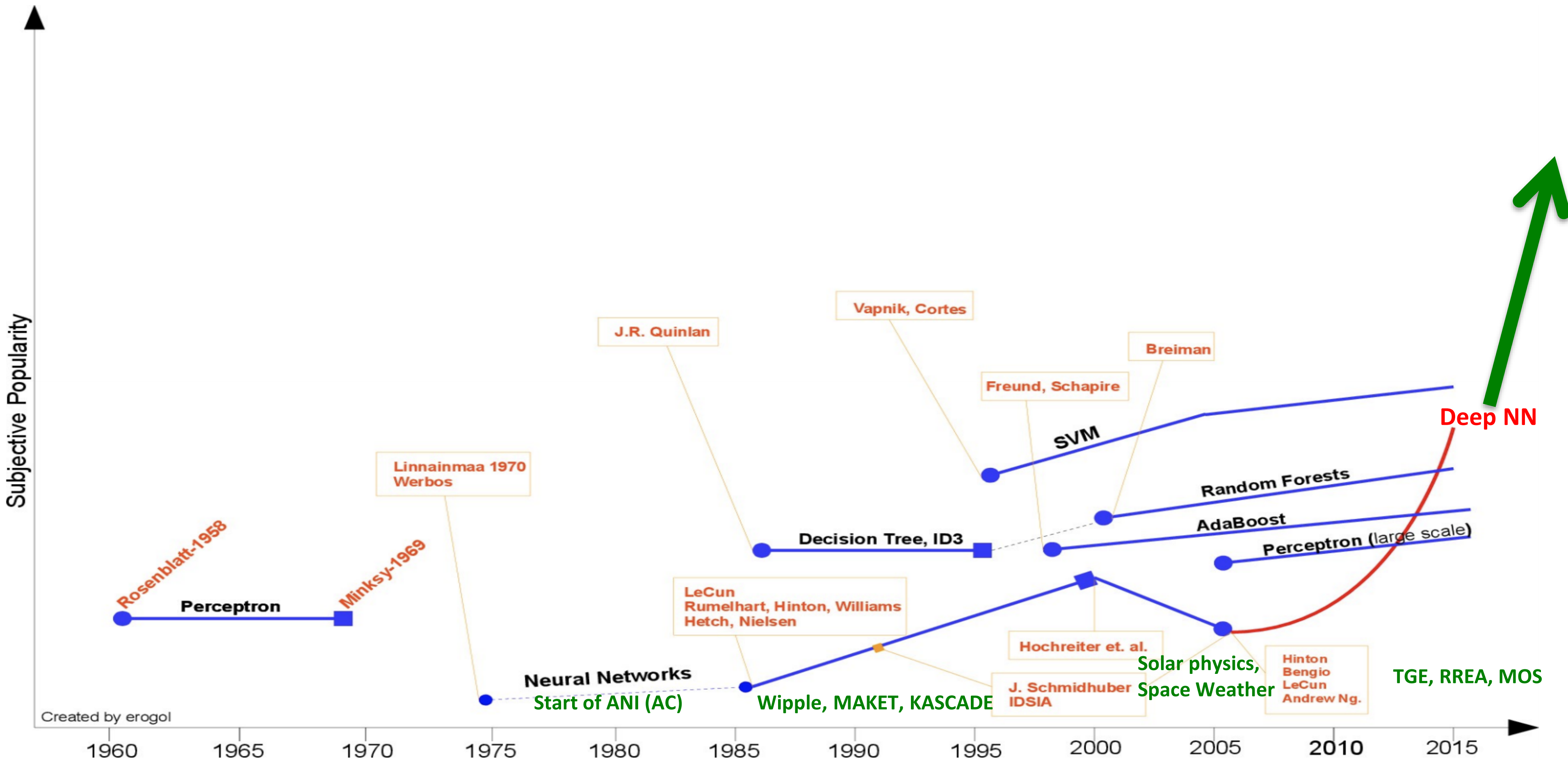
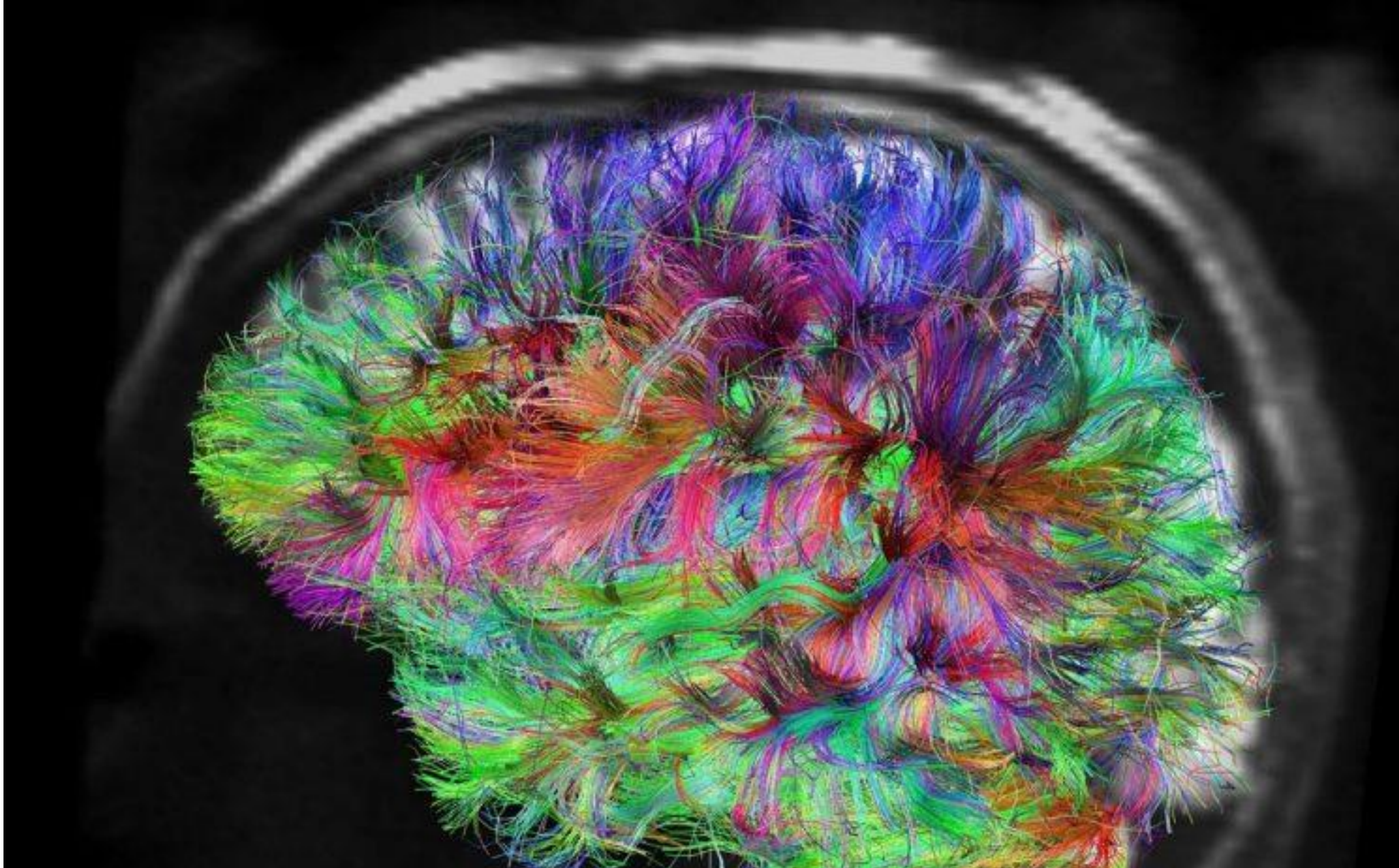
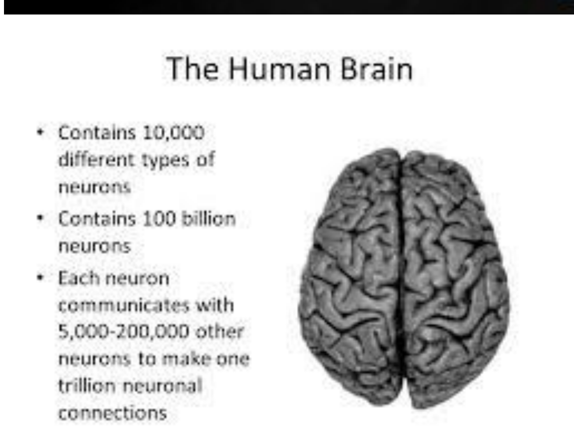
Machine Learning not only in High-Energy Astrophysics (Astroparticle Physics)

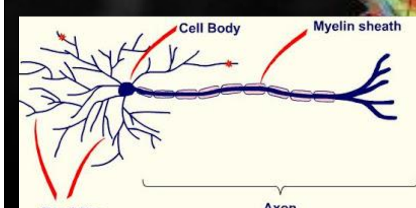
Ashot Chilingarian

Cosmic Ray Division, Yerevan Physics Institute, Armenia









https://www.creativitypost.com/article/to_boost_brain_health_and_performance_harness_neuroplasticity_the_right_way

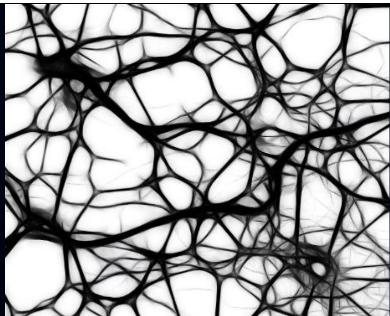
Human brain is a highly dynamic and constantly reorganizing system.

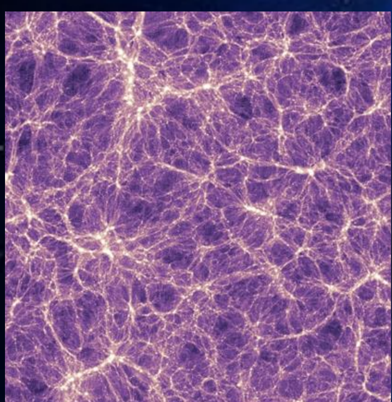
1. The central concept is the brain's lifelong capacity to change and rewire itself in response to the stimulation of learning and experience (neuroplasticity);
2. By solving problems (in physics) one can repeatedly stimulate the same area of the brain, which strengthens existing neural connections and creates new ones. Over time, the brain can become more efficient, requiring less effort to do the same job;
3. Continuous learning helps to fight against age-related decline and potential dementia pathology by increasing the connections between neurons, increasing cellular metabolism, and increasing the production a substance produced by the body to help maintain and repair neurons;
4. Do you know that London taxi drivers have a larger hippocampus than London bus drivers? Ask why or find yourself!

мозгом Больцмана называется гипотетически возможная субстанция (объект), возникающая в результате флуктуаций физического вакуума, осознавшая себя и способная наблюдать за окружающим миром.

только наличие наблюдателя (мозга Больцмана) упорядочивает окружающий мир. До появления стороннего наблюдателя Вселенная в обязательном порядке является хаотической.

Deep Density Displacement Model (D3M), which learns from a set of prerun numerical simulations, predicts the nonlinear large-scale structure of the Universe. Note the connections between galaxies: are they of the same purpose as in brain?





We show that D3M is able to accurately extrapolate far beyond its training data and predict structure formation for significantly different cosmological parameters.

https://doi.org/10.1073/pnas.1821458116

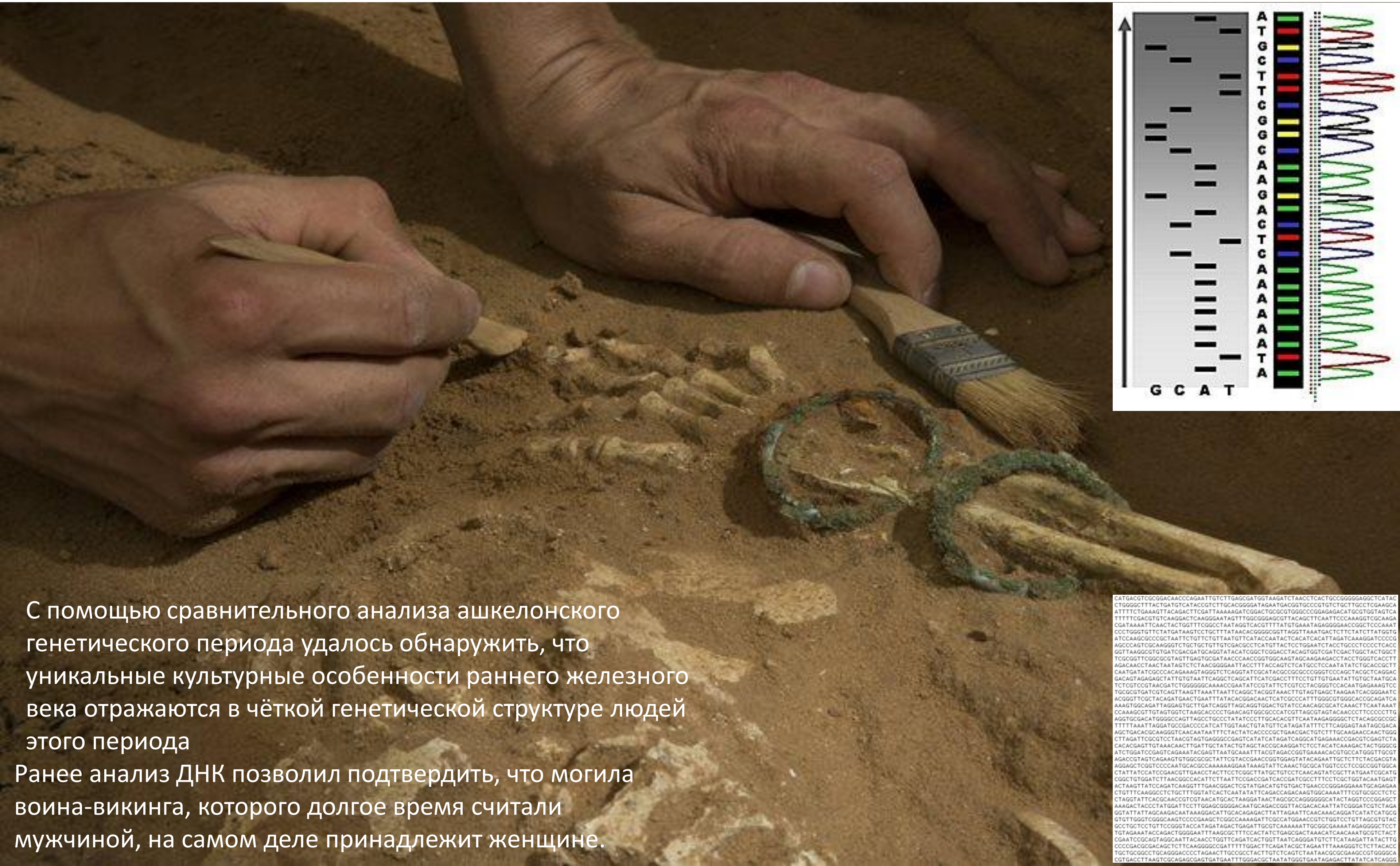
Machine learning: Measurements, Decisions and Predictions

1. **Decision making in random noisy environments: driver on the street, investor in the market, customer in the supermarket; student choosing profession, traveler choosing the rout, medical diagnostics, etc…**
2. **Pattern recognition: N x N picture; binary (black and white) or vector (color) in each of thousands (millions) of pixels: monitor crime on the streets, find a person; read a drawing, etc…**
3. **ML is not magic producing results on any input! you have to study math and statistics!**
4. **New algorithms and expert knowledge of the domain are in great need!**
5. **Maximal salary in IT sector: no problem with jobs in coming decades;**
6. **“Future of Armenia depends on ML” Armen Sargsyan!**

Domains

1. Speech recognition: automatic answering services;
2. Automatic control in mass production: check wrong pizza; and many others;
3. AI algorithms scanning data bases of all patients died from any diseases and finding optimal strategy of curing patient;
4. Optimal strategy of getting profit on highly violent stock markets: what and when to by, when to sell, how to put margins!
5. Transport: optimal routs in city, fast response to traffic jams, navigators, safety…
6. Big brother watching streets to prevent crime!
7. Education: personalize learning programs;

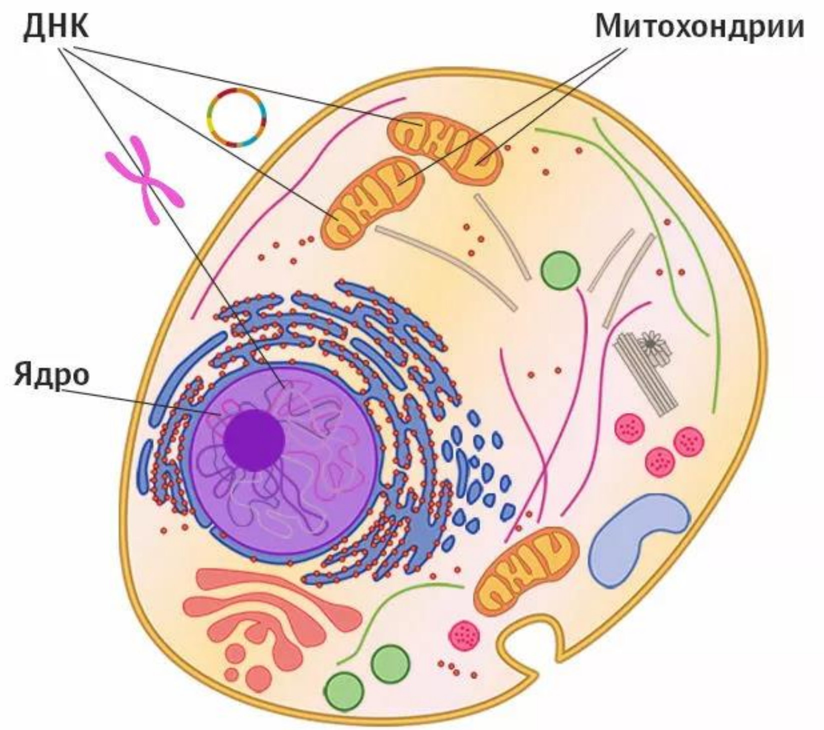
Paleo - genomics: Archeology



С помощью сравнительного анализа ашкелонского генетического периода удалось обнаружить, что уникальные культурные особенности раннего железного века отражаются в чёткой генетической структуре людей этого периода

Ранее анализ ДНК позволил подтвердить, что могила воина-викинга, которого долгое время считали мужчиной, на самом деле принадлежит женщине.

У человека один из самых маленьких митохондриальных геномов, всего 16,5 тысячи пар нуклеотидов, 37 генов. Для сравнения: у наземных растений — сотни тысяч пар. мтДНК наследуется только от матери к дочери.



Human genome contains 3 billion coding elements

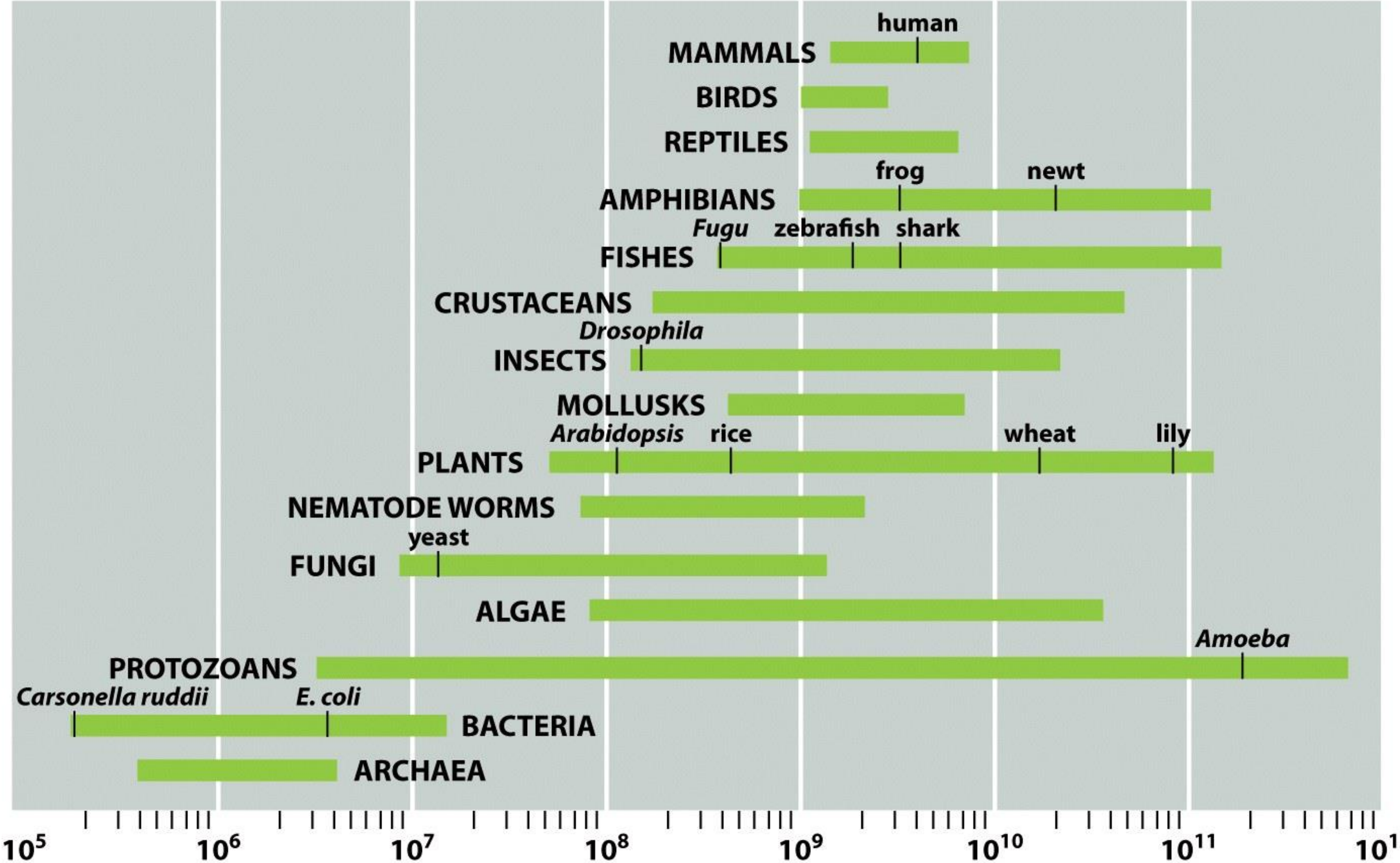
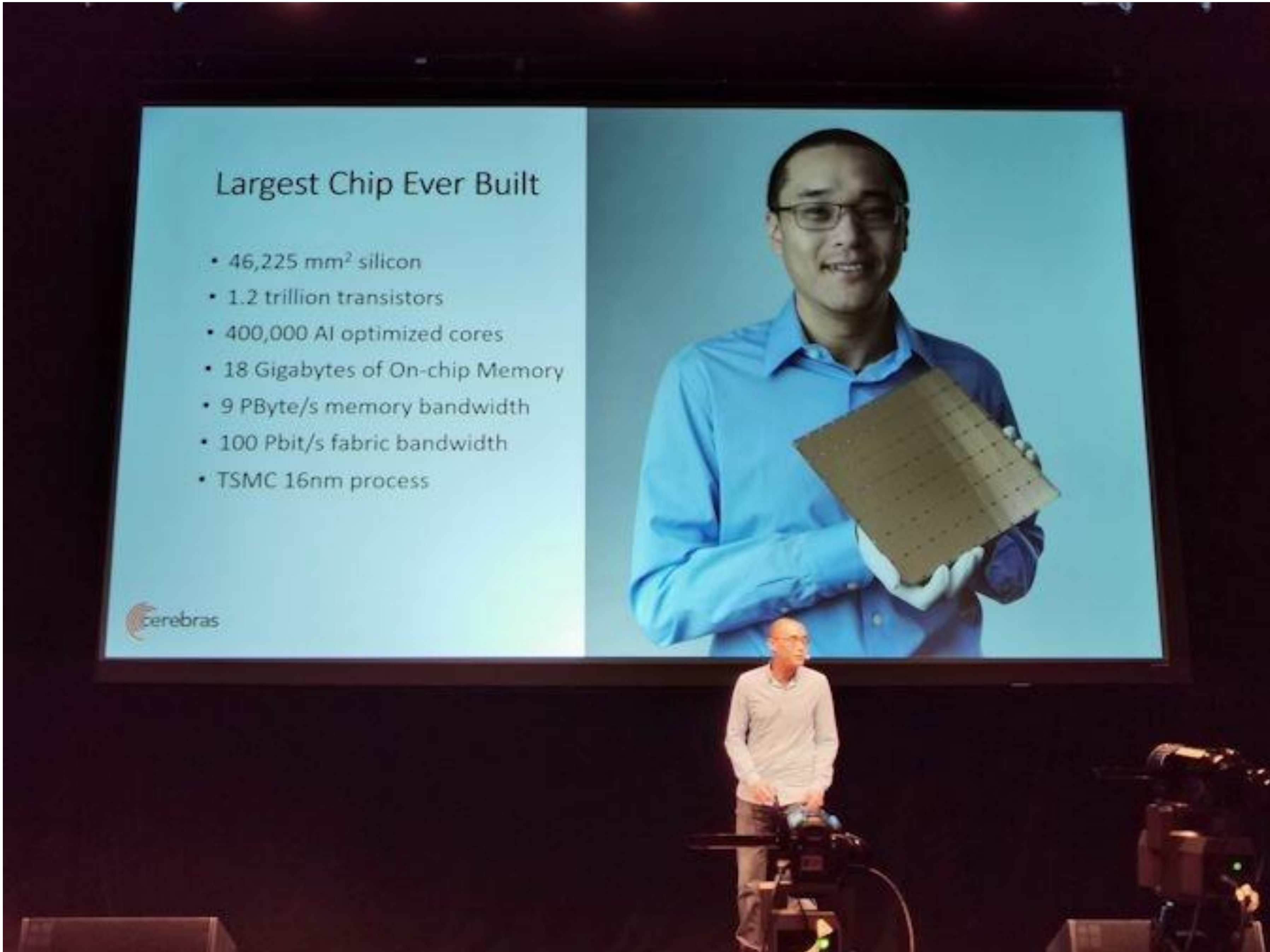


