Machine Learning
DA220

Soumitra Samanta
soumitra.samanta@gm.rkmvu.ac.in
Office: PB405

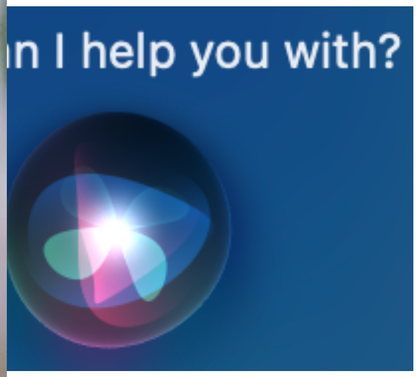
ML Class schedule

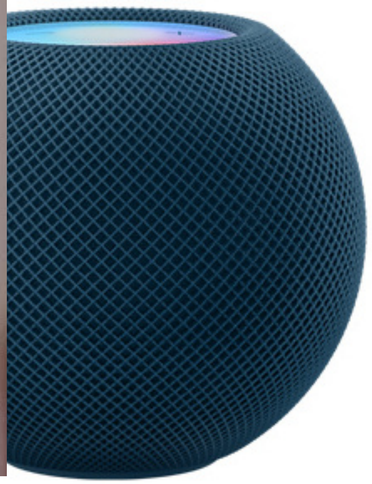
- Tuesday - 10:00 AM - 12:00 PM (Arya lab)
- Thursday - 10:00 AM - 12:00 PM (Arya lab)
- Programming lab
- TA:
  - Rajdeep Mondal (2nd yr. PhD student)
    - Room no.- PB412 or MB214
  - Suvajit Patra (1st yr. PhD student)
    - Room no.- MB214

What is machine learning ?

Some applications and motivation

- Virtual a





Images: Wikipedia, Amazon, Apple and Google
Video: YouTube

Language processing

- Question-answering system in natural language
  - IBM Watson
    - Won first prize in Jeopardy, 2011

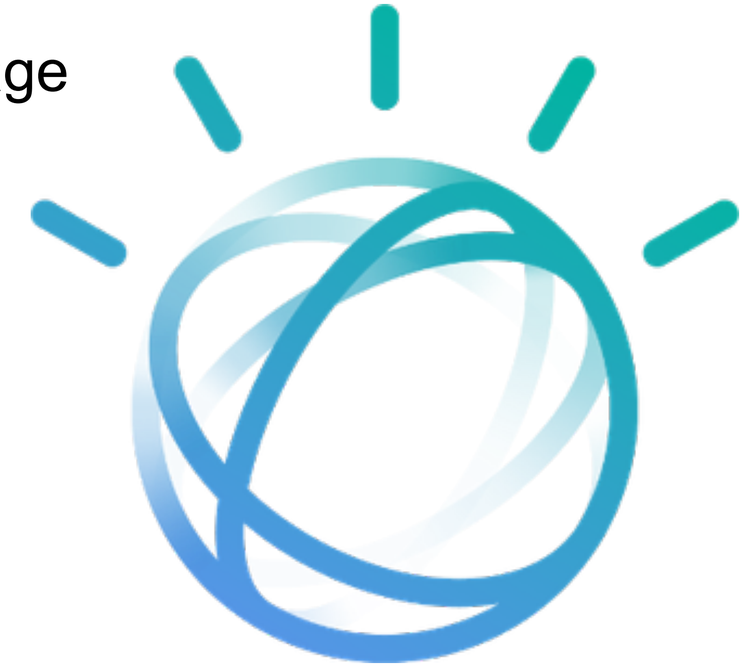
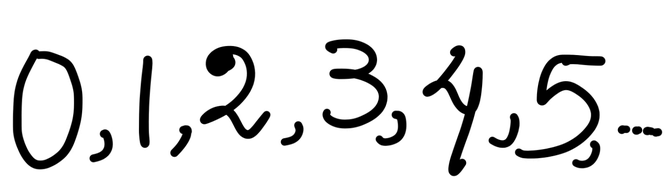


Image: Wikipedia

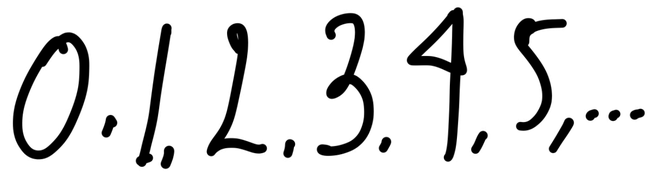
Some applications and motivation (Cont.)

Video: YouTube

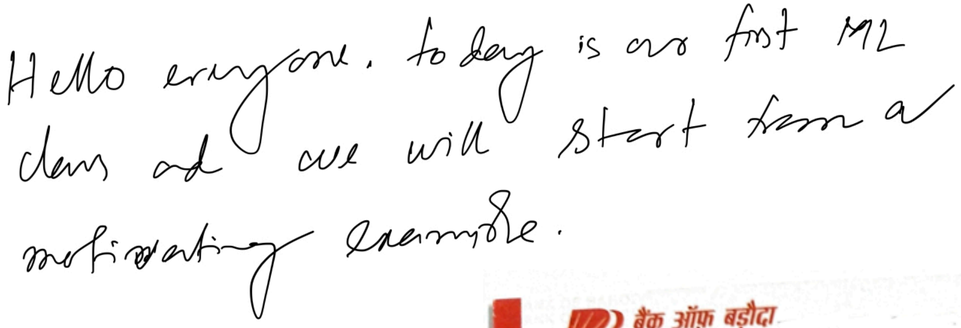
Hand-written character recognition



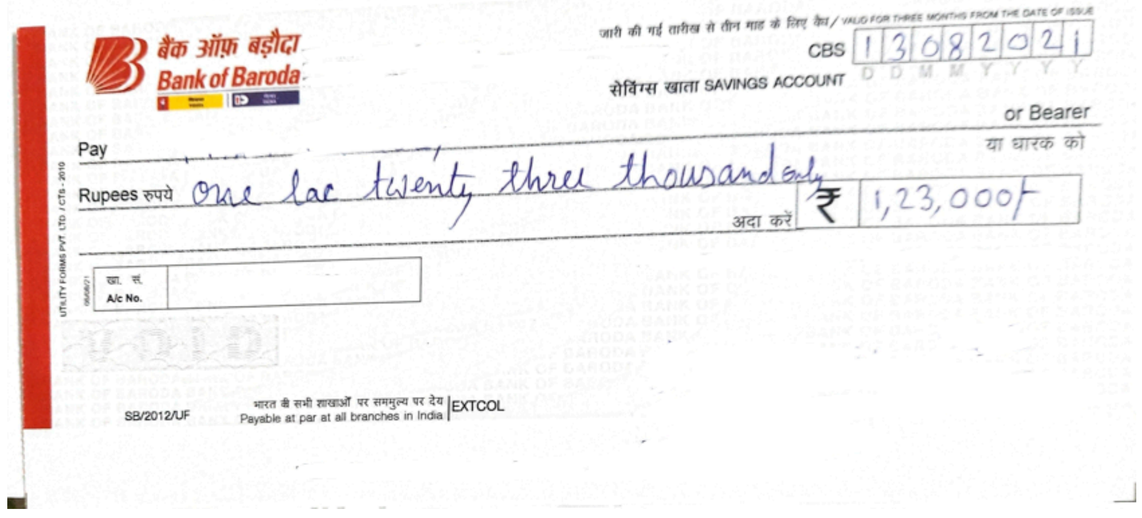
Person-1



Person-2



Hello everyone, today is our first ML class and we will start from a motivating example.



Images: Soumitra

Hand-written digit recognition

- MNIST digit dataset[1]
  - Image: 28×28

THE MNIST DATABASE

of handwritten digits

Yann LeCun, Courant Institute, NYU
Corinna Cortes, Google Labs, New York
Christopher J.C. Burges, Microsoft Research, Redmond

*Please refrain from accessing these files from automated scripts with high frequency. Make copies!^*

The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.

It is a good database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on preprocessing and formatting.

Four files are available on this site:

```
train-images-idx3-ubyte.gz:  training set images (9912422 bytes)
train-labels-idx1-ubyte.gz:  training set labels (28881 bytes)
t10k-images-idx3-ubyte.gz:   test set images (1648877 bytes)
t10k-labels-idx1-ubyte.gz:   test set labels (4542 bytes)
```

please note that your browser may uncompress these files without telling you. If the files you downloaded have a larger size than the above, they have been uncompressed by your browser. Simply rename them to remove the .gz extension. Some people have asked me "my application can't open your image files". These files are not in any standard image format. You have to write your own (very simple) program to read them. The file format is described at the bottom of this page.

The original black and white (bilevel) images from NIST were size normalized to fit in a 20x20 pixel box while preserving their aspect ratio. The resulting images contain grey levels as a result of the anti-aliasing technique used by the normalization algorithm. the images were centered in a 28x28 image by computing the center of mass of the pixels, and translating the image so as to position this point at the center of the 28x28 field.

