Fundamentals of Computational Forensics:

Machine Learning and Predictive Analytics

Carl Stuart Leichter PhD
carl.leichter@ntnu.no
NTNU Testimon Digital Forensics Group

NTNU

Norwegian University of Science and Technology

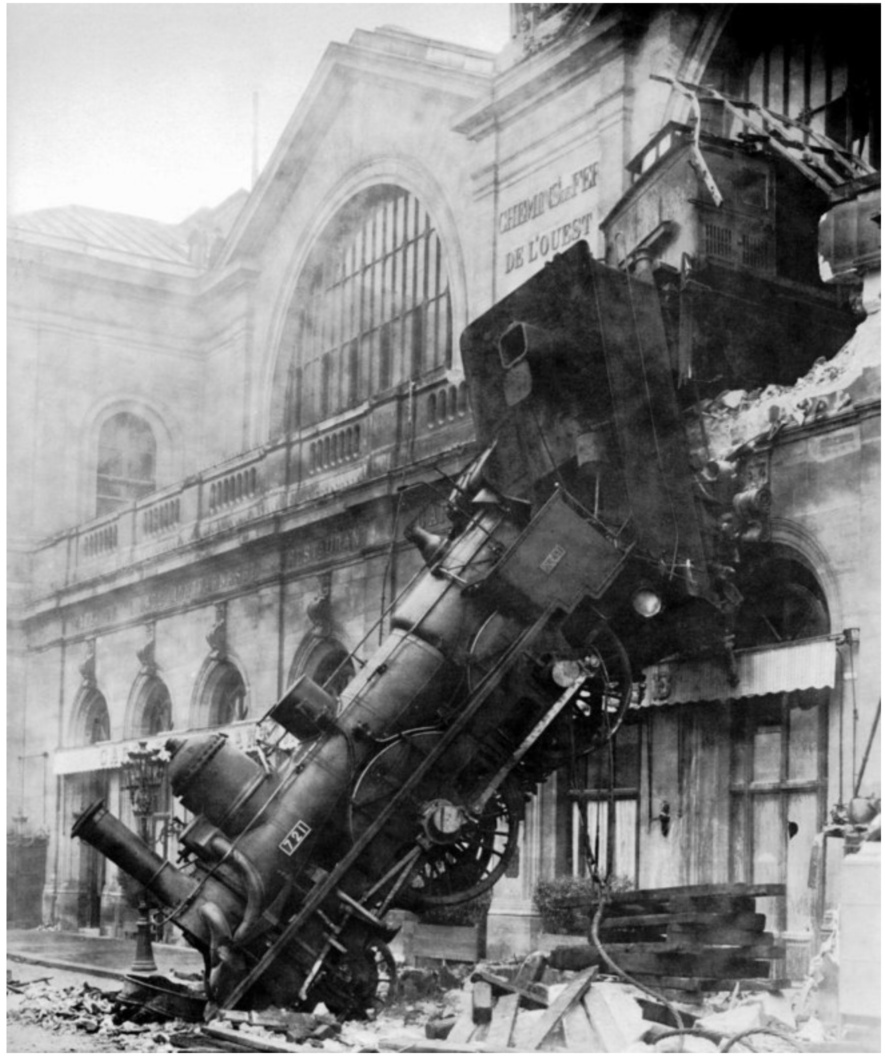
NTNU Testimon Digital Forensics Group

- Cyber Threat Intelligence and Security Operations
  - Malware, IDS, etc
- Digital Evidence Analysis and Linkages
  - Digital Forensics, Network Analysis, Big Data, Simulations, etc
- Public Sector partners

  ØKOKRIM, KRIPOS, CYFOR, etc
- Private Sector partners

  Telenor, NorSIS, mnemonic, KMPG, PWC, etc

Avoid “Push Button” Forensics



https://en.wikipedia.org/wiki/Montparnasse_derailment#/media/File:Train_wreck_at_Montparnasse_1895.jpg

Machine Learning Basics

1. Digital Forensics Motivation
2. Building Models of Systems Under Study
3. Attributes as Features/Feature Space
4. Different types of ML approaches
5. Advanced Topics

MLB-0

Models 1

- Models To Explain the Structure in the Data

M1-0

What Are Our Assumptions?

- We ASSUME there is a hidden structure in our data
  - Exploratory Data Analysis (EDA)
  - Confirmatory Data Analysis (CDA)
- We ASSUME the structure in our data is a reflection of that data's origin (what we are examining)
- We ASSUME that the structure revealed by our data analysis ***is*** the hidden structure we are seeking
- Sometimes, our assumptions are wrong….

M1-1

Building Models

It doesn't matter how beautiful your theory is,

it doesn't matter how smart you are.

