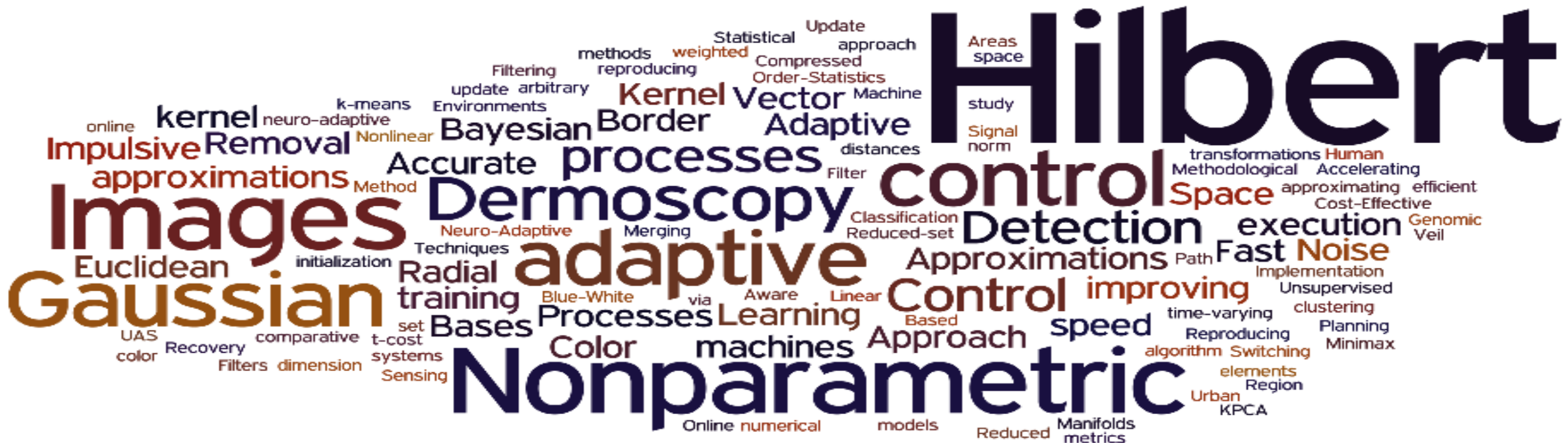


By Praseed Pai K.T.



- Co-Author of Books, “.NET Design Patterns” and “C++ Reactive Programming” (writing) for Packt Publishing
- Has Written a university level accredited paper on “Ontology”
- Has created a course on “Philosophical Tools for Software Engineering” (Presented @ Rubyconf India – Preconference)
- Presented in more than 200 events
- “Father” of SLANG Compiler Infrastructure
Slang4.net, Slang4Jvm, Slang4CPP (LLVM), SlangJs, Slang4Py, Slang4VB.net
- A budding expert on comparative philosophy (Indian/Western)
- A Critique of Digital Solutioning and Technology Fads
- Sr. Solutions Architect @ UST Global

Who is an Architect?

Architect (n) – Any person who has “fooled” around in the Software Industry for sizeable time (ever shrinking span) who is past his prime, as a Programmer Or Engineer, Systematically moved up in the hierarchy to obey “Peter Principle”.

-

Why this session?

- The ML is not a “silver bullet”
- Where it works, it works for everybody
- Not a Viable career option
- There are other worthwhile things to do in this world.

An Indian Style Negation (Neti Neti aka Via Negativa)



What is Machine Learning? (Tom Mitchell Definition)

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”

- Dissecting Analytics

The Art, Craft and Science of Analysis

Analysis/Synthesis Model of Problem Solving

A Top Down Decomposition of the Problem into Parts to a granular level , until we have reached a state where we cannot decompose parts further or it has become fine-grained to be amenable for studying it.

A Bottom up process of Synthesis

In Western Philosophy and Science, Rene Descartes is regarded as the father of modern Analysis

Reductionism vs Holism – Analytic Thinking vs System Thinking

Assumption of Independence of Variables and Interdependence of Variables

Additive factors (Linear) vs Non Linear Factors

I will try to make it Simple!