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

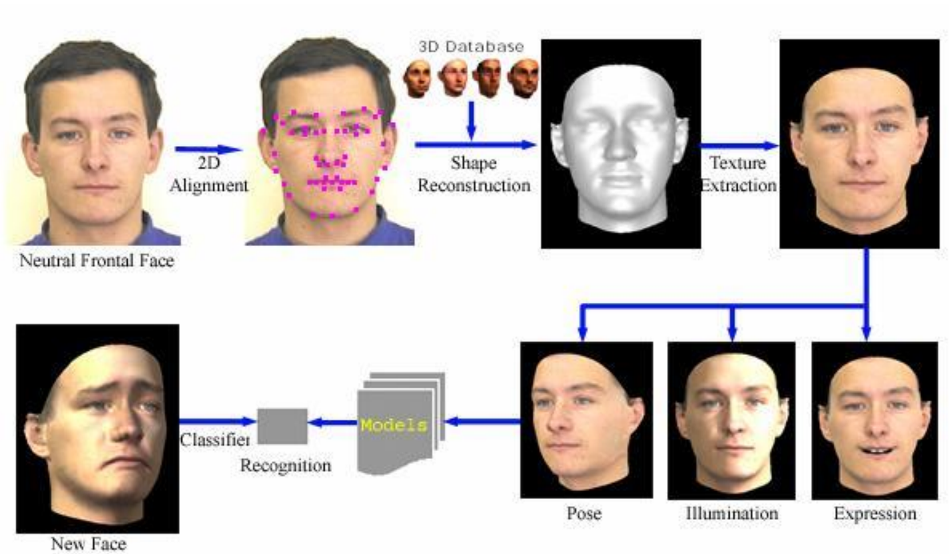

Largest Chip Ever Built
• 46,225 mm² silicon
• 1.2 trillion transistors
• 400,000 AI optimized cores
• 18 Gigabytes of On-chip Memory
• 9 PByte/s memory bandwidth
• 100 Pbit/s fabric bandwidth
• TSMC 16nm process
cerebras

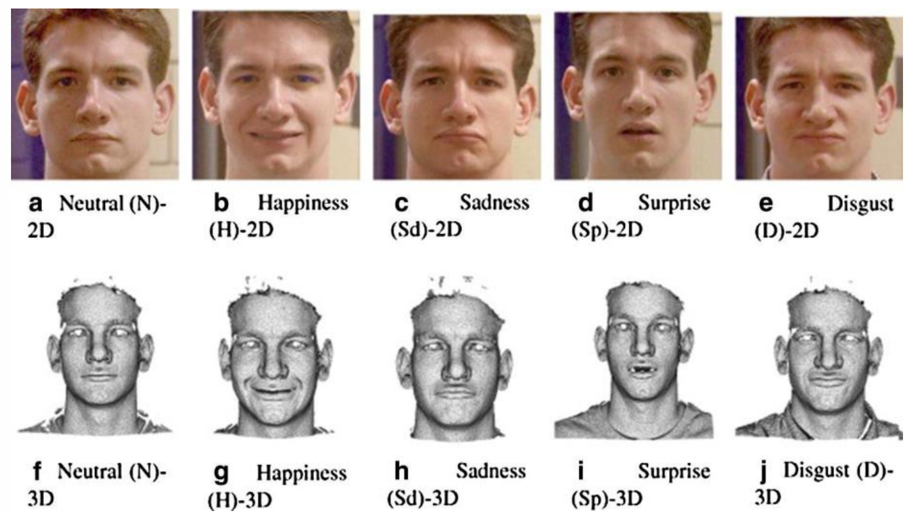
Cerebras — Cerebras Wafer Scale Engine

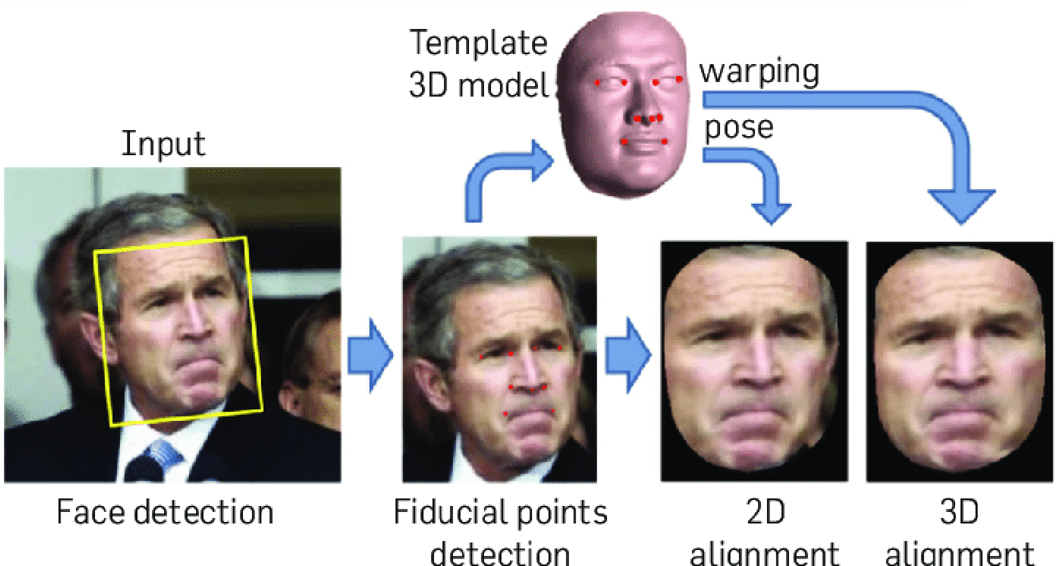
1. Каждый кристалл Cerebras WSE содержит 1,2 трлн транзисторов, организованных в 400 000 ИИ-оптимизированных вычислительных ядер и 18 Гбайт локальной распределённой памяти SRAM;
2. Всё это связано ячеистой сетью с общей производительностью 100 петабит в секунду. Пропускная способность памяти достигает 9 Пбайт/с;
3. В сравнении с самыми современными графическими ядрами чип Cerebras обеспечивает в 3000 раз больший объём памяти на кристалле и в 10 000 большую скорость обмена с памятью;
4. Вычислительные ядра Cerebras — SLAC (Sparse Linear Algebra Cores) — полностью программируемые и могут быть оптимизированы для работы с любыми нейронными сетями.;
5. процессор Cerebras оказывается в сотни или даже тысячи раз эффективнее для ИИ и машинного обучения.

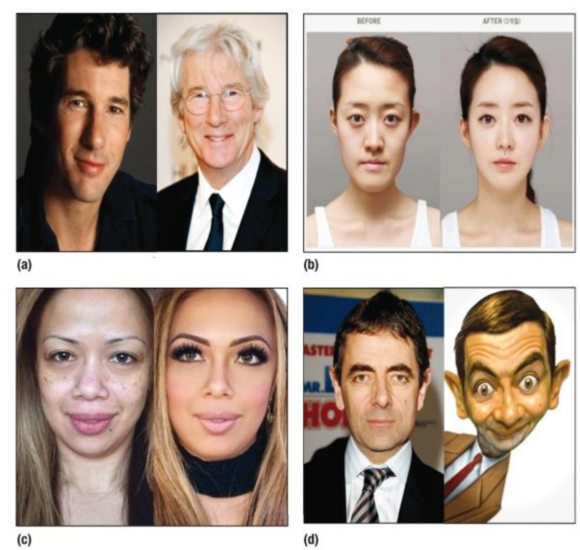
Images are not static!

Automatic alignment and pose estimation









Preprint ЕФИ-992(42)-87

ԵՐԵՎԱՆԻ ՖԻԶԻԿԱՅԻ ԻՆՍՏԻՏՈՒՏ
ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ
YEREVAN PHYSICS INSTITUTE

A.A.CHILINGARYAN,N.O.KHUDONYAN,

D.B.SAAKYAN,G.Z.ZAZYAN

RECOGNITION OF CORRELATED PATTERNS WITH SPIN GLASS-LIKE MODELS

ЦНИИатоминформ
ЕРЕВАН — 1987

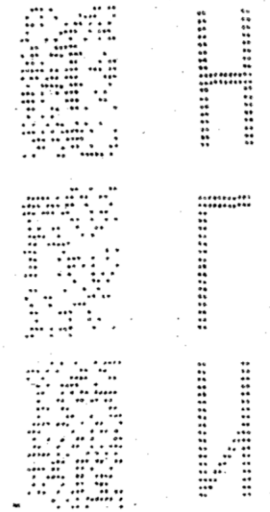


Fig.2

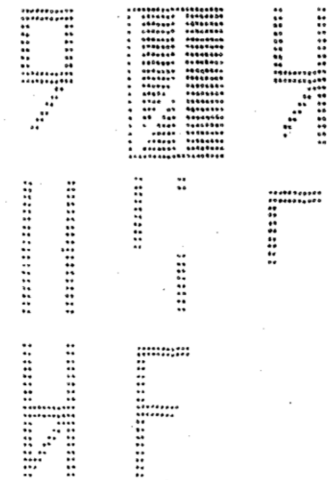
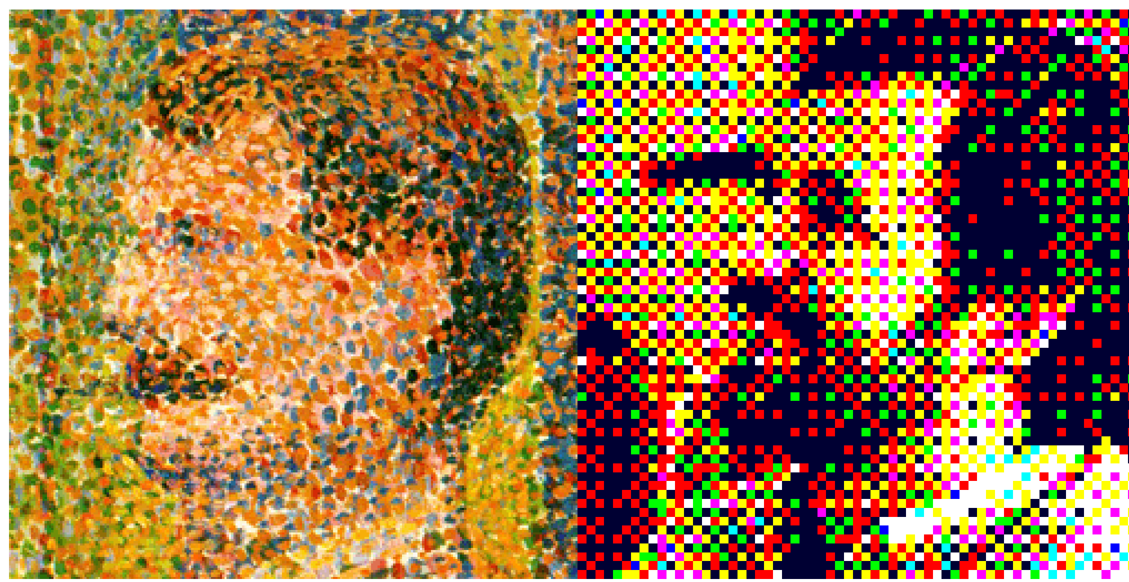
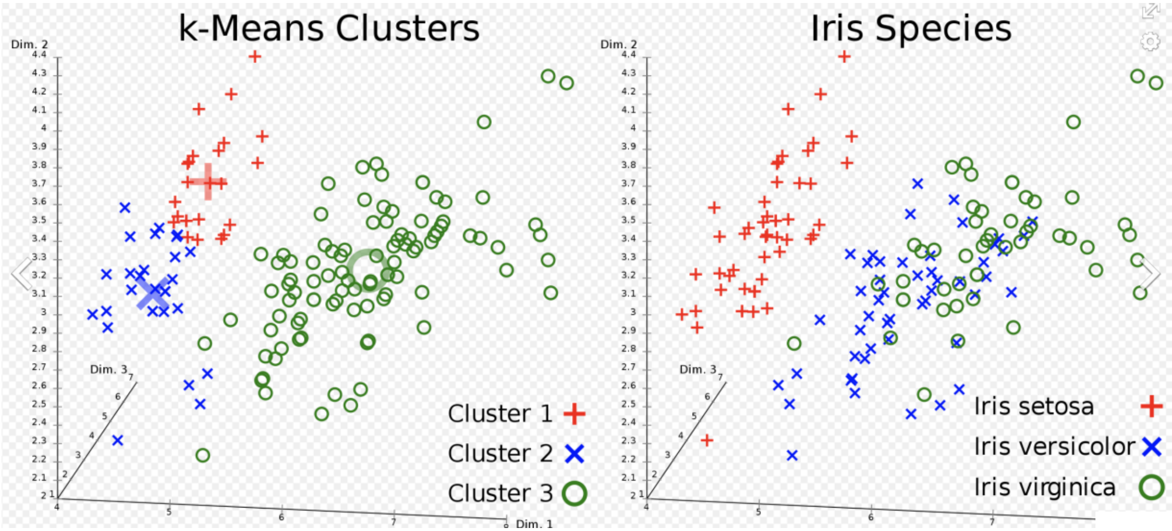


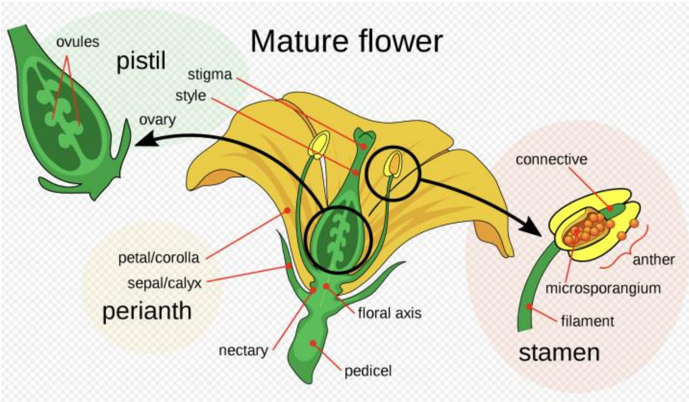
Fig.1

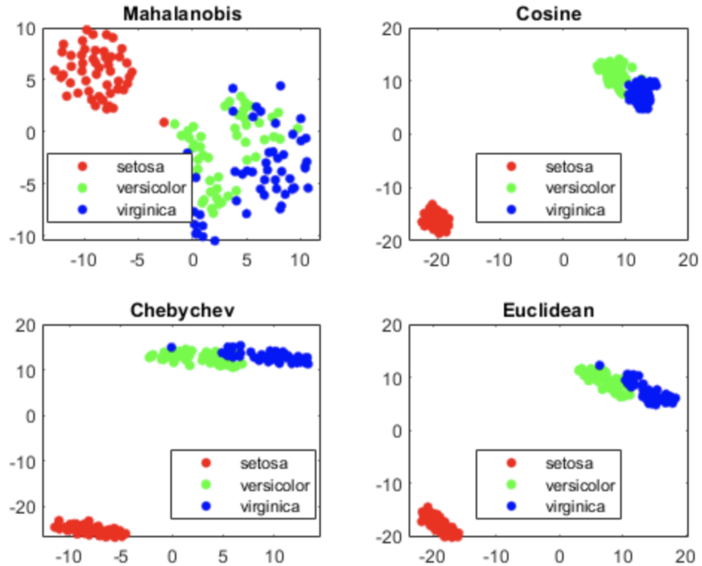


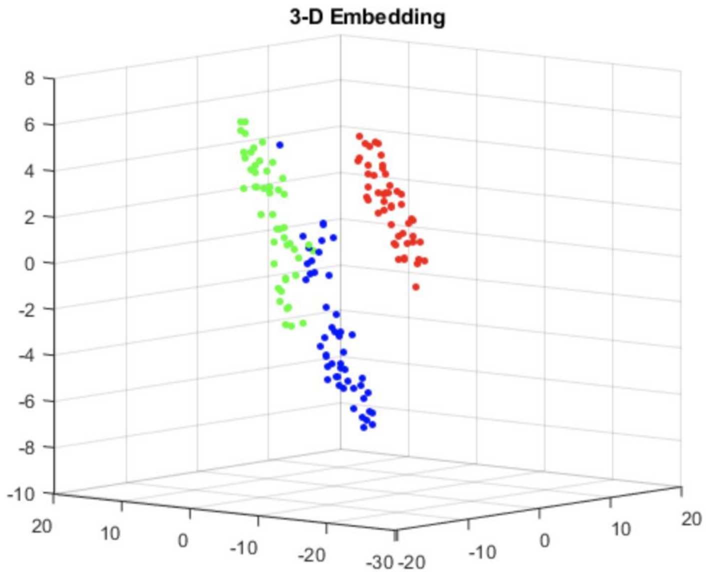
Stock Markets: forecasting; regression; optimal strategy: Stocks, Commodities Tumble as China Strikes Back

Fisher IRIS flower dataset (1936)









https://archive.ics.uci.edu/ml/datasets/iris

UC Irvine Machine Learning Repository

https://archive.ics.uci.edu/ml/index.php

UCI

About Citation Policy Donate a Data Set Contact

Search

Repository Web Google

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Browse Through: **481** Data Sets

Table View List View

Default Task
- Classification (356)
- Regression (103)
- Clustering (88)
- Other (55)

Attribute Type
- Categorical (38)
- Numerical (318)
- Mixed (55)

Data Type
- Multivariate (367)
- Univariate (24)
- Sequential (49)
- Time-Series (94)
- Text (55)
- Domain-Theory (23)
- Other (21)

Area
- Life Sciences (109)
- Physical Sciences (49)
- CS / Engineering (177)
- Social Sciences (26)
- Business (31)
- Game (10)
- Other (74)

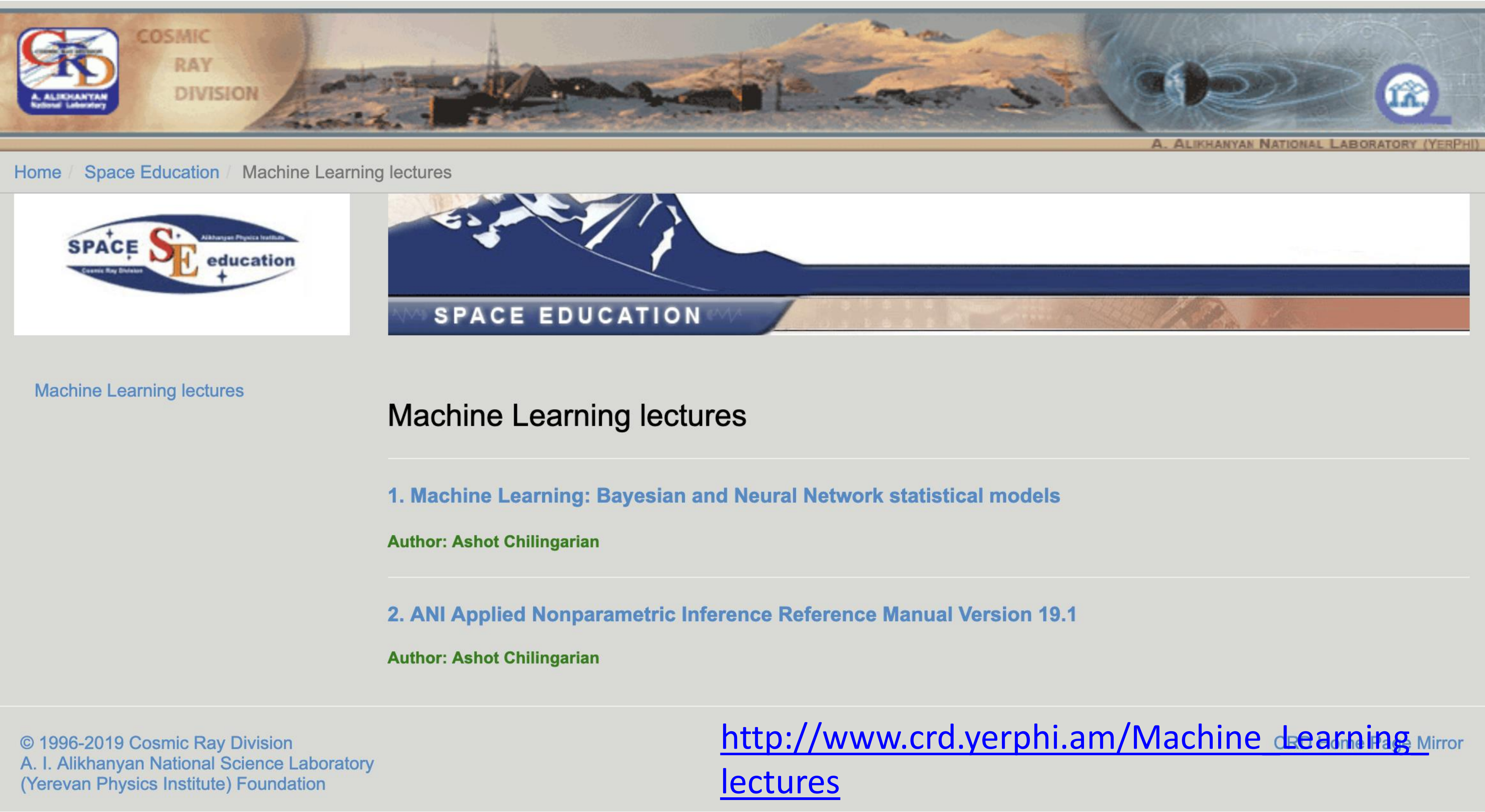
 # Attributes
- Less than 10 (117)
- 10 to 100 (217)
- Greater than 100 (85)

 # Instances
- Less than 100 (27)
- 100 to 1000 (166)
- Greater than 1000 (254)

Format Type

Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998
Artificial Characters	Multivariate	Classification	Categorical, Integer, Real	6000	7	1992
Audiology (Original)	Multivariate	Classification	Categorical	226		1987
Audiology (Standardized)	Multivariate	Classification	Categorical	226	69	1992
Auto MPG	Multivariate	Regression	Categorical, Real	398	8	1993