If it doesn't agree with experiment,

<u>it's wrong</u>.

-Richard P. Feynman

Why Build Models?

- Suspect used computer to engage in illegal activity.
- Incriminating files were deleted
  - HDD file space is now unallocated
  - Unallocated space partially over-written
    - Traces can still be found.

- Want a ML to recover partially deleted files that are missing headers.

- Each **target** file type has a characteristic ***structure***
  - HTML files
    “<“ “>”
  - JPGs
    - Higher information entropy

- We have a mental ***model*** of the **targets**
- Want the ML algorithms to learn and build internal models of the targets.
  - they build internal models of the data

M1-3

Some Principles of Model Building

1. Observation (Data Input)

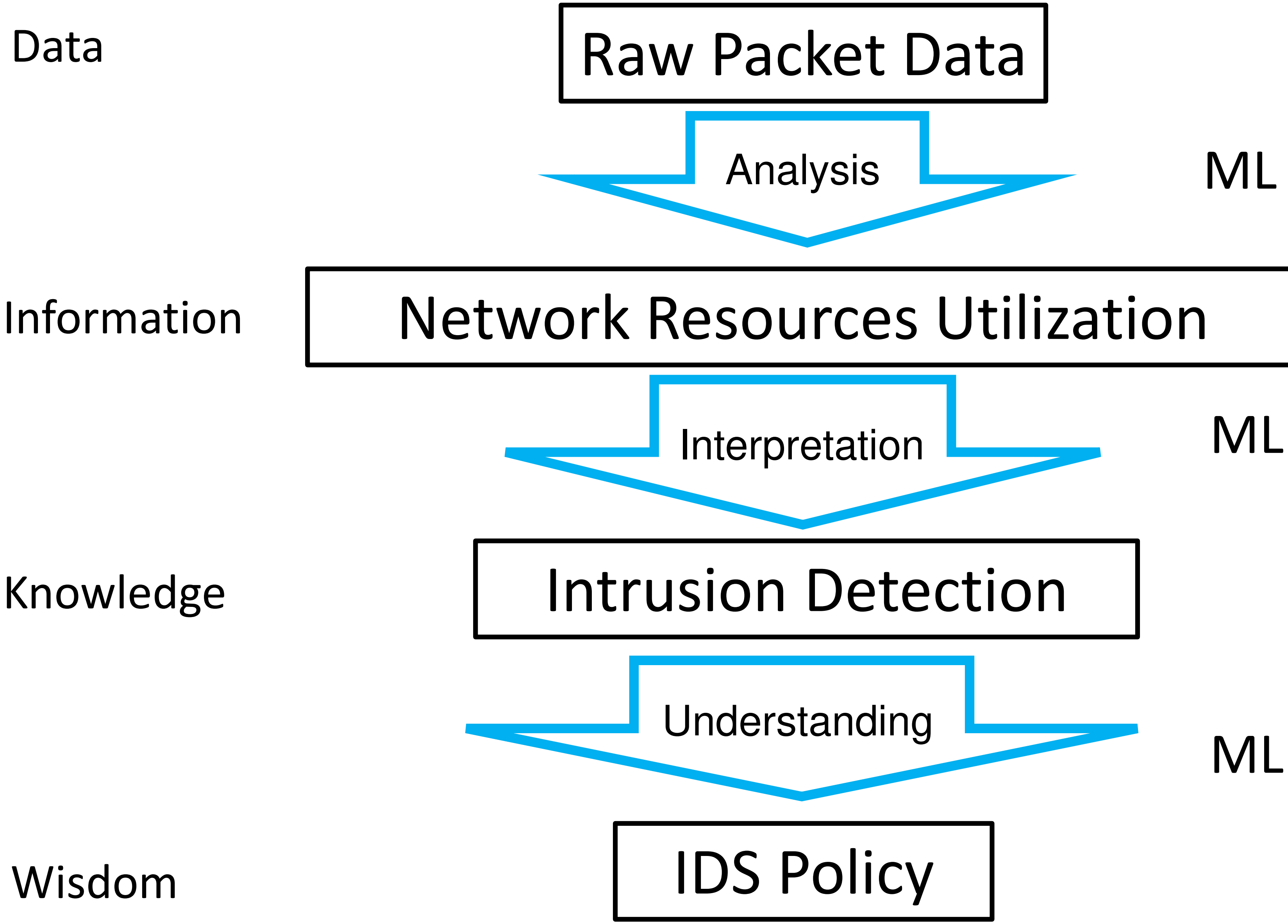
2. Generalization (Model Construction)

3. Application (Model Utilization)

- The choices made for #1 and #2 are driven by #3:
  - It. Depends. Upon. Your. Application.
  - (IDUYA)

M1-4

DIKW Progression



M1-5

Attribute Data (1)

Model of Attribute's Source (2)

Useful Output (3)

M1-6S

Data:
Attributes and Features

A&F-0

Why is The Feature Space So Important?