Prathikriya vaadhigalum
bhoorshvasigalum pradhama drishtya
agalchayyil aayirunenkilum avar
thammillulla andhardhara
kuravayyirunnu ennu venam karudhan.
adhannu nammal thottathu.

"MANASILLAYILLA"

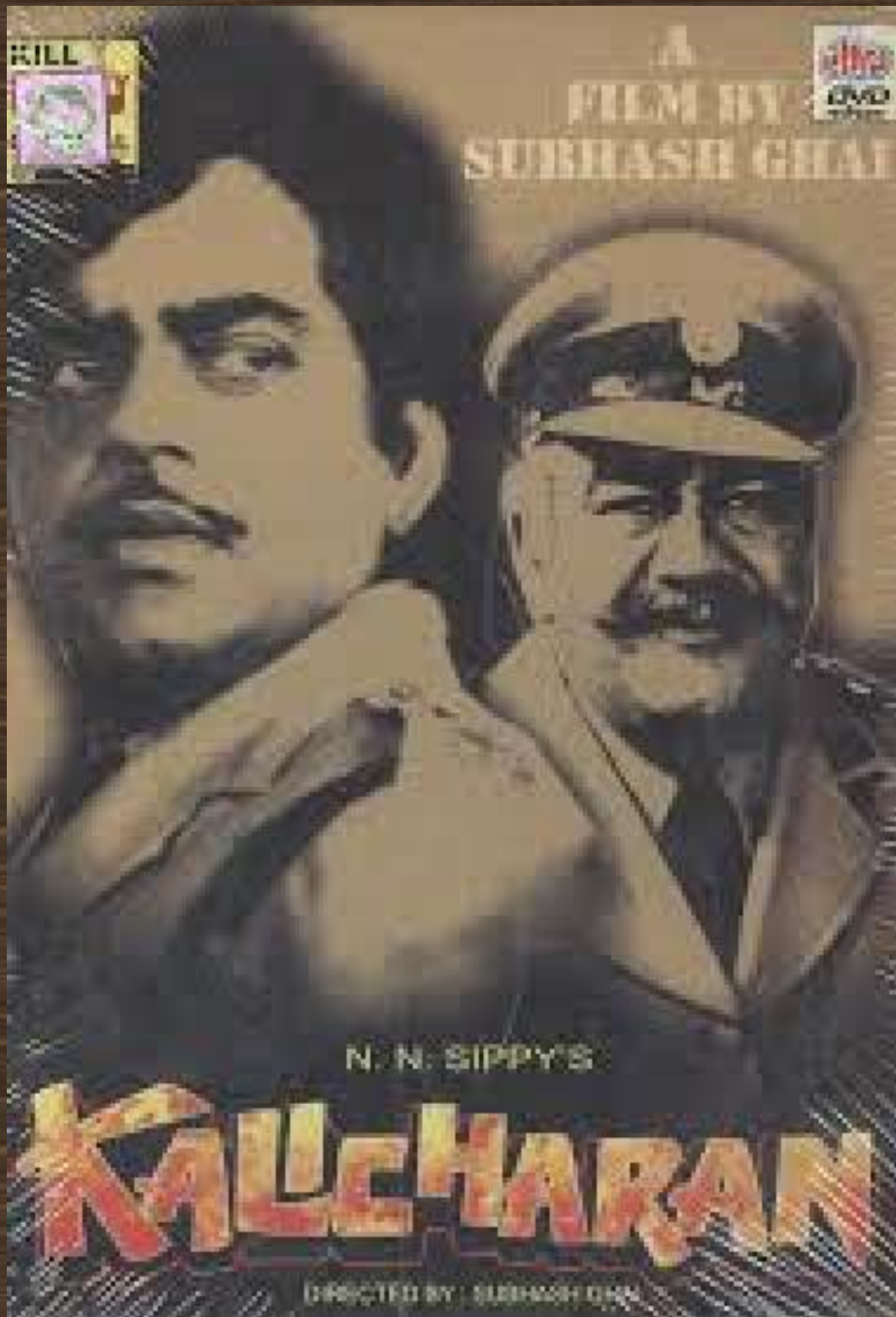
Athayyathu vargadhipathyavum
colonialist chinthasaranigalum
radicalayya matamalla, ippol
manasillayyo.