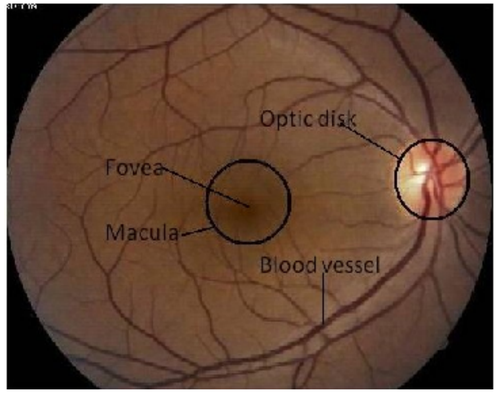
- Extended MNIST[2]

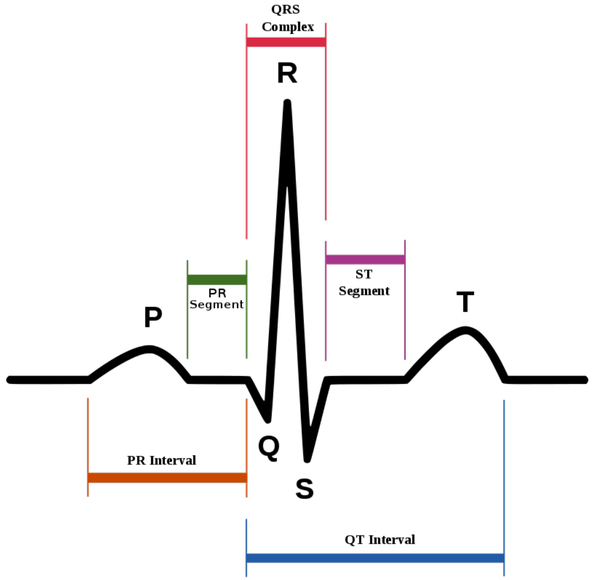
[1]http://yann.lecun.com/exdb/mnist/
[2]https://www.nist.gov/itl/products-and-services/emnist-dataset
https://cs.stanford.edu/people/karpathy/convnetjs/demo/mnist.html

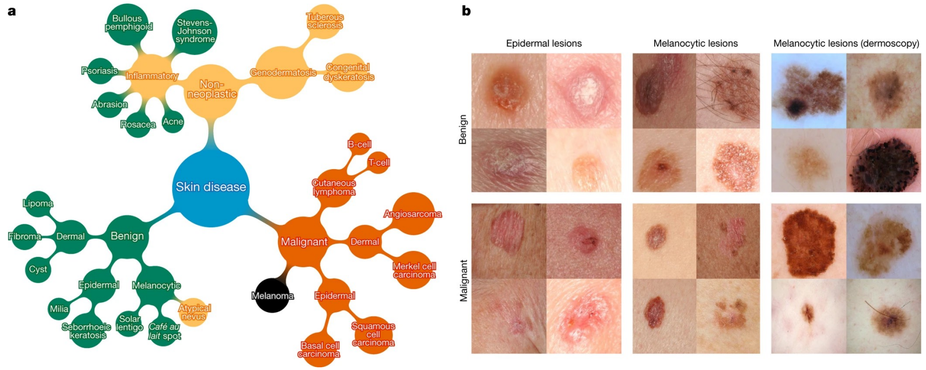
Image: Wikipedia

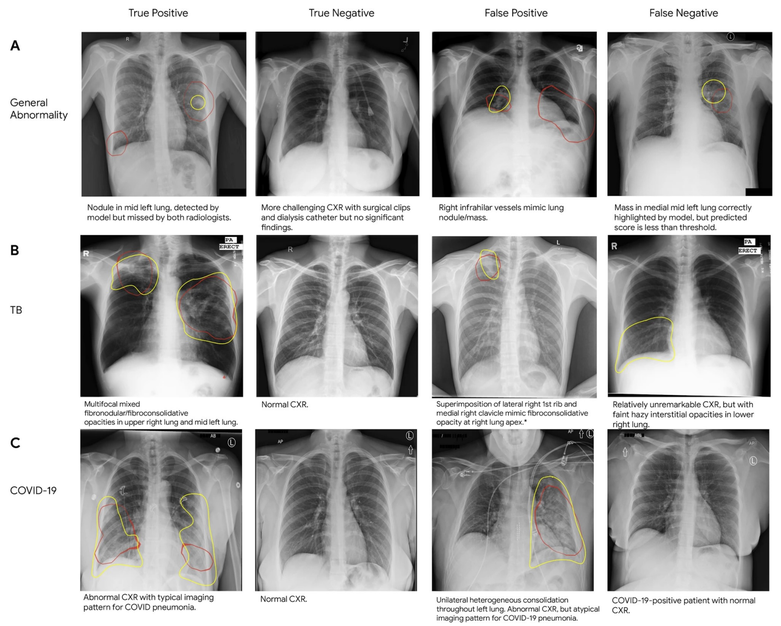
Computer aided diagnosis

- Automation understanding of biosignals
  - EEG, ECG, EGG etc.
- Diabetic retinopathy
- Skin cancer detection
- Abnormal chest x-ray detection
- Desiged to assist (not replace) physicians





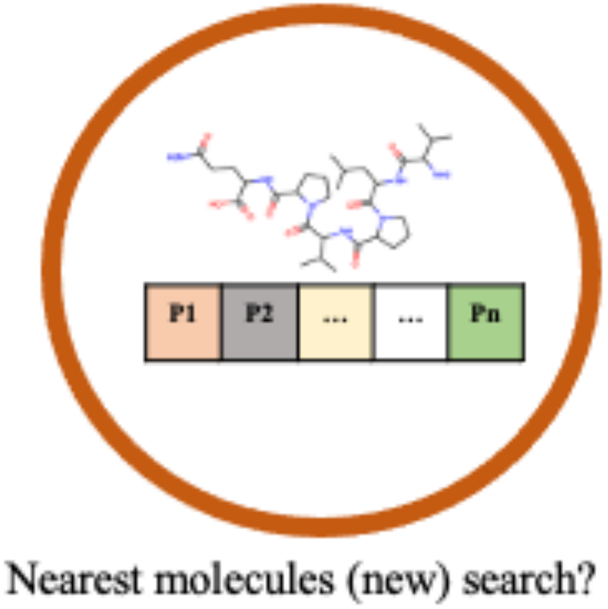


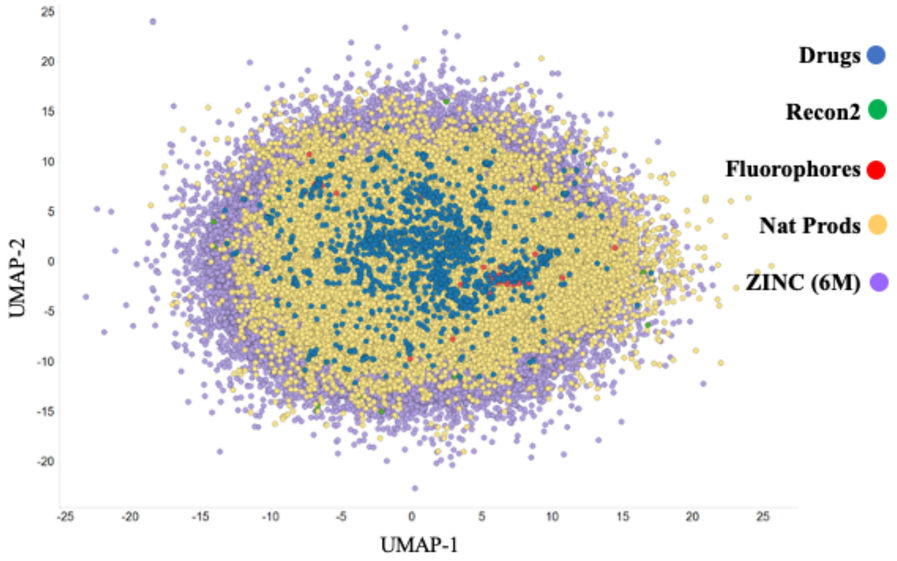


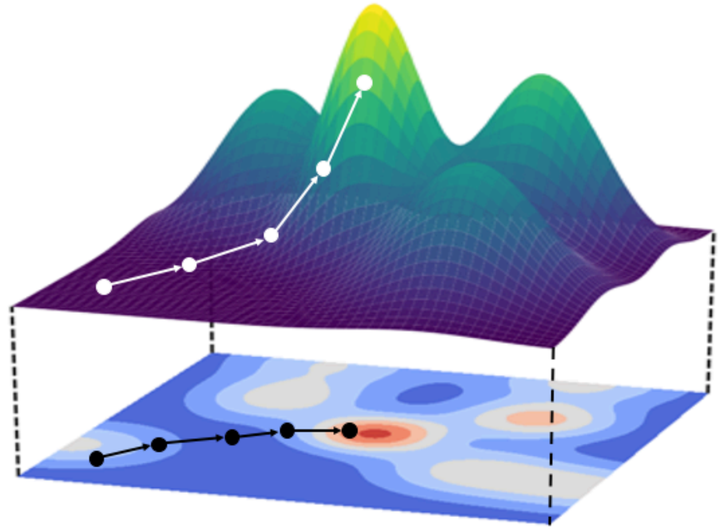
Images: Wikipedia, Esteva et al. Nature 2017, Nabuli et al., Sci. rep. 2021, Soumitra

