ED5340 - Data Science: Theory and Practise

L16 - Introduction To Machine Learning

Ramanathan Muthuganapathy (https://ed.iitm.ac.in/~raman)
Course web page: https://ed.iitm.ac.in/~raman/datascience.html
Moodle page: Available at https://courses.iitm.ac.in/

ML - Definition?

Tom Mitchell

- A computer program is said to learn from Experience E wrt some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
- Broadly - Using Data to answers certain questions.

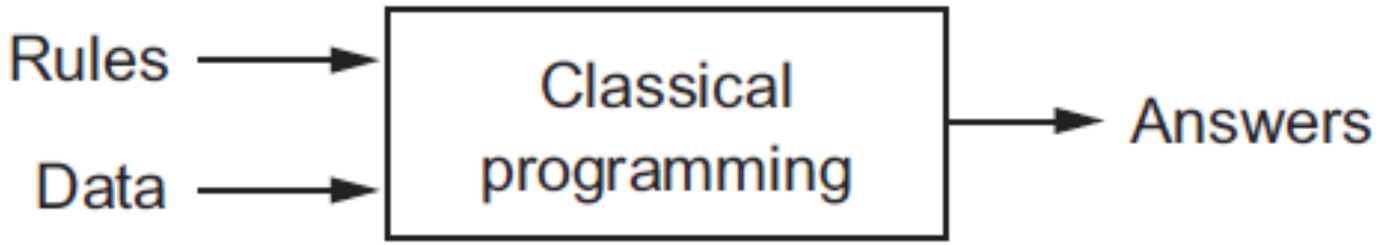
ML - Definition?

Email spam classifier

- A computer program is said to learn from Experience E wrt some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
  - T - Classifying emails as spam or not spam
  - E - Labelling email as spam or not (Data)
  - P - The number (fraction) of emails correctly classified as spam or not.

