Introduction to Data Analysis with AI Algorithms

Part - II

Dr. Nikolaos Bakas

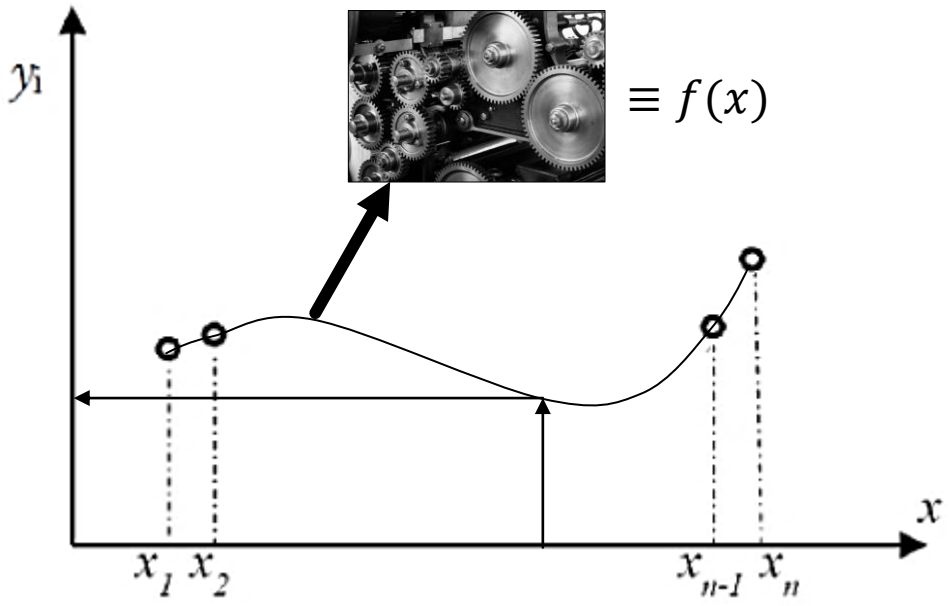
- **Theoretical Aspects**
  - What is an Artificial Neural Network (ANN)
  - How we train an ANN
- **Parallelism in Deep Learning**
  - How Parallel Stochastic Gradient Descent works
  - Results with Distributed GPUs
- **Applications**
  - Real Estate
  - Dentistry
  - Civil Engineering
  - Literature Reviews
  - Product Recommendation
  - Object Detection

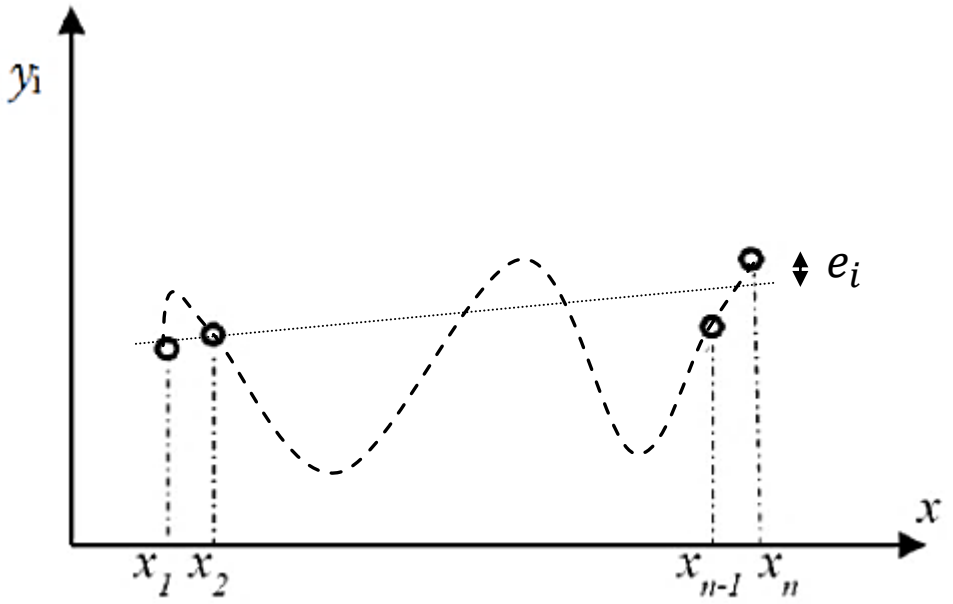
https://castorc.cyi.ac.cy/

Introduction to Data Analysis with AI Algorithms

- **Theoretical Aspects**
- Parallelism in AI
- Applications

Introduction to Data Analysis with AI Algorithms





Develop a **Machine**, which will **Learn,**
to Predict the Output, for new Input

A fundamental problem

Taylor Polynomials

$$f(x_N + h_j) = f(x_N) + h_j f'(x_N) + \frac{h_j^2}{2} f''(x_N) + \ldots + \frac{h_j^n}{n!} f^{(n)}(x_N)$$

Fourier Series

$$s_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi n x}{P}\right) + b_n \sin\left(\frac{2\pi n x}{P}\right) \right)$$

Radial Basis Functions

$$f(\mathbf{x}) = \sum_{k=1}^{N} \left(\sum_{j=1}^{n} w_{jk} \varphi_j(\mathbf{x}) \right) v_k + b_0$$

Artificial Neural Networks

$$y_i \cong f_i(x_{i1}, x_{i2}, \ldots, x_{in}) = \sum_{k=1}^{N} v_k \sigma \left(\sum_{j=1}^{n} w_{jk} x_{ij} + b_k \right) + b_0$$

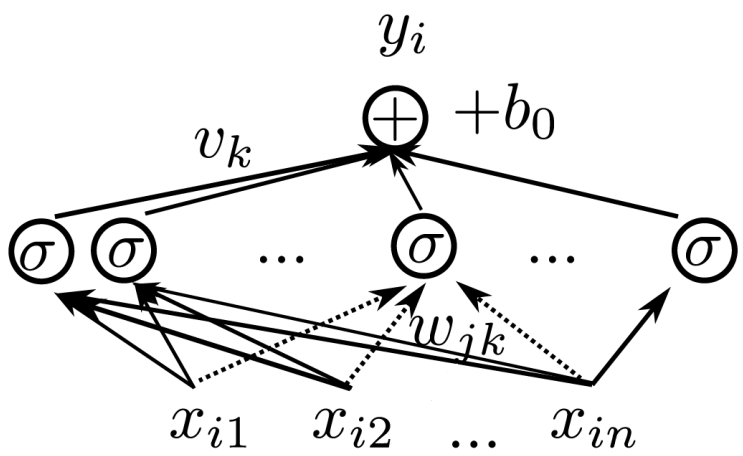
Function Approximation

Let x_{ij} be some given data of $j \in \{1, 2, \ldots, n\}$ input variables in $i \in \{1, 2, \ldots, m\}$ observations of y_i responses.