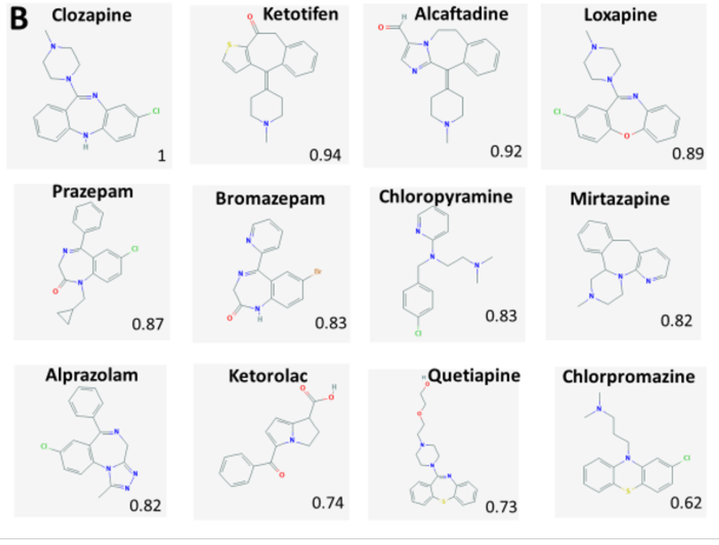
Drug discovery

- Search (drug) molecule with desired properties









Images: Soumitra

Protein-folding

- Prediction of protein structure
  - CASP: https://predictioncenter.org/casp15/index.cgi
  - AlphaFold2 (DeepMind)- CASP-14, 2020

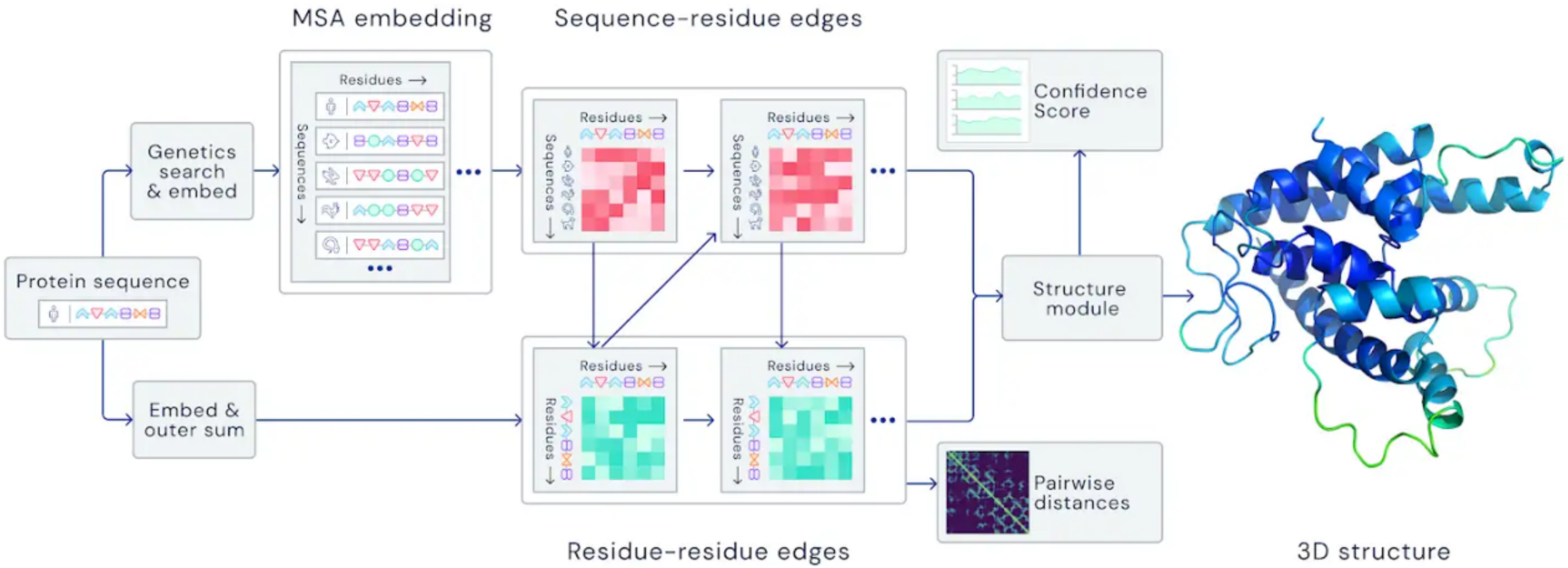


Image: Wikipedia

Bioinformatics

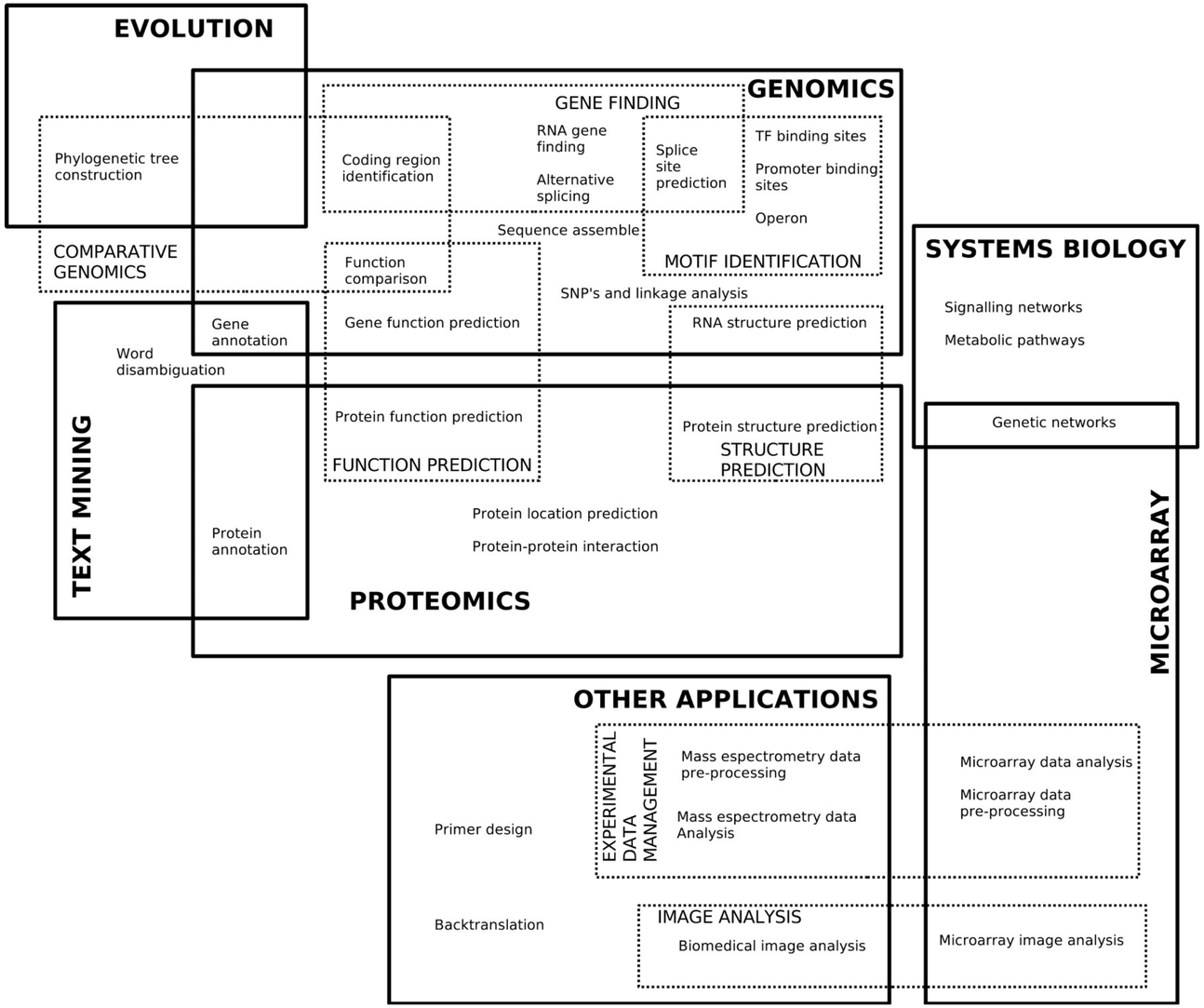
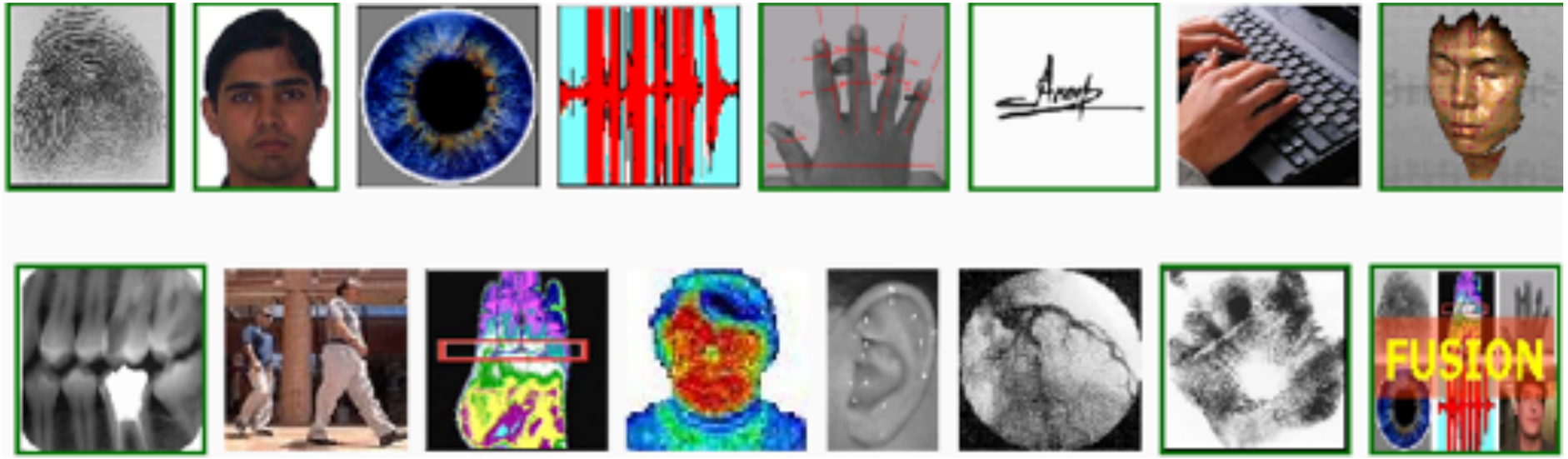