- Machine Learning isn't magic
- A trained ML algorithm builds an internal model of the feature space.

- SPEND MORE TIME ON THIS
- Features vs attributes

A&F-1

From Attributes to Feature Spaces

Wood Classification Example

- Have a big pile of mixed wooden blocks
- 2 different kinds of wood
  - Ash
  - Pine

- Want to be able to measure a wooden block's attributes and use them to determine the type of wood

- Decided on two optical attributes
  1. Overall brightness
  2. Wood grain prominence (peak to peak variation)

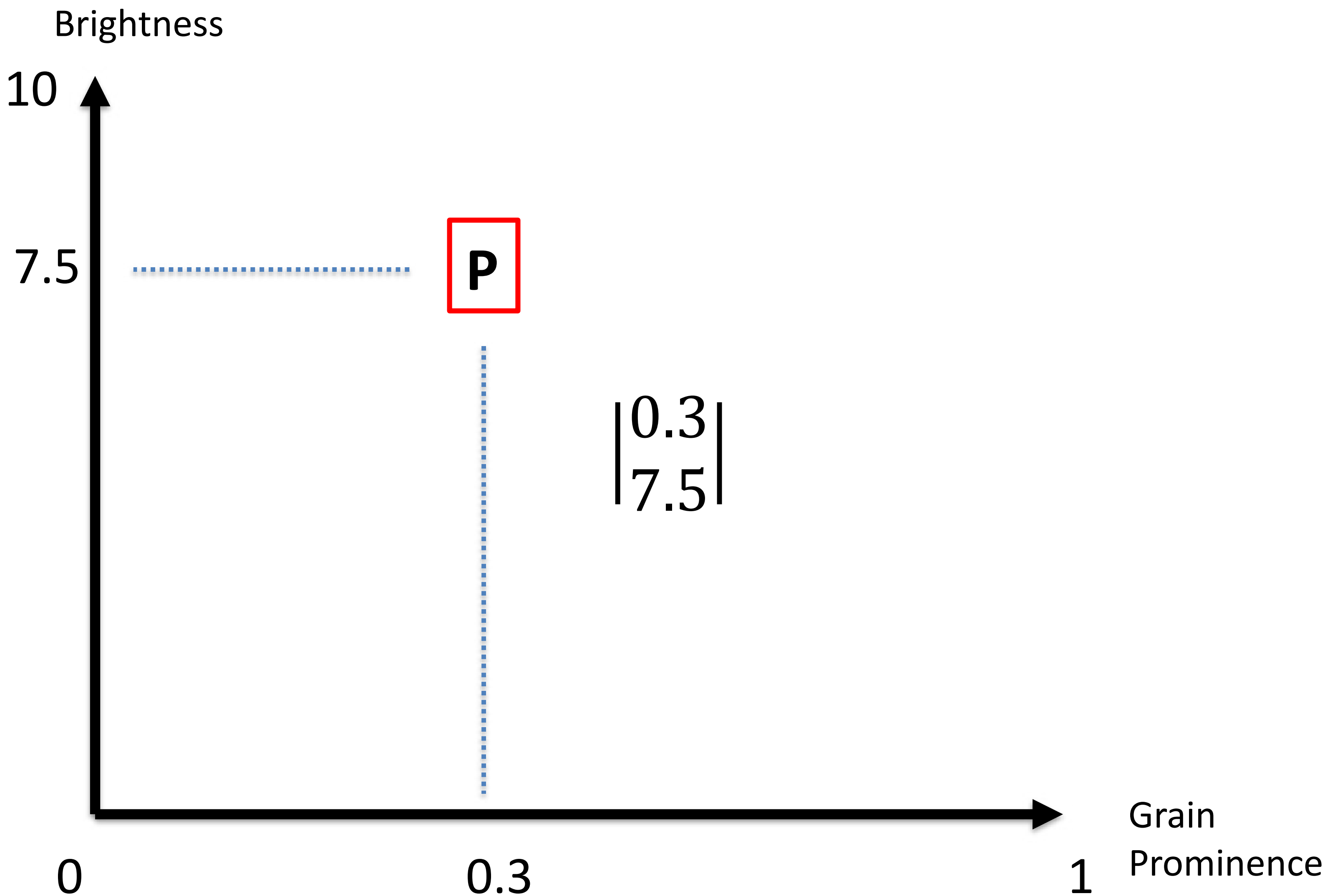
A&F-2

Wood Brightness and Grain Prominence

http://www.dannerscabinets.com/blog/mn-custom-cabinet-shop-custom-cabinets/

A&F-3

Attributes Form Feature Vectors



$$\begin{vmatrix} 0.3 \\ 7.5 \end{vmatrix}$$

A&F 4

Important Aspect of Feature Spaces

If a feature space is a vector space,

=> ***<u>All the tools of Linear Algebra can be utilized!</u>***

A&F 5

What Does Your Data Represent?