എന്ത് കൊണ്ട് നമ്മൾ തോറ്റു, എന്ത് ലളിതമായി
പറഞ്ഞാലെന്നാ? വെറുതെ confusion ഉണ്ടാക്കുന്നതെന്തിനാ ?

What is Machine Learning? (Tom Mitchell Definition)

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .”



“Akalmānd apni akal se daulatmand ban sakta ha, magar daulatmand apni daulat se akalmand nahi ban sakta hai”

“An Intelligent person can become a machine learning expert, but, by doing machine learning one is not going to be intelligent”

Machine Learning - a Deep Structure approach

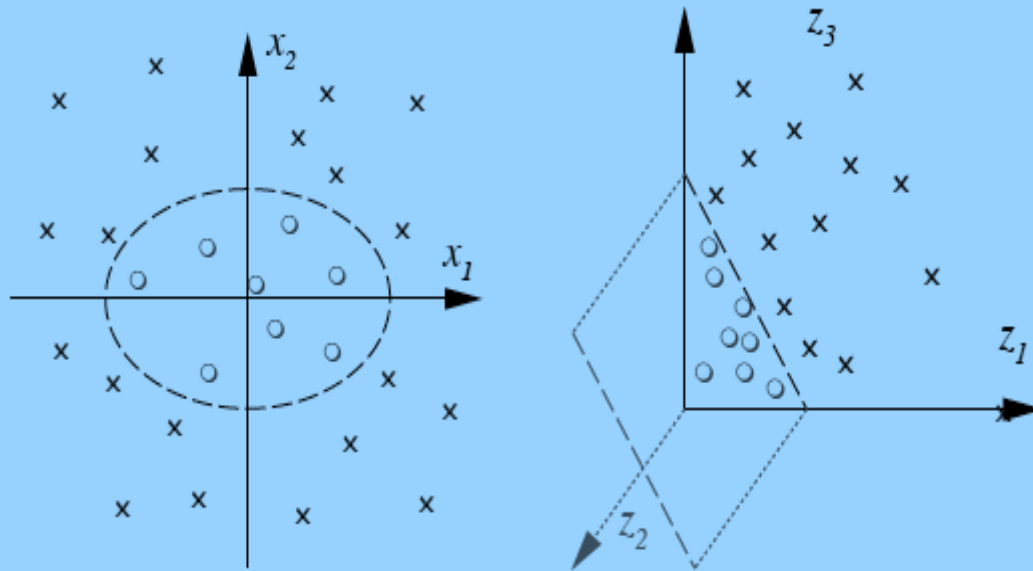
Algorithmic Techniques

- Hilbert Space Methods
- Statistical Learning
- Deep Learning (a Form of statistical learning)

Hilbert Space Methods

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$(x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$



Hilbert Spaces

A real Hilbert Space X is endowed with the following operations:

1. Vector addition: $x + y$
2. Scalar multiplication: ax , $a \in \mathbb{R}, x \in X$
3. Inner product $\langle x, y \rangle \in \mathbb{R}$, with properties:

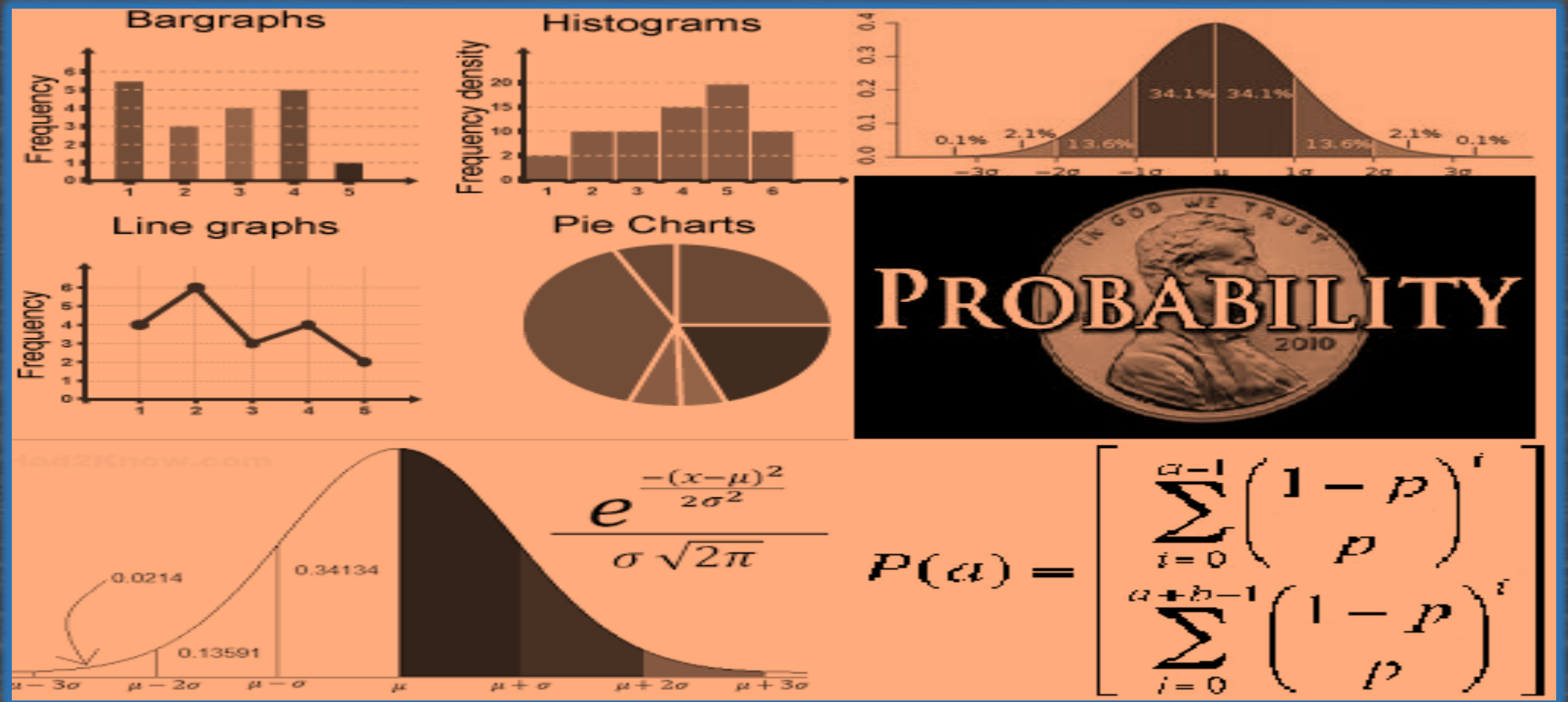
$$\langle x, y \rangle = \langle y, x \rangle \quad \langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle \quad \langle x, x \rangle \geq 0$$