Image: Larranaga et al. BiB, 2006

Biometric recognition

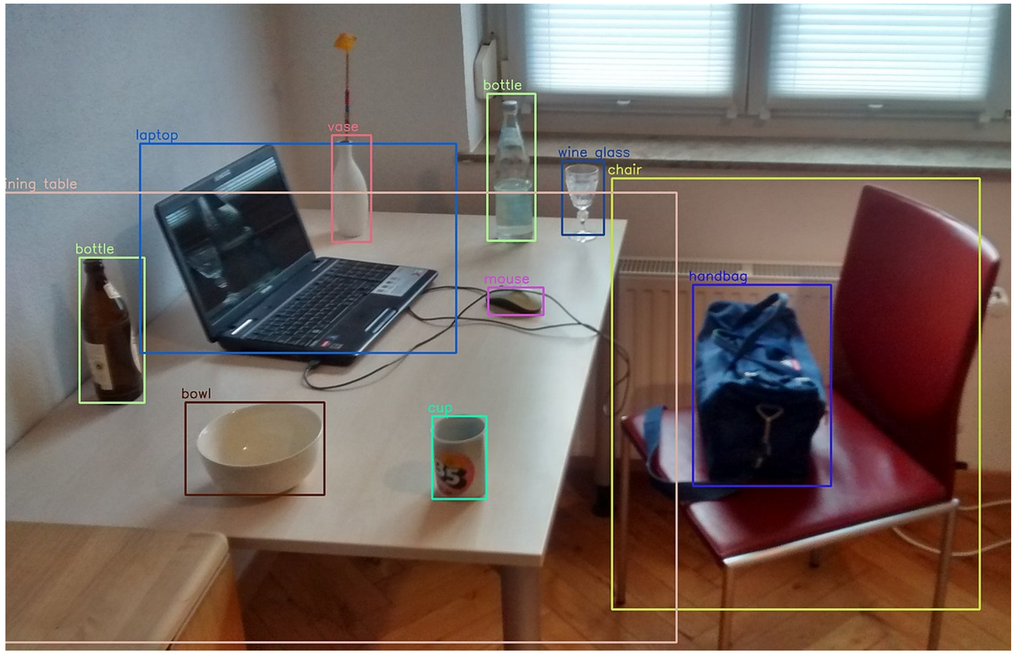
Is this Osama's fingerprint



Multimodal biometric

Images: Google

Object detection



PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

Started in 2005, 4-object class

IMAGENET

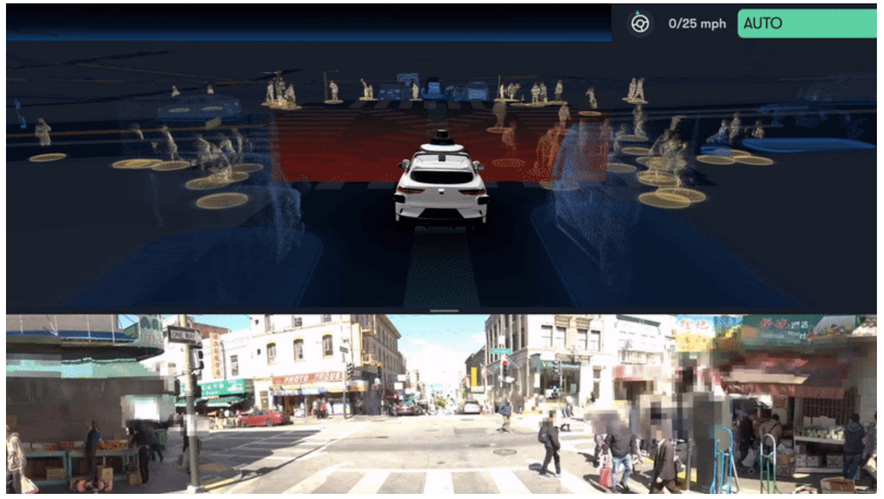
Started in 2010, 1000-object class

Images: Github, Wikipedia, imagenet, Pascal

Self driving car

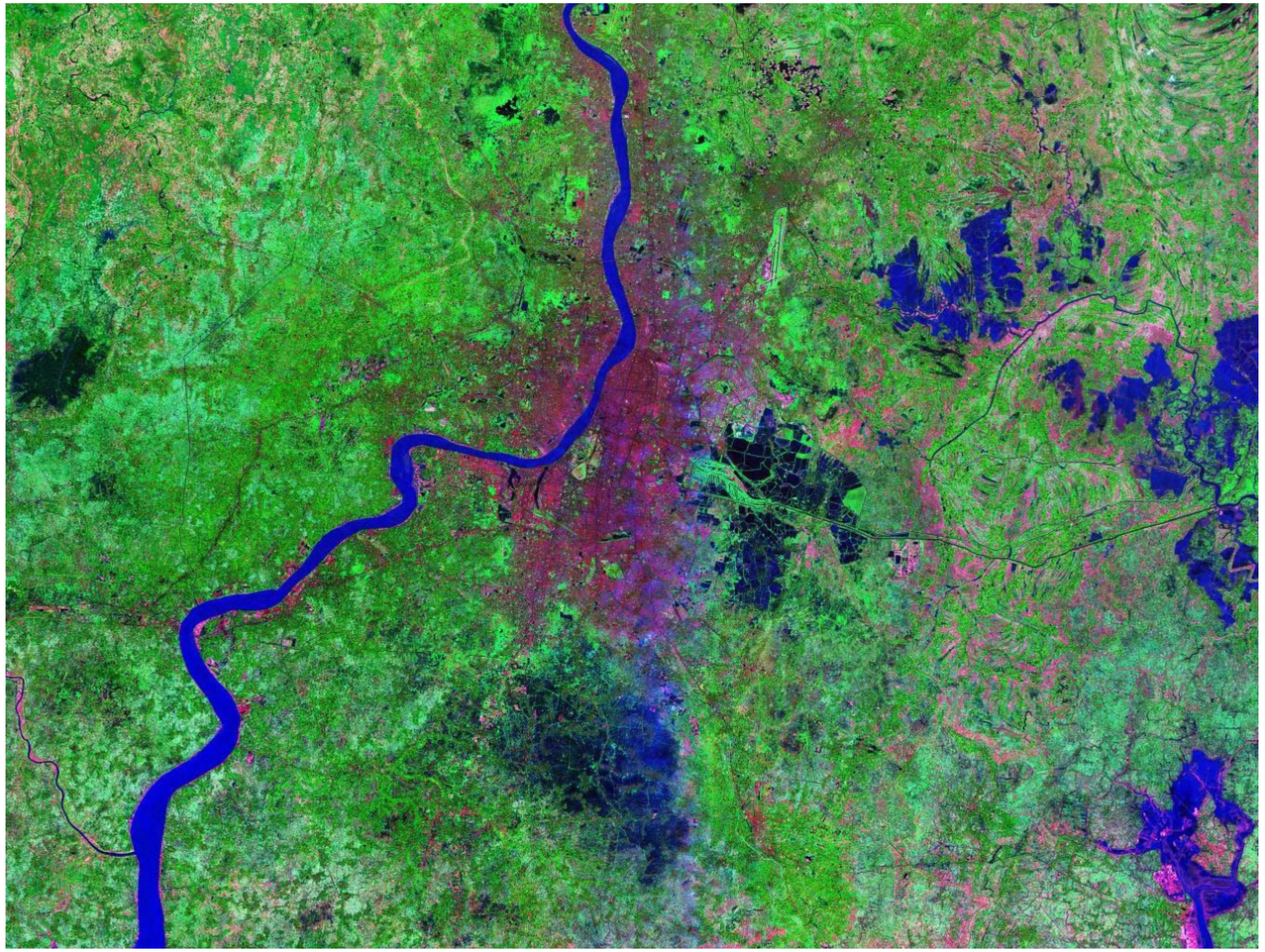
- DARPA grand challenge
  - 2005- Stanley

- Every car company jump into self driving car
  - Tata, BMW, GM, Nissan, Mercedes-Benz, etc.
- Big tech company
  - Google (Waymo), Apple, Uber, Tesla etc.



Images: Wikipedia, Waymo

Satellite image analysis



Where are water body, concrete, and green area?