- The **attributes** of what you are studying/modelling:
  - Length (meters, inches, light years)
  - Weight (grams, pounds, carats)
  - Time (seconds, years)
  - Money
  - Number of Packets
  - Number of Bytes
  - Etc

Can All Be Combined into Feature Vector

Enables Data Fusion

A&F-4

Some Digital Forensics Attributes

- Intrusion Detection Packet Structure
  - Packet Size
  - Data Size
  - TTL Time
  - ACK Sequence
- Malware File Structure
  - File Size
  - Data Section Size
  - Data Entropy
  - API Calls
- Crime Investigation
  - Character distribution
  - Data Entropy
    - 80% - Compression
    - ~100% - Encryption

A&F-5

Data Collection (Observation)

- What attributes are *important*?
- Are there *redundancies* we can exploit?
  - – Fewer attributes required
    - Reduce data dimensionality
    - Reduce model complexity

Attribute Data Preprocessing

- Prepare the data for use in ML
- Clean the data
  - Remove outliers
  - Reduce noise

- Feature Extraction
  - Spectral Analysis
  - Principal Component Analysis
  - Independent Component Analysis

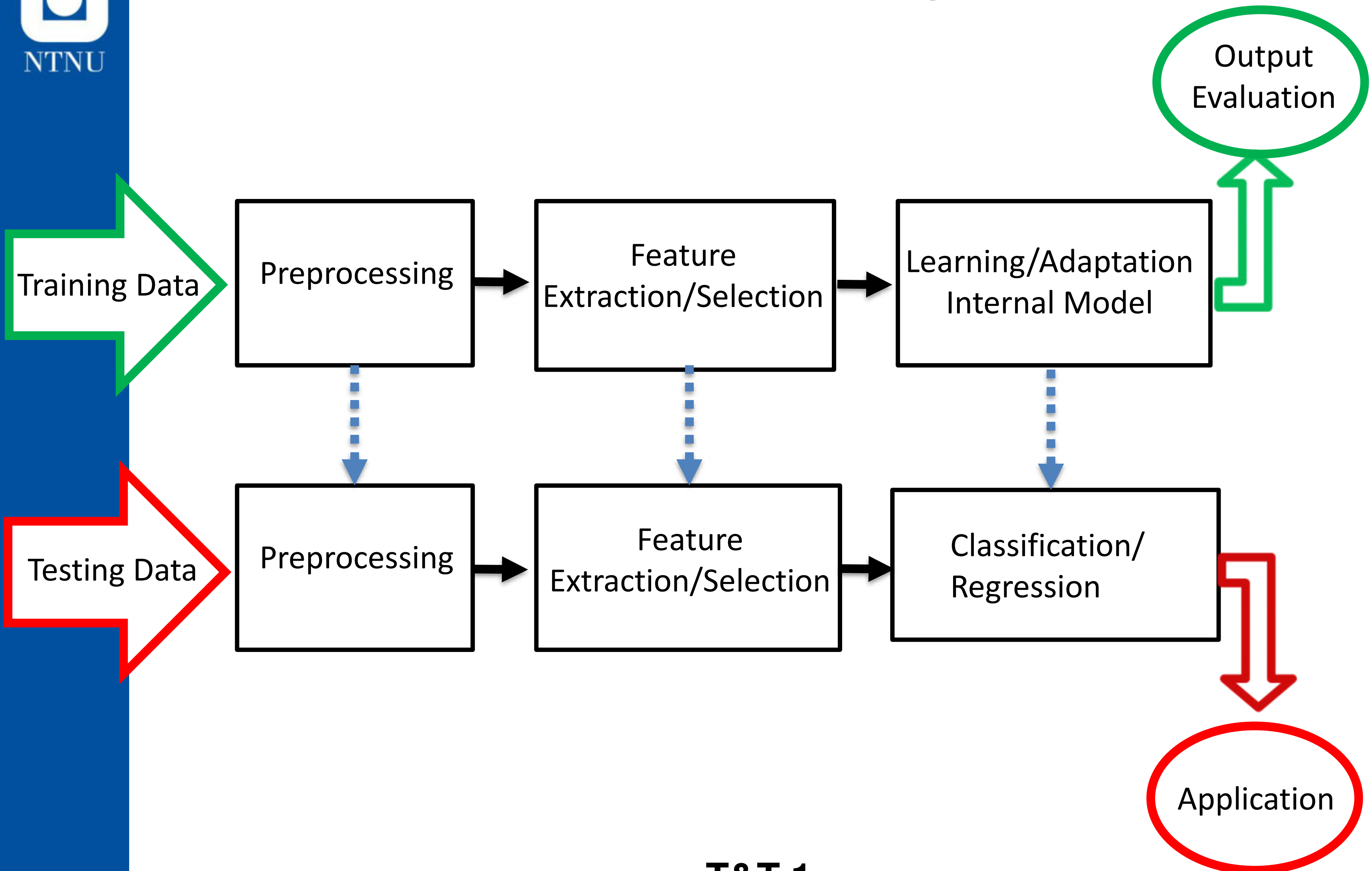
- Feature Selection
  - Remove redundant features (CFS)