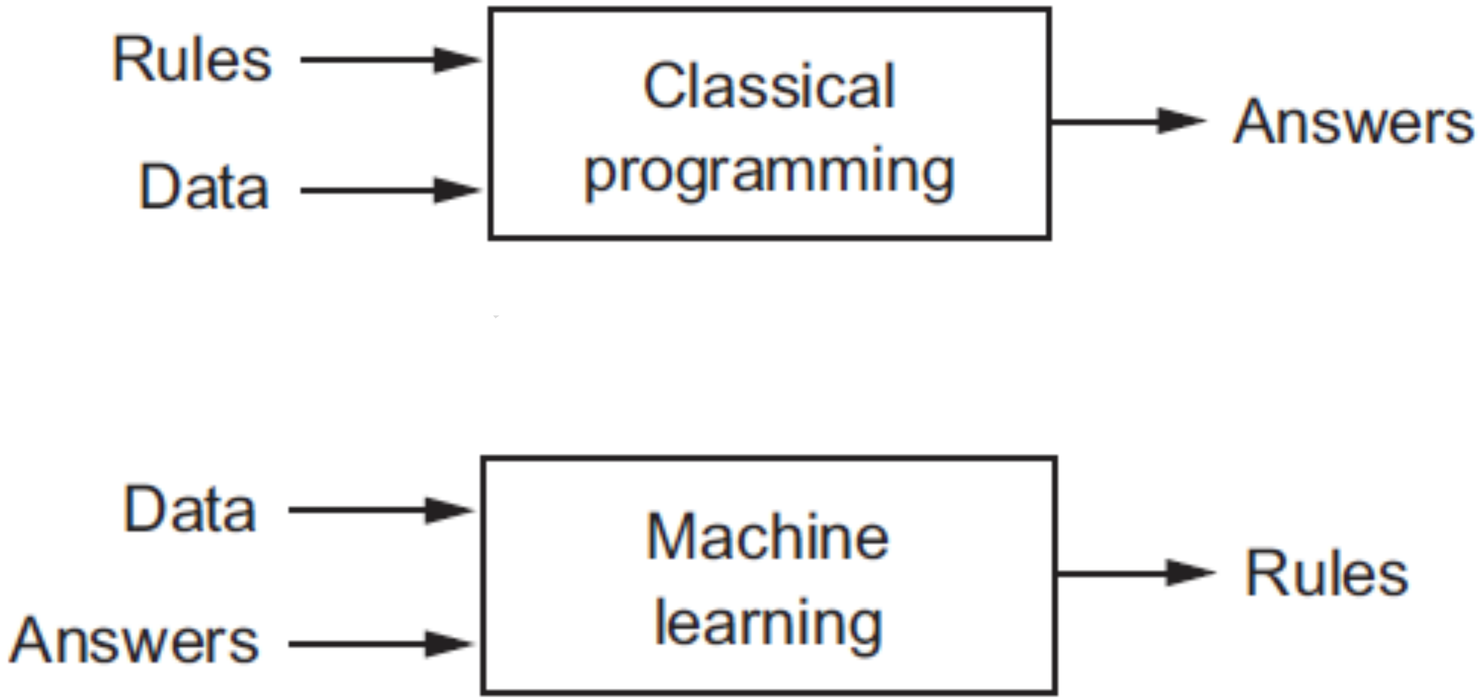
ML - A new programming paradigm

DL using Python - Franchois Chollet et al.



ML - A new programming paradigm

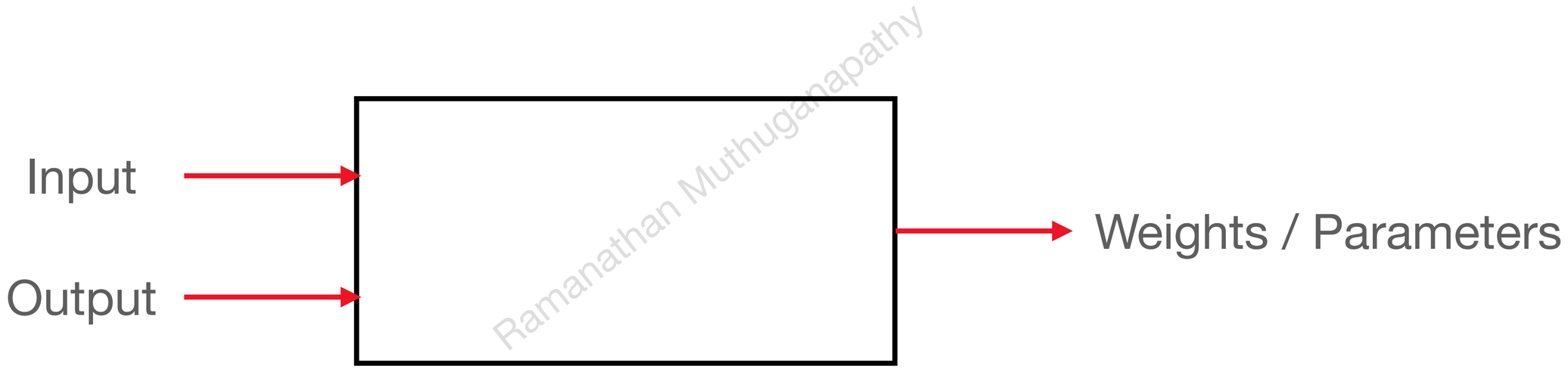
DL using Python - Franchois Chollet et al.