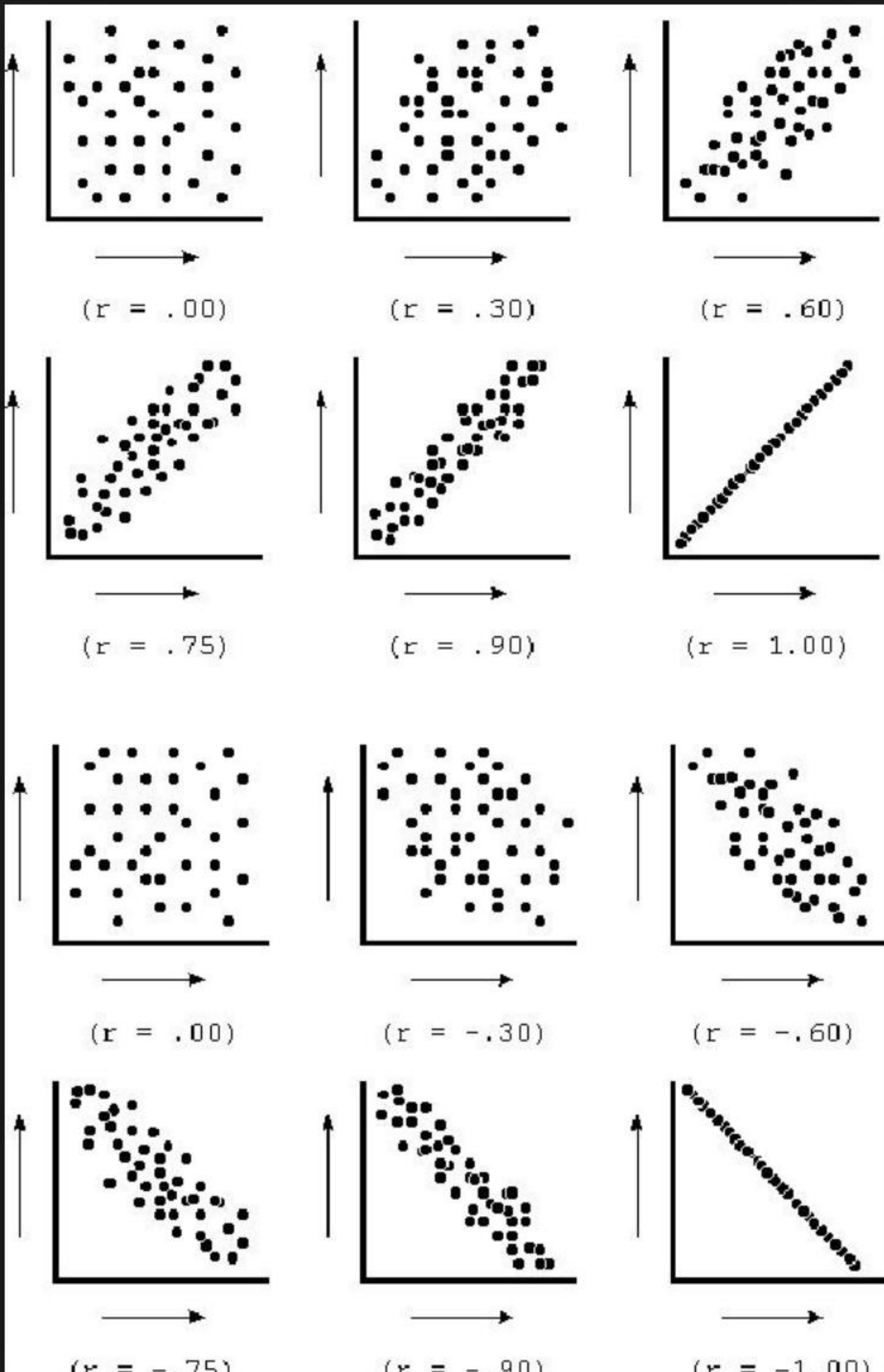
CRD Machine Learning resources

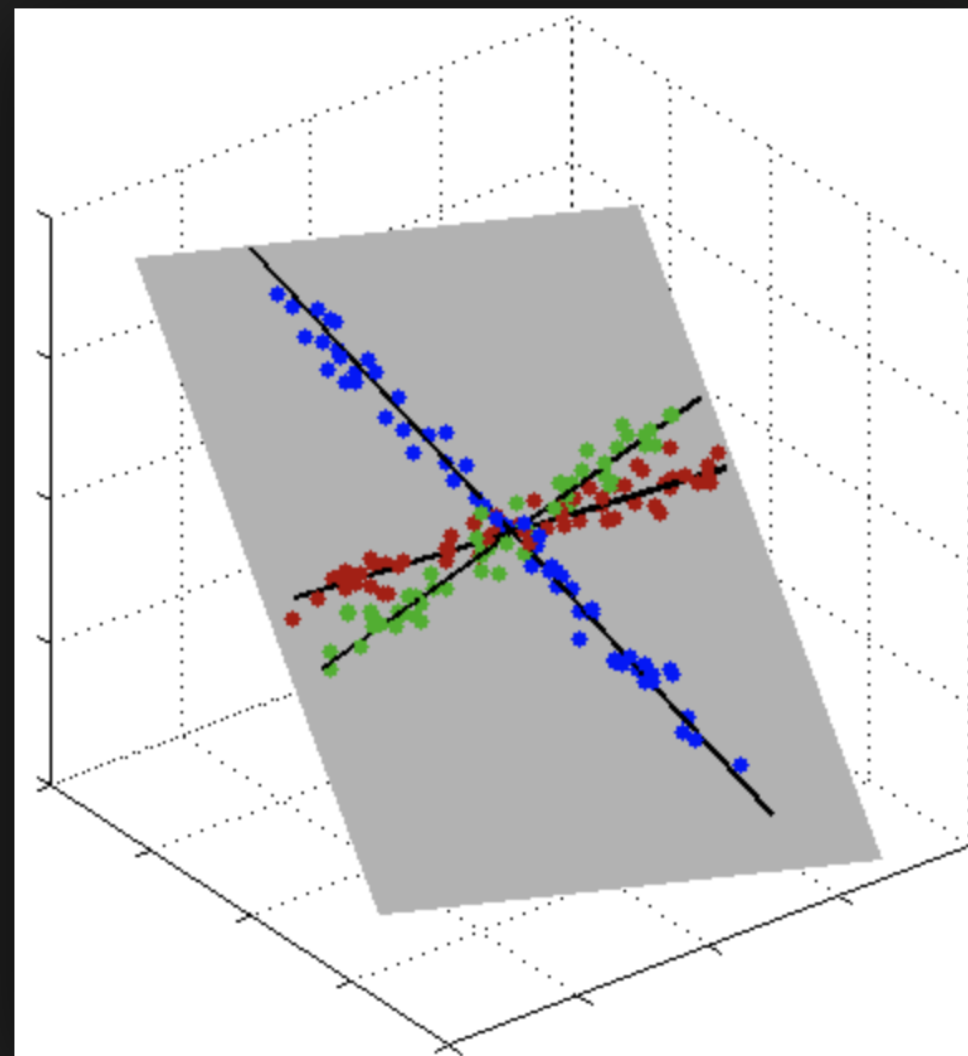


http://www.crd.yerphi.am/Machine_Learning_lectures

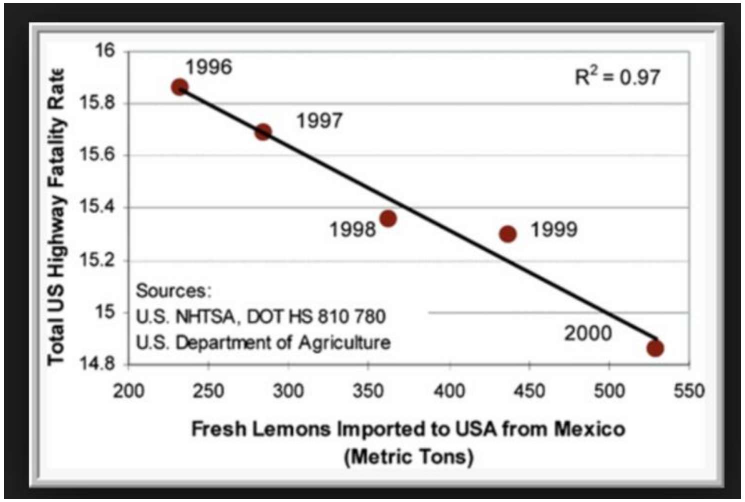
Correlation

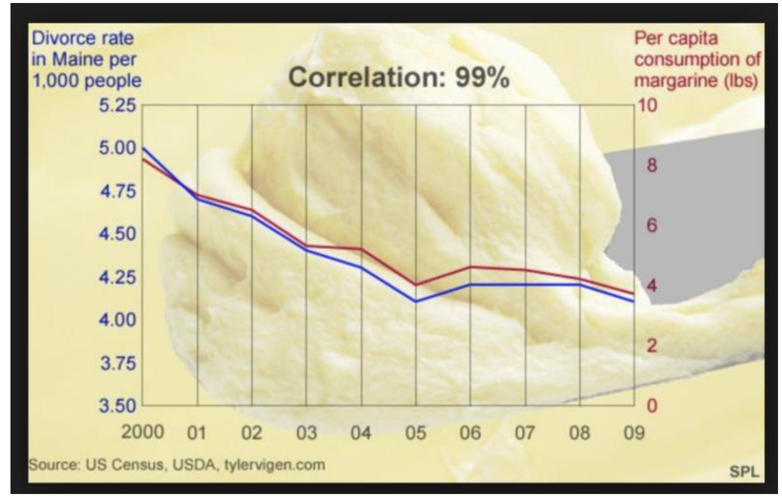
$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\,n\Sigma x^2 - (\Sigma x)^2\,][\,n\Sigma y^2 - (\Sigma y)^2\,]}}$$

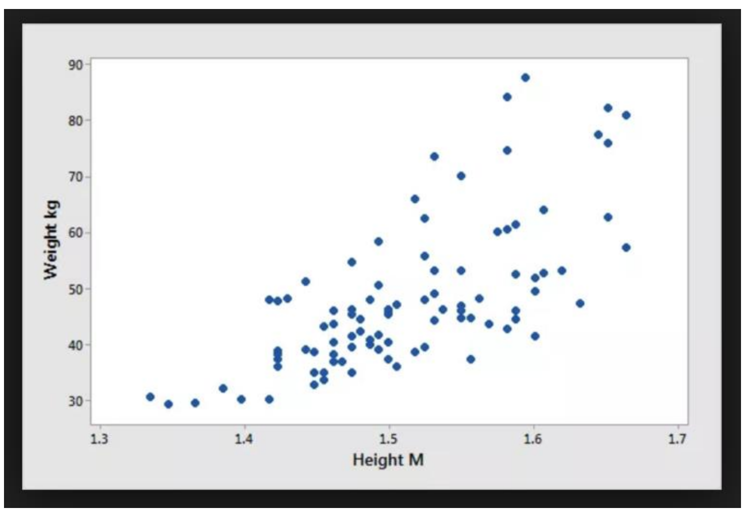




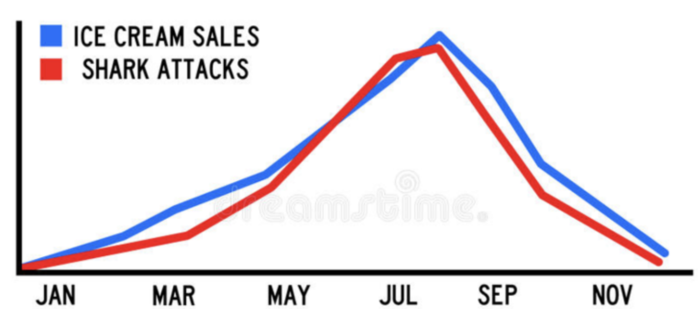
"False" Correlations no causal relation







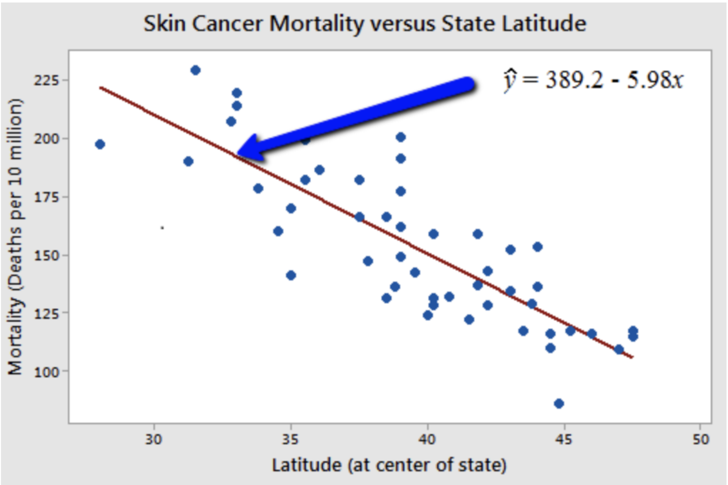
CORRELATION IS NOT CAUSATION!

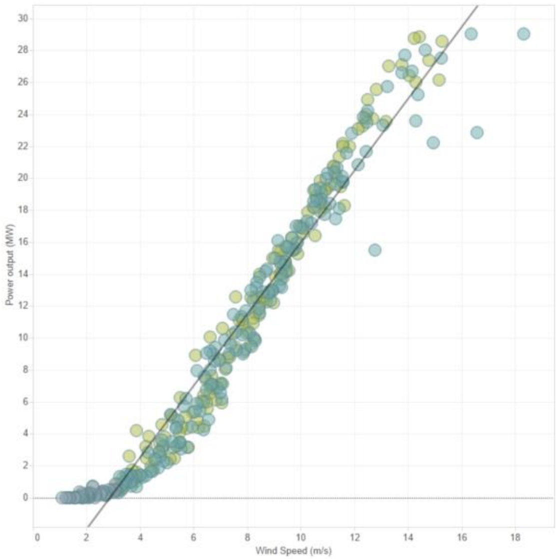


Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

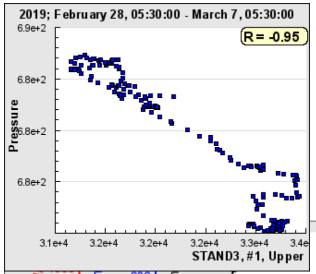
Observer make correlation: number of umbrellas – intensity of rain and get highly positive; thus we can make rain by taking umbrellas!

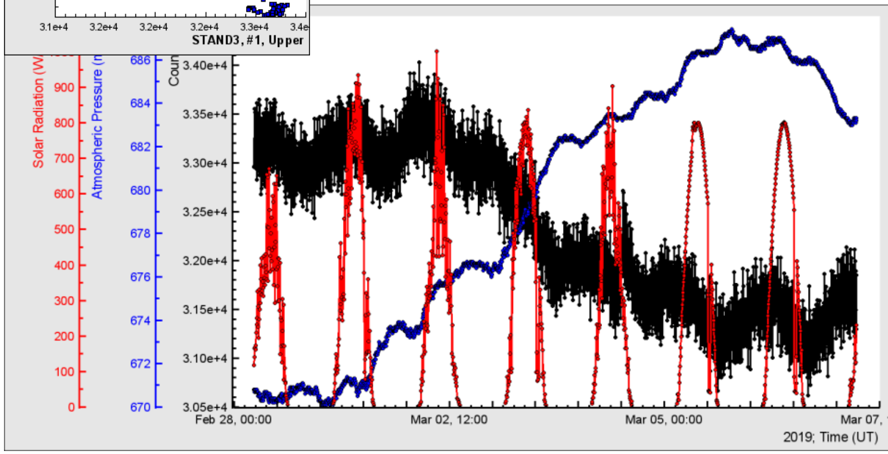
True correlation – causal relations

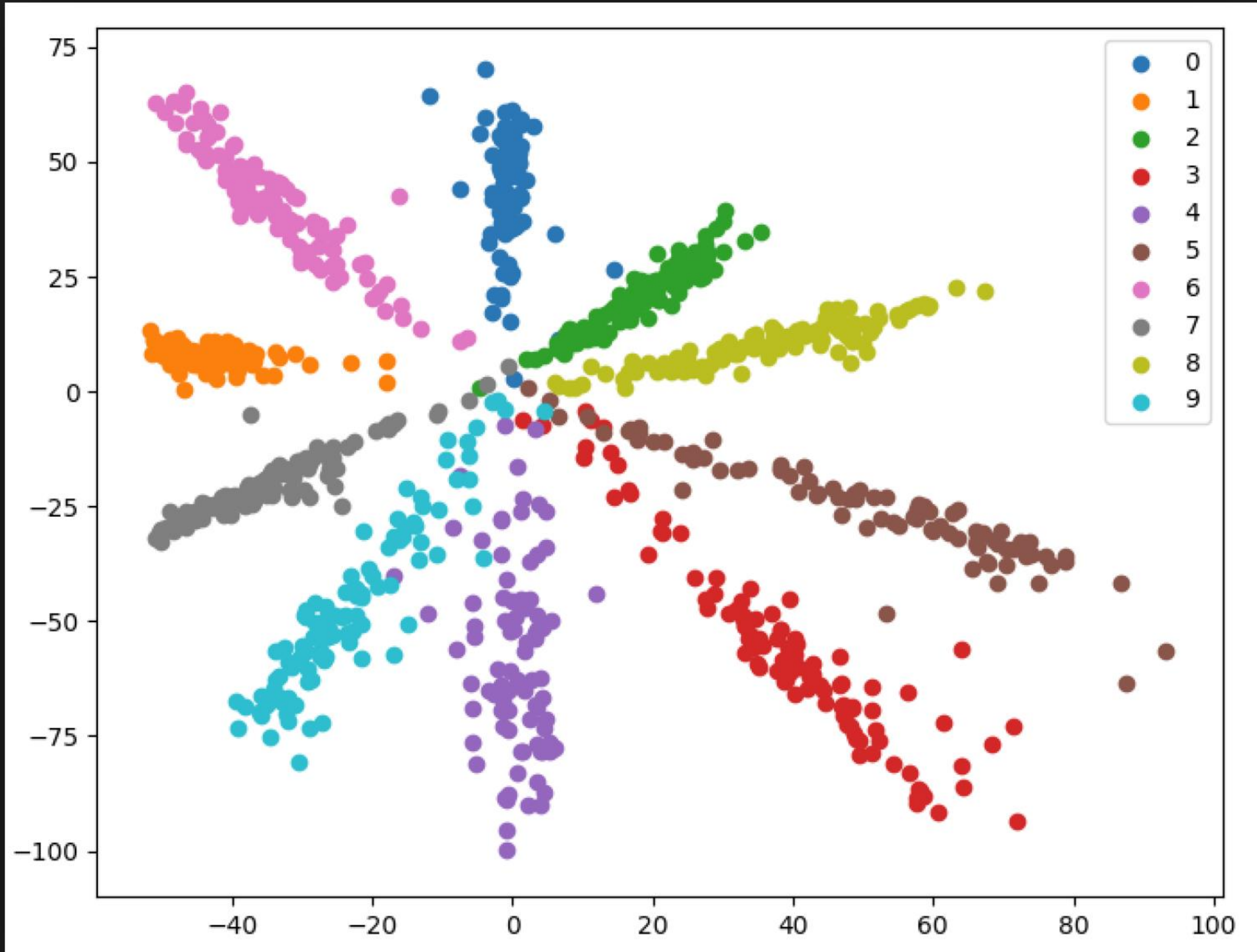




Atmospheric pressure vs count rate



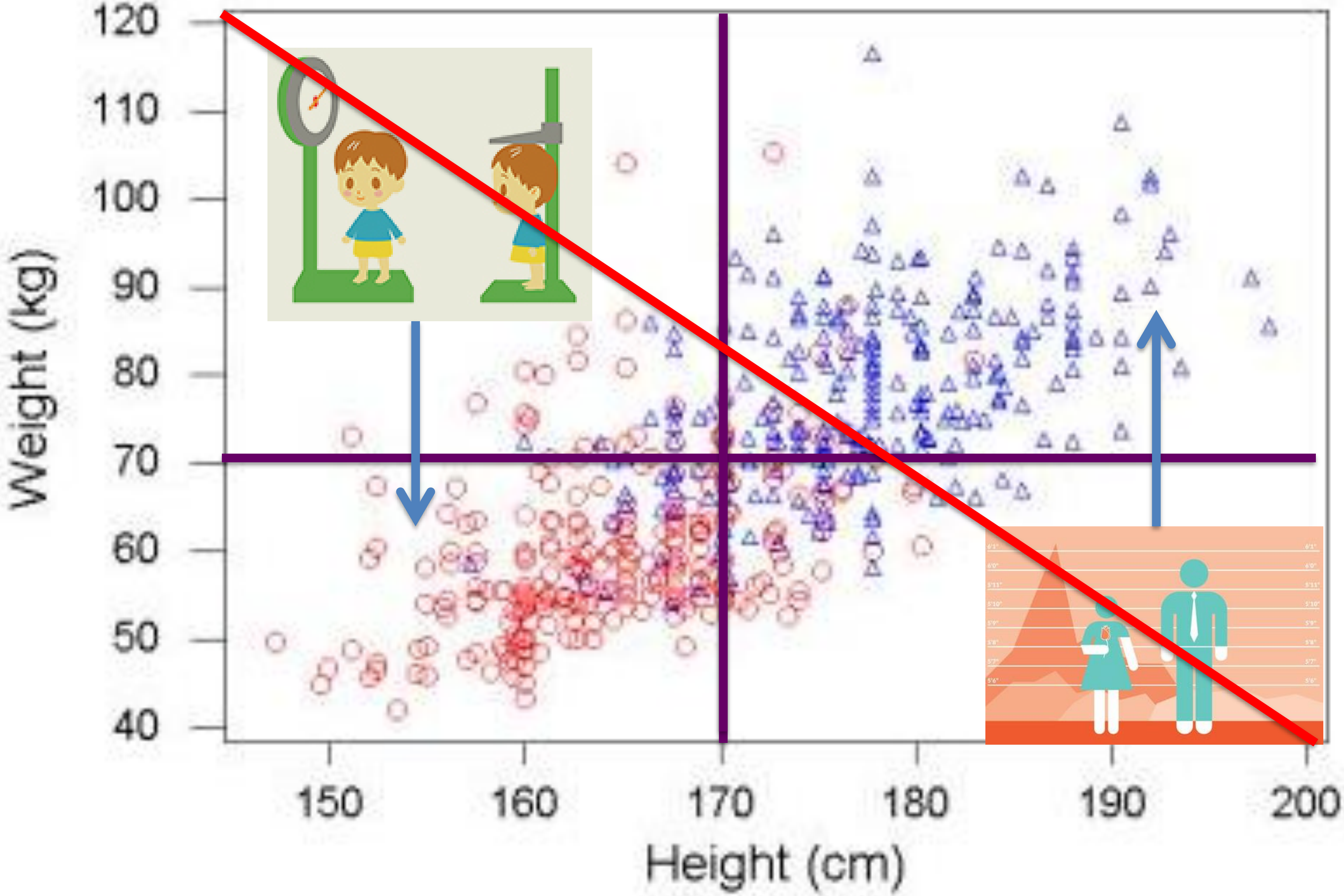




75
50
25
0
−25
−50
−75
−100
−40
−20
0
20
40
60
80
100
0
1
2
3
4
5
6
7
8
9

Example: automatic selection of child and adult

1. Goal – cinema: 12+, or 16+; Speed of autopilots on the roads in presence near schools; airplane passengers;
2. Select variables: measurements of weight and height (no problem);
3. Domain – direct product of x and y spaces R1 x R1;
4. How to deal with random variables? The mean and variance: Gaussian distribution! 2 main parameters!
5. 2 different variables: deviation from mean; %, or N of sigma, range.
6. Classification problem as mapping problem



Weight (kg)
120
110
100
90
80
70
60
50
40
150
160
170
180
190
200
Height (cm)

Mapping feature space to decision space!

1. **f(x, y)) = x+y :** What will be if we add kilograms and meters?
2. **Percent of mean value:** Mean weight 70 kg, mean height – 170 cm;
3. 100 kg (100-60)/60 ~ 66.7%; 30 kg (30-60)/60=-50%;
4. Now we can add percent! F(x+y)=66.7%+ -50% = 17.7%
5. Mapping of 2 dimensional future space – to one dimensional decision space!
6. f(x,y) = 1, if x+y < 50%; f(x,y)=2, if x+y > 50%
7. But why 50? Maybe 45 will be better?
8. Training with teacher to find best mapping function!