A&F-7

Basic Machine Learning: Testing & Training Data

T&T-0

The Machine Learning Process



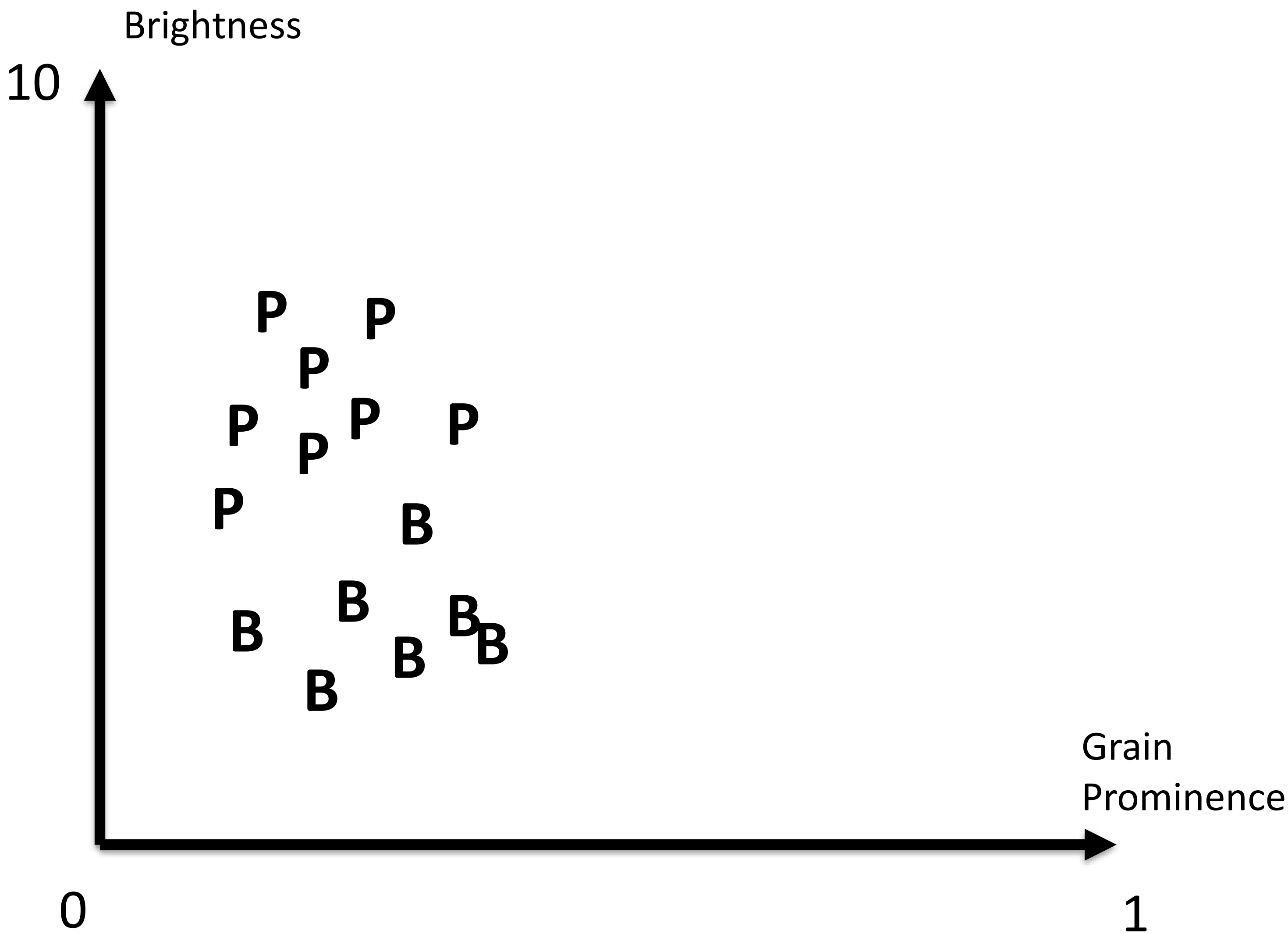
T&T-1

Training/Testing Data Partition

- Not all of the available data is used in **training**
- Some of the data is held back, to **test** the model that was created by the ML adaptation to the training data
- A good model with sufficient data will learn to "generalize"
  - During training, it will adapt to the hidden structure in the data
  - If the data contains a good representation of the system under study (by implication, the structure in the system) then it will recognize the **test data as new data samples** from the system