4. Norm $\|x\| = \langle x, x \rangle^{1/2}$ $\|x\| = 0 \Leftrightarrow x = 0$

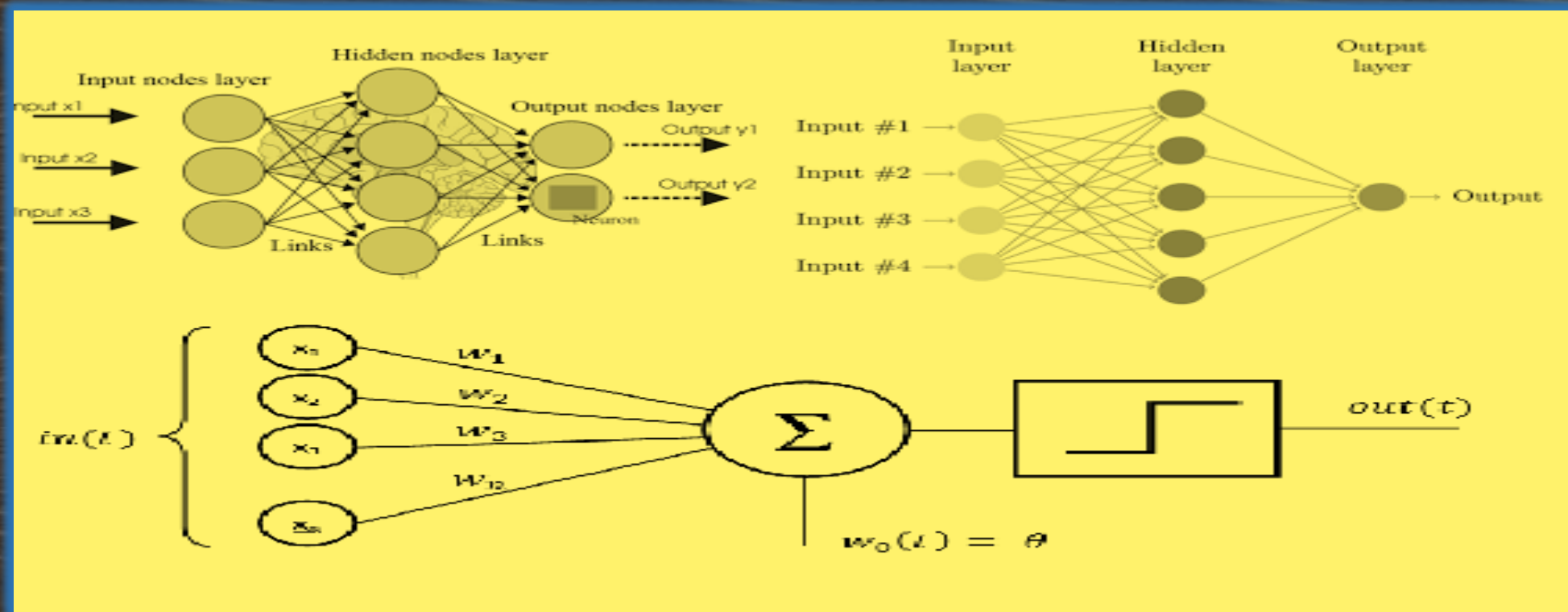
Basic facts of a Hilbert Space X

1. X is complete
2. Cauchy-Schwarz inequality $|\langle x, y \rangle| \leq \|x\| \|y\|$ where the equality holds if and only if $x = \lambda y$

Statistical Methods



Deep Learning Methods



$$X = (x_1, x_2, x_3, \dots, x_n)$$

$$W = (w_1, w_2, w_3, \dots, w_n)$$

$$\text{Sum}(w_i, x_i) = w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \dots + w_n * x_n;$$