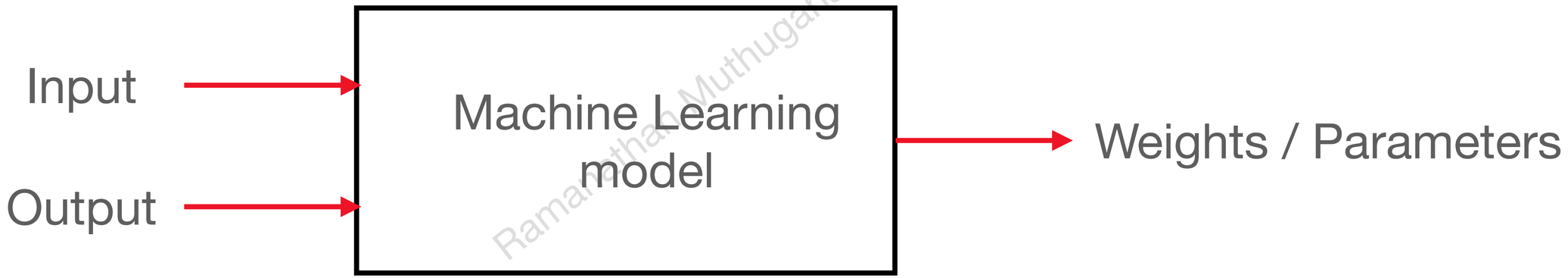
ML - Requirements

- Input - Data (usually in numbers form, E.g. point-sets, images, coordinates)
    - Features - One or many (single / multiple)
- Output - Target values (Answers - already known values)
- Learning - Mapping of input to output.

ML - Overall Paradigm



ML - Overall Paradigm



ML - Overall Paradigm

Supervised learning

- Model / Hypothesis Function: E.g. $h_w(x) = w_0 + w_1 x$,
  $h_w(x) = w_0 + w_1 x + w_2 x^2$, $h_w(x) = w_0 + w_1 x_1 + w_2 x_2$

ML - Overall Paradigm

Supervised learning

- Model / Hypothesis: E.g. $h_w(x) = w_0 + w_1 x$, $h_w(x) = w_0 + w_1 x + w_2 x^2$, $h_w(x) = w_0 + w_1 x_1 + w_2 x_2$
  - Form - Linear, Quadratic etc.
  - Weights / Parameters - $w's(w_0, w_1, \ldots\ldots)$

ML - Overall Paradigm

What is that we do in ML?

- Weights / Parameters - $w's(w_0, w_1, \ldots\ldots)$ are the unknown
- In ML, form is given by the user and the ML predicts the weights based on the data
- Ultimately, the weights are identified (Learning the weights).
- Machine does not change the form, it is the user who can change the form.

ML - Overall Paradigm

How are the weights identified?

- Ground truth data - Input feature / output $(\mathbf{x}, \mathbf{y})$ are the knowns
- Use a model / hypothesis as $h(w)$
- Develop an error / cost / loss function $J(w) = J(\mathbf{y}, \bar{\mathbf{y}}) = J(\mathbf{y}, h(w))$
- The weights are identified by
  - min $J(w)$
- Essentially, ML problem is now reduced to an optimization problem.
- Weights are identified using Optimization.

ML - Algorithms

- Supervised learning
- Unsupervised learning
- Semi-supervised learning

ML - Algorithms

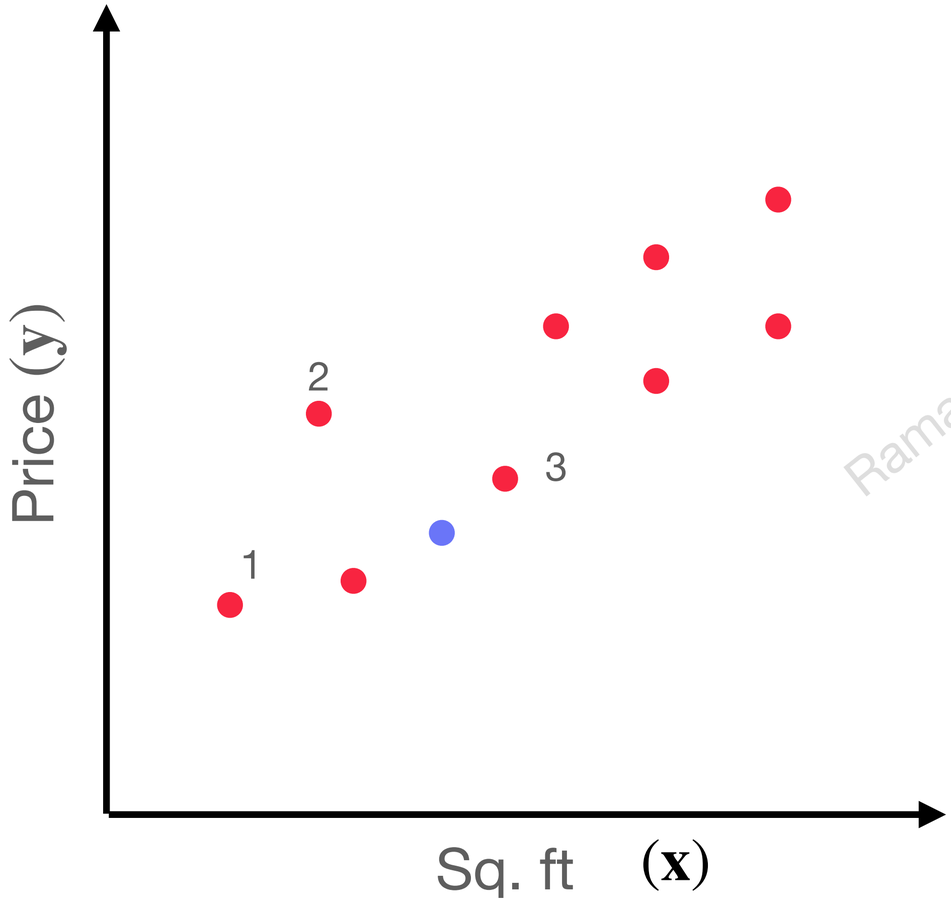
- Supervised learning
- Unsupervised learning
- ~~Semi-supervised learning~~