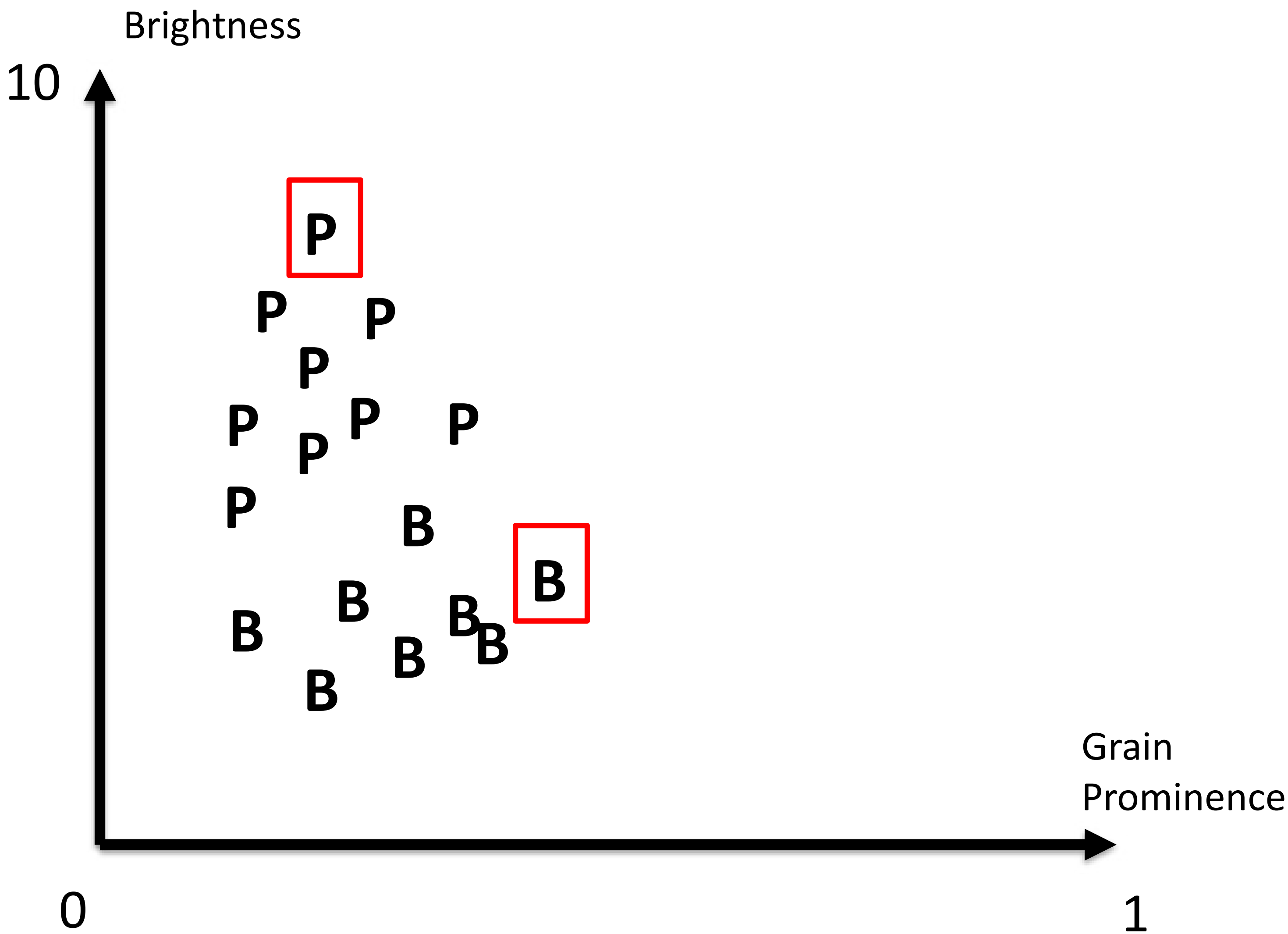
T&T-2

Training the Wood Classifier

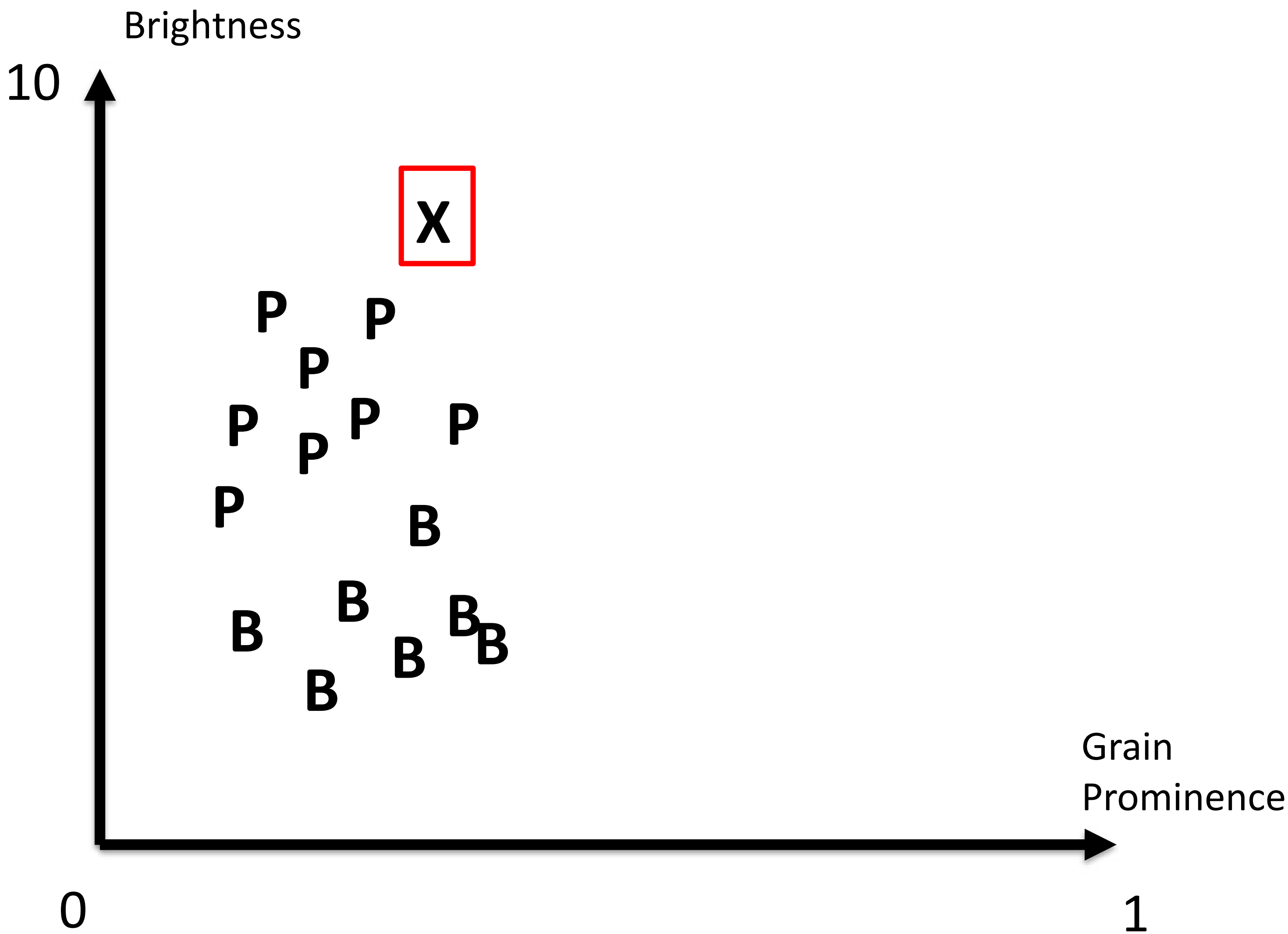


T&T-3