In Statistical Methods

$(w_1 \dots w_n)$ might be the probability of each coefficient

In Hilbert Space Methods

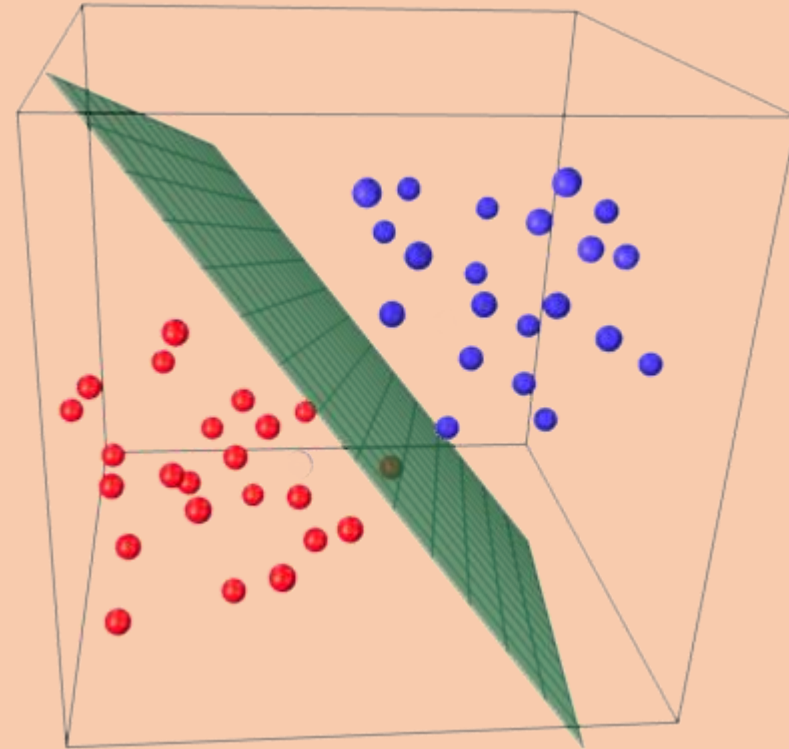
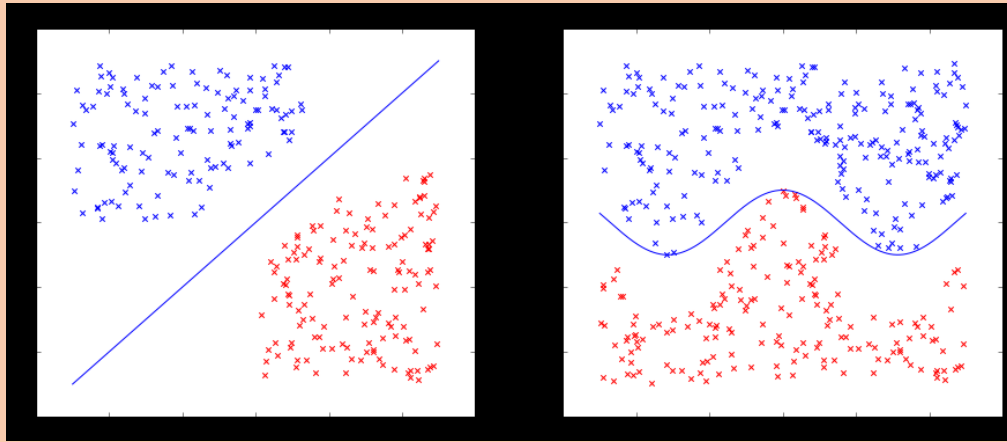
$(w_1 \dots w_n)$ defines a hyperplane which partitions data