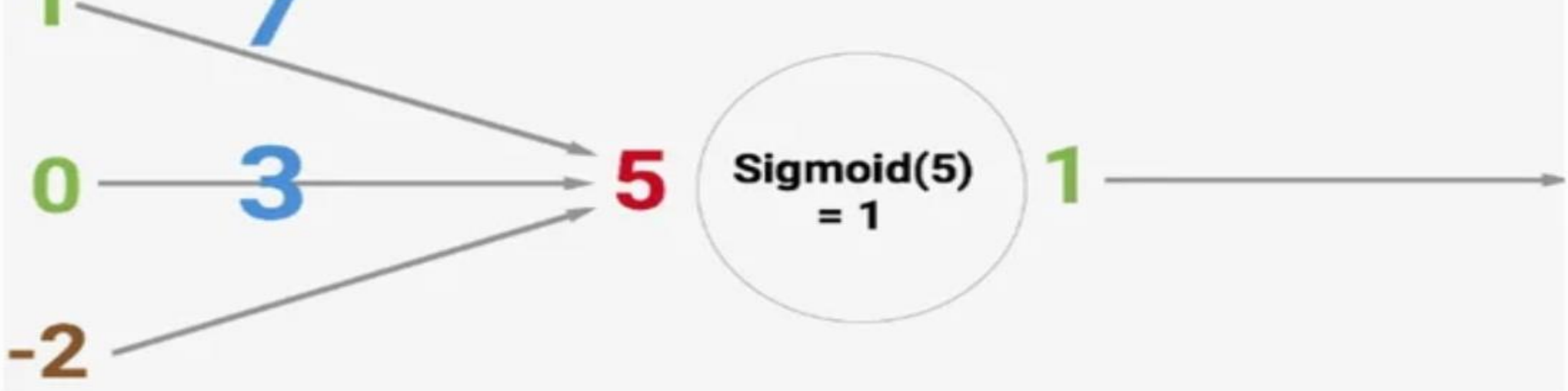
Metrics to Evaluate your Machine Learning Algorithm (we need training sample TS – measurements with known classification)

$$Accuracy = \frac{Number\ of\ Correct\ predictions}{Total\ number\ of\ predictions\ made}$$

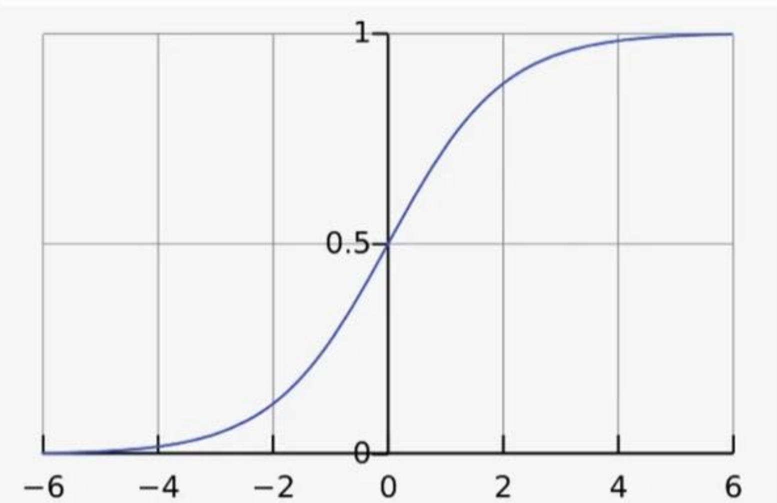
n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

$$MeanAbsoluteError = \frac{1}{N}\sum_{j=1}^{N}|y_j - \hat{y}_j| \qquad MeanSquaredError = \frac{1}{N}\sum_{j=1}^{N}(y_j - \hat{y}_j)^2$$

How meth neuron works?



$$1 * 7 + 0 * 3 - 2 = 5$$

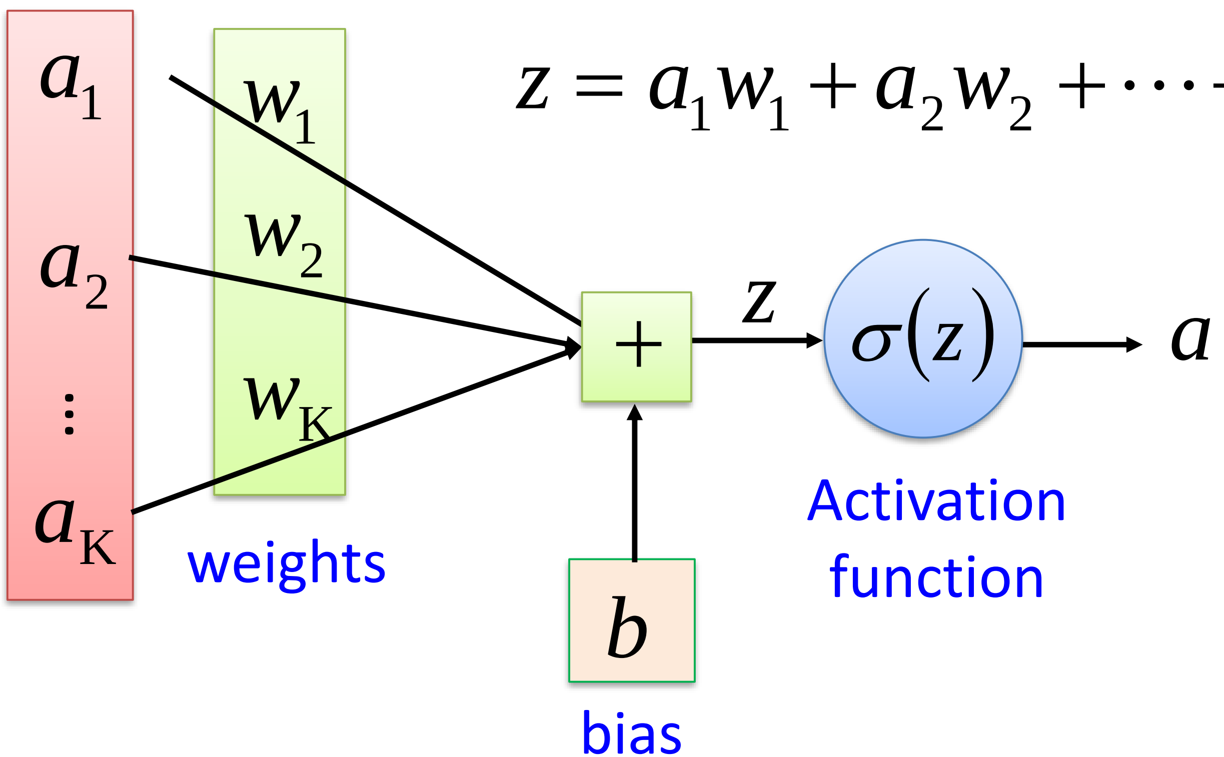


https://zen.yandex.ru/media/id/5bbcbc1ba5bd5400a990e7d9/izuchaem-neironnye-seti-kak-sozdat-neiroset-za-4shaga-5cda8a1a5631d800b3136f0f

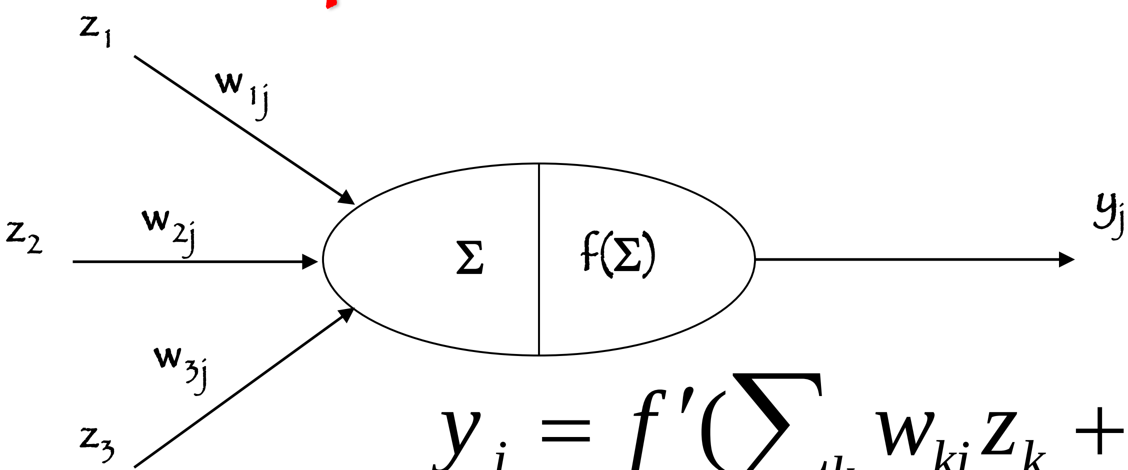
Element of Neural Network

Neuron $f: R^K \rightarrow R$

$$z = a_1 w_1 + a_2 w_2 + \cdots + a_K w_K + b$$

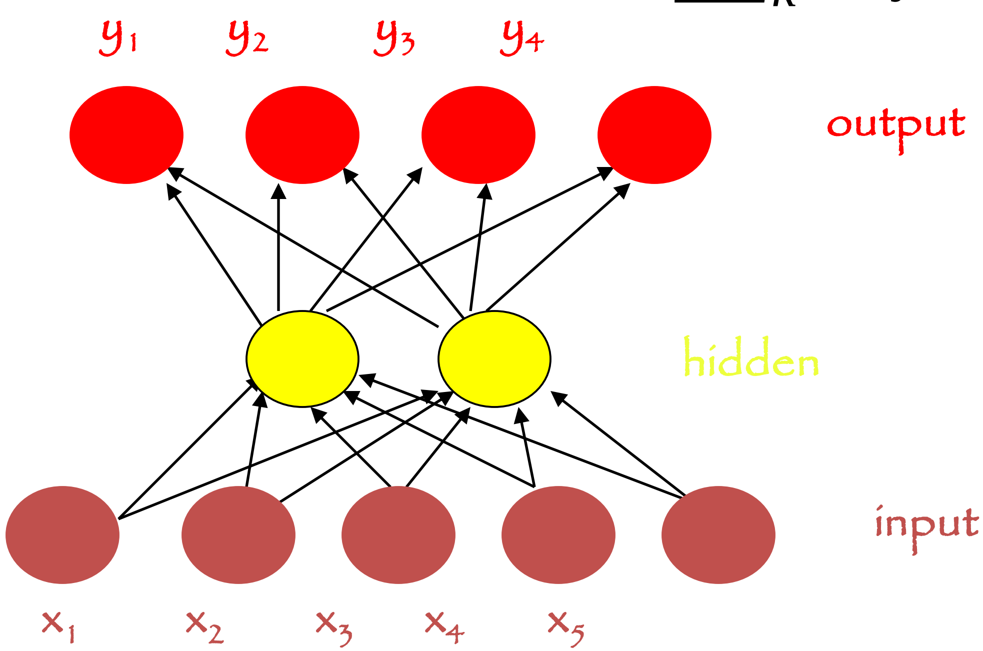


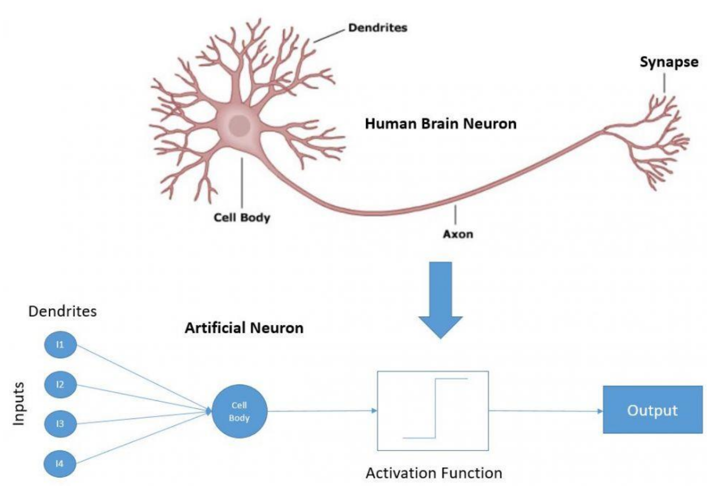
The artificial neuron: how to train?



$$y_j = f'\left(\sum_k w_{kj} z_k + w_{0j}\right) =$$

$$= f'\left(\sum_k w_{kj}\left(f\left(\sum_i w_{ik} x_i + w_{0k}\right)\right) + w_{0j}\right)$$





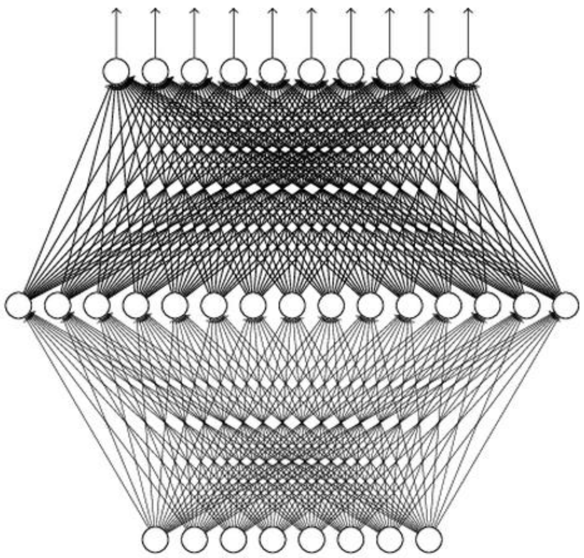
Universality Theorem

Any continuous function f

$$f : R^N \rightarrow R^M$$

Can be realized by a network with one hidden layer

(given **enough** hidden neurons)! So, what for we need deep NN with many-many layers?



Reference for the reason: http://neuralnetworksanddeeplearning.com/chap4.html

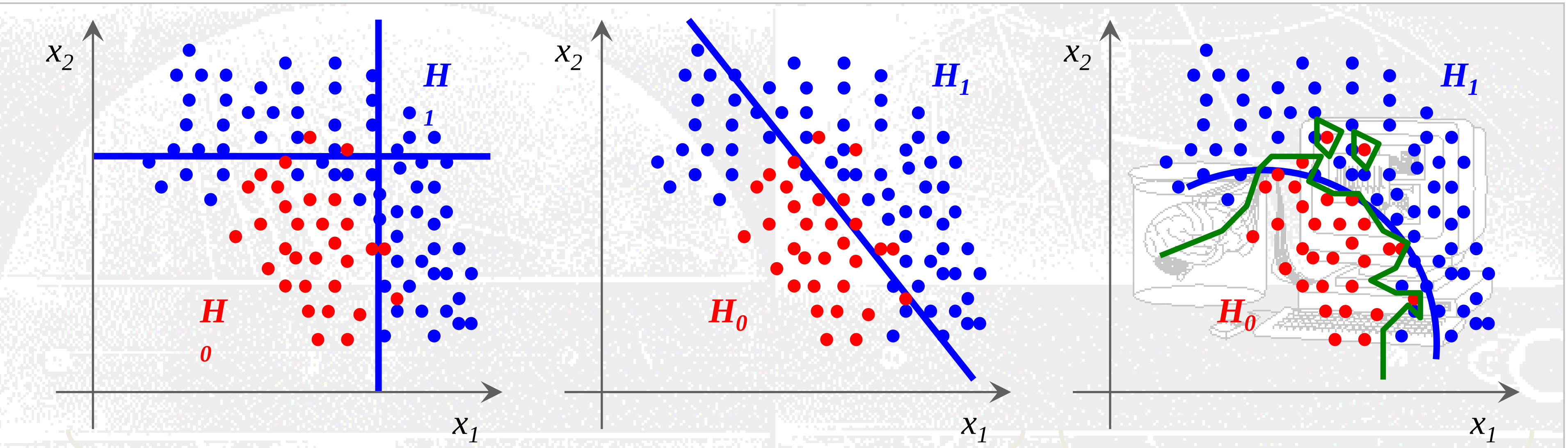
Why "Deep" neural network not "Fat" neural network?

Classification of primary hadron to light and heavy nuclei

Suppose data sample with two types of events: Protons and Iron nuclei obtained from the simulation – solving direct problem of CR!

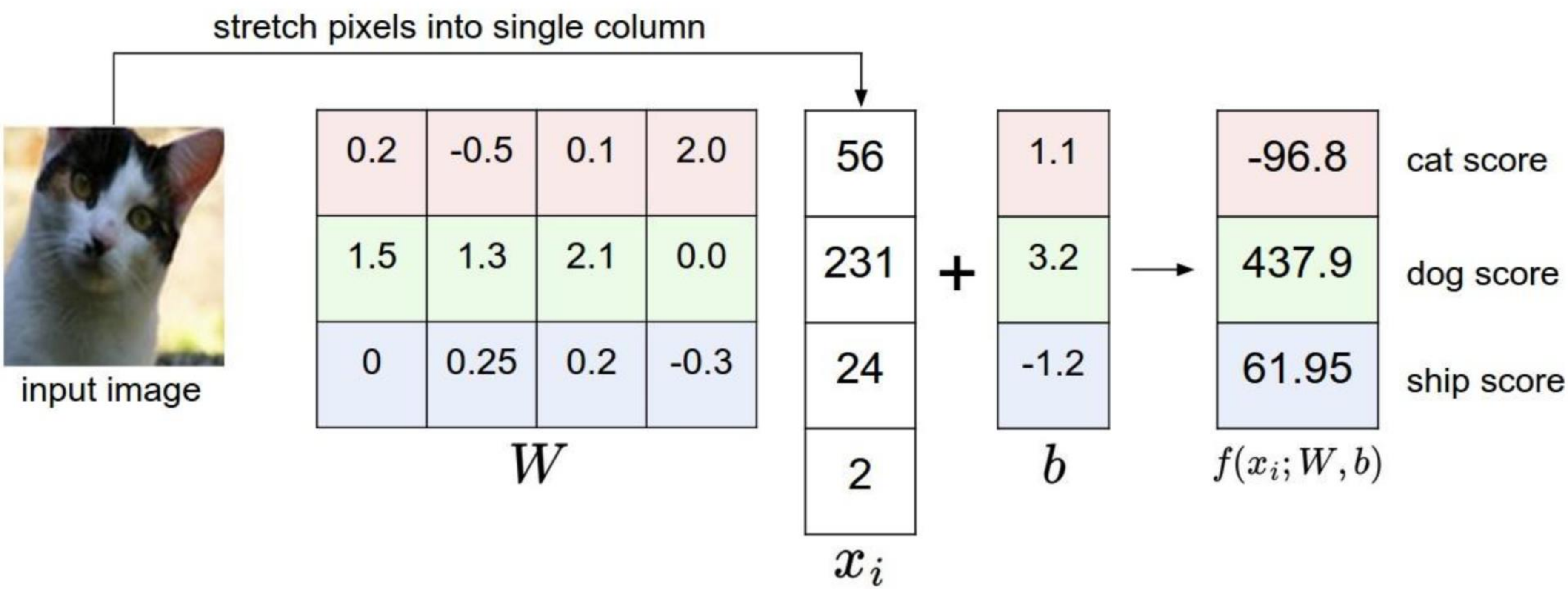
- *We have found discriminating input parameters Ne and Nmu*
- *What decision boundary should we use to select Iron nuclei ?*

Rectangular cuts? *A linear boundary?* *A nonlinear one?*

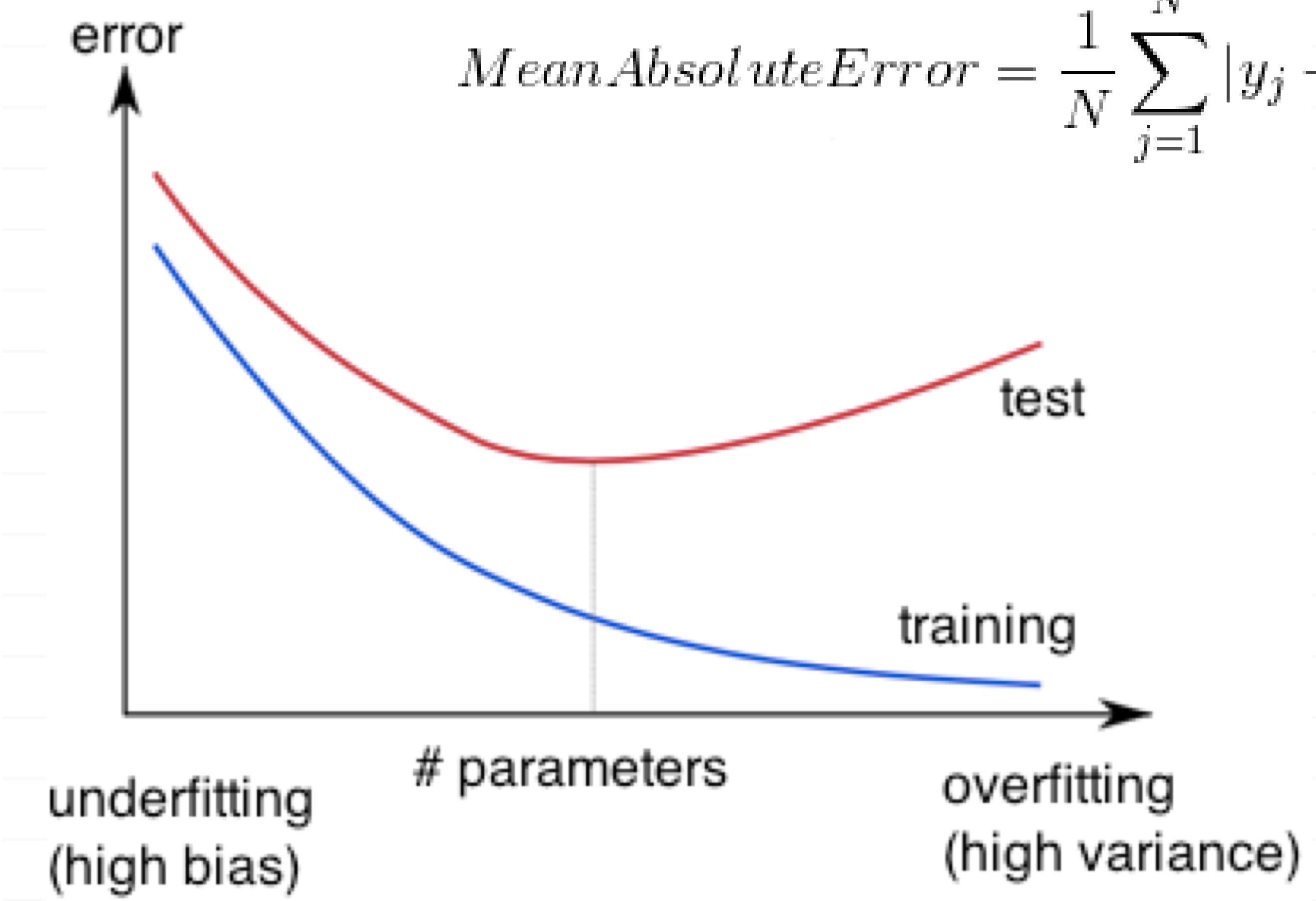


- How can we decide this in an optimal way ? → Let artificial neural network do it!
- We need training samples – training with teacher – simulations, direct CR problem