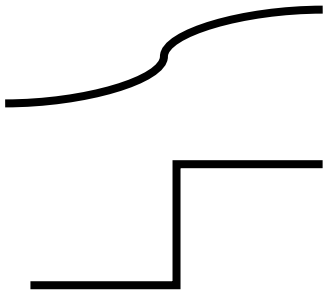
The Universal Approximation Theorem ensures the existence of an integer N, such that

$$y_i \cong f_i(x_{i1}, x_{i2}, \ldots, x_{in}) = \sum_{k=1}^{N} v_k \sigma \left(\sum_{j=1}^{n} w_{jk} x_{ij} + b_k \right) + b_0,$$

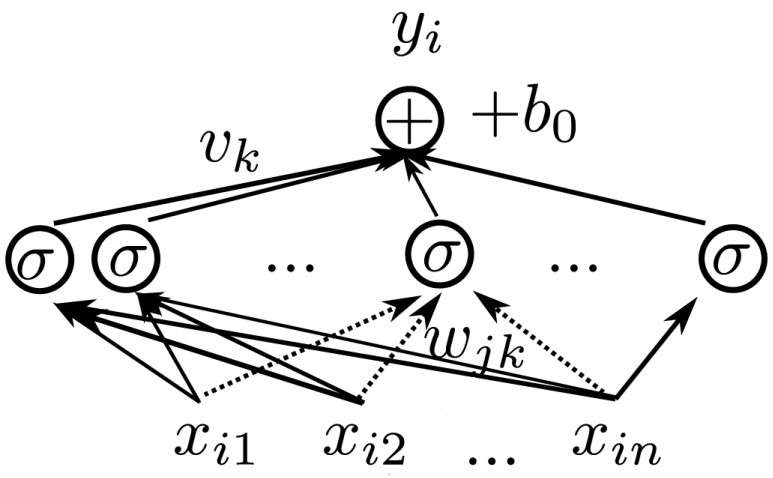
with approximation errors $\epsilon_i = y_i - f_i$ among the given response y_i and the corresponding simulated f_i, arbitrarily low.



$$\sigma(x) = \frac{1}{1+e^{-x}}$$



Approximation with Neural Networks



$$y_i \cong f_i(x_{i1}, x_{i2}, \ldots, x_{in}) = \sum_{k=1}^{N} v_k \sigma \left(\sum_{j=1}^{n} w_{jk} x_{ij} + b_k \right) + b_0,$$

$$L(w_{jk}, b_k, v_k) = \sum_{i=1:m} (f_i - y_i)^2$$

Loss Function

Depending on the data x_{ij} and the weights w,v,b

For simplicity, we use the symbol w, $\forall w := \{w_{jk}, b_k, v_k\}$

unknown

Gradient Descent

$\forall$epoch n, do:

$$w_{n+1} = w_n - \gamma \frac{\partial L}{\partial w}$$

This is performed by Machine Learning Frameworks, such as **Tensorflow, PyTorch, etc.**

Depends on the weights AND the data points x_{ij}

Stochastic Gradient Descent

- At each iteration, pick a random data-point *i*
- Differentiate the *Loss Function for this data-point*
- Update the weights for the entire network

i	$x_{i,1}$	$x_{i,2}$	…	$x_{i,N}$	y_i
1					
2					
…					
m					

$$\forall \text{epoch } n, \text{do:}$$

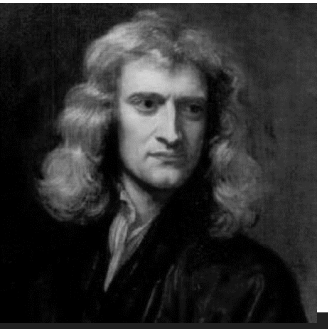
$$\forall \text{point } i, \text{do:}$$

$$w_{n+1,i} = w_{n,i} - \gamma \frac{\partial L_i}{\partial w}$$

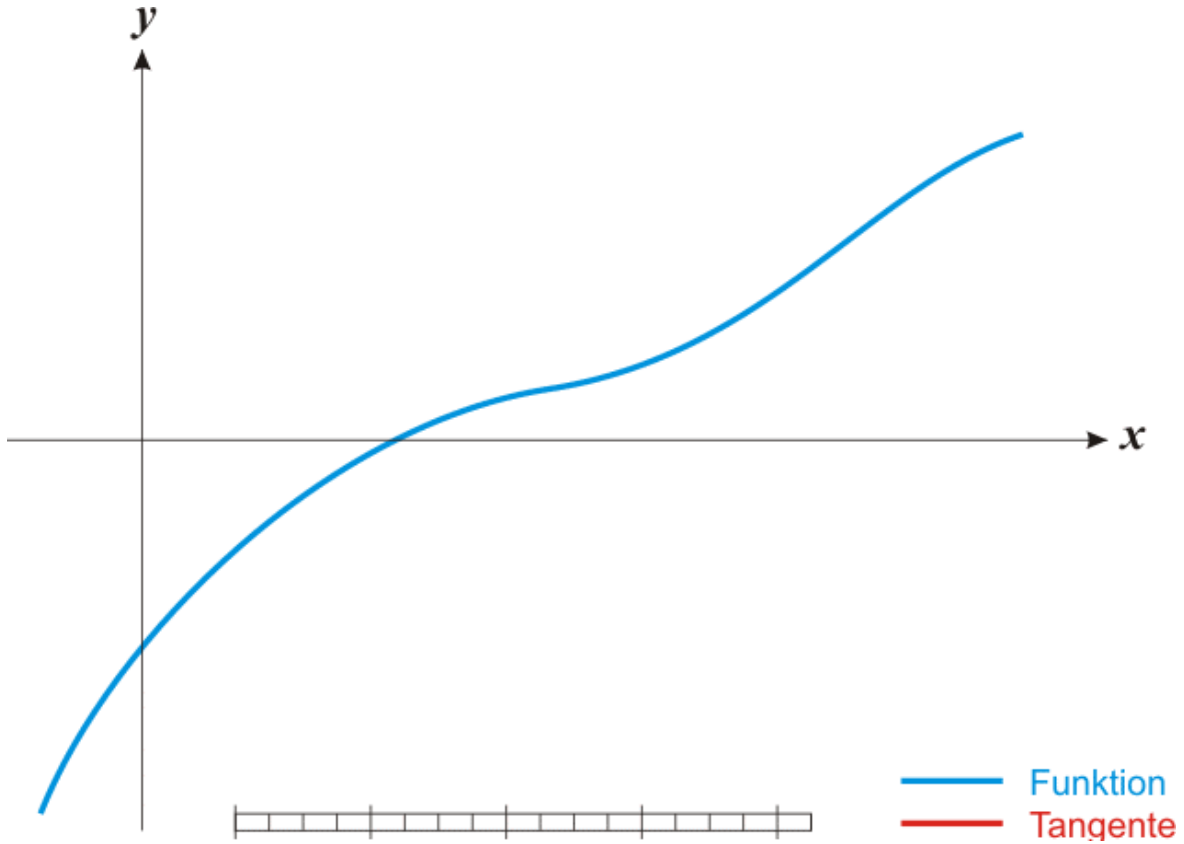
- ✓ In **mini-batch Stochastic Gradient Descent**, *i* as the current "batch" of data, that is to say subset of indices $\in \{1,2,...,m\}$.
- ✓ **Parallelization** is being performed there, by updating the weights for all batches in parallel.