In Deep Learning

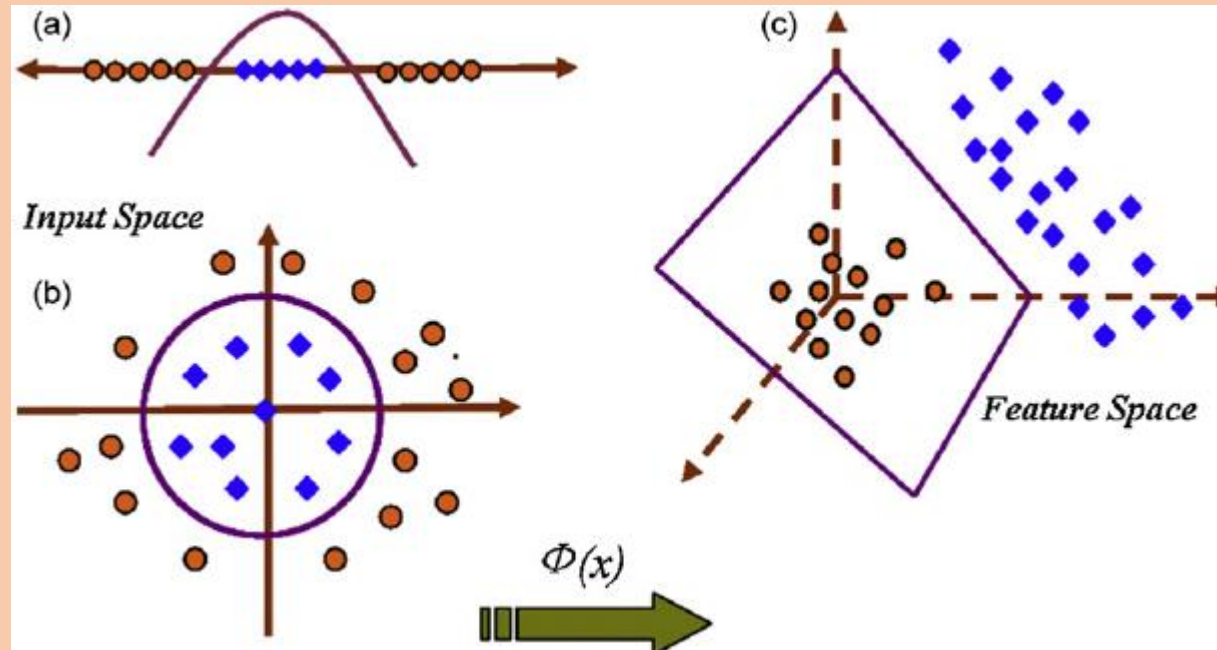
$(w_1 \dots w_n)$ defines the Weights of Input neuron

In a way, the weight is considered as the degree of influence of each variable on the target. Learning is about finding the weights of each co-efficients of equation

Linear Separability vs Absence of it



Linear Separability can be achieved by transformation to Higher dimensional space



Examples of Value Functions

- Linear Regression

- Input: feature vectors

- Output:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$
$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b = \sum_{i=1}^n w_i x_i + b$$

- Logistic Regression

- Input: feature vectors $\mathbf{x} = (x_1, x_2, \dots, x_n)$

- Output: $f(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w} \cdot \mathbf{x} - b}}$

sNo	Diameter (inches)	Number of toppings	Price (\$)
1	6	2	7
2	8	1	9
3	10	0	13
4	14	2	17.5
5	18	0	18

$$\begin{bmatrix} 1 & 6 & 2 \\ 1 & 8 & 1 \\ 1 & 10 & 0 \\ 1 & 14 & 2 \\ 1 & 18 & 0 \end{bmatrix} \xrightarrow{\text{Transpose}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 6 & 8 & 10 & 14 & 18 \\ 2 & 1 & 0 & 2 & 0 \end{bmatrix} \quad Y = X\beta$$

$$\beta = (X^T X)^{-1} X^T Y$$

from numpy.linalg import inv

from numpy import dot, transpose

X = [[1, 6, 2], [1, 8, 1], [1, 10, 0], [1, 14, 2], [1, 18, 0]]

y = [[7], [9], [13], [17.5], [18]]

print dot(inv(dot(transpose(X), X)), dot(transpose(X), y))

Non Linear Regression



Machine Learning “anomalies”

- Algorithmic Intractability with Turing Machine/Lambda/Predicate Logic
- Linearity assumption with Hilbert space method
- Inductive errors inherent in statistical methods

Algorithmic Intractability – A simple example

- The notion of NP-Hard and NP-Complete
- $(26 \text{ factorial divided by } 1 \text{ million}) / (3600 * 24 * 365) / 100000000$

Problems with Hilbert space method

- It assumes that Variables independently act on the output
- The above assumption does not hold in most real life situations
- The variables are inter-dependent
- Some Anecdotes explaining this
 - Combination of people give different results

Statistical methods - anomalies

- The problem of induction
- Changes in the environment
- Effect of Exogeneous data
- Lack of pattern in data
 - Meta Trader Example

Who does ML in a viable manner?