ML - Algorithms

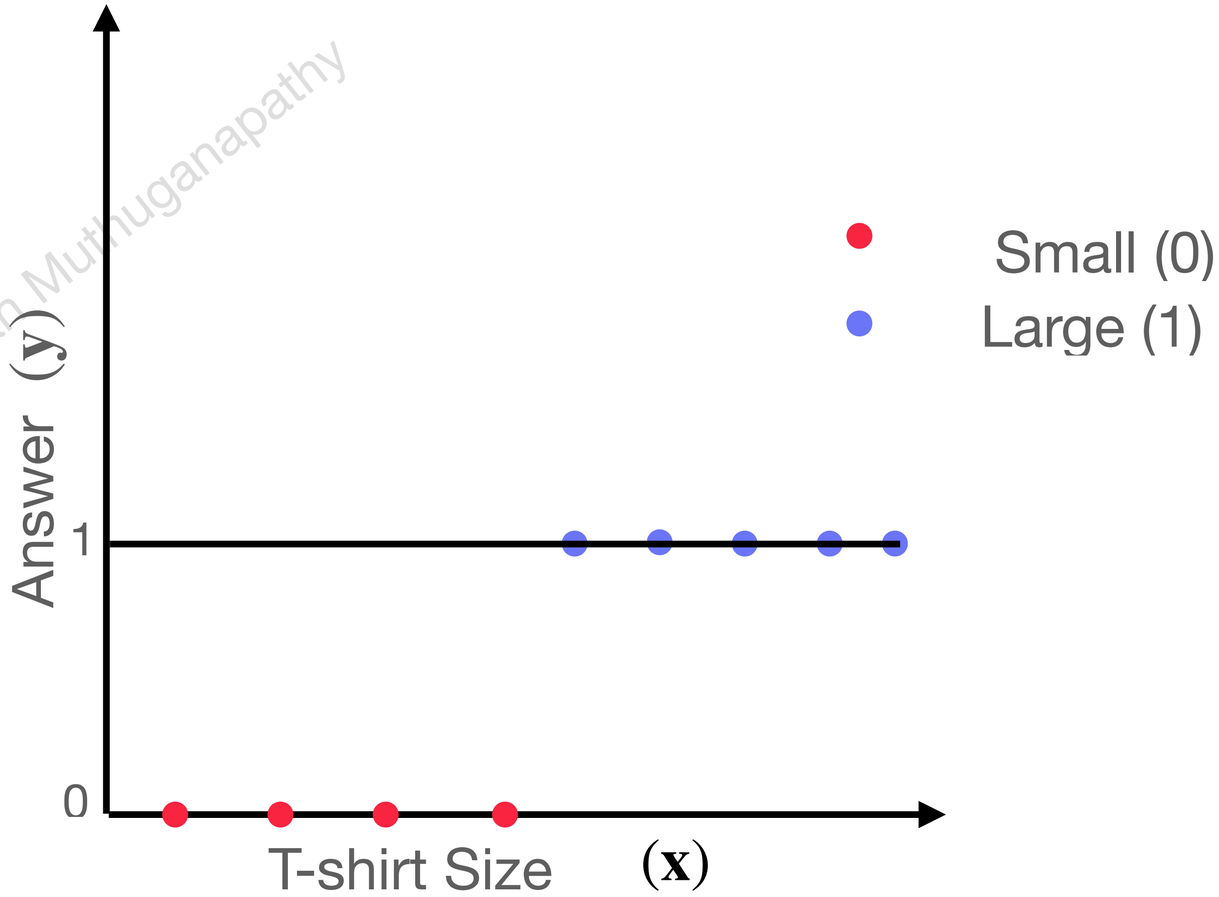
Supervised learning

- Labelled Data (Right answers already given)
  - House price / feet (common example)
  - Experimental data (loading, thermal etc.)
  - Cancerous (Benign or malignant)
  - Geometry Data classifications (MCB, CADNet, CADSketchNet etc.)
  - AlexNet, ImageNet
  - ShapeNet, ModelNet (Mostly using Deep Networks)
- Prediction / Classification problems