Image: http://geology.com/world-cities/

AlphaGo

- Computer program defeat professional human Go player
  - Beat
  - Beat world no 1 Go player Ke Jio, 2017

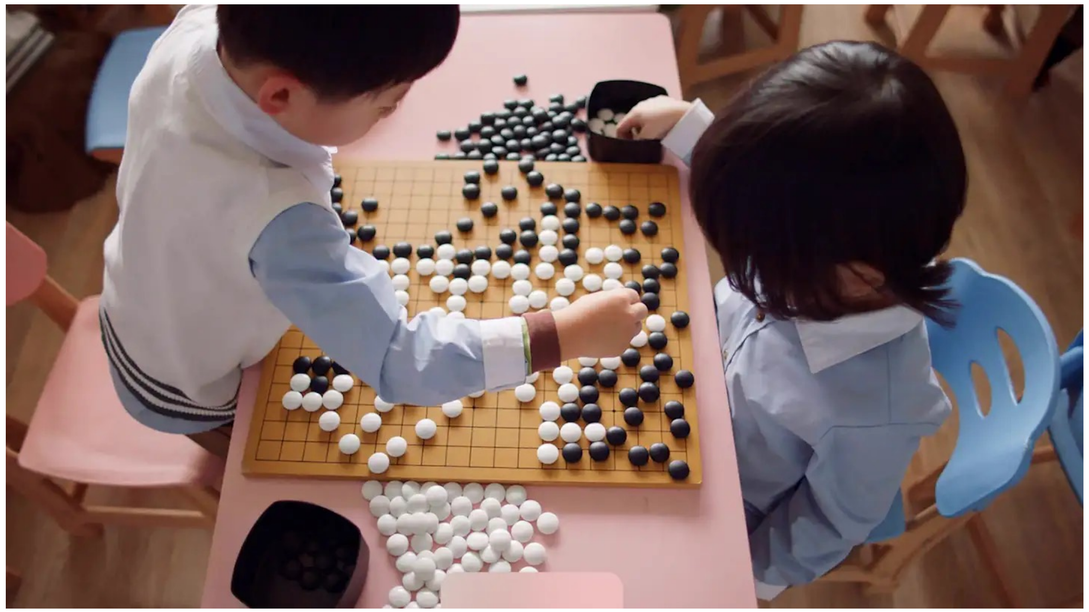
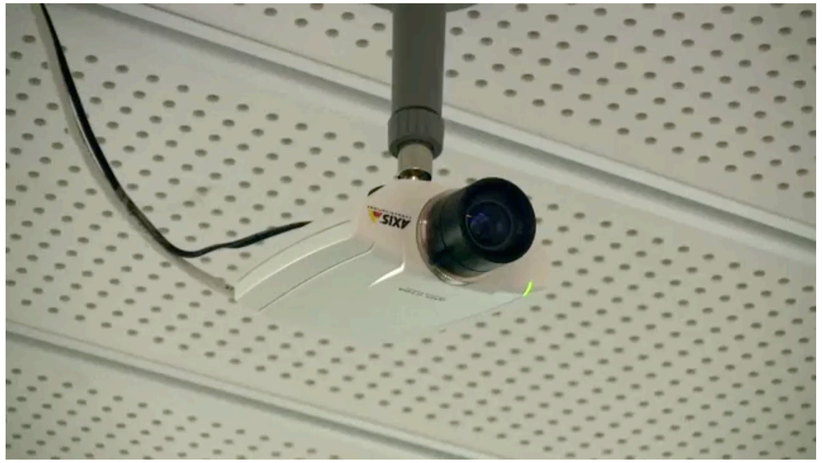
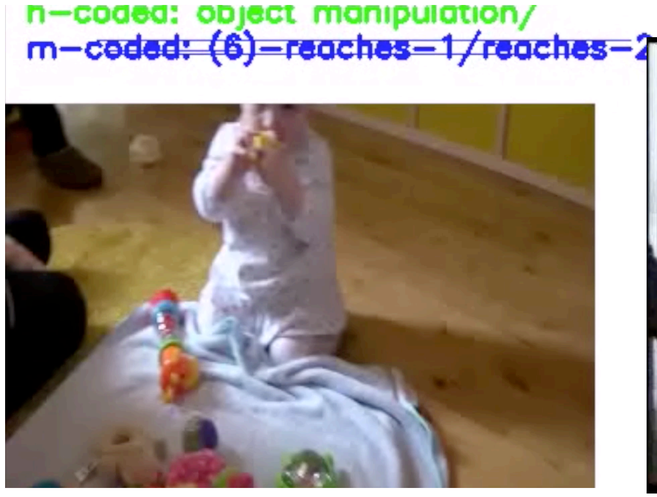


Image: DeepMind
Video: YouTube

Human action detection

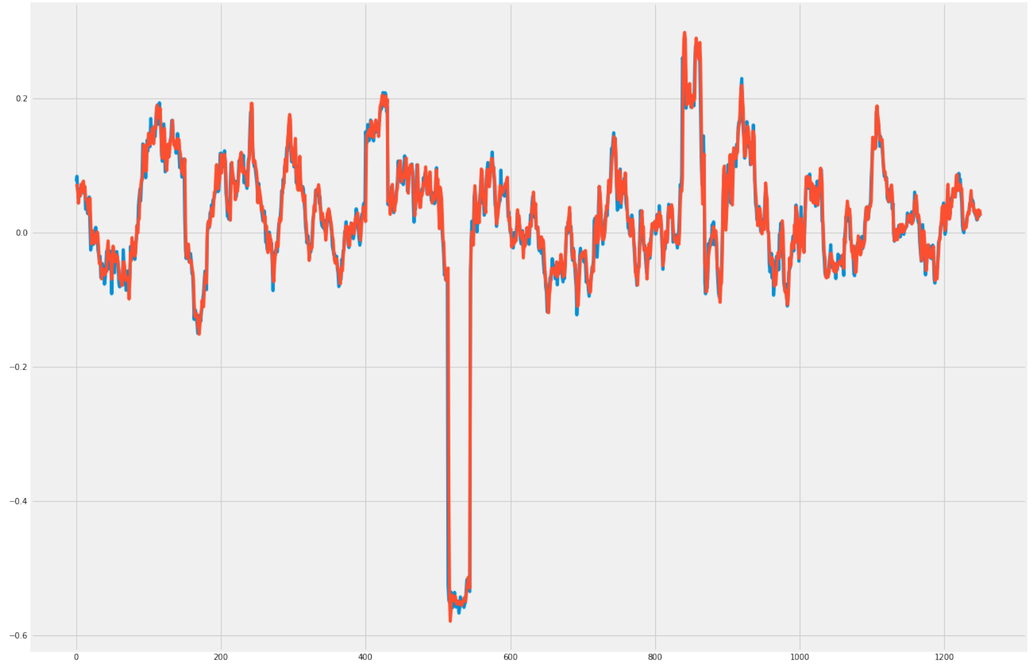




Videos: Youtube, LuCiD

Financial forecasting

- Stock market prediction



- Unwanted bank transaction (fraud) detection

Application of ML

- Applications
  - Image processing
  - Computer vision
  - Speech recognition
  - Natural language processing
  - Optical character recognition
  - Financial forecast
  - Medical diagnosis
  - …

What is ML?

Video: https://www.youtube.com/watch?v=QHH3iSeDBLo

What is machine learning?

- Extract (learn) meaningful information from the examples (training) and answer the query for unseen examples
- Specific task
- How can we allow computer to extract meaningful information from the examples?
  - Design general rules
  - Development of algorithms

Slide credit: Luxburg, 2020

Inference
deductive vs inductive

- Deductive
  - Process of reasoning from one or more statements (premises) to reach a logical conclusion
- Example-1:
  - Premises-1: All RKMVERI students are excellent
  - Premises-2: Amol study in the RKMVERI
  - Conclusion: Amol is an excellent student
- In general:
  - If $A \implies B$ and $B \implies C$
  - Then $A \implies C$
- If premises are correct, then the conclusion is certain

Inference
deductive vs inductive

- Inductive
  - Method of reasoning in which a body of observation is synthesised to come up with a general principle
- Example-1:
  - Very often, we drop lots of things
  - All the times, the things fall downwards, but not upwards.
  - We can conclude that, if we drop things, likely they always fall downwards
- The truth of the conclusion is probable, based on the evidence given so far
- So, conclusion is not certain!
  - Example?