- The Biggies , their ridiculous budget and the resources
- Why they can afford it?
- “We all can drive a rolls Royce, but Ambani’s son can own and operate it”

Will more data helps us make NP problems tractable?

- No....”I have discovered a wonderful proof, where this session is too short for the explanation”
- More data won't produce better result (Central Limit theorem)

When should we use ML?

After you have tried

- If/else/while/select/update programming
- Linear/Quadratic complexity algorithms
- Patented algorithms
- Heuristics based solutions
- Approximate Solutions
- Stochastic solutions

All of the above are control path programming. When these fail you can opt for Machine learning based solutions which will reason based on data.

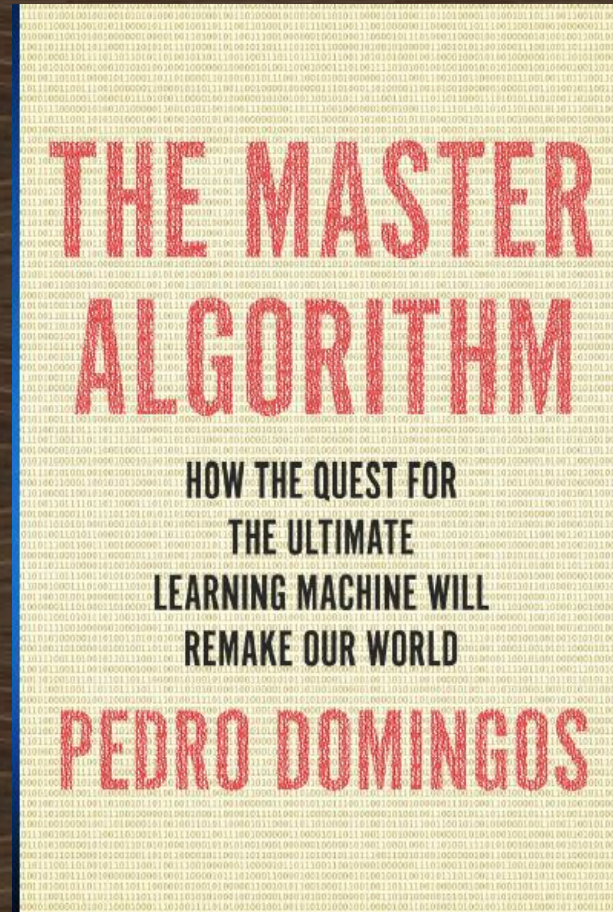
Which is more important?



Aaj mere paas building(ML) hai,
property(lot) hai ,
bank balance (Blockchain) hai,
bungla(Appsec) hai,gaadi hai(DevOps)
Kya hai, Kya hai tumhare paas?

....Mere Paas Maa (Angular) hai

How to proceed from here?



The Five Tribes of Machine Learning

Tribe	Origins	Master Algorithm
Symbolists	Logic, philosophy	Inverse deduction
Connectionists	Neuroscience	Backpropagation
Evolutionaries	Evolutionary biology	Genetic programming
Bayesians	Statistics	Probabilistic inference
Analogizers	Psychology	Kernel machines

The Big Picture

Tribe	Problem	Solution
Symbolists	Knowledge composition	Inverse deduction
Connectionists	Credit assignment	Backpropagation
Evolutionaries	Structure discovery	Genetic programming
Bayesians	Uncertainty	Probabilistic inference
Analogizers	Similarity	Kernel machines

But what we really need is
a single algorithm that solves all five!

● Representation

- Probabilistic logic (e.g., Markov logic networks)
- Weighted formulas \rightarrow Distribution over states

● Evaluation

- Posterior probability
- User-defined objective function

● Optimization

- Formula discovery: Genetic programming
- Weight learning: Backpropagation

Q&A

- If any!