Testing the Wood Classifier



Using the Wood Classifier



T&T-4

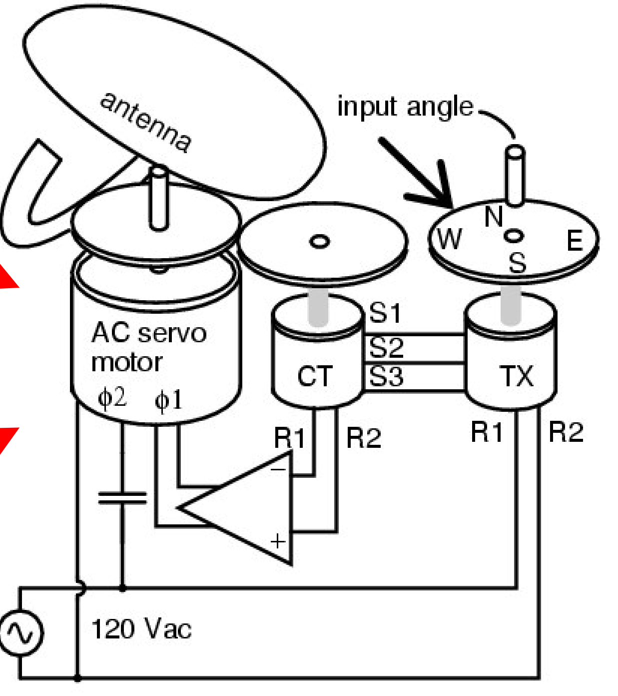
The Internal Model