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

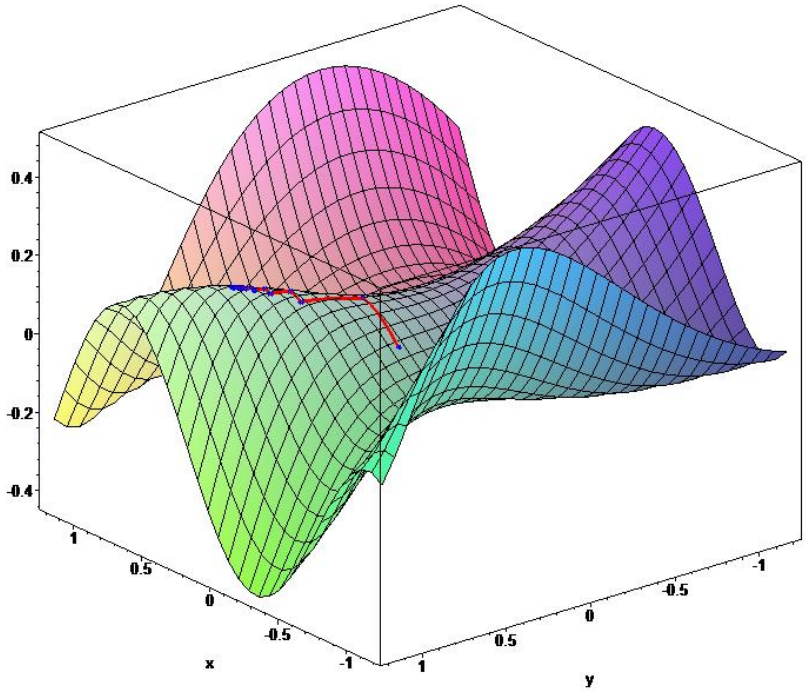
$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

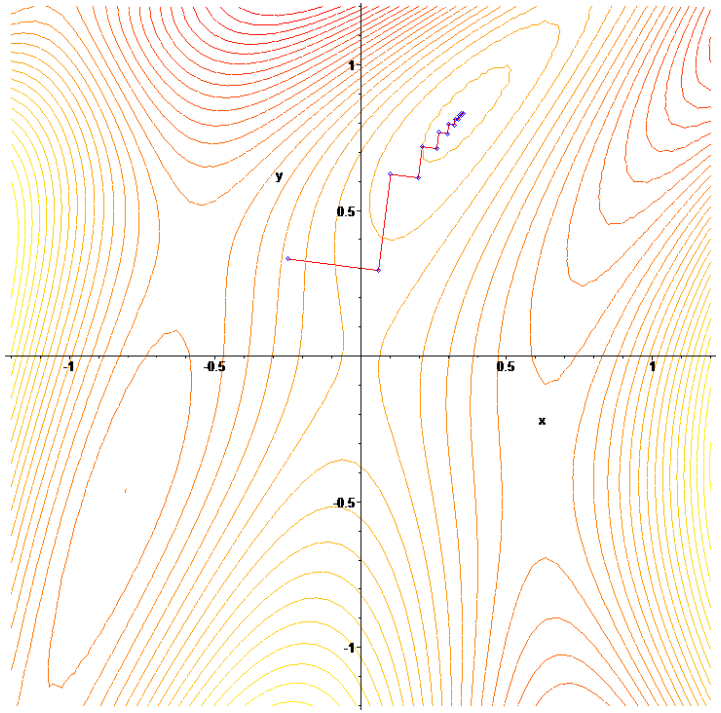


On analysis by infinite series
Isaac Newton, 1669

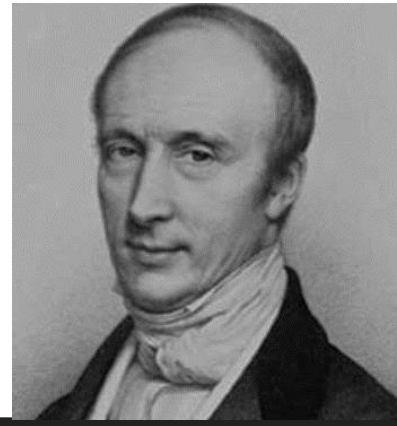


Newton's Method





Methode generale pour la resolution des systemes d'equations simultanees
Augustin-Louis Cauchy
1847



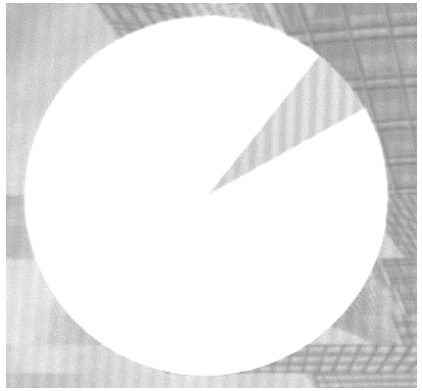
$$\mathbf{x}_{n+1} = \mathbf{x}_n - [\mathbf{H}f(\mathbf{x}_n)]^{-1}\nabla f(\mathbf{x}_n)$$

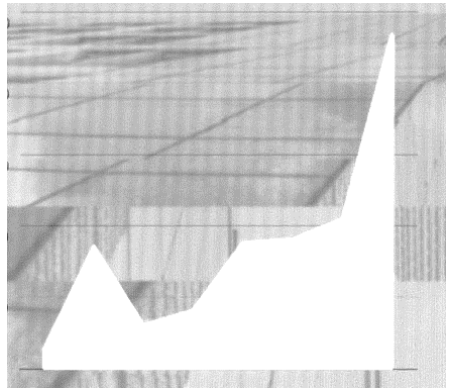
$$\mathbf{a}_{n+1} = \mathbf{a}_n - \gamma\nabla F(\mathbf{a}_n)$$

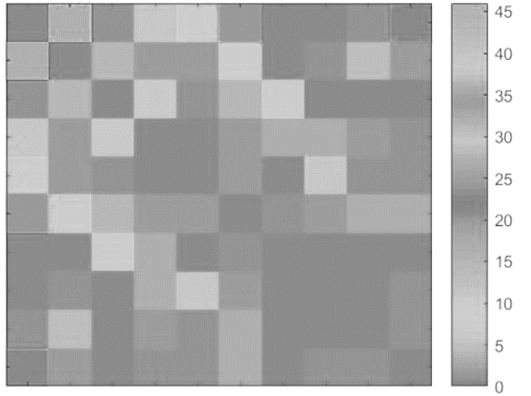
Newton's Method and Gradient Descent

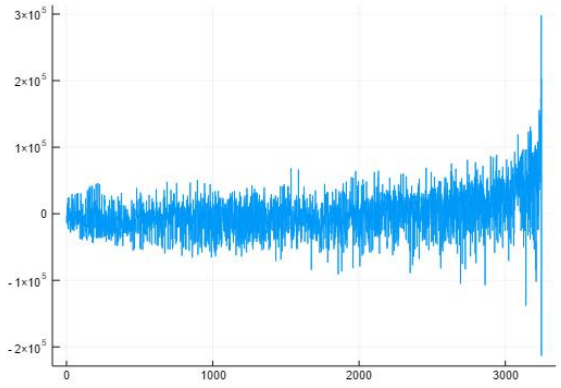
figures (left to right):
Joris Gillis - Created with Maple 10
Public Domain
Public domain - Library of Congress Prints and Photographs Division.

1. Exploring the Data-Set
   - Distributions
   - Multiple Correlations
3. Error Analysis
   - Analysis of Residuals
   - Outliers
4. Sensitivity Analysis
5. Interpretation

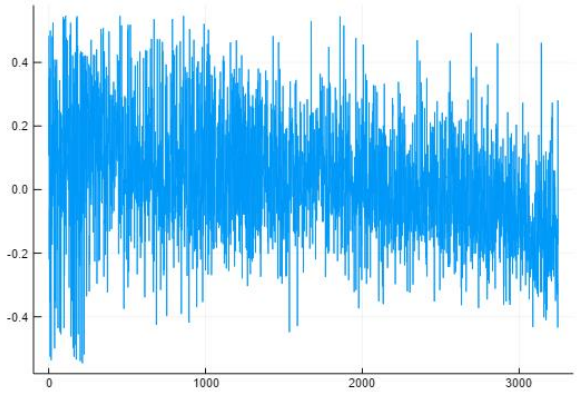




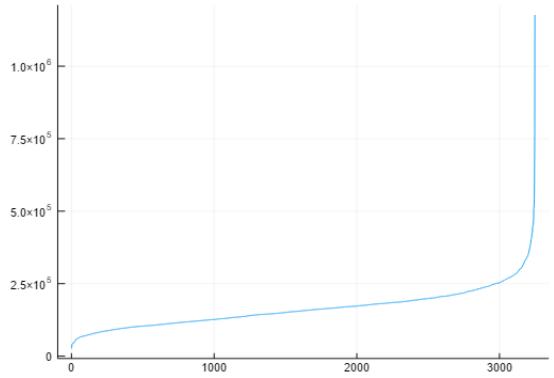




Heteroscedastic



Biased

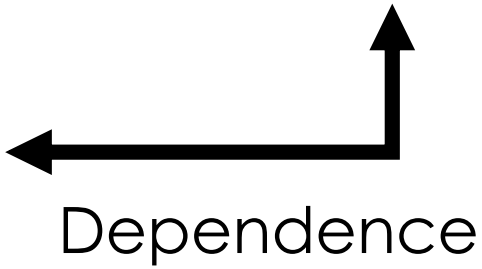


Outliers

2. Predictive Modelling

More of an Art than a Science

- Approximation Methods
- Optimization



A big variety of

- Methods
- Algorithms
- Theories

Hutson, M., 2018. AI researchers allege that machine learning is alchemy. *Science*, *360*(6388), p.861.