Example: credit: Luxburg, 2020

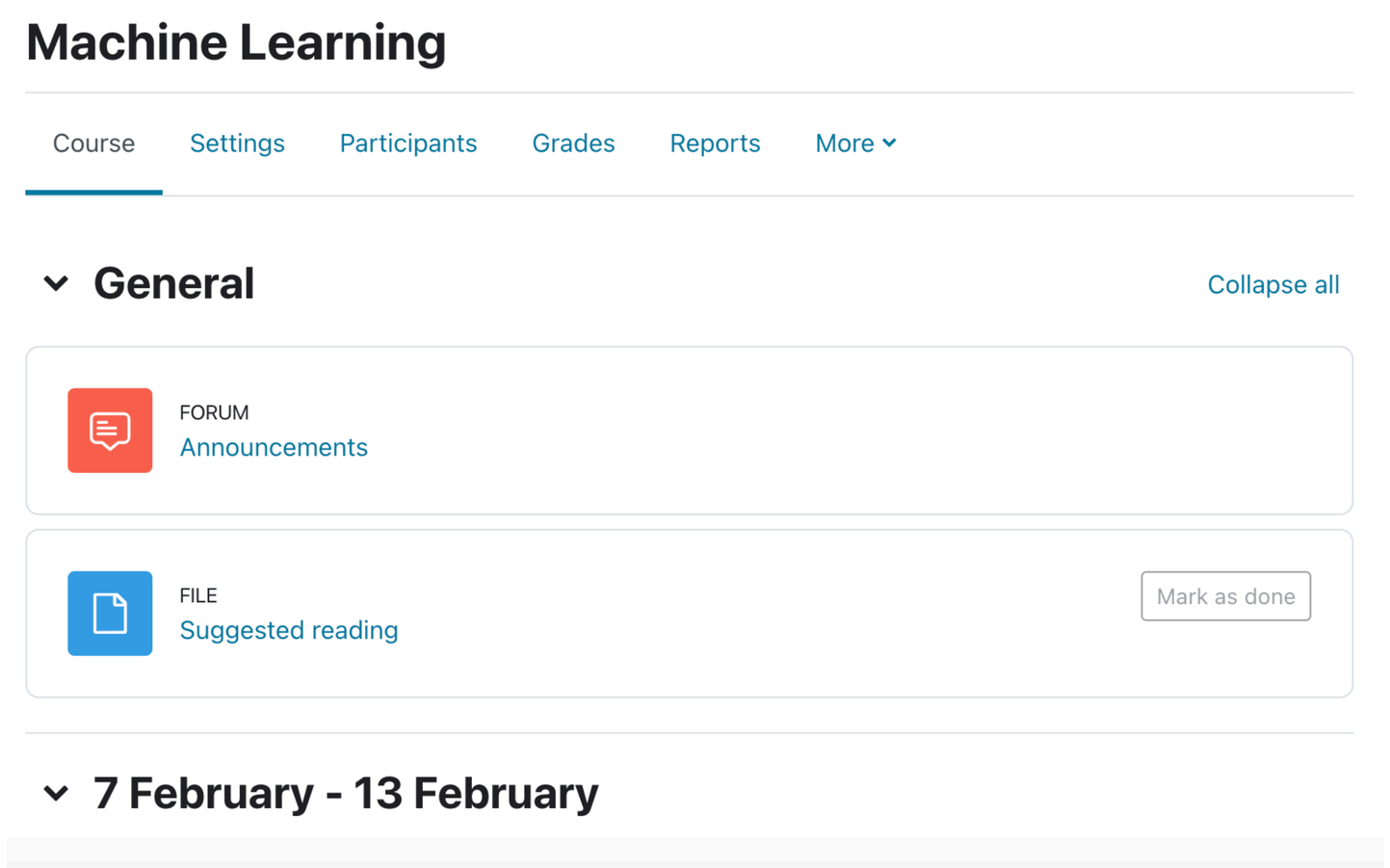
What is machine learning?

- Extract (learn) meaningful information from the examples (training) and answer the query for unseen examples
- Specific task
- How can we allow computer to extract meaningful information from the examples?
  - Design general rules
  - Development of algorithms
- Automate the process of inductive inference

Slide credit: Luxburg, 2020

Course logistics

Course page in xlms



https://xlms.rkmvu.ac.in/course/view.php?id=47

Prerequisites

- Mathematics
  - Linear Algebra, Multivariate Calculus, Basis Optimisation and Basic probability
  - No worries, we will touch some background when we need
- Computer programming: Any one from C/C++/Python (recommended for the class project and assignments)/MATLAB/Octave
- Basic concept in Algorithms and Data Structure

Linear Algebra

- Vector space
  - Definition, Basis, Dimension, Eigen value and Eigen vectors etc.
- Matrix
  - Addition, Multiplication, Trace, Inverse etc.
  - Positive definite matrices, Singular value decomposition etc.

Multivariate Calculus

- Derivative, Partial derivative, Taylor series expansion, Chain rules etc.

Basic Optimisation

- Convex set, Convex hull, Convex function
- Gradient of a function, Hessian
- Constrained and Unconstrained optimisation problem, Optimality condition

Probability

- Definition, Random variables, Distribution function and their different variants
- Conditional probability, Independence, Expectation, Variance, Moments, Entropy
- Low of large numbers, Central limit theorem

Evaluation

- Mid-term: 15%
- End-term: 50%
- Assignments: 15%
- Project: 20%

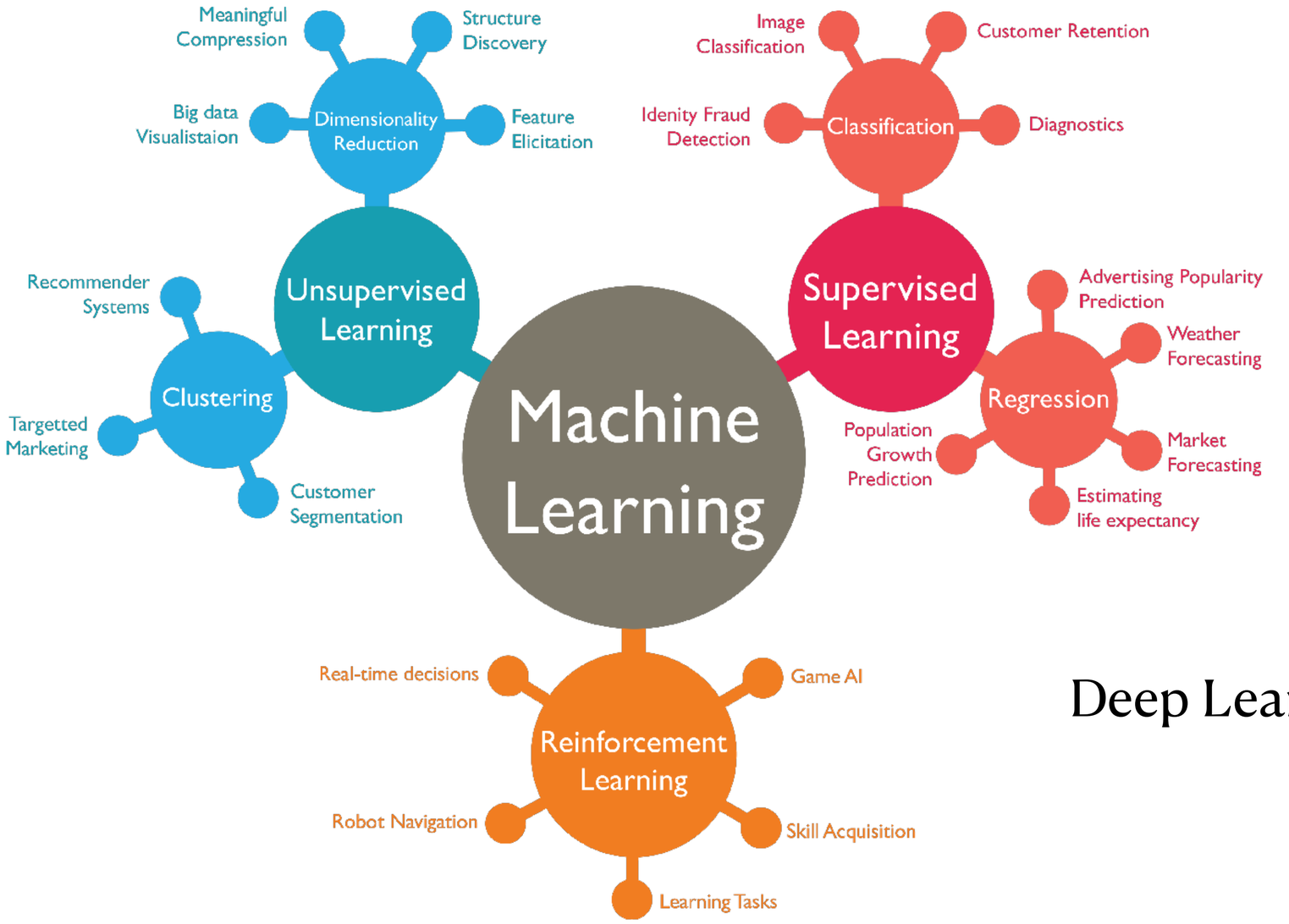
Assignments

- Mostly implementation of different ML algorithms
  - Python (preferable)/C/C++/MATLAB/Octave
- There might be one-two theoretical assignments
- Submission deadline is strict and weightage reduce to 40%/day after the deadline
  - We will consider 11.59PM as our day end

Projects

- Can be done in a group (max two students)
- Define your own project
- Submit a one page project proposal- within fixed time (first six weeks) 21-03-2023?
- Finished the work within the time-line
- Report submission
  - Submission deadline: 5th June, 2023
  - We will consider 11:59PM as our day end
- Final presentation
  - 20 min (divided into group members)/poster presentation
  - 7th June, 2023

Course syllabus



Deep Learning

Image: https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png

Course materials

- Books: We will follow multiple books for different topics
  - Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, Foundations of Machine Learning, MIT Press, Second Edition, 2018
  - Shai Shalev-Shwartz and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014
  - Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006
  - Kevin Patrick Murphy, Machine Learning: a Probabilistic Perspective, MIT Press, 2012
  - Trevor Hastie, Robert Tibshirani and Martin Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, 2015
  - R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, 2nd ed., Wiley, New York, 2000
  - S. Theodoridis and K. Koutroumbas, Pattern Recognition, Academic Press, San Diego, 1999
  - K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, New York, 1990
  - L. Devroye, L. Gyorfi and G. Lugosi, A probabilistic Theory of Pattern Recognition, Springer, 1996
  - I. Goodfellow, Y Bengio, and A. Courville, Deep Learning, MIT Press, 2016