Regression vs Classification



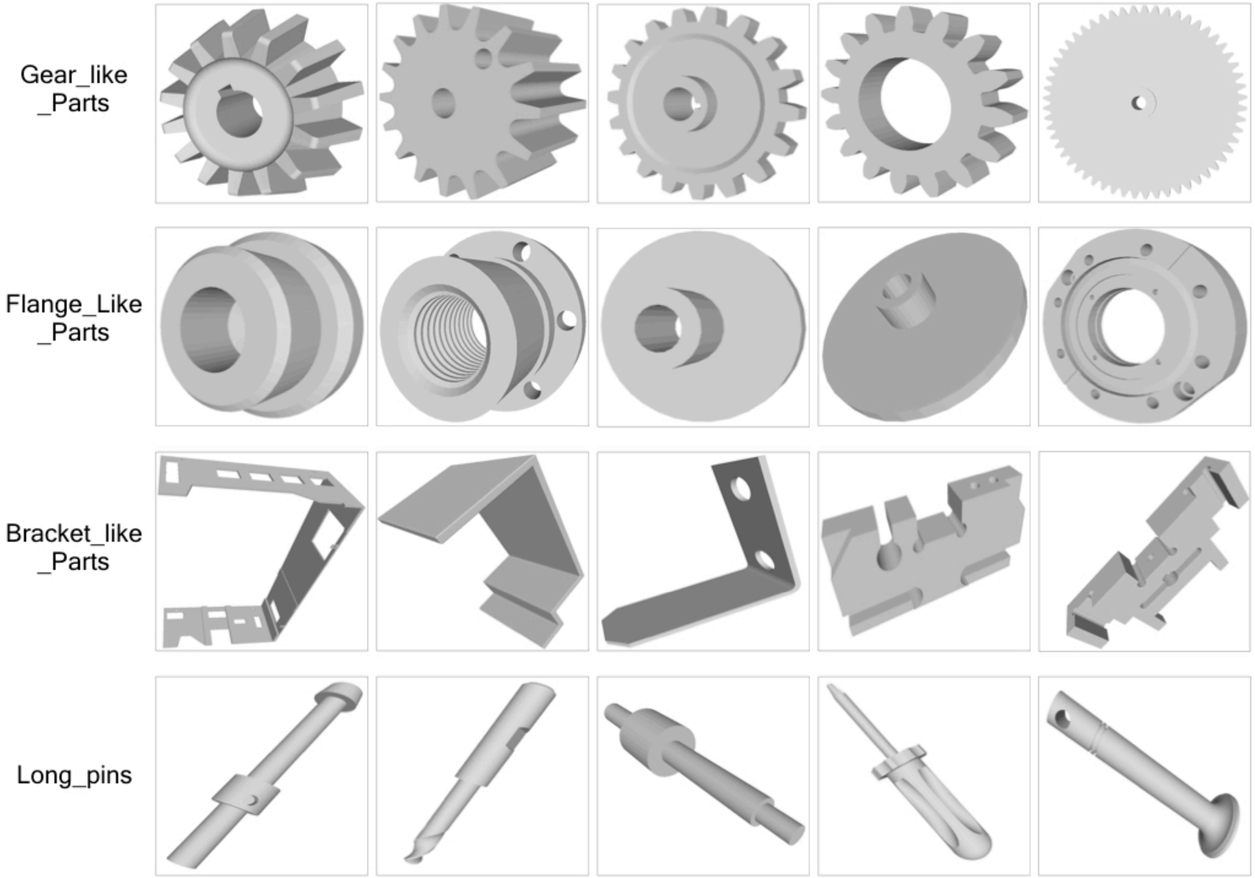
Continuous value o/p



Discrete value

3D Model Classification

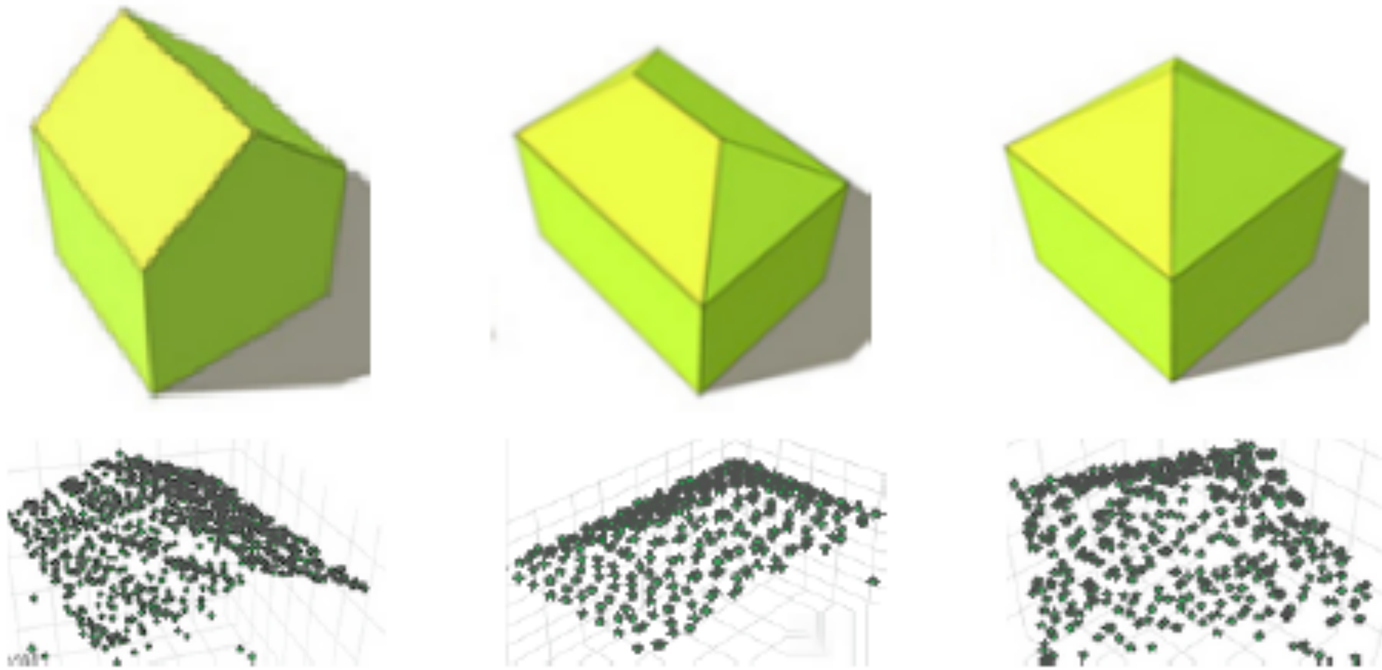
CADNet



A Convolutional Neural Network Approach to the Classification of Engineering Models, https://ed.iitm.ac.in/~raman/agcl/access_r1_web.pdf

