabilistic framework** to choose the right model/cost function.

Course Overview

K. N. Toosi
University of Technology

- Part I: Linear Algebra and Matrix Analysis
- Part II: Multivariate Calculus
- Part III: Optimization
- Part IV: Probability and Statistics