Academic ethics

- Your grade should reflect your own work
- Copying or paraphrasing someone's work (code included), or permitting your own work to be copied or paraphrased, even if only in part, is strictly forbidden, and will result in an automatic grade of zero for the entire assignment or exam in which the copying or paraphrasing was done.
- So, ask yourself before copying from others
- If you are going to have trouble completing an assignment, talk to the instructor and TA before due date

Data

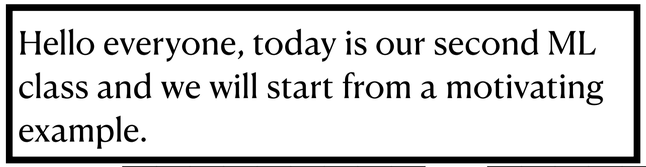
Recap

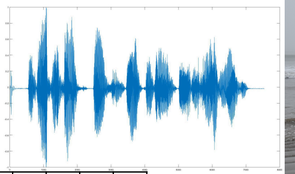
- What is ML?
- Some applications of ML
- Deductive vs Inductive inference

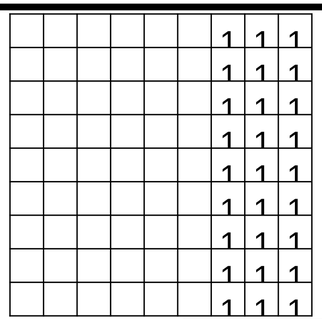
Data

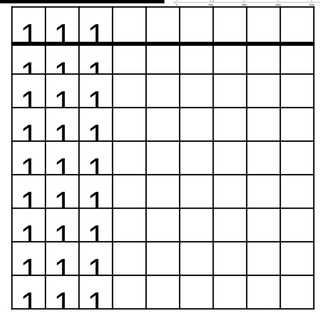
- Different types of data
  - Text, Audio, Images, Videos, etc.
- Representation
  - Example:
    - Hand-written digit recognition
    - Cat vs Dog
    - Stock (particular) price prediction
  - Feature
    - Any distinct aspect, quality or characteristic

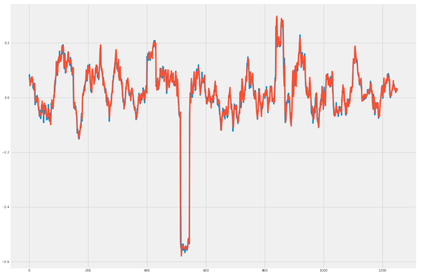
      Ex. numeric (height)
    - Combination of d features is d-dims column vector, we will call it as feature vector
    - d-dims space defined by the feature vectors is called feature space
    - Data are represented as a point in the feature space











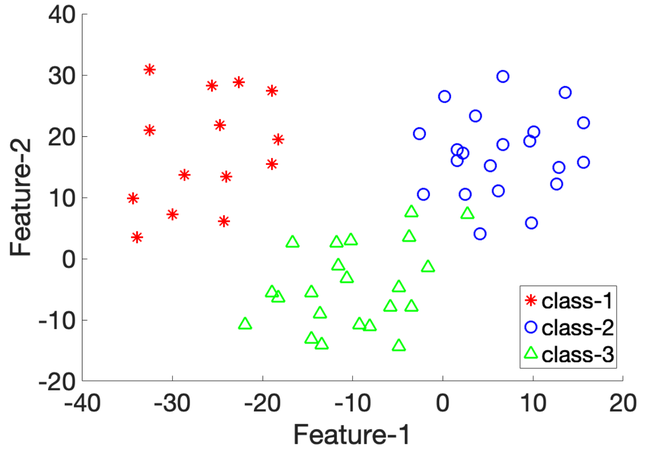
Images: Kaggle, Stackexchange
Video: LuCiD

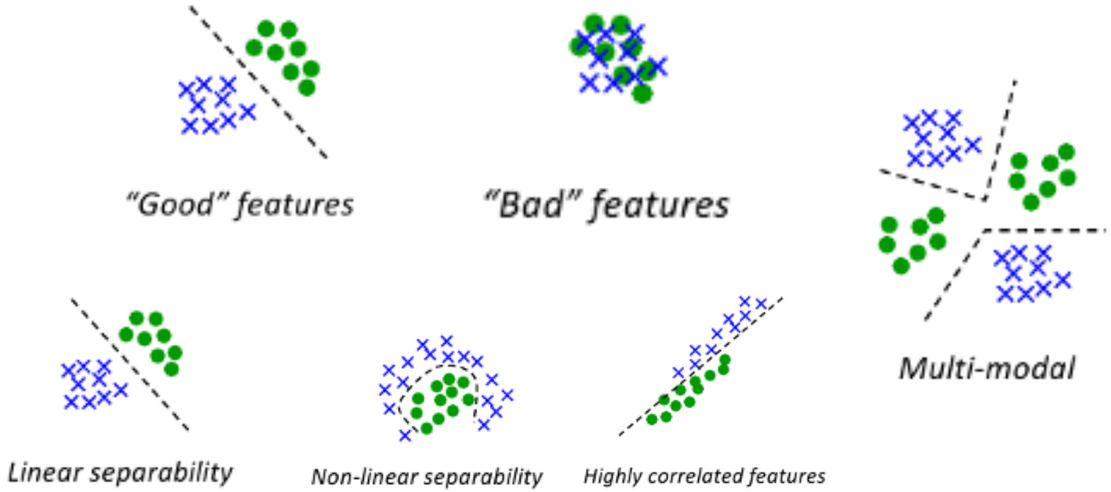
Data (cont.)

- Feature
  - Any distinct aspect, quality or characteristic

    Ex. numeric (height)
  - Combination of d features is d-dims column vector, we will call it as feature vector
  - d-dims space defined by the feature vectors is called feature space
  - Data are represented as a point in the feature space ($\mathcal{X}$)

$$X = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^d \end{bmatrix}$$





Images: Google

Data (cont.)

- Syntax:
  - Data: $(X_i, Y_i); i = 1,...,n$
    - Where X_i is the i^{th} data representation (features) and $X_i \in R^d$
    - Y_i is the i^{th} data label (class) and $Y_i \in R$
  - Let f be a model/algorithm for a particular task
  - $\bar{Y}_i$ is the model/algorithm output: $\bar{Y}_i := f(X_i)$
  - How do you evaluate your model/algorithm?
    - Loss: $\ell(X_i, Y_i, \bar{Y}_i)$

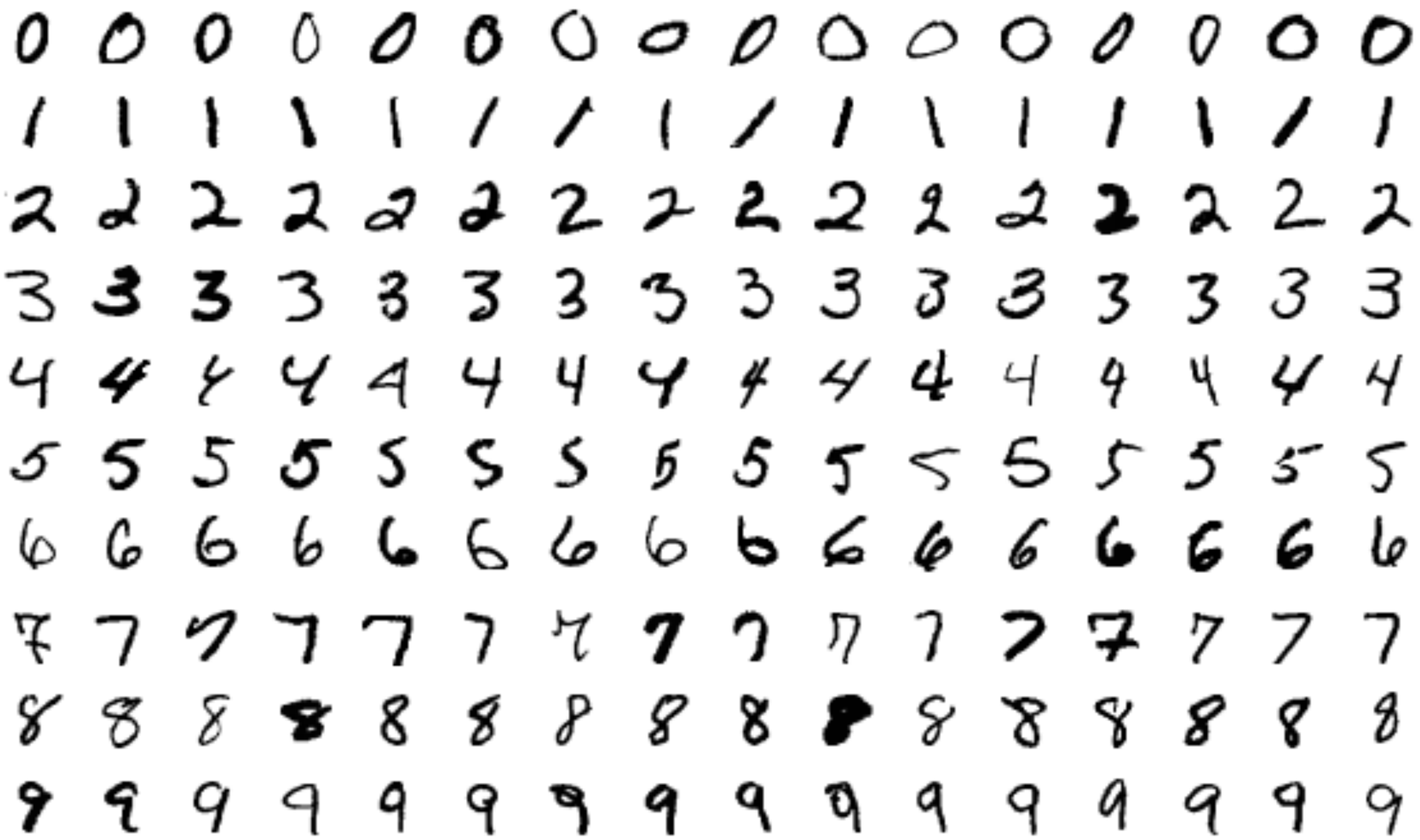
Data similarity

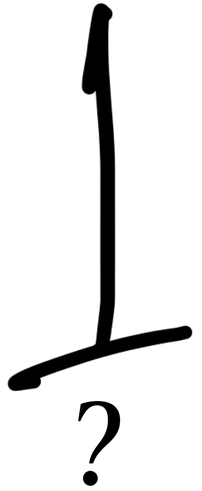
- How do we compare two objects?

Images: Wikipedia

Data similarity (cont.)

- Why we need comparison?





?

?

Image: Wikipedia

Similarity measure

- Distance between two points $X_1 \in R^d$ and $X_2 \in R^d$: $\mathscr{D}(X_1, X_2)$?
  - $\mathscr{D}(X_1, X_2)$ considered a metric if it satisfies the following properties:
    - $\mathscr{D}(X_1, X_2) \geq 0$, $\mathscr{D}(X_1, X_2) = 0$ iff $X_1 = X_2$
    - $\mathscr{D}(X_1, X_2) = \mathscr{D}(X_2, X_1)$
    - $\mathscr{D}(X_1, X_3) \leq \mathscr{D}(X_1, X_2) + \mathscr{D}(X_2, X_3)$
- Commonly used metrics are:
  - Minkowski distance/metric of order p:
    - $$\mathscr{D}^p(X_1, X_2) = \left\{ \sum_{i=1}^{d} \left| x_1^i - x_2^i \right|^p \right\}^{\frac{1}{p}}$$
  - Manhattan or city-block distance
    - $$\mathscr{D}(X_1, X_2) = \sum_{i=1}^{d} \left| x_1^i - x_2^i \right|$$

Similarity measure (cont.)

- Commonly used metrics are:
  - Minkowski distance/metric of order p:
    - $\mathscr{D}^p(X_1, X_2) = \left\{ \sum_{i=1}^{d} \left| x_1^i - x_2^i \right|^p \right\}^{\frac{1}{p}}$
  - Manhattan or city-block distance:
    - $\mathscr{D}(X_1, X_2) = \sum_{i=1}^{d} \left| x_1^i - x_2^i \right|$
  - Euclidean distance:
    - $\mathscr{D}(X_1, X_2) = \left\{ \sum_{i=1}^{d} \left| x_1^i - x_2^i \right|^2 \right\}^{\frac{1}{2}}$
  - Chebyshev distance:
    - $\mathscr{D}(X_1, X_2) = \max_i \left\{ \left| x_1^i - x_2^i \right| \right\}$

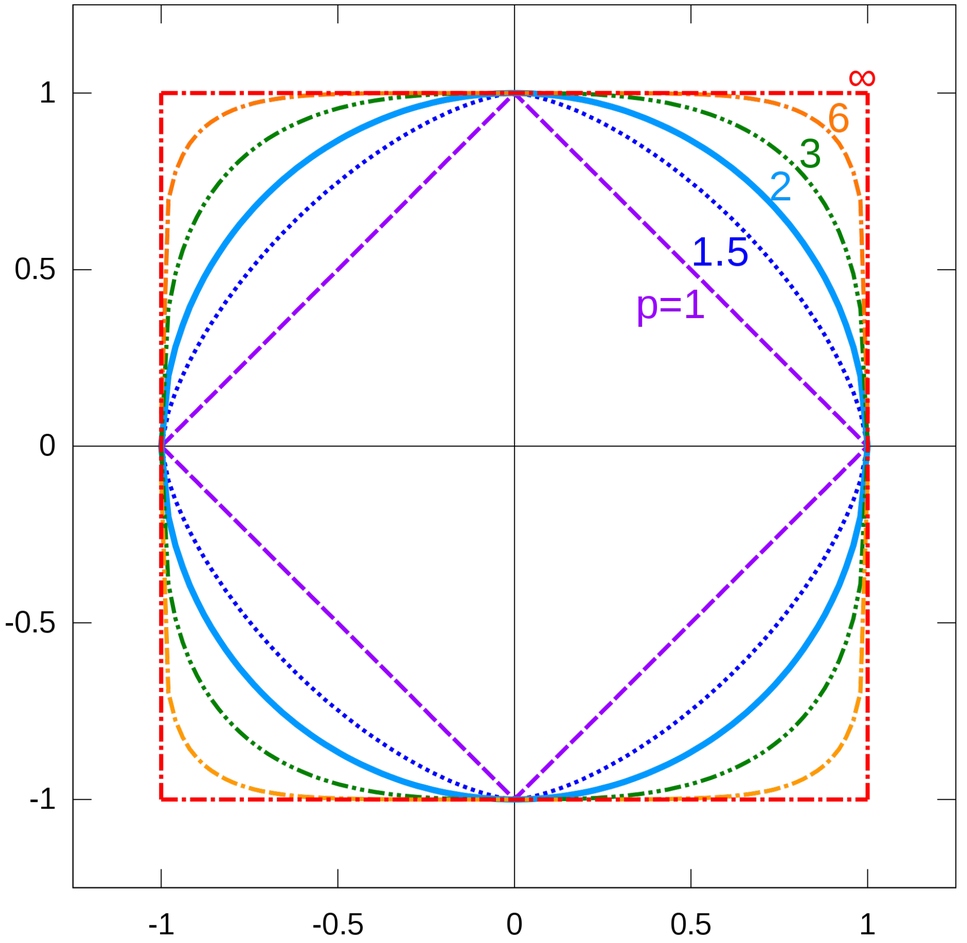


Image: Wikipedia

Homework-1

- Can you draw this type of figure in 2D & 3D?
  - Use Minkowski distance for $p = -2, -1, 0, 1, 1.5, 2, 3, 6, \infty$
  - Draw all the points within the the intervals has unit distance from the origin for all p's:
    - 2D: $x \in [-1, 1]$ and $y \in [-1, 1]$
    - 3D: $x \in [-1, 1]$; $y \in [-1, 1]$ and $z \in [-1, 1]$

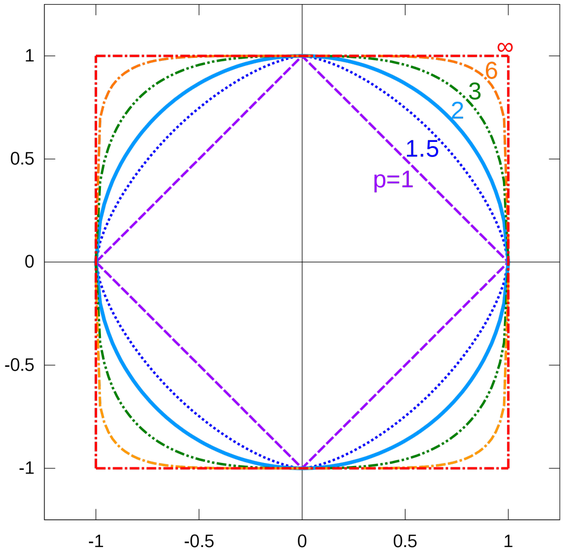


Image: Wikipedia

Which data to evaluate the model?

- Data partition/division:
  - Training, Validation and Testing
    - Training- 50%
    - Validation - 20%
    - Testing - 30%

Which data to evaluate the model?

- Model:
  - Let f be a model/algorithm for a particular task
    - Example: Regression $f : X \to \mathbf{R}$
      - $f(X) = w^0 + w^T X$
  - Model parameters: w^0, $w = \begin{bmatrix} w^1 \\ w^2 \\ \vdots \\ w^d \end{bmatrix}$
  -

Which data to evaluate the model?

- Data partition/division:
  - Training, Validation and Testing
    - Training- 50%
    - Validation - 20%
    - Testing - 30%

Homework-2

- Create a random dataset in $\mathbf{R}^{100}$ of size 50000 with random class labels from $\{1,2,3,4\}$. Now partition the data into the following subsets:
  - Training- 50%
  - Validation - 20%
  - Testing - 30%
  - Plot (bar) the frequency of each class label for each subset.

How do you evaluate your model/algorithm?

- How do you evaluate your model/algorithm?
  - Accuracy (%) - classification
  - Mean square error (MSE) - regression
  - …