Recognizing cat: multiple feature
N x N pixels

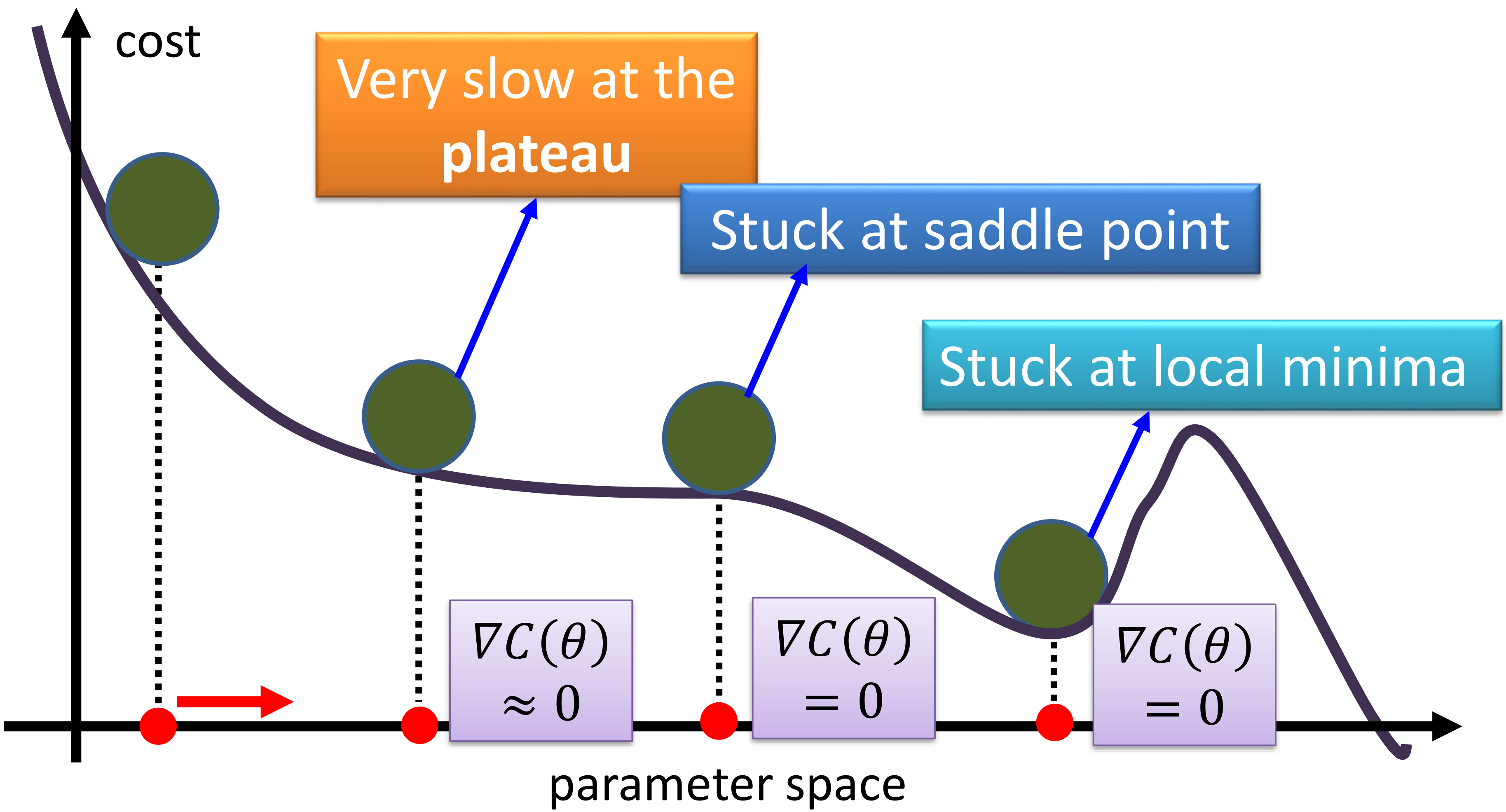


Overtraining and generalization

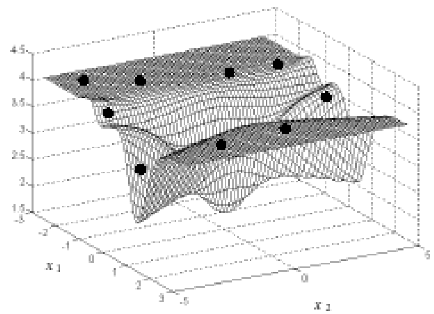


$$MeanAbsoluteError = \frac{1}{N}\sum_{j=1}^{N}|y_j - \hat{y}_j|$$

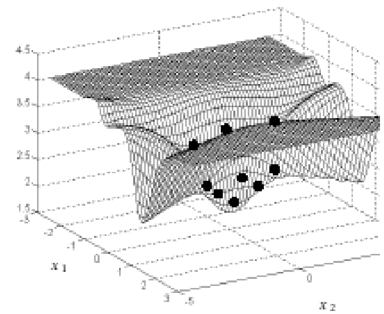
Local and global minimum



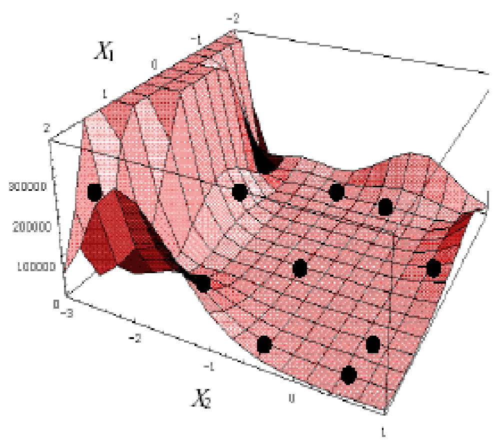
Searching of the Extremum



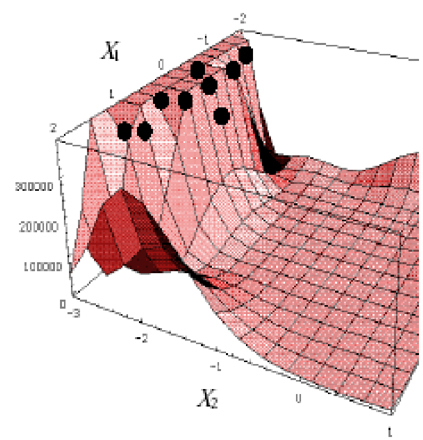
minimisation problem: $t = 1$



$t = 10$

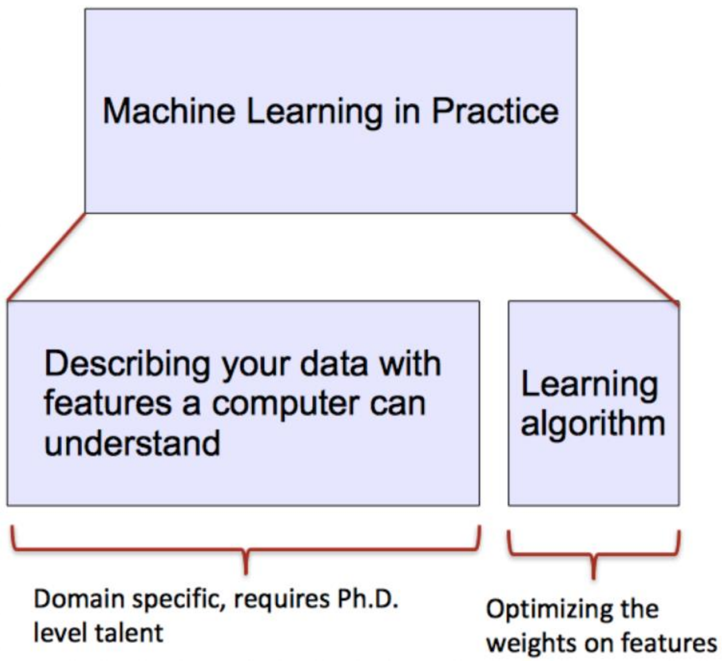


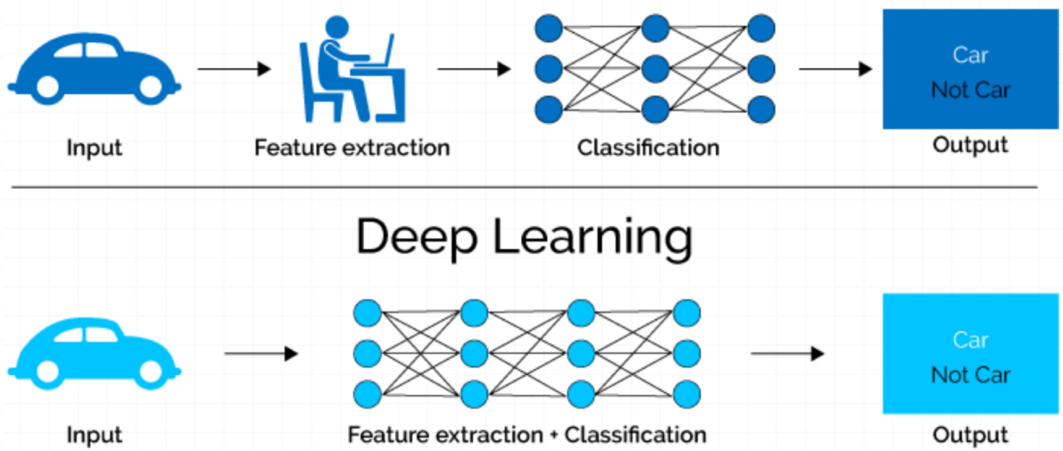
maximisation problem: $t = 1$

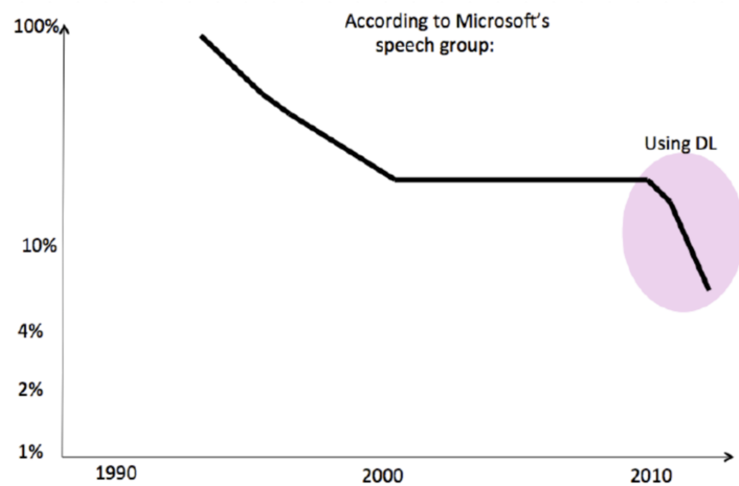


$t = 1000$

Deep learning (with and without teacher)

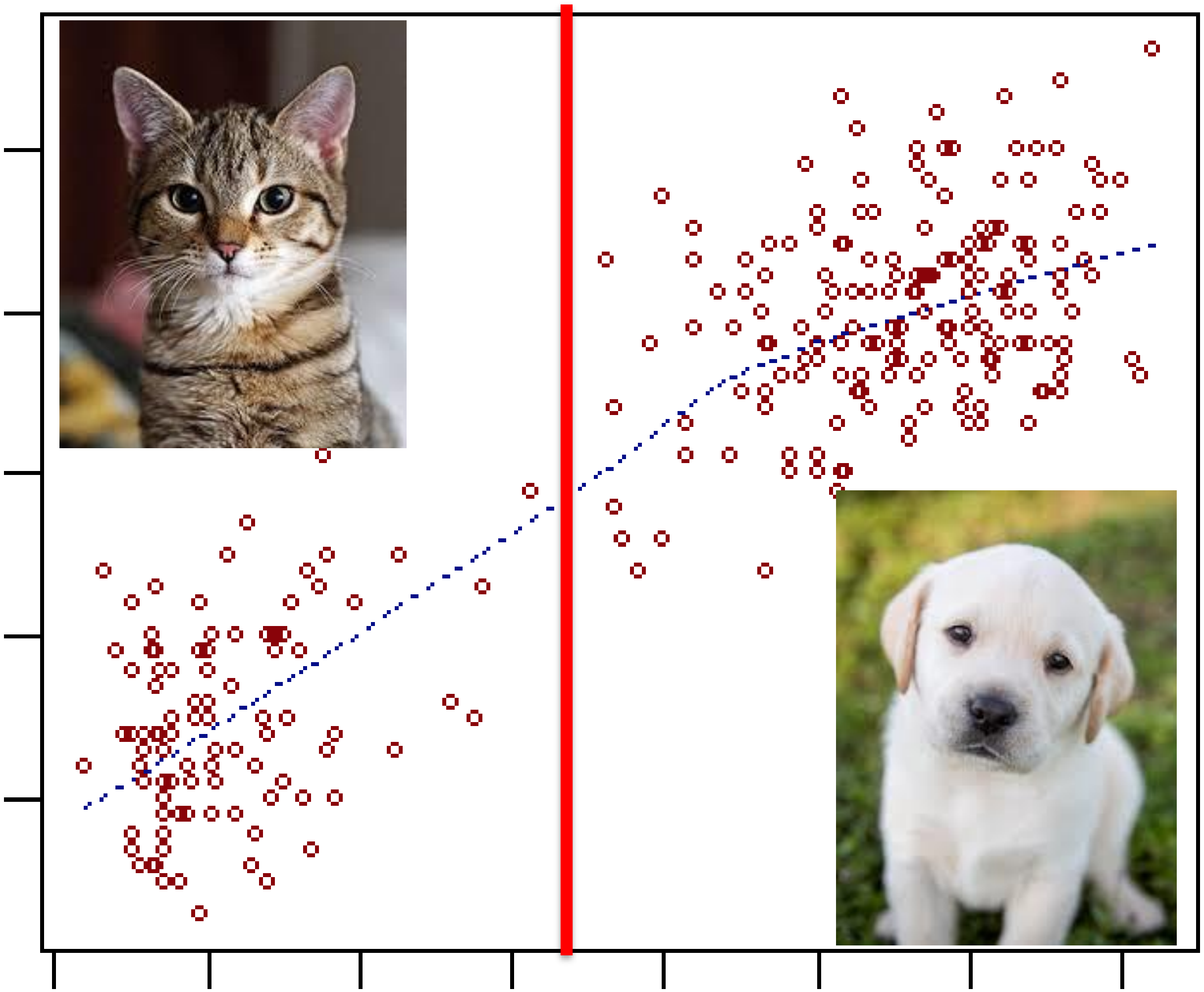






Deep Learning in Speech Recognition

Feature extraction: clustering training without teacher



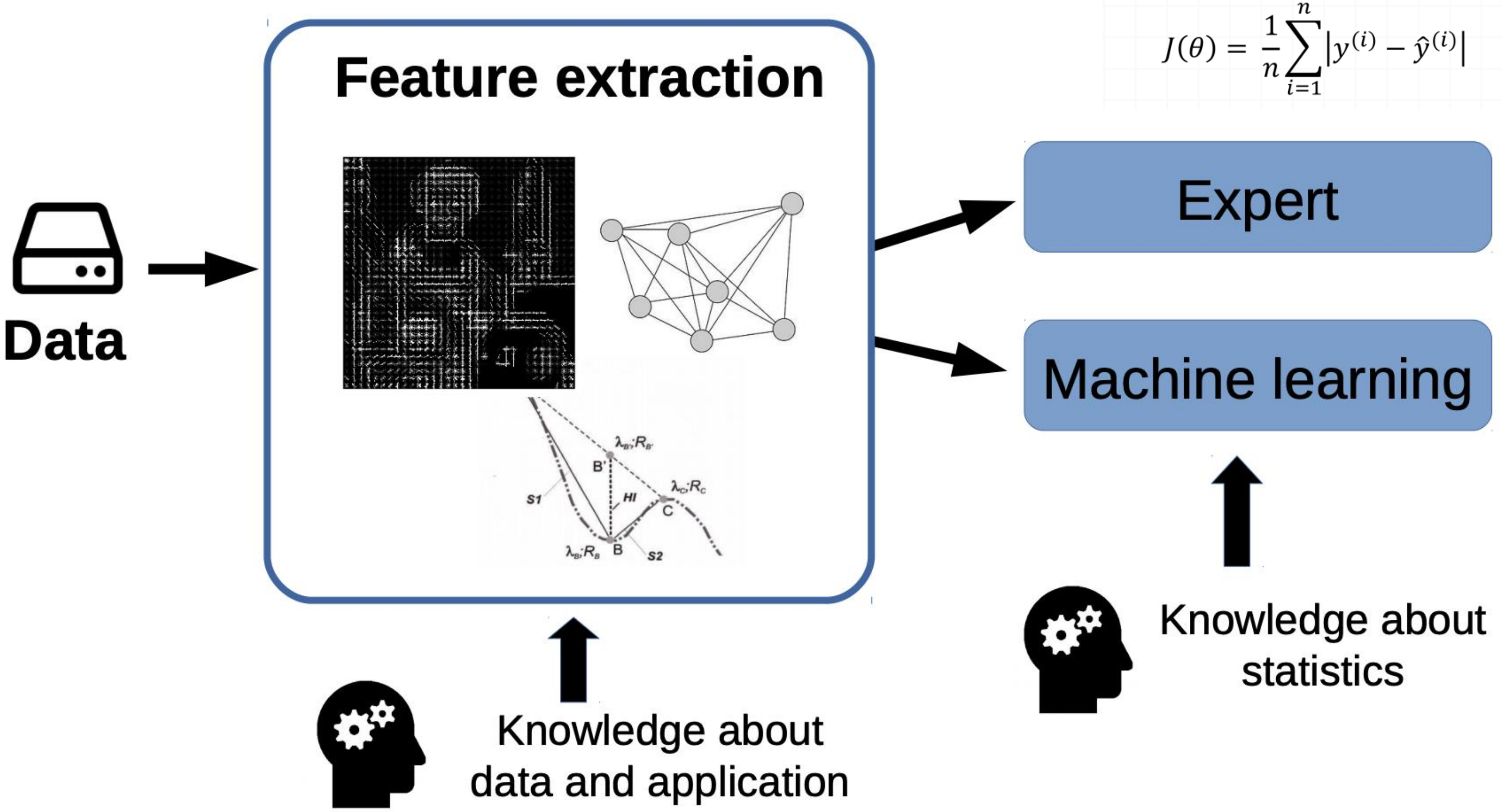
Deep learning

Mean Squared Error

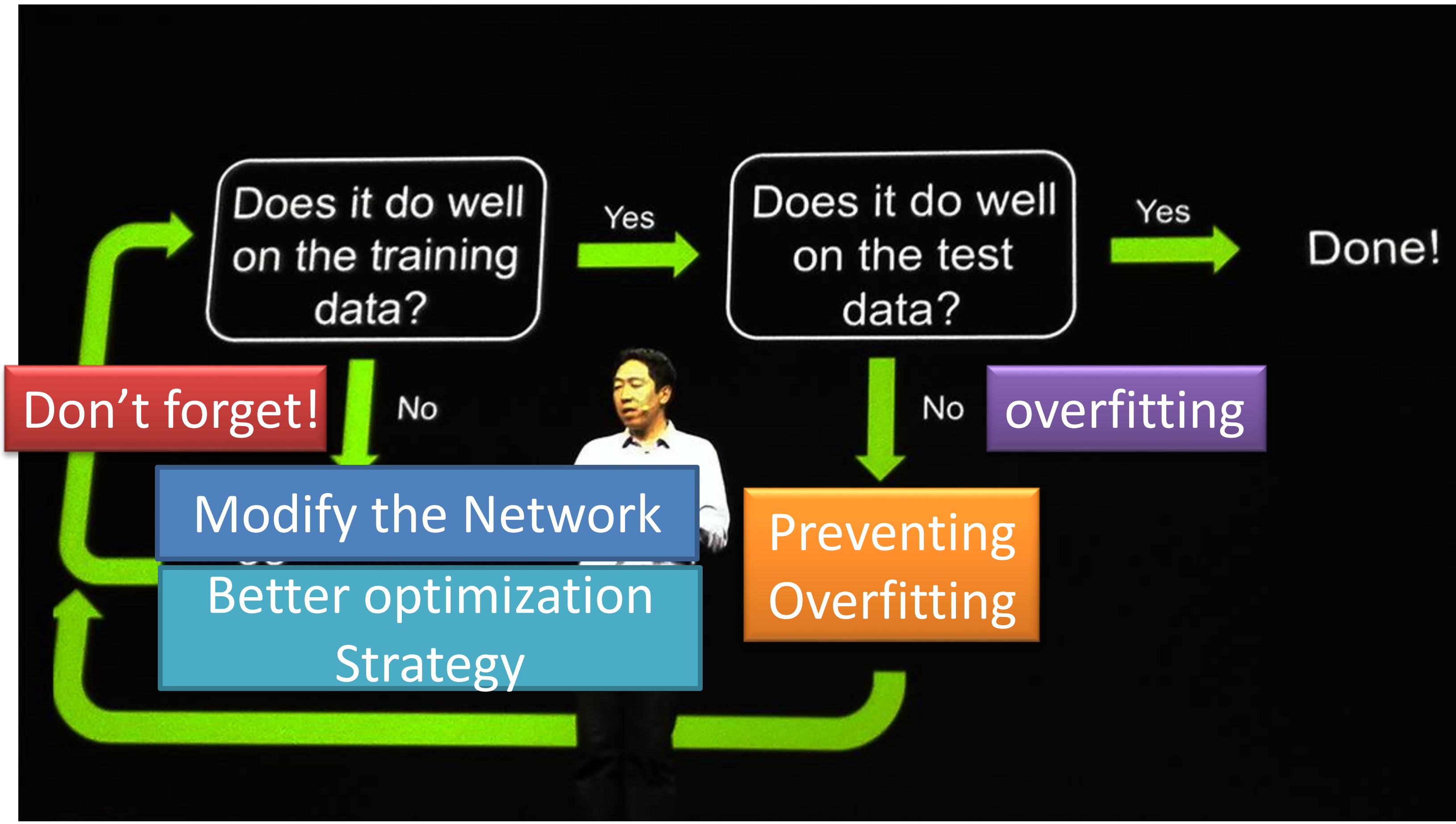
$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

Mean Absolute Error

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left|y^{(i)} - \hat{y}^{(i)}\right|$$



Recipe for Learning



http://www.gizmodo.com.au/2015/04/the-basic-recipe-for-machine-learning-explained-in-a-single-powerpoint-slide/

What tasks we want to solve in Astroparticle Physics and Genome analysis?

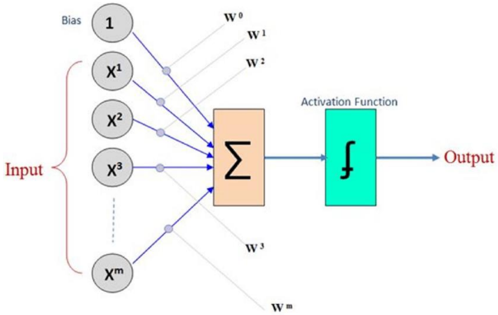
Inverse problems can be solved after solving direct problems!

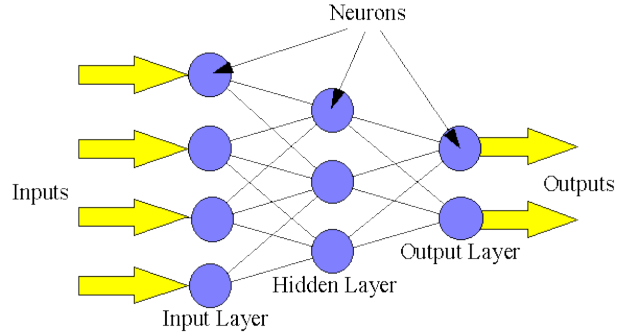
Experimental data *Simulated data*

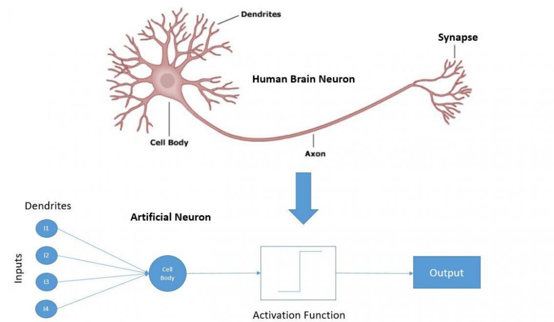
?,?(N_e,N_μ,N_h,S…) $\Leftarrow$ E,A(N_e,N_μ,N_h,S…)

1. Green and blue are not the same!
2. Identification of primary particle type
3. Estimation of primary particle energy
4. Background rejection
5. Genome analysis

Why Neural Networks?







1. Neural Networks, like a Bayesian models, belong to the general class of nonparametric methods that do not require any assumption about the parametric form of a statistical model they use;
2. Are appropriate technique for classification and estimation tasks;
3. Are able to treat multidimensional input data;
4. Neural networks can easily deal with large amounts of training data experimental (millions of events);
5. Robustness and quality of prediction : Neural network methods if properly used approach Bayesian methods!
6. After troublesome training the implementation phase is very fast and can be made with hardware accelerators!
7. Retraining with new acquired knowledge is easy and contain in the new coefficient of NN, can be easily and save distributed among world (Armenian network project!)

Training and Implementation phases

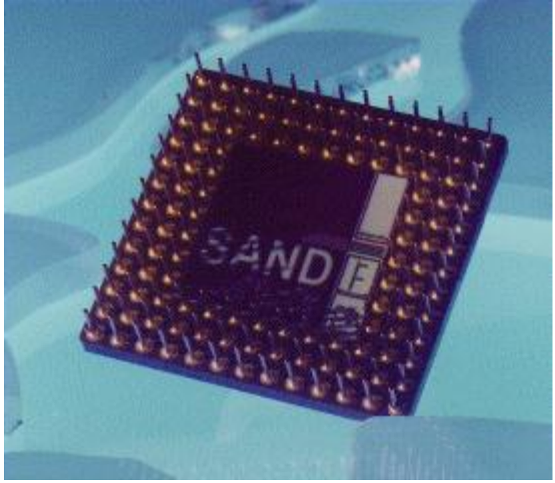
- Step 1: Train some model on classified data- MC images

Data analysis technique
+
Classified data- γ; p
→
Classification model- trained NN- slow

- Step 2: Apply trained model to unclassified data to obtain signal

Classification model
+
Unclassified – exp. data
→
Classification iBgr. Rejection – very fast

Neural Chip SAND in online data processing of extensive air showers



- The neural chip SAND (Simple Applicable Neural Device) was designed to accelerate computations of neural networks at a very low cost basis, due to the fact that only few peripheral chips are necessary to use the neural network chip in applications. Four SAND-chips were implemented on one PCI-board. The board is highly usable for hardware triggers in particle physics. The performance of a SAND-PCI-board is 800 Mega Connections per Second due to four neuro-chips, each with four parallel 16 bit multipliers and 40 bit adders. SAND is able to implement feedforward neural networks with a maximum of 512 input neurons and three hidden layers. Kohonen feature maps and radial basis function networks may be also calculated. The application of the SAND-PCI-board is proposed for cosmic ray physics to allow online analysis of extensive air showers.
- Computer Physics Communications 126 (2000) 63–66, 2000 Elsevier Science

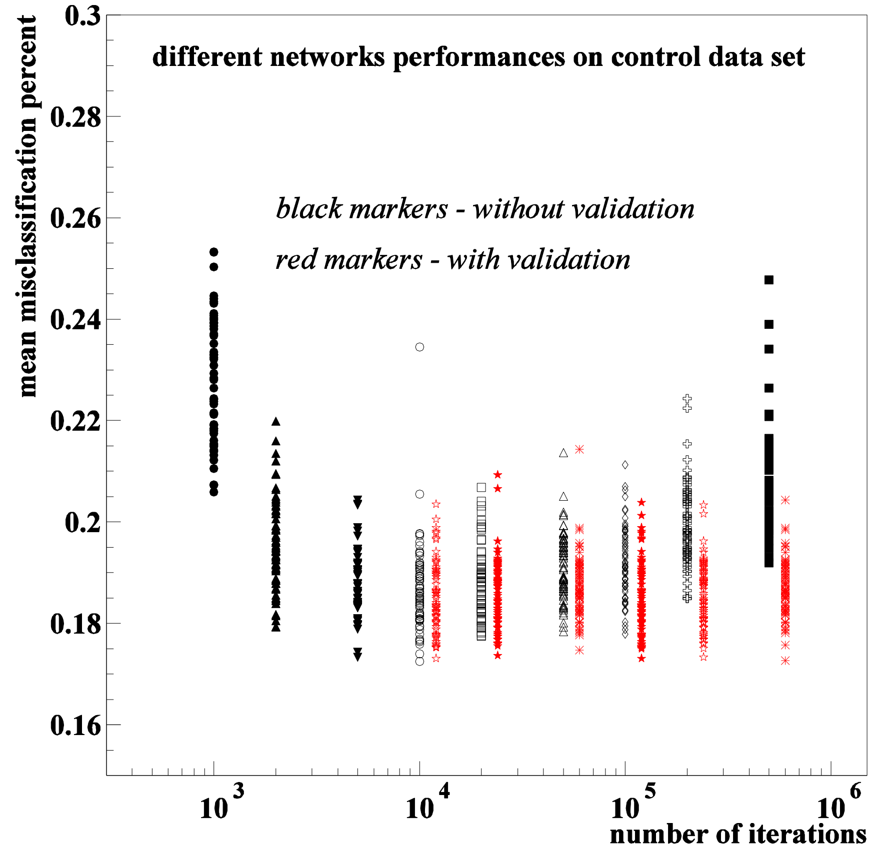
Common Drawbacks in NN Training Process

1. The central issues in an ANN implementation remain the net training algorithms and strategies, which, in general, should not depend on the particular problem specifics to be solved and on data samples generated for the training purpose.
2. The main aim of the NN training is to learn the general rule for the problem solution by processing finite samples available for training and to achieve acceptable (reasonable) performance (generalization) when applied to the control (independent) events not used for the NN training.
3. In practice, the network which performs best on training sample, is not obligatorily the best on control sample, even more, if the overfitting occurs, such a network will fail processing control events. These are the questions about robustness and reliability of NN training which have to be addressed by techniques developed to overcome such kind drawbacks as overtraining control and ”best” NN selection.