Internal Model Principle

Θ

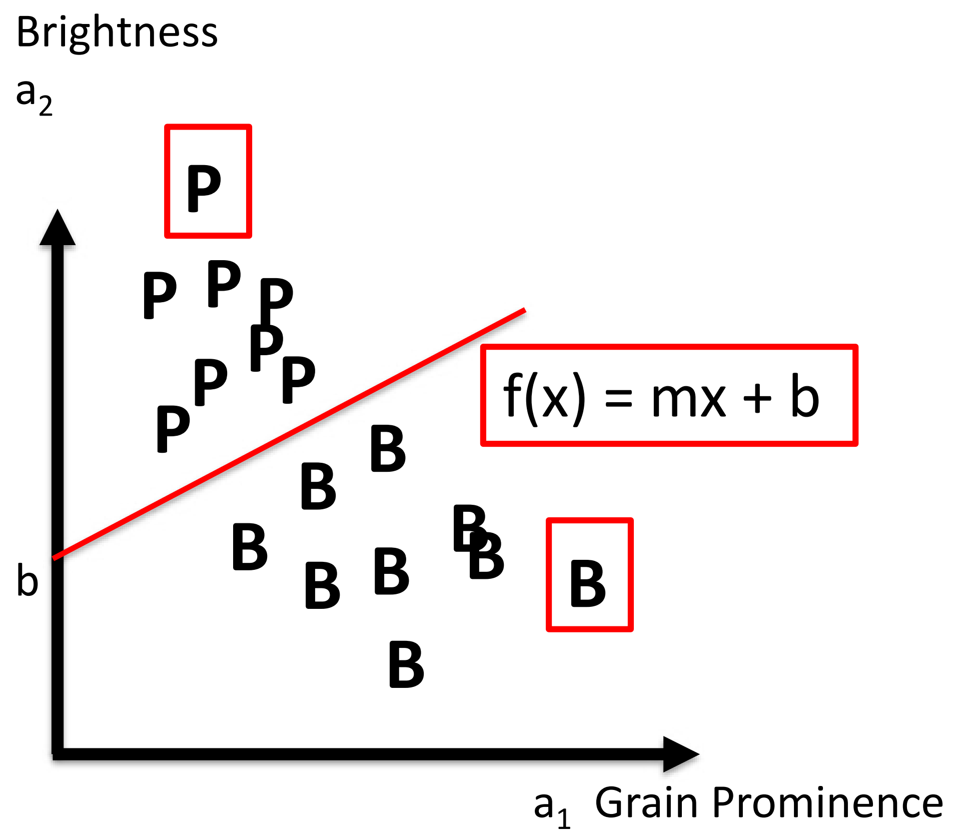
ω_Φ

α_Φ



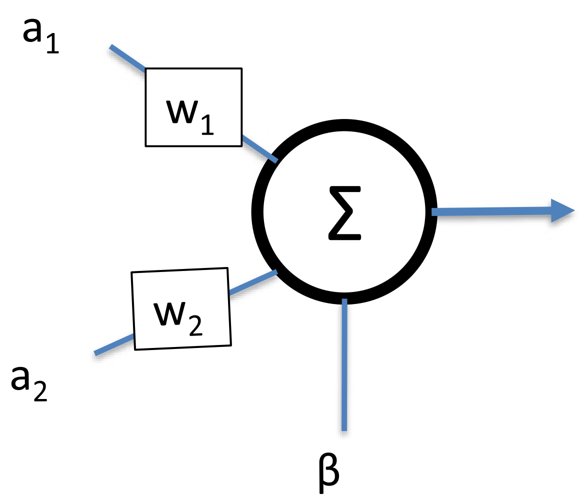
IM-1