Ananthaswamy, A., 2019. From counting with stones to artificial intelligence: the story of calculus. *Nature*, *568*(7750), pp.32-33.

AI Algorithms

- Theoretical Aspects
- **Parallelism in AI**
- Applications

Introduction to Data Analysis with AI Algorithms

Parallel Stochastic Gradient Descent

- For each rank
  - Pick a random data-point
  - Compute the gradient and update the weights

on i^{th} GPU ←

$$w_{n+1,i} = w_{n,i} - \gamma \frac{\partial L_i}{\partial w}$$

on j^{th} GPU ←

$$w_{n+1,j} = w_{n,j} - \gamma \frac{\partial L_j}{\partial w}$$

- ✓ Larger batch size is faster
- ✓ Smaller batch size more accurate
- ✓ Batch must fit into memory
- ✓ More nodes if GPU cannot handle size

- Mix the weights
  - e.g. average, ensembles, vector summation, etc.
  - accuracy loss

we need a framework (MPI, Horovod), to do this operation in parallel ←

$$w_{n+1} = \frac{1}{h} \sum_{k=1:h} w_{n+1,k},$$

where h is the number of GPUs!

Cassava Leaf Disease Classification

- **Scope**: To distinguish between several diseases that cause material harm to the **food supply of many African countries**, such as Uganda, Kabul, etc. In some cases the main remedy is to burn the infected plants to prevent further spread, which can make a rapid automated turnaround quite useful to the farmers.
- **Process**: Feed a picture with a cassava plant using a photo from a relatively inexpensive camera (e.g. from mobile phone) and get the type of disease, or healthy.

CLASSES

3	13158
4	2577 → Healthy
2	2386
1	2189
0	1087

"4":"Healthy"

"0":"Bacterial Blight (CBB)"

"1":"Brown Streak Disease (CBSD)"

https://www.kaggle.com/c/cassava-leaf-disease-classification/data

Cassava Leaf Disease Classification

800

600

Each Pixel:

3 Colors (RGB)

3 Float Numbers (0.0 – 1.0)

Tot. 21397 images

- 800*600*3 = 1.44×10^6 features
- 5.76 MB each, in float precision
- + Augmentation on-the-fly

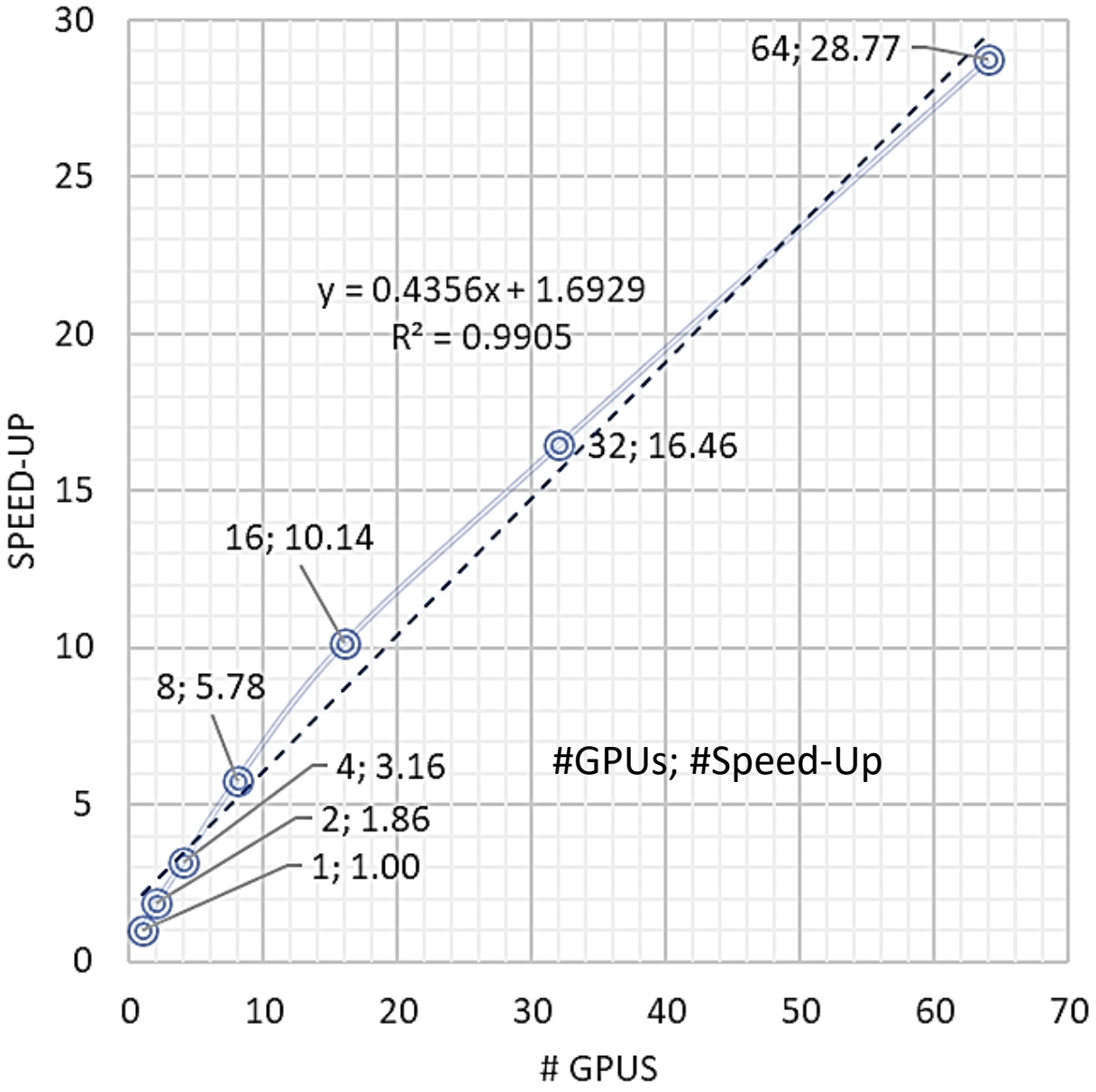
Pretrained Networks

- Started with **ResNet50**
- Finally, calculations with **EfficientNet**
- Pretrained – open source
- Pytorch + Horovod to parallelize

19M parameters to optimize

https://arxiv.org/pdf/1905.11946.pdf
https://www.kaggle.com/khyeh0719/pytorch-efficientnet-baseline-inference-tta

Speed-Up on Cyclone: 1 – 64 GPUs



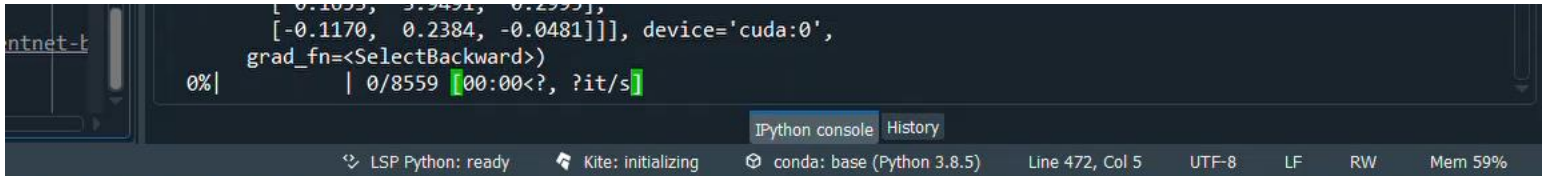
# GPUs	train minutes (10 epochs)	Speed-Up	% valid. Accuracy	Accuracy Loss
1	44.82	1.00	89.00	0.00%
2	24.07	1.86	89.14	0.16%
4	14.18	3.16	89.11	0.12%
8	7.76	5.78	88.90	-0.11%
16	4.42	10.14	88.46	-0.61%
32	2.72	16.46	88.69	-0.35%
64	1.56	28.77	86.40	-2.92%

- image size: 512x512
- train batch size: 52 (per GPU)
- learning rate: 1e-4
- weight decay:1e-6

https://www.meetup.com/PyData-Cyprus/events/276154247/

PyTorch + Horovod

1



Local Machine
NVIDIA GeForce GTX 1650
4GB RAM

2

Cyclone
8 GPU nodes
4 NVIDIA Tesla V100-SXM2
32GB each

3

Cyclone
16 GPU nodes
4 NVIDIA Tesla V100-SXM2
32GB each

Real-Time Experiments

- Theoretical Aspects
- Parallelism in AI
- **Applications**

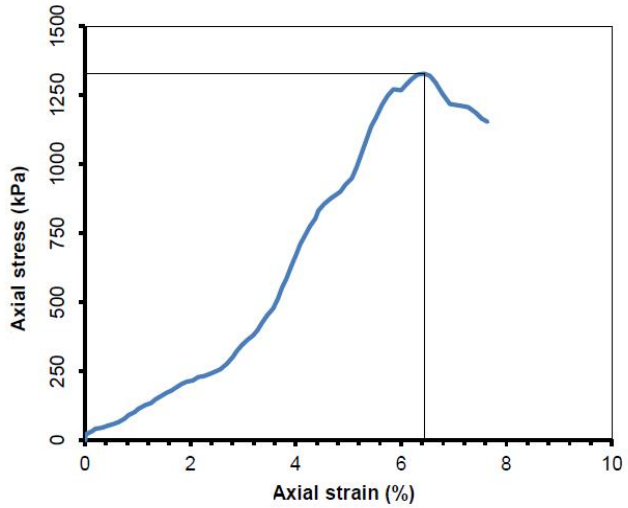
Introduction to Data Analysis with AI Algorithms

- **Methods**
  - Predictive Modelling
  - Unsupervised Learning
  - Optimization
- **Data Types**
  - Structured Data
  - Text Analysis
  - Computer Vision

Introduction to Data Analysis with AI Algorithms

- Biomass
- Urea
- Cement Kiln Dust
- Rice Husk Ash
- Sand
- Water

Cement Industry Wastes



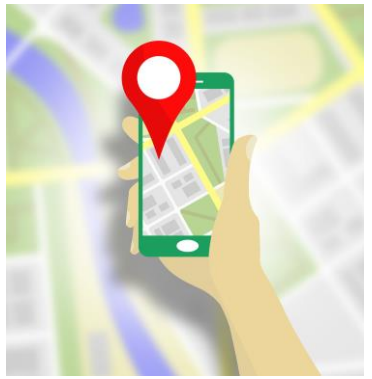
- Ingredients vs Compressive strength
- Simulate behaviour
- Identification of An Optimal Recipe

EcoCement

A Novel Composite Material for the Construction Industry

- Town / Quarter
- Enclosed extent
- Covered extent
- Uncovered extent
- Built years
- Parcel Extent
- Share Percent
- Density
- Coverage
- Condition code
- View code
- Class code
- Shape Code
- Road Code

+



➡ **Predicted Price (€)**

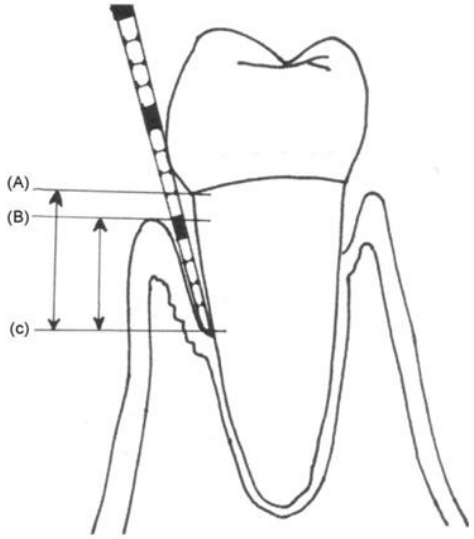
Beautiful home in the fast growing ... This well maintained home consists of 3 bedrooms, 2 bathrooms, full walk in shower and ... Laminate and carpet throughout home...

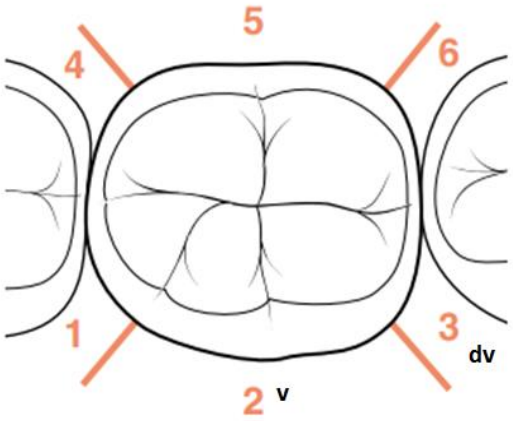
	Coefficient	p-Value
WONDERFUL^2	0.2077	3.6680e-17
log(BEAUTIFUL)	0.0400	1.7865e-05
log(m2)	1.0006	6.3105e-144

Real Estate

text mining in advertisements

- Prediction of tooth loss (periodontitis)
- Accuracy Importance
  - Expensive Treatment
  - Decision for therapy or extraction





- Factors:
  - age
  - smoking
  - diabetes
  - oral hygiene, etc

- Logistic Regression 43%
- Deep Learning 87%

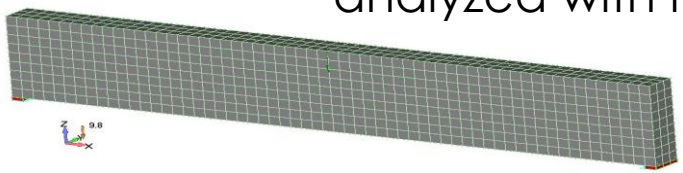
Dentistry

Vary for each beam

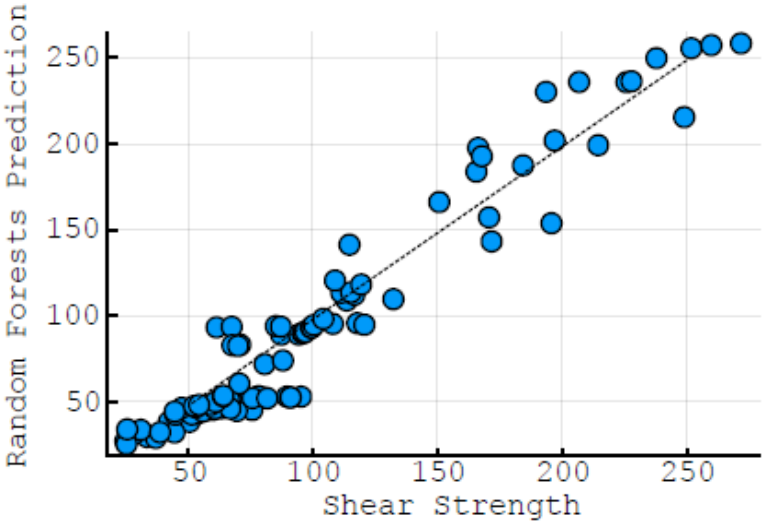
- span *L*
- width *b*
- effective depth *d*
- modulus of Elasticity *E*
- reinforcement ratio

Predict the Ultimate load Capacity

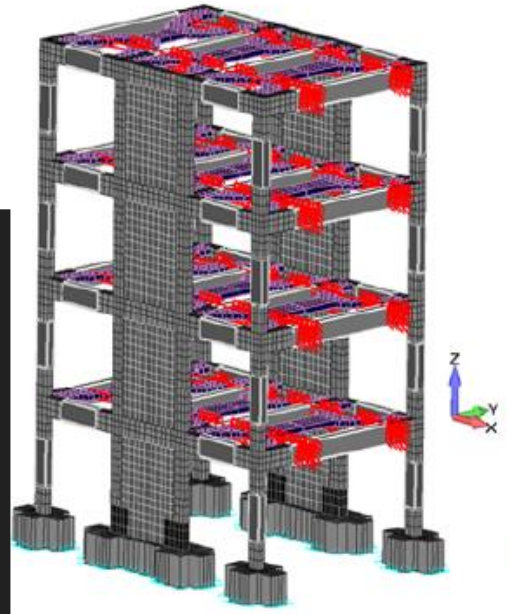
35000 beams
analyzed with non-linear FEM



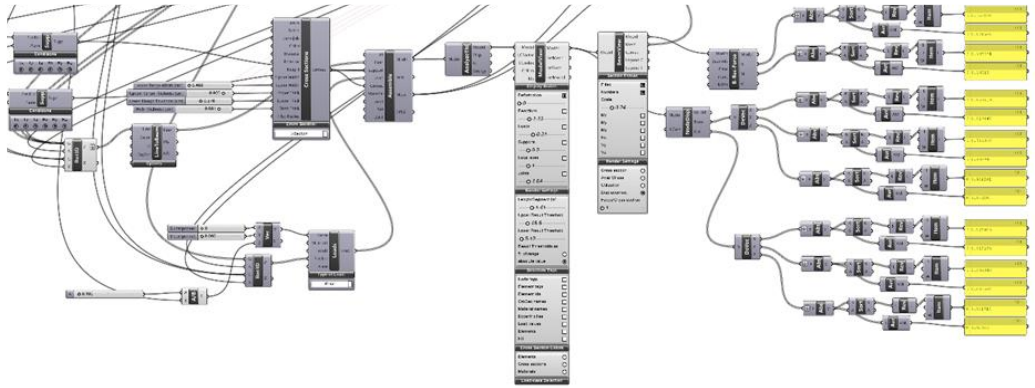
- **Simulate** Finite Element Modelling
- Develop **new Formula** for Building Codes
- Predict out of sample **Experimental** Data

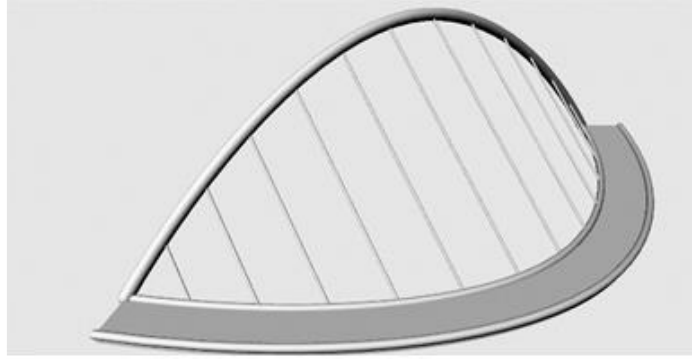


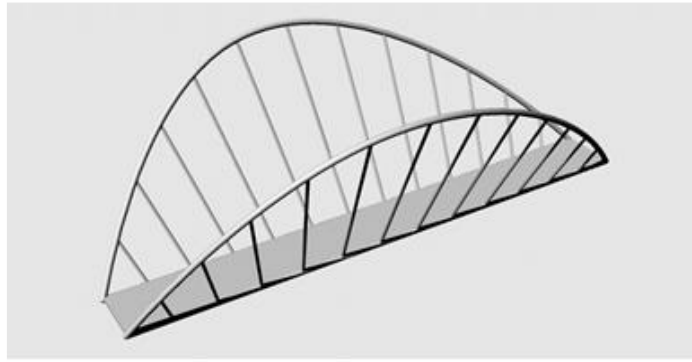
Finite Element Modelling

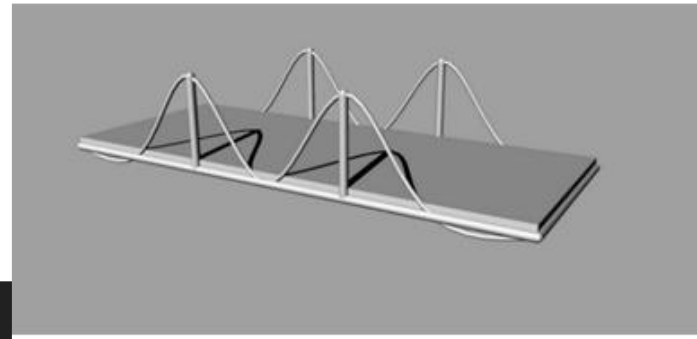


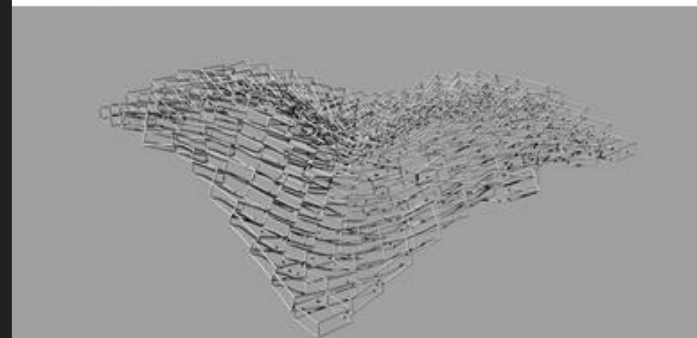
- Database of Parametric Structures
- Analyze with Grasshoper3D & Karamba3D
- Questionnaires to Vote for Aesthetic Value
- Identify Links with Structural Response



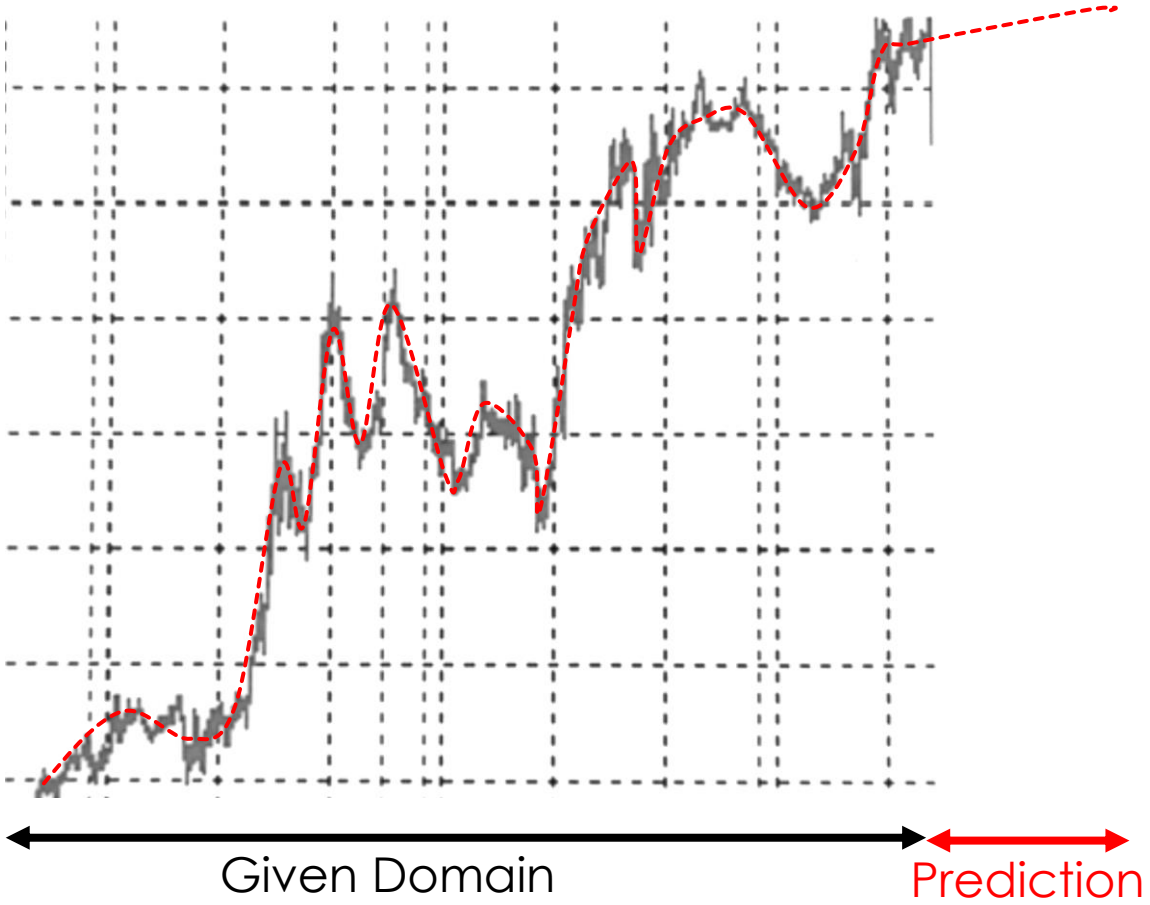








Architecture



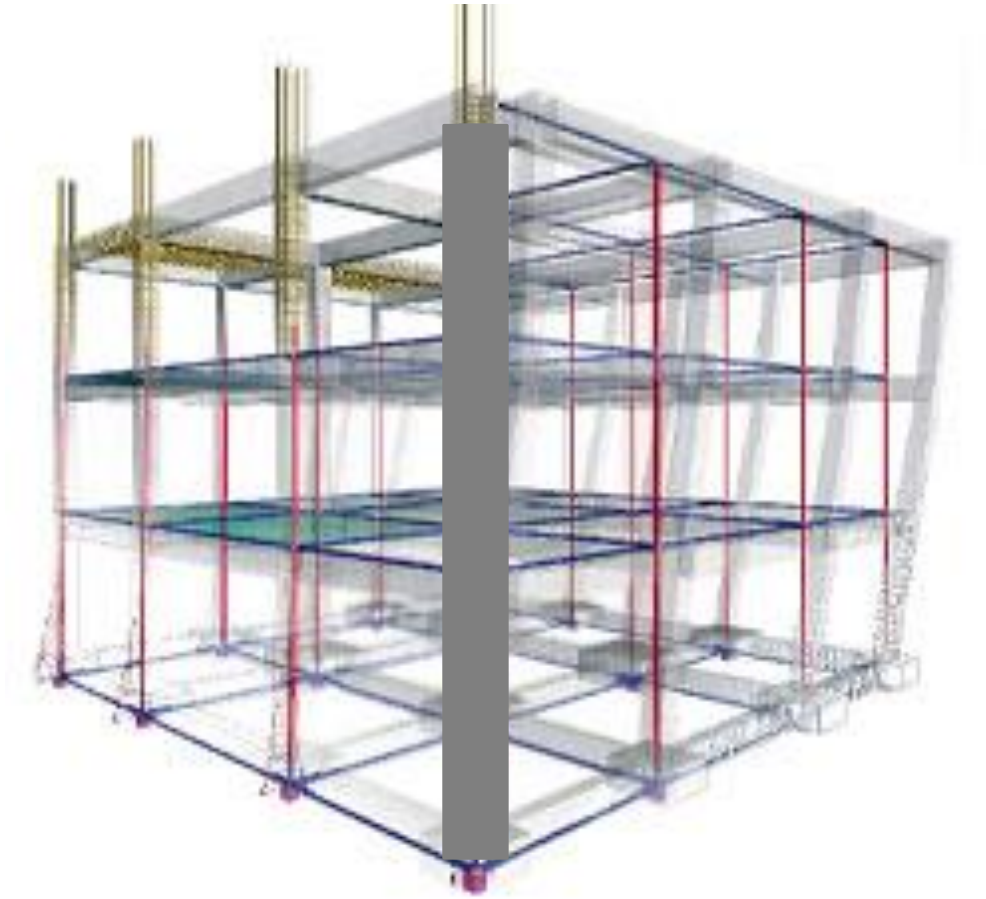
- Timeseries Forecasting
- Auto-Regressive Models
- Convert to Multivariate
- Transformer Networks
- Noise

Timeseries

- Transactions Data
- User – Item Matrix
- Modell users' preferences
- Suggest new movies, products, etc

Recommender Systems

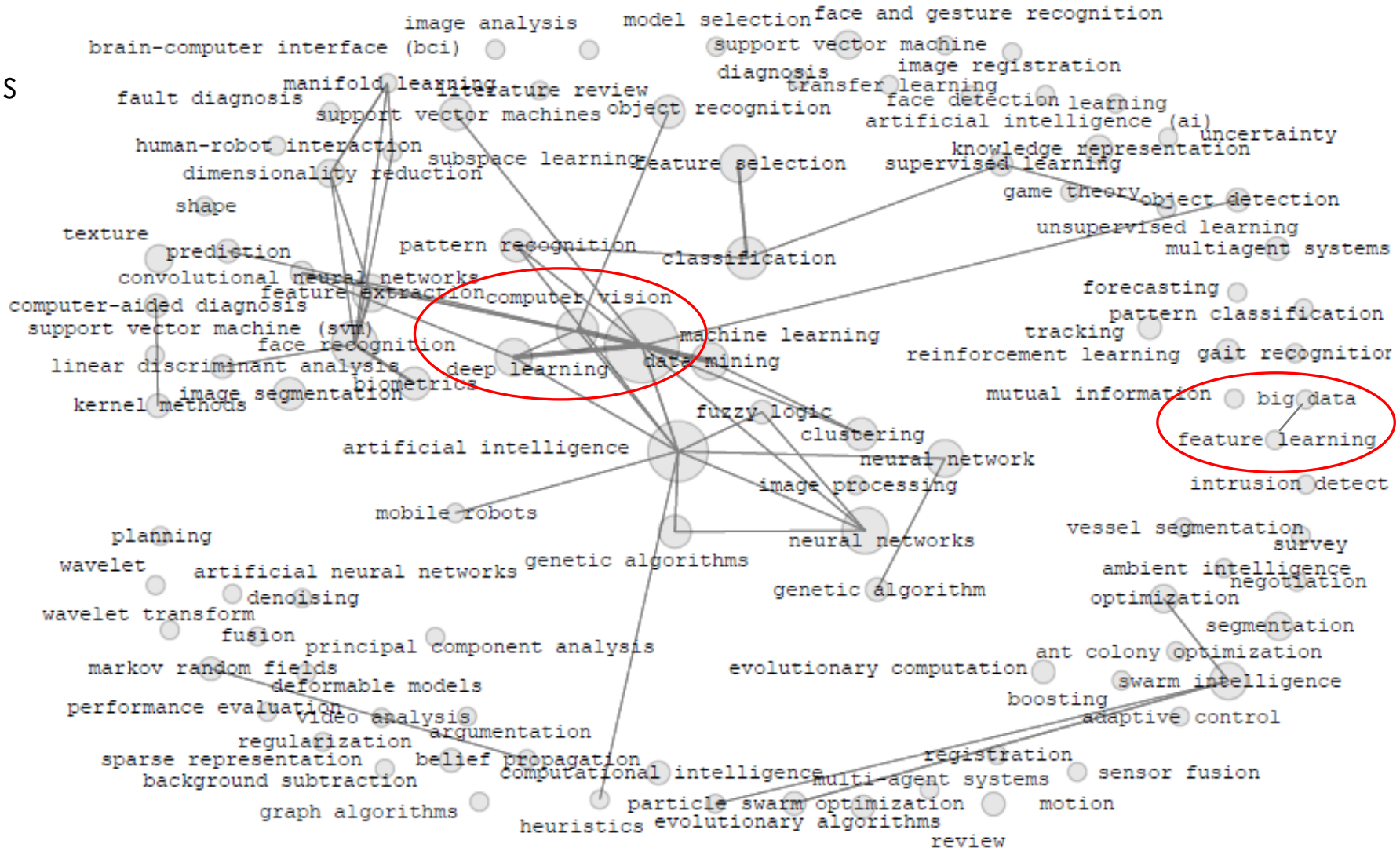
- Optimization of Structural Response
- Varying the dimensions of columns & beams
- Enhancement of Earthquake Response
- Cost reduction



Engineering

- Co-Occurrences of Keywords
- Convert to a Map:
  - Distances reflect the Dissimilarities
  - Unsupervised Learning

1. $c_{ij} :=$ *contingency table* (co-occurrence of objects)
2. $s_{ij} :=$ similarity
3. $ds_{ij} := \frac{1}{s_{ij}}$ (dis-similarity)
4. $d_{ij} = \|x_i - x_j\|$ (distance on map)
5. $f_{ij} := |ds_{ij} - d_{ij}|$ (objective function)
6. Optimality criteria satisfied?
   - *NO* ⇦
   - ⇩ *YES*
7. End => drawing of the bibliometric map



Computational Analysis of Scientific Literature

ZOOM

Interpolation

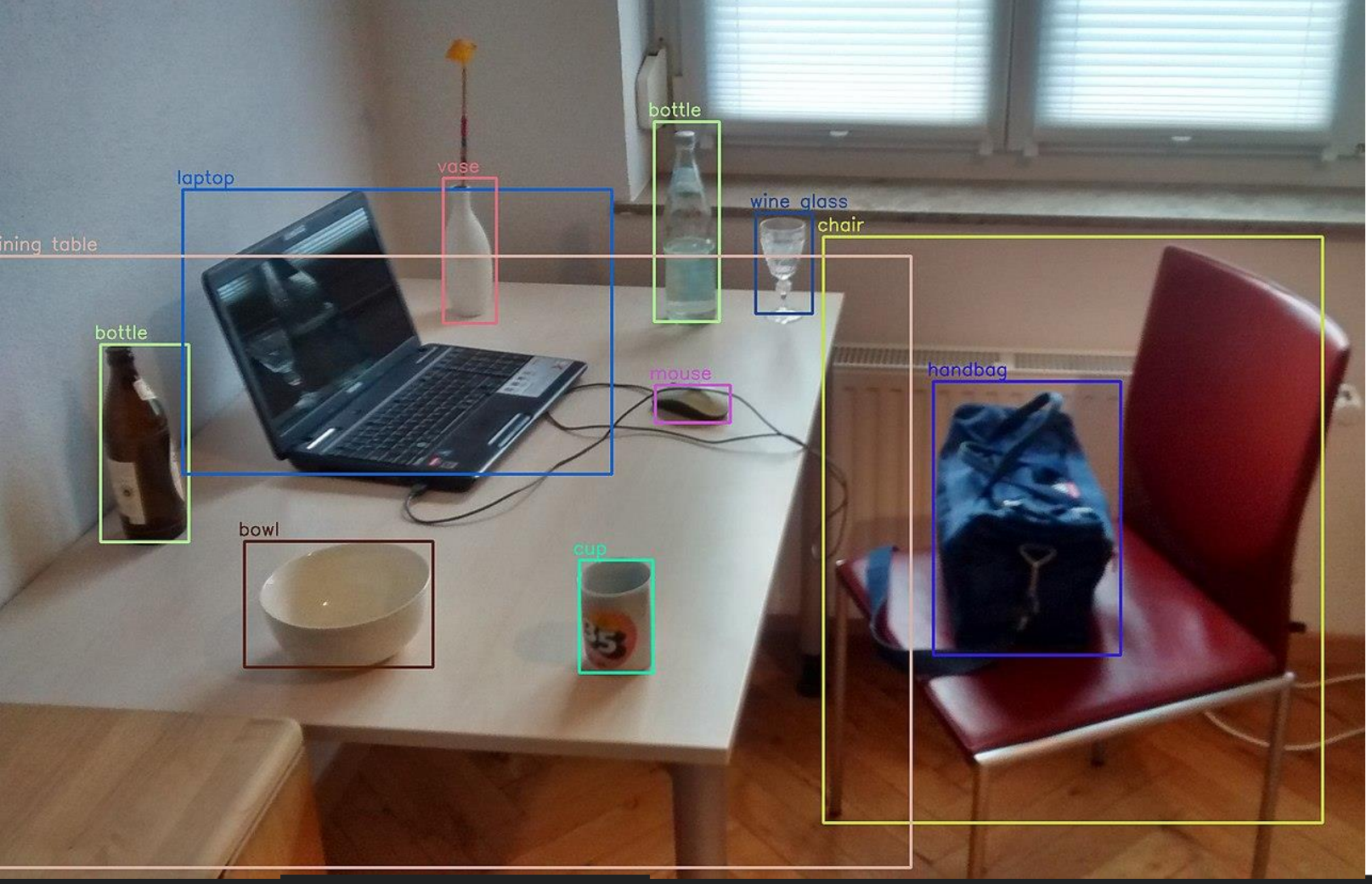
Extend

Extrapolation

Image Processing

A few samples from the MNIST test dataset.
14 December 2017
Josef Steppan

Handwritten Digits Recognition



- YOLOv5[1] model
- trained on COCO[2] dataset
  - 330K images
  - 1.5 million object instances
  - 80 object categories
- Detect Bounding Box

[1]https://github.com/ultralytics/yolov5
[2]https://cocodataset.org/

Object Recognition

Original image: Schreibtisch-mit-Objekten.jpg, MTheiler

Object Recognition

THE CYPRUS INSTITUTE
RESEARCH·TECHNOLOGY·INNOVATION

EURO

Introduction to Data Analysis with AI Algorithms

Dr. Nikolaos Bakas