ANI package strategies of NN training http://www.crd.yerphi.am/Download_ANI#Download (cmz and exe files, manual)

1. Learn a general rule and not particular training sample: continuous overtraining control;
2. MRSES(Multistart Random Search with Early Stop) and RSV (Random search with validation – cure for overtraning). A committee of several networks is used to generate a voting procedure for choosing best for each event, training started from the random points in the space of net parameters;
3. RSV strategy: train single network with runtime validation for controlling overtraining, use not training sample error but error your network made on independent control sample!
4. Special type of “mapping networks” for selected best signal domain; using signal-to-noise ratio in NN training as quality function;
5. US Patent application 38509/0003: MULTIVARIATE RANDOM SEARCH METHOD WITH MULTIPLE STARTS AND EARLY STOP FOR IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES BASED ON MICROARRAY DATA

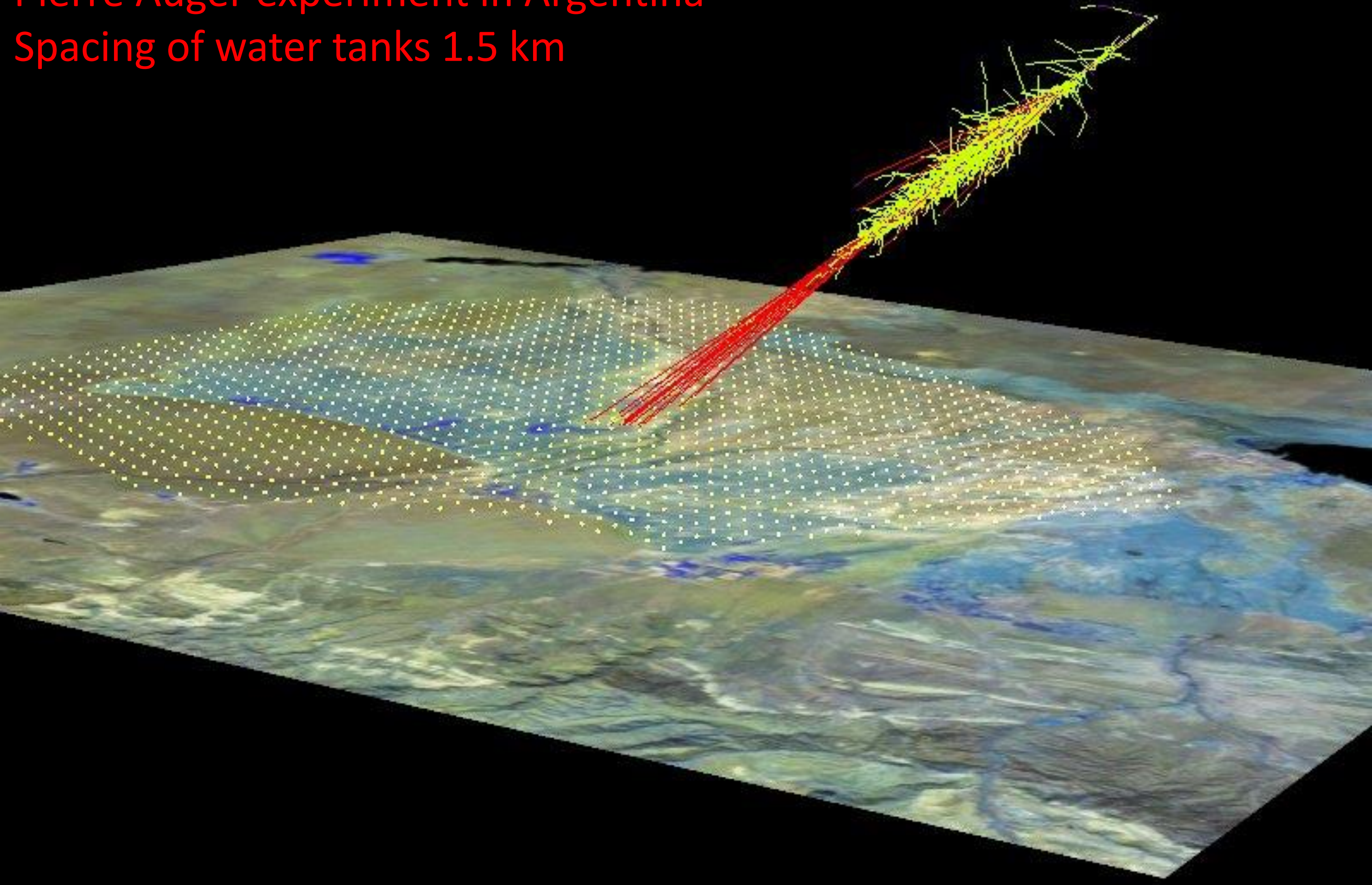
The Efficiency of Run-Time Validation



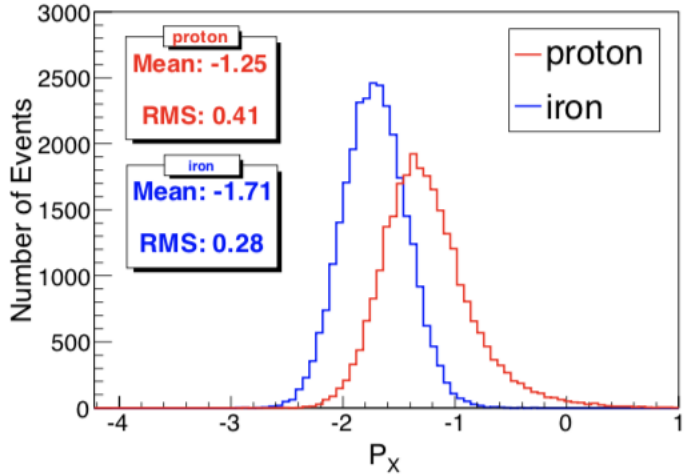
To understand what NN is doing physicist have start with Feature Selection

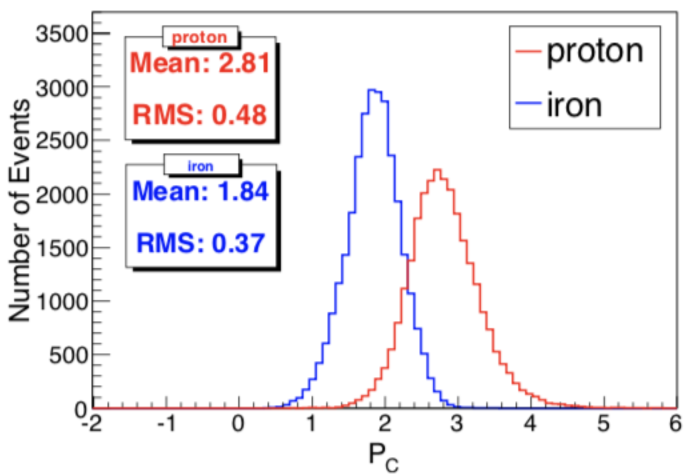
1. Binominal NP non tractable problem;
2. Selecting best single features;
3. Correlation analysis – selecting best pairs;
4. Test best pair + best singles;
5. Random search in multivariate feature space with return at “bad” step and tuning of step size;
6. Multiple start and early stop of random search;
7. Bayesian risk estimate on each step: Empirical error calculation with independent control sample;
8. Comparisons of Bayesian error and mpirical error of Nntraining.

Pierre Auger experiment in Argentina
Spacing of water tanks 1.5 km



Proton and Helium Event Selection by Large High Altitude Air Shower Observatory(LHAASO) - consisting of three detector arrays: kilometer square array (KM2A) which includes the electromagnetic detector array and muon detector array, water Cherenkov detector array (WCDA) and wide field of view Cherenkov telescope array (WFCTA).

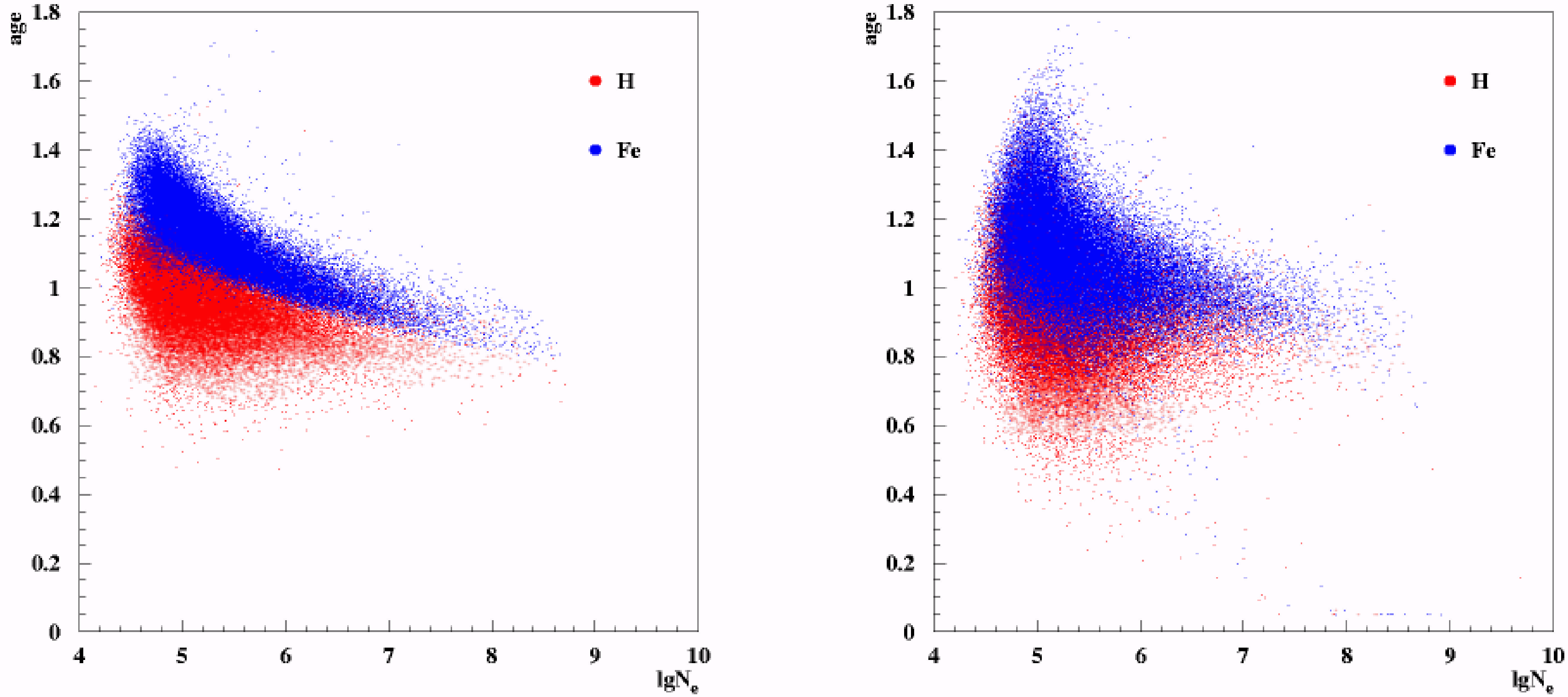




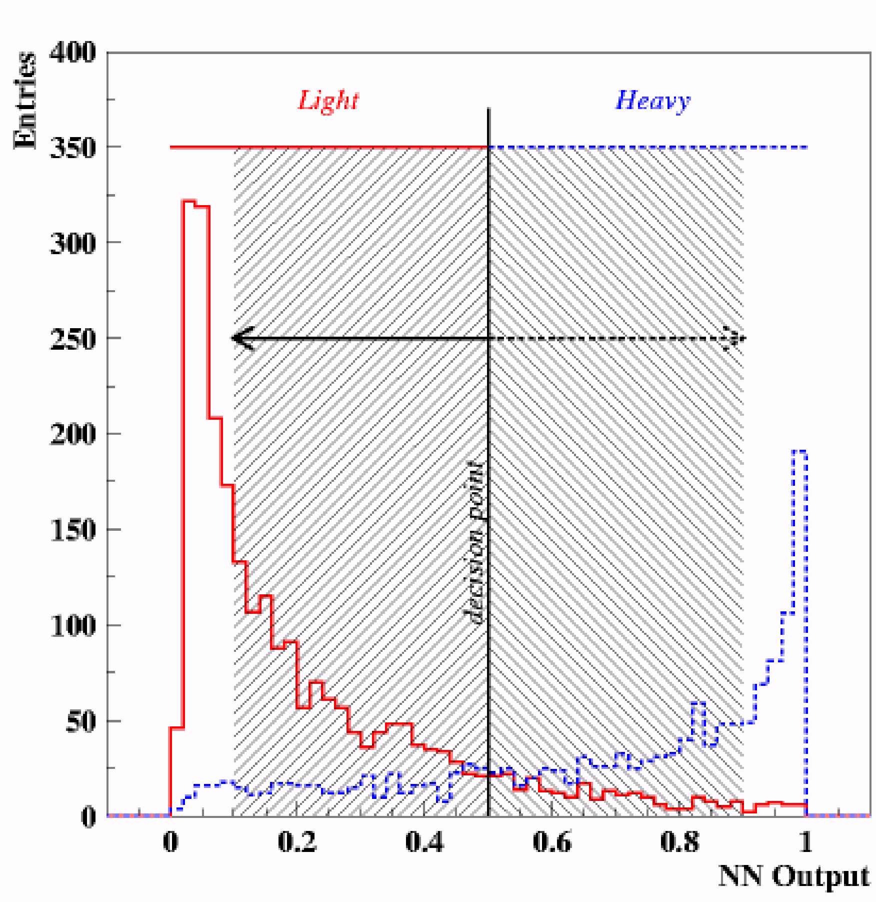
The distributions of mass sensitive parameters P_X (left) and P_C (right) for proton (red line) and iron (blue) initiated showers. Δθ, the exactly parameter used to reconstruct the Xmax, is applied instead of the reconstructed Xmax. Δθ is the angular distance between the shower arriving direction and the gravity center of the Cherenkov image. After normalization, the structure of the mass sensitive parameter P_X is as follows: $P_x = \Delta\theta - 0.0103 \times R_p - 0.404 \times N_0^{pe}$

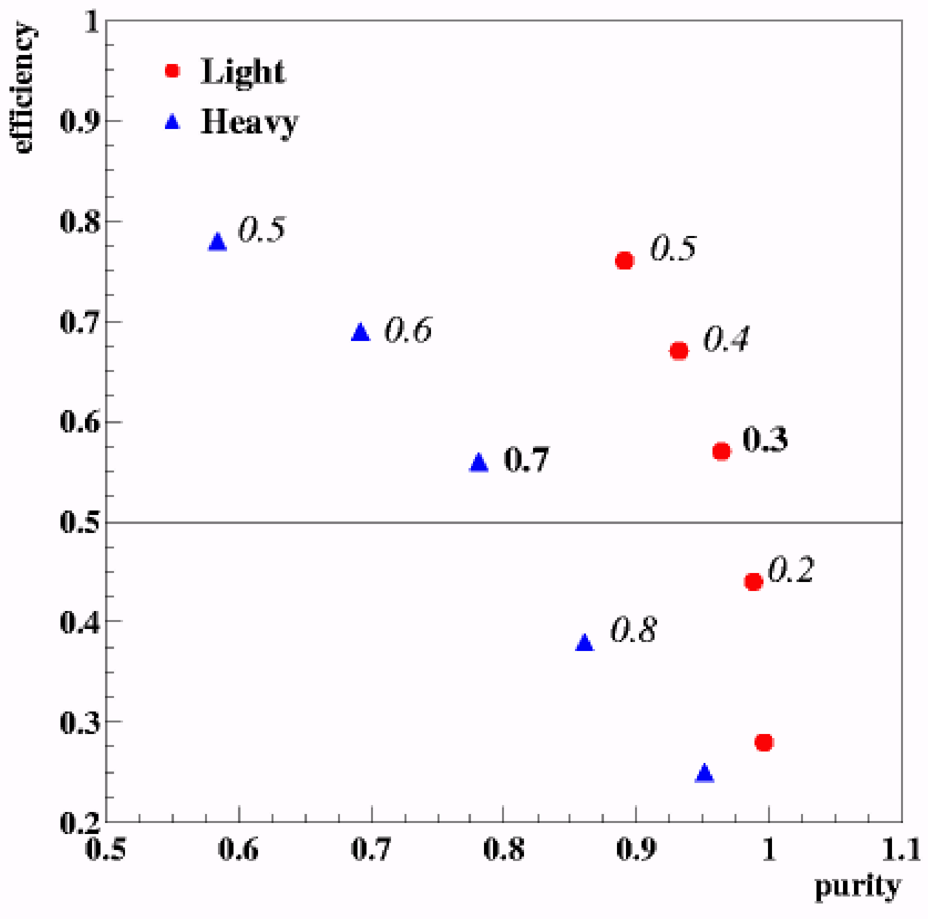
R_p is the perpendicular distance between the telescope and the shower axis; and N_0^{pe} is shower energy. Moreover, the ratio of length and width of the Cherenkov image is also a traditional and effective parameter: $P_c = L/W - 0.0137 \times R_p + 0.239 \times N_0^{pe}$

Scatter plot of Shower Size and Shape Parameters for light and heavy nuclei

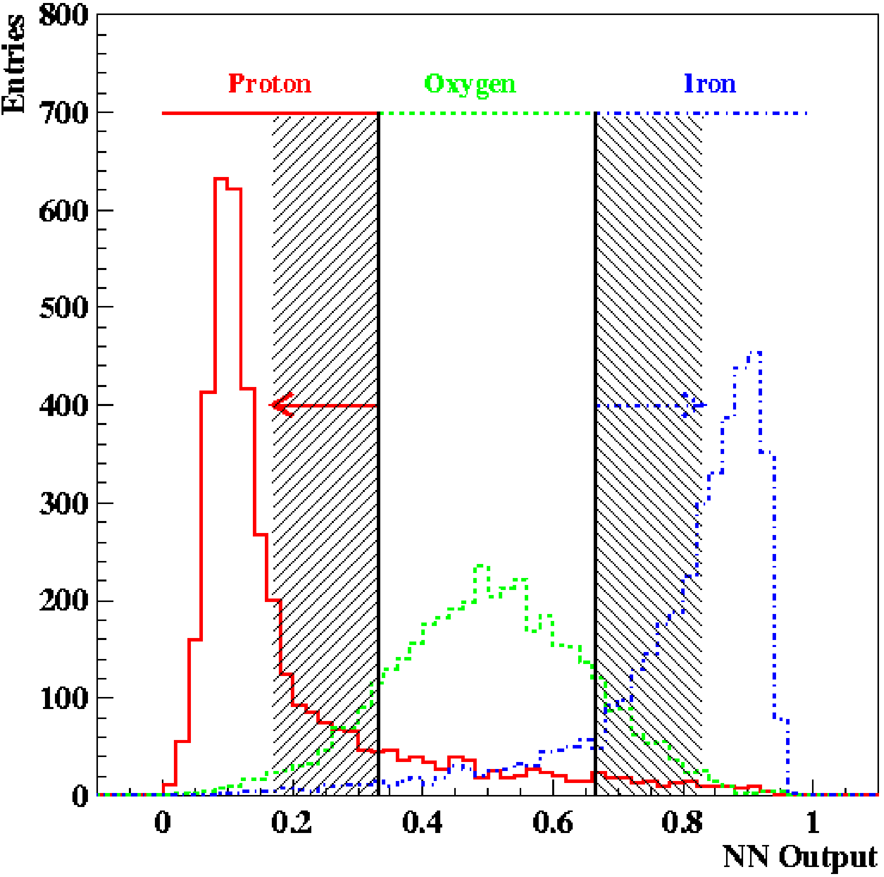


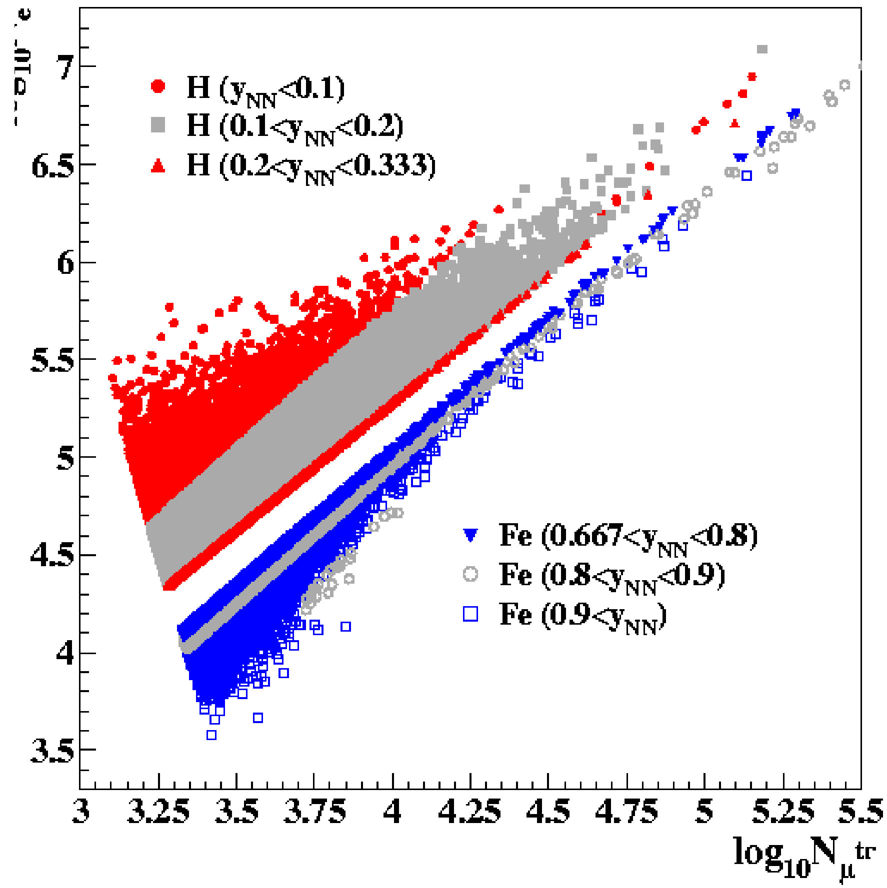
Partial spectra: light and heavy nuclei:2-way Classification of MAKET-ANI data