Roof classification

Point cloud data



Roof Classification from 3D LiDAR Point Clouds using Multi-view CNN with Self Attention, https://ed.iitm.ac.in/~raman/agcl/GRSL.pdf

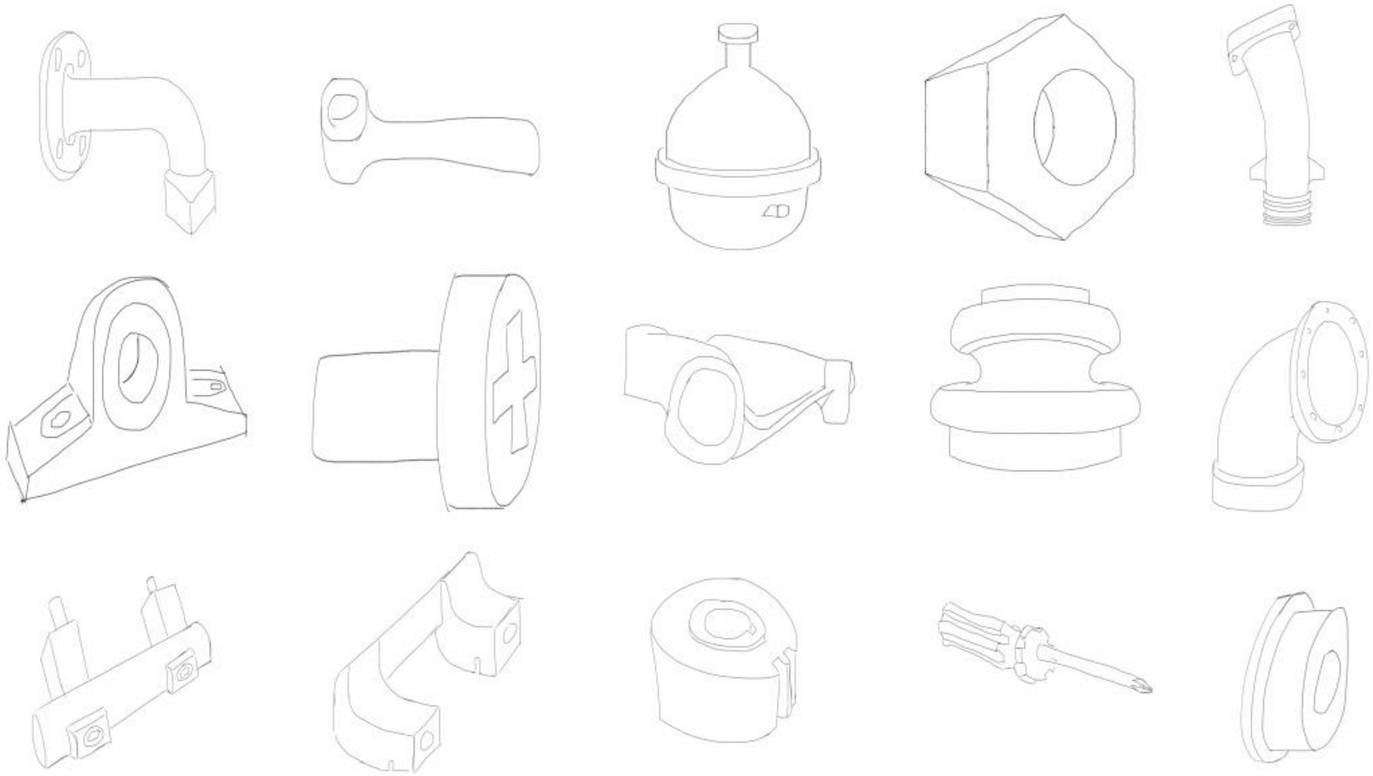
ML - Algorithms

Unsupervised learning

- Data without labelling
  - Market analysis
  - News article analysis
  - Grouping of 3D parts
  - Partial search and retrieval
  - Geometric Classification (ABC Dataset)
  - Social network analysis