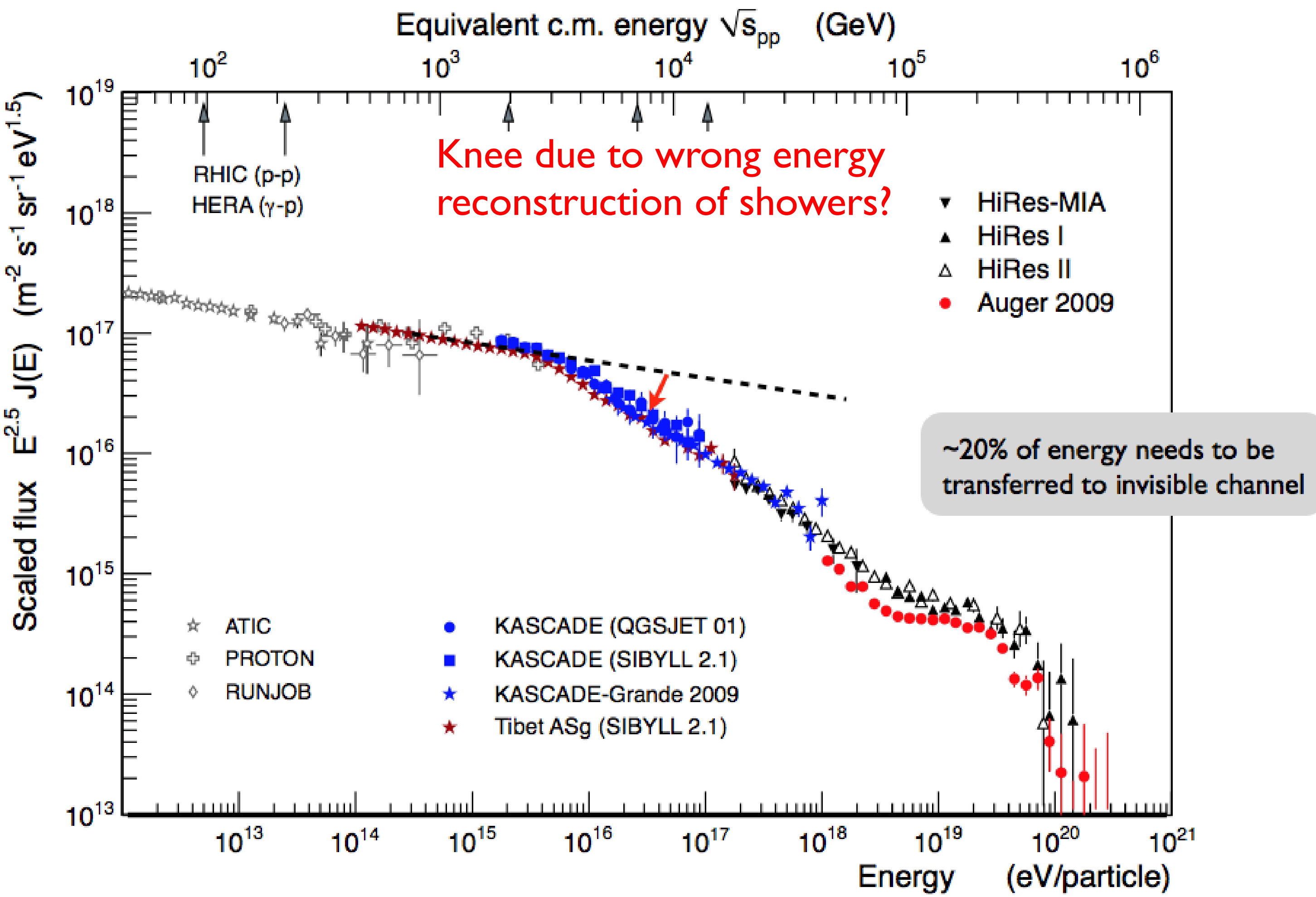


3-way KASCADE Data Classification

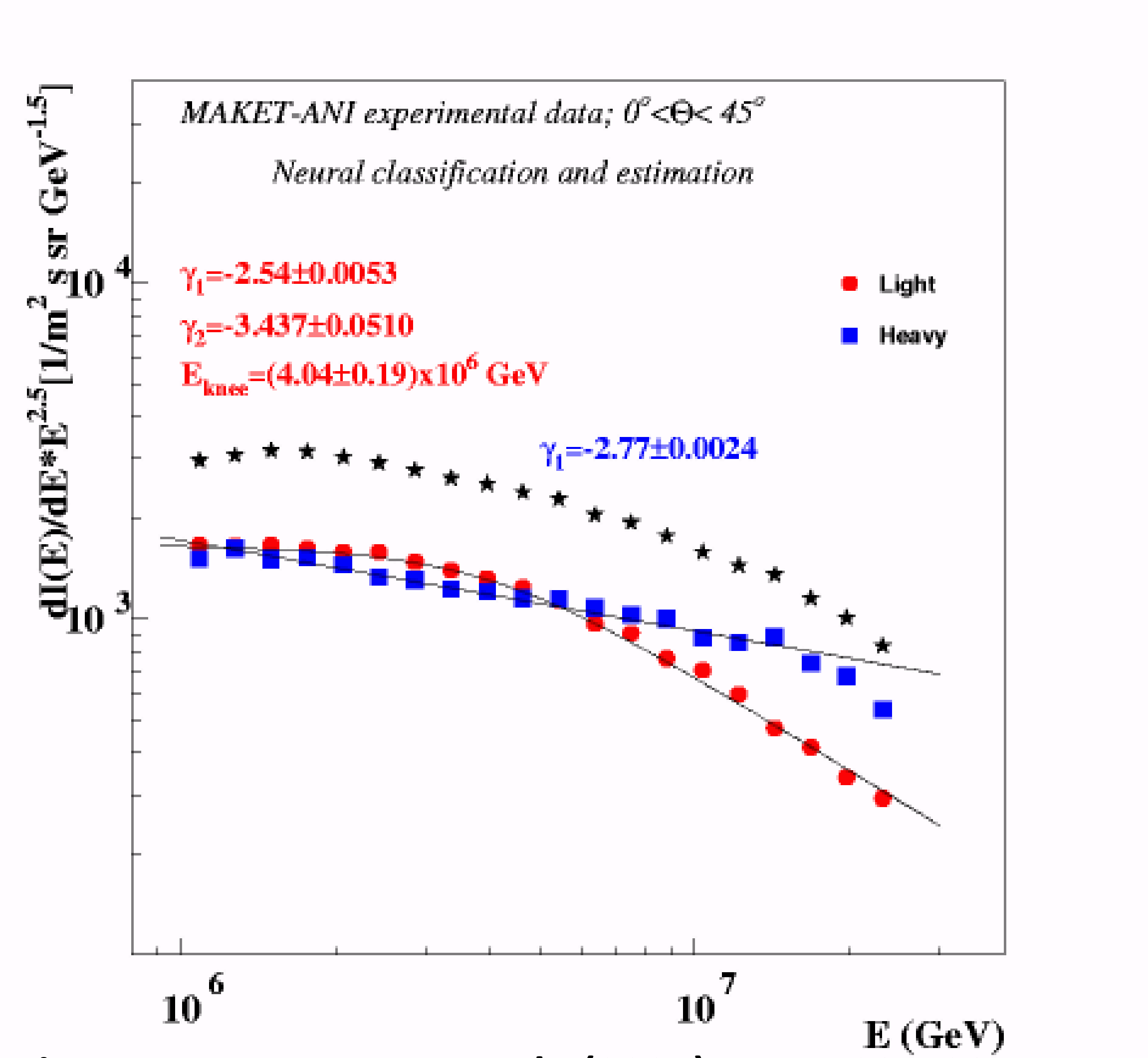




“Knee” limit of space accelerators or escaping energy?



Neural classification of the all particle energy spectra to Light and Heavy Nuclei



Credit, Chilingarian A., Hovsepyan G. et.al., (2004), ApJ, 603, L29

On 22 September 2017, IceCube detected a neutrino that was special: Its very high energy (roughly 290 teraelectronvolt) indicated that the particle might have originated from a distant celestial object. Scientists were also able to identify its incoming direction with high precision.

In fact, Fermi-LAT, a space observatory that conducts all-sky surveys, reported that the direction of the neutrino was in line with a known gamma-ray source in an active state: the blazar TXS 0506+056. What is more, MAGIC, a 17-meter twin telescope that probes high energy

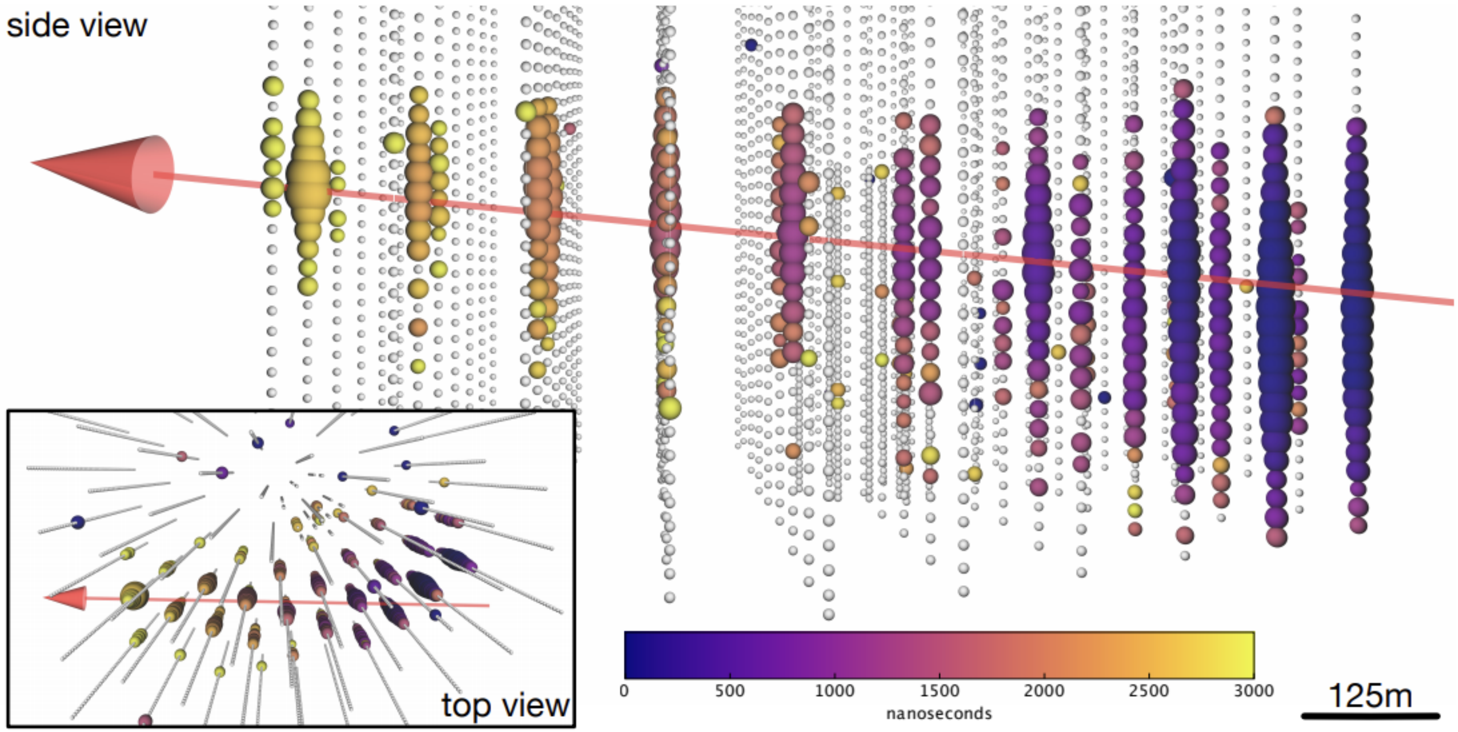
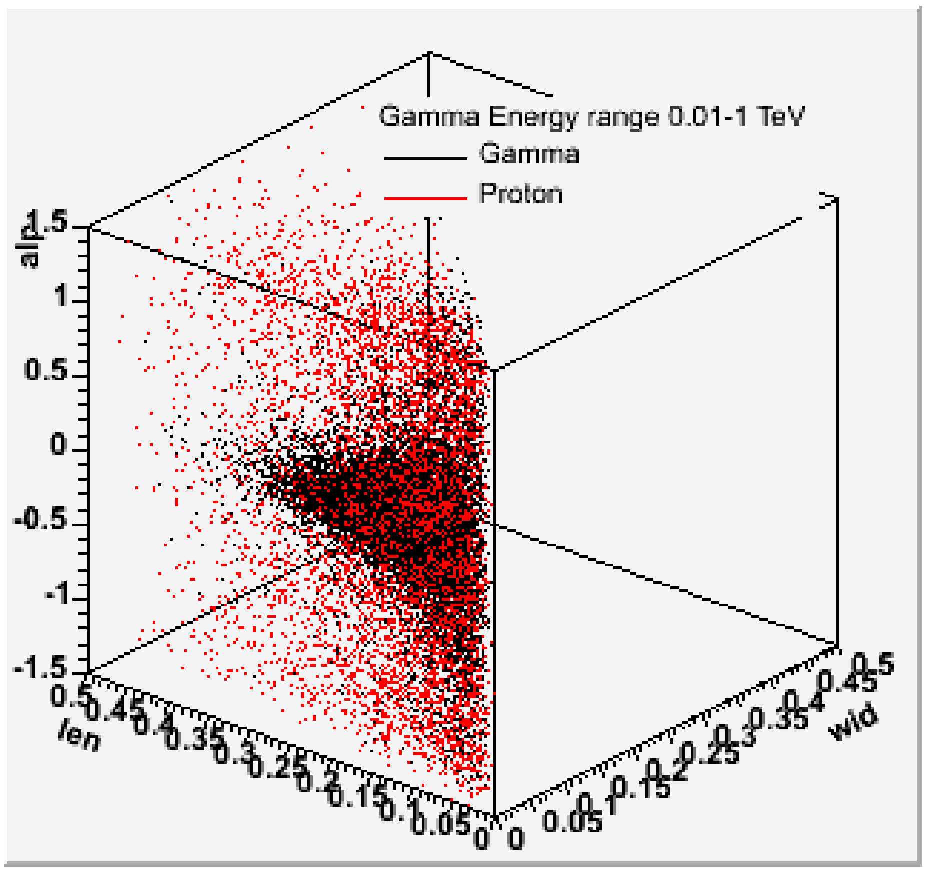
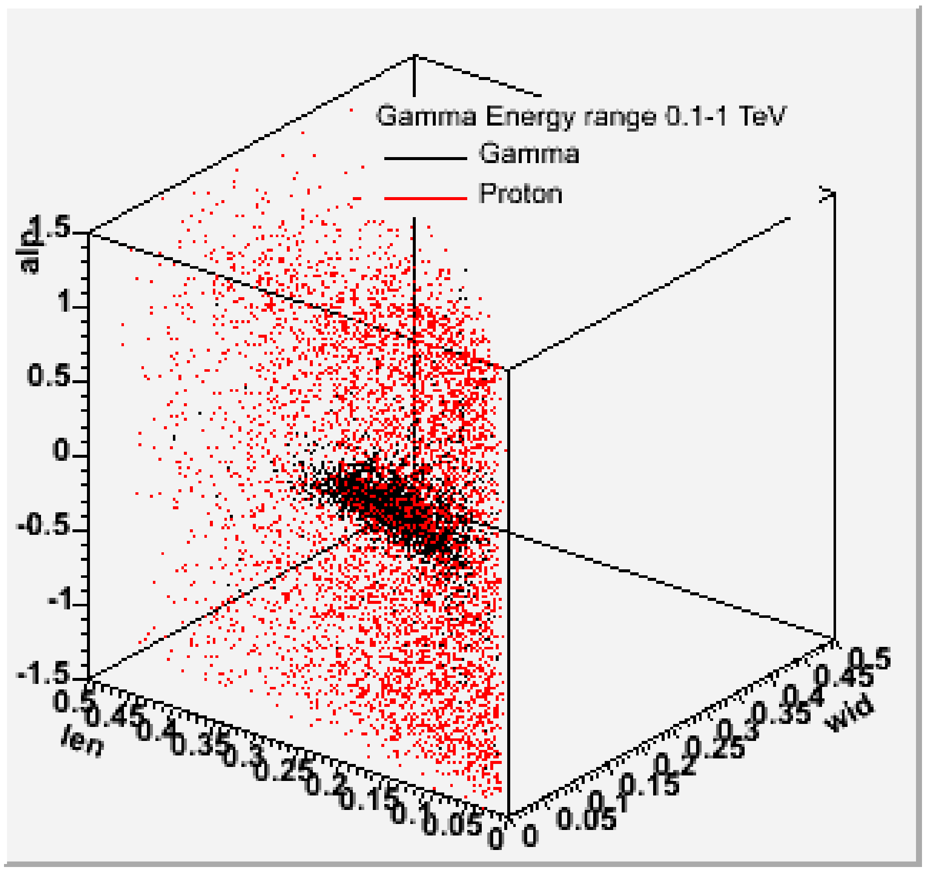


Figure 1: **Event display for neutrino event IceCube-170922A.** The time at which a DOM observed a signal is reflected in the color of the hit, with dark blues for earliest hits and yellow for latest. Time shown are relative to the first DOM hit according to the track reconstruction, and earlier and later times are shown with the same colors as the first and last times, respectively. The total time the event took to cross the detector is $\sim$3000 ns. The size of a colored sphere is proportional to the logarithm of the amount of light observed at the DOM, with larger spheres corresponding to larger signals. The total charge recorded is $\sim$5800 photoelectrons. Inset is an overhead perspective view of the event. The best-fitting track direction is shown as an arrow, consistent with a zenith angle $5.7^{+0.50}_{-0.30}$ degrees below the horizon.

Combining Shape and Orientation we'll achieve large Q-factors (σ)





The Wedge cut (supercut0
Chilingarian, A.A. and M.F. Cawley , Multivariate analysis of Crab Nebula data, Note to Whipple collaboration from July 5, 1990.

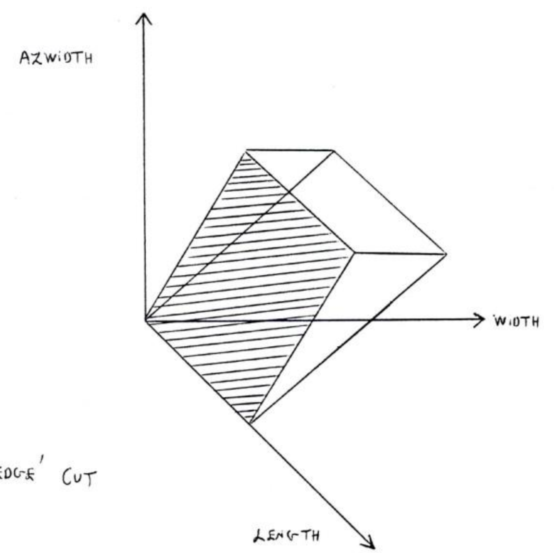


FIG. 5
THE 'WEDGE' CUT

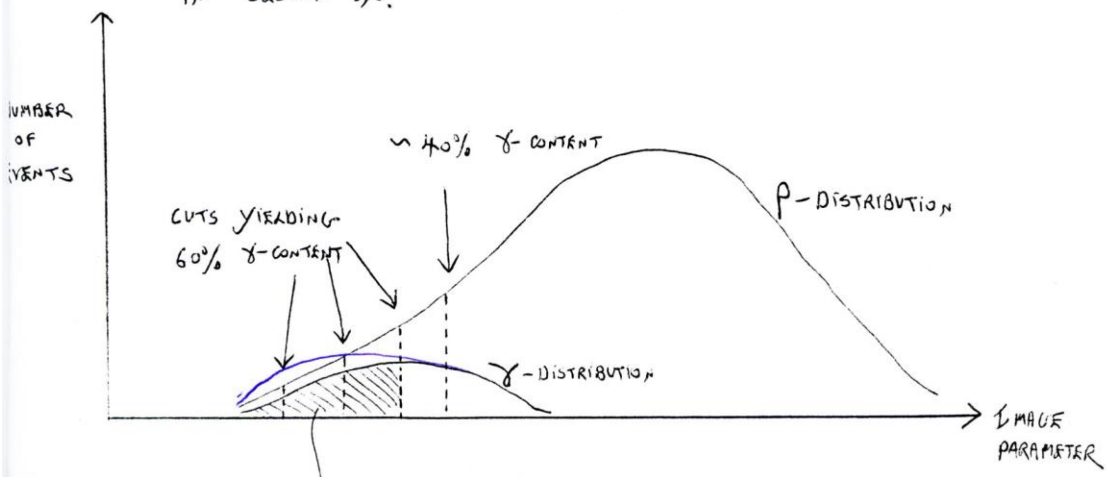
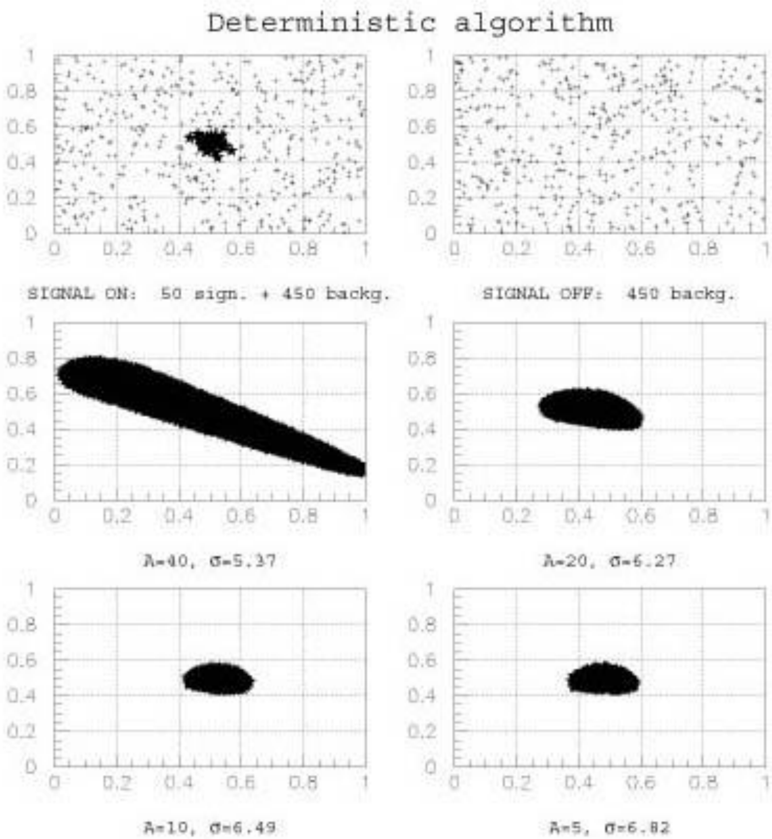
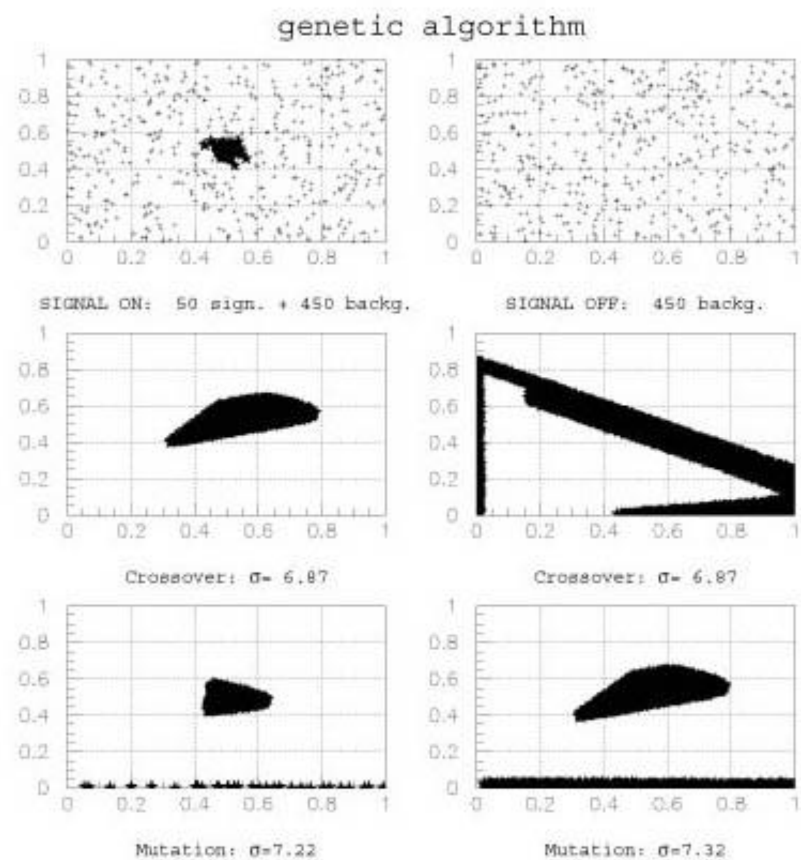


FIG. 6 1-D ILLUSTRATION OF IDEAS DISCUSSED IN SECTION 4.

First detection of Crab nebula by the Whipple collaboration, contained in the famous 1988-1989 data files, consisting of 65 ON-OFF pairs, more than 1 million images, σ = 27

Artificial intelligence(machine learning) approach: New Type of NN – Mapping Networks -Maximizing Signal Significance – by optimizing the shape of Gamma Domain





Comparison of the different background rejection methods WIPPLE detection of CRAB, 1988-1989

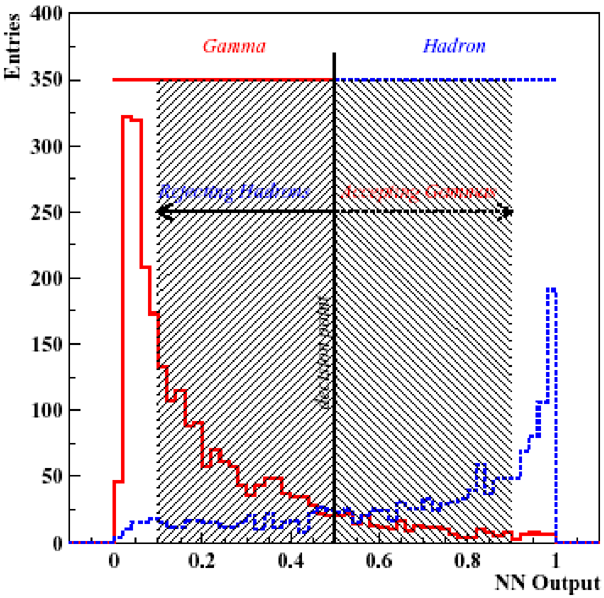
	N^*_{on}	N^*_{off}	σ	DIFF	DIFF/N^*_{off}	N^*_{off}/ N_{off}
Raw	506255	501408	4,8	4847	0.01	
Azwidth	14622	11389	20.4	3233	0.28	0.0227
Wedgecut*	6017	3381	27.2	2636	0.78	0.0067
Supercut **	4452	1766	34.3	2686	1.52	0.0035
Neural*** 4::5::1	6278	2858	35.8	3420	1.20	0.0057

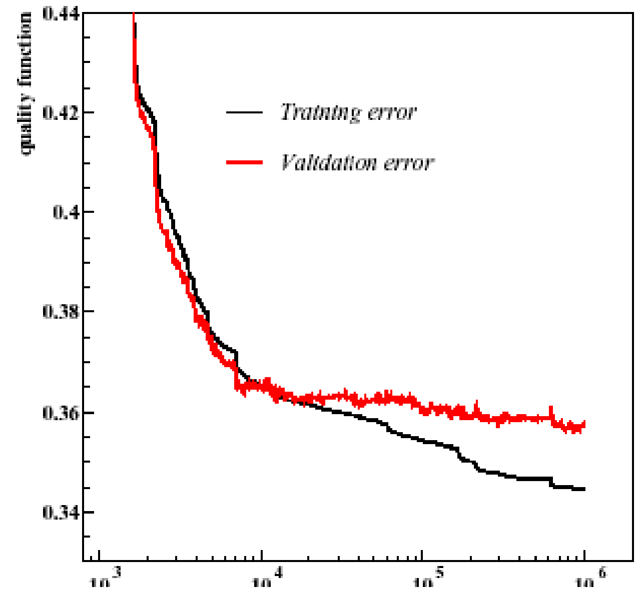
*Chilingarian, A.A. and M.F. Cawley. Application of multivariate analysis to atmospheric Cherenkov imaging data from the Crab nebula. Proc. 22 ICRC, 1, 460-463, Dublin, 1991.

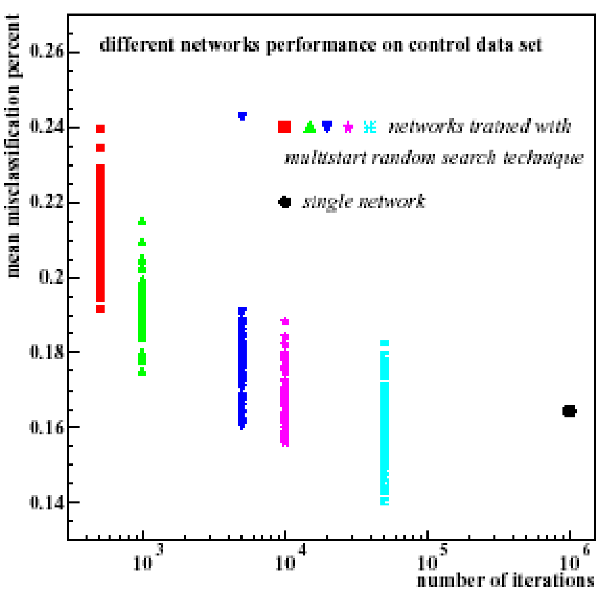
**Punch, M., C.W. Akerlof, M.F. Cawley, et. al.. Supercuts: an improved method of selecting gamma-rays. Proc. 22nd Internal. Cosmic Ray Conf., Dublin, 1, 464-467, 1991

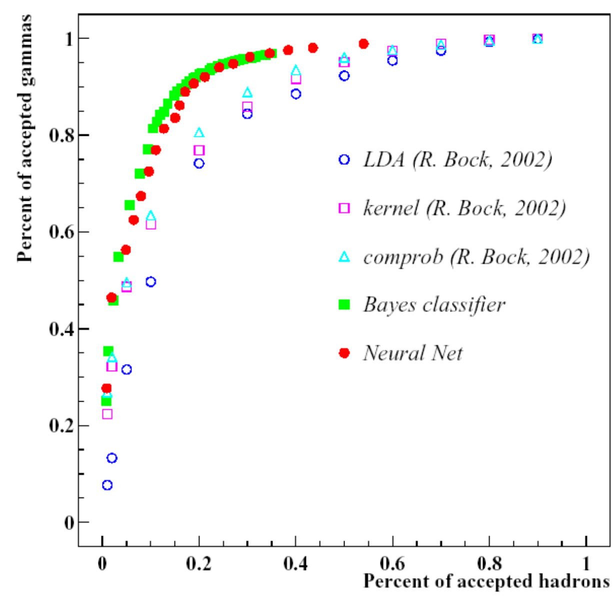
***Chilingaryan A. A., Neural classification technique for background rejection in high energy physics experiments, Neorocomputing, 6, 497, 1994.

MRSES Training of NN to Reject Hadrom Background in Detecting AGN and SNR





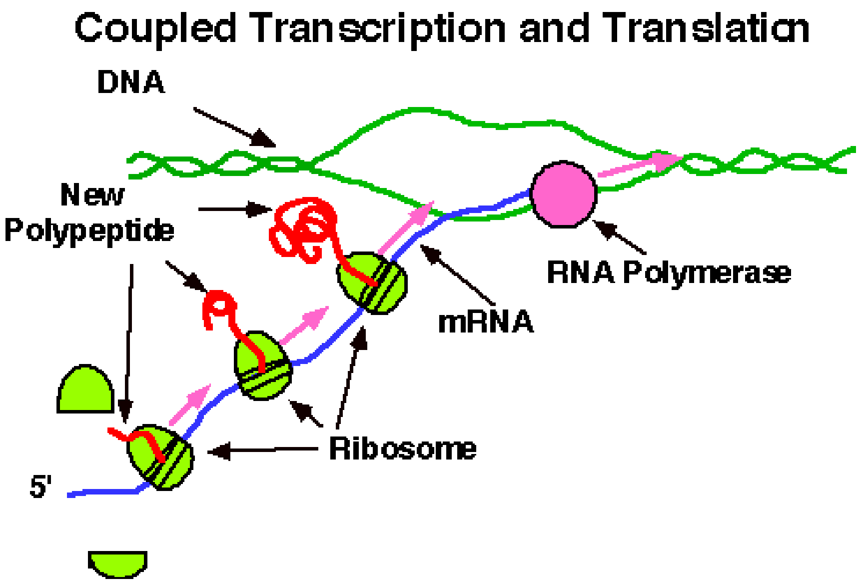


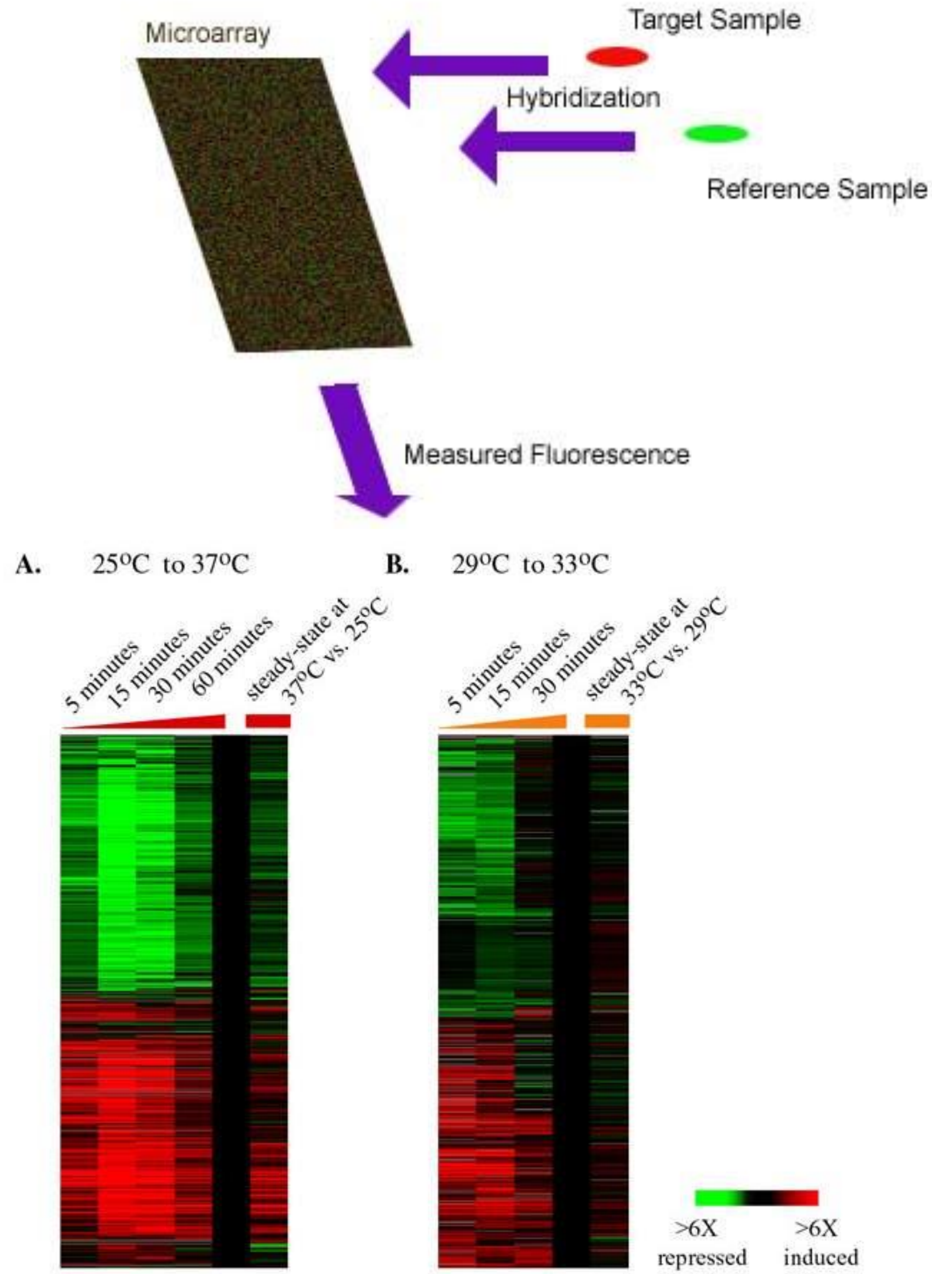


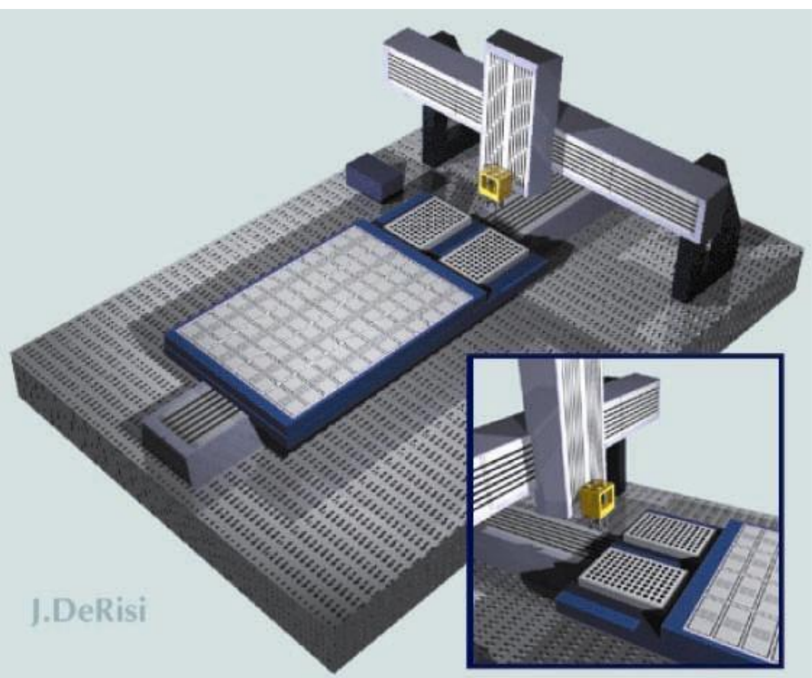
Why is so important to reject background

- "ON" sample- obtained with telescope axes oriented in direction of the putative gamma-ray source and
- "OFF" – sample, obtained by pointed telescope axes in direction of the same celestial co-ordinates, but after the source already leave the destination;
- If the field of view of the telescope is enough large it is possible to take "ON" and "OFF" scans simultaneously, selecting within field of view samples pointed to source and to "empty" space;
- Best estimate of SIGNAL = ON – OFF = 10100-10000=100;
- SIGNIFICANCE OF SIGNAL DETECTION
- SIGNAL-TO-NOISE RATIO = SIGNAL/OFF = 100/10000=0.01;
- Let's assume that we succeed to reject 99% background keeping 50% signal!
- ON=150; OFF = 100; SIGNAL = 150 – 100 = 50
- SIGNAL-TO-NOISE RATIO = SIGNAL/OFF = 50/100=0.5;

Microarray Technique







Selection of the best subset of genes (20 best from 1000) NP-complete problem

Class 1

	S1	S2	S3						S*n*
gene1									
gene2									
. . .									
gene *k*									
. . .									
. . .									
gene *p*									

Class 2

	S1	S2	S3				S*m*
gene1							
gene2							
. . .							
gene *k*							
. . .							
. . .							
gene *p*							

f(20 genes) ~ $f(R_m)* C_{20}^{1000} = 1000!/20!/980! \sim 10^{60}/10^{18} = 10^{42}$ operations; One element of covariance matrix calculate in 1 mcsec, then time required on 10 Pfl CPU ~ 10^{19} sec ~ 10^{12} years;

Age of Universe is $1.37 * 10^{10}$ years

MRSES Heuristic: f(MRSES – 5genes) ~ f(Rm)* $N_{iterat} * N_{cycles}$ ~$25*500*10{,}000 = 1.25*10^8$ operations: 12,5 seconds;

MRSES NET Training 1

1. split available sample in to training and validation samples by the ratio of 2:3 and 1:3 respectively.

2. fix the number of networks to be trained N_{cycle}

3. fix the number of iterations N_{iter}.

4. select the initial weights of the network randomly from gaussian distribution with prechosen mean μ and σ variance (σ is a small number).

5. set the initial quality function to a very large value.

6. select the iteration step size randomly from the prechosen $\Delta v \in R$ interval.

7. perform a random iteration step in the multidimensional space of NN weights from the initial point to modify all weights of randomly chosen neuron, the alternation of weights is performed according to the following:

MRSES NET Training 2

$$\vec{V_i} = \vec{V_{i-1}} + Q^p_{i-1} \cdot (RNDM - 0.5) \cdot \Delta v; \qquad i = 1, N_{iter} \tag{1}$$

were $\vec{V_i}$ is the vector of NN weights obtained at i-th iteration step, Δv is step size and RNDM is a random number from [0-1] interval. The term Q^p_{i-1} introduces dependence of value of random step on the already achieved quality function and controls the strength of this dependence, $p = 0, 1, 2$.

8. calculate the objective (quality) function at each step by presenting all the training events to Neural Network:

$$Q_i = \frac{1}{M_{events}} \sum_k (OUT_k - TRUE_k)^2 * W_k, \qquad k = 1, M_{events}, \tag{2}$$

were OUT_k is the actual output of NN, $TRUE_k$ is the corresponding goal value of k-th input vector from the training set, W_k is event weight, M_{events} is the number of training events.

Mrses training 3

9. If $Q_i \leq Q_{i-1}$, then the vector $\vec{V}_i$ is kept as new weights of NN and the next step is initializing from point $\vec{V}_i$ in multidimensional space of NN weights, otherwise - return to the previous point is implemented and a new random step is performed.

10. if $i = N_{iter}$, stop random iterations, save the obtained NN weights, otherwise go to 7,

11. repeat steps 4-10 N_{cycle} times (train N_{cycle} networks),

12. apply each trained network for the validation sample classification,

13. select the network best performed on the validation set.

YerPhI publications on ML (1)

- A.Chilingarian**, Analysis of Data Interpretation Methods for ANI Experiment**, Problems of Nuclear Science and Technology, techniques of physics experiment series, **3**/12/1982
- A. Chilingarian, **Identification of High Energy Particles as Pattern Recognition Problem,** Problems of Nuclear Science and Technology, techniques of physics experiment series, **4**/12/1982
- A. Chilingarian, **On the Effectiveness of Statistical Methods of High Energy Particle Identification,** Preprint YerPhi-582(69), 1982
- A. Chilingarian**, Development of Statistical Methods in Cosmic Ray Physics** Proceedings of the 18th ICRC, Vol. **5**, Bangalor, India, 1983
- N. Akopov, S. Harutunyan, A. Chilingarian, Sv. Harutunyan, et al., **The Applied Program Package for Multivariate Statistical Analysis**, Proceedings of the 6th School of Physics Calculation Methods, (52-55), 1985
- A. Chilingarian, S. Galfayan**, Calculation of Bayesian Risk by Means of KNN Probability Density Estimation ,** " Problems of Control", issue. 66, Vilnius, Litvania, (64-78), 1985.
- V. Avakyan, V. Matevosyan, A. Chilingarian, N. Chilingarian, et al., **The Bayesian Identification of Cosmic Ray Hadrons,** Preprint YerPhi -933(84), 1986
- A. Chilingarian, N. Khudonyan, D. Sahakyan, G. Zazyan, **Recognition of Correlated Patterns with Spin Glass Like Models,** Preprint YerPhi-992(42), 1987

YerPhI publications on ML (2)

- A. Chilingarian, G. Zazyan, **A Classification Method of Determination of Mass Composition of Primary Cosmic Rays,** Preprint YerPhi -1210(87), 1989
- A. Chilingarian**, Statistical Conclusions under Nonparametric A Priori Information,** Computer Physics Communication, Vol.**54**, (381-390), 1989.
- A. Chilingarian, G. Zazyan, The Bootstrap Method of Determining the Coefficients of Distribution Mixture, Pattern Recognition letters II, (781), 1990
- F. Aharonyan, A. Konopelko, A. Chilingarian, A. Plyasheshnikov, **The Multivariate Analysis of Cherenkov Images of EAS from γ-Quanta and Protons**, Nuclear Instruments and Methods, A302, (522), 1991.
- A. Chilingarian, G. Zazyan, **The Analysis of Multiparticle Production at High Energies with Renie Dimensions,** Computer Physics Communication 56, 1991.
- A. Chilingarian, **Investigation of Mass Composition and Energetic Spectrum of Cosmic Ray Protons in the Energy Ranges 1015-1017 eV according to the Data of Cosmic Ray Complex Facilities,** IL Nuovo Cimento, 14C(6), (555), 1991.
- A. Chilingarian, **Development of the Data Processing Methods in High Energy Physics. From a Data Base to a Knowledge Base**, Preprint YerPhi-1327(22), 1991.
- A. Chilingarian, H. Zazyan, **Particle Beam Experiments in Cosmic Ray Physics. Strong Interaction Parameters Determination by Statistical Pattern Recognition,** Preprint YerPhi-1345(40), 1991.

YerPhI publications on ML (3)

- A. Chilingarian, **The Mathematical Models of Neural Networks As Pattern Recognition Systems,** Preprint YerPhi-1350(45), 1991.
- F. Aharonyan, A. Chilingarian, A. Plyasheshnikov, et al., **Multivariate Correlation Analysis for the System of γ-Telescopes,** Proceedings of USSR Academy of Science, physics series, Vol. **55**, 1991.
- A. Chilingarian, G. Zazyan, **Experiments with Particle Beams in Cosmic Rays. Determining Strong Interaction Parameters by means of Pattern Recognition methods**, Nuclear Physics, 54, (128-136), 1991.
- A. Chilingarian, **Neural Net Classification of the γ and p Images Registered with Atmospheric Cherenkov Technique,** Proceedings of the 22th ICRC, OG-4.7.23 , vol.1, Dublin, Ireland, 1991.
- S. Galfayan, A. Chilingarian, **Median Estimators of Probability Density Multivariate Function**, Proceedings of the 4th Seminar on Multivariate Statistical Analysis, Tsakhkadzor, Armenia, 1991.
- A. Chilingarian, R. Shaghoyan, **Mathematical Models of Neural Networks, part 2**, Preprint YerPhi-13590(1), 1993.
- F. Aharonyan, R. Mirzoyan, A. Chilingarian et.al, **The System of Optical Cherenkov Telescopes, A New Approach to High Energy Gamma Astronomy,** Experimental Astr., **2**, (331-344), 1993.

YerPhI publications on ML (4)

- A. Chilingarian, **Neural Classification Technique for Background Rejection in High Energy Physics Experiments,** Neurocomputing, Vol. l6, (497), 1994.
- M. Cawley, A. Chilingarian, **Optimizing the Non-Linear Gamma-Ray Domain in {VHE} Gamma-Ray Astronomy using Neural-Network Classifier,** Proceedings of the 24th International Cosmic Ray Conference (ICRC),Vol.**3,**(742), Rome, Italy, 1994.
- A. Chilingarian, **Detection of Weak Signals Against Background using Neural Network Classifiers**, Pattern Recognition Letters, vol. 16, (333), 1995.
- A. Chilingarian, **The Non-linear Signal Domain Selection using a New Quality Function in Neural Net Training, Nuclear Instruments and Methods (NIM),** 389A, (242), 1997.
- A. Chilingarian et al., **On the Nonparametric Classification and Regression Methods for the Multivariate {EAS} Data Analysis**, Nuclear Physics B, 52B, (237), 1997.
- **KASCADE** collaboration, **Estimate of the Cosmic Ray Composition by a Pattern Analysis of the Core of PeV EAS,** Proceedings of the 26th ICRC, Vol. **1,** (329), Salt- Lake City, Utah, USA, 1999.
- A. Vardanyan, H. Gemmeke, W. Eppler, T. Fischer, A. Chilingarian, **The SAND/1 Neurochip as Fast "Intelligent" Trigger for the MAGIC Experiment,** Proceedings of the 2nd Irano-Armenian Workshop on Neural Networks, (107), Tehran, Iran, 1999.
- M. Roth, A. Chilingarian, A. Vardanyan , **A Nonparametric Approach for Determination of Energy Spectrum and Mass Composition of Cosmic Rays from EAS Observables**, Nuclear Phys. B, 75A, (302), 1999.
- A. Chilingaryan, N. Gevorgyan, A. Vardanyan, D. Jones and A. Szabo, **Multivariate Approach for Selecting Sets of Differentially Expressed Genes,** Math. Biosciences, Vol.**176**(1), (59-69), Elsevier Science Inc. PII: S0025-5564(01)00105-5, 2002.

YerPhI publications on ML (5)

- T.Antoni, A. Chilingarian, et.al., for the **KASCADE** collaboration, **A Non- Parametric Approach to Infer the Energy Spectrum and the Mass Composition of Cosmic Rays**, Astroparticle Physics, Vol.16, (245-263), 2002.
- T.Antoni, A. Chilingarian, et al., for the **KASCADE** collaboration, **Preparation of Enriched Cosmic Ray Mass Groups with KASCADE,** Astroparticle Physics Vol.**19**, (715-728), 2003.
- A. Chilingarian, G. Gharagyozyan, G. Hovsepyan, S. Ghazaryan, L. Melkumyan, A. Vardanyan, **Light and Heavy Cosmic-Ray Mass Group Energy Spectra as Measured by the MAKET-ANI Detector,** The Astrophysical Journal, **603**, (L29-L32), 2004.
- A. Chilingarian, G. Hovsepyan, G. Gharagyozyan, G. Karapetyan, **On the Statistical Methods of the Signal Significance Estimation in the Detection of the Signal from the Point Sources of Cosmic Rays**, International Journal of Modern Physics A, Vol. 20, No.29, (6753-6765), 2005.
- A. Chilingarian, V.Babayan, N.Bostanjyan, G.Karapetyan, H.Martirosyan, M.Zazyan,
- **Multivariate Correlation Analysis of Transient Solar Events by the Facilities of Aragats Space Environmental Center (ASEC),** Proceedings of 29th International Cosmic Ray Conference, Vol. **2**, S.H. 3.5, (281-284), Pune, India, 2005.
- A. Konopelko, A Chilingarian and A Reimers, **Study on cosmic ray background rejection with a 30 m stand-alone IACT using non-parametric multivariate methods in a sub-100 GeV energy range,** J. Phys. G: Nucl. Part. Phys. 32, (2279 -2291), 2006.
- A. Chilingarian, G. Gharagyozyan, G. Hovsepyan, G. Karapetyan, **Statistical Methods for Signal Estimation of Point Sources of Cosmic Rays,** Astroparticle physics **25,** (269-276), 2006.