Sketches as input

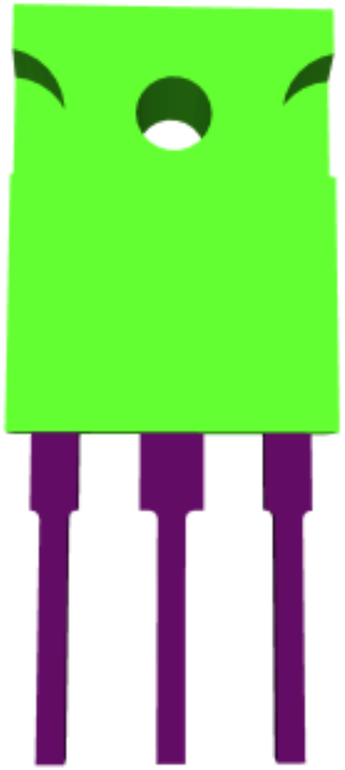
CADSketchNet

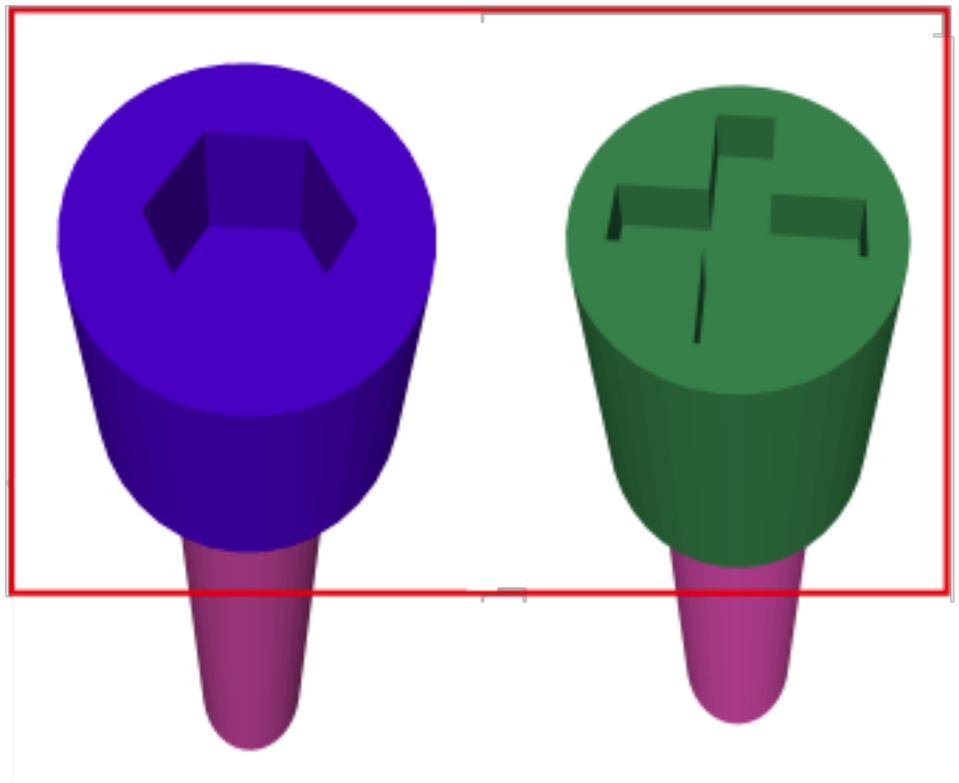


CADSketchNet- An Annotated Sketch dataset for 3D CAD Model Retrieval with DeepNeural Networks, https://ed.iitm.ac.in/~raman/agcl/CADSketchNet_3DOR.pdf

Clustering

Parts of 3D Models





Autoencoder-based part clustering for part-in-whole retrieval of CAD models, https://ed.iitm.ac.in/~raman/agcl/CADG_segmentation_R2.pdf

ABC dataset

https://cs.nyu.edu/~zhongshi/publication/abc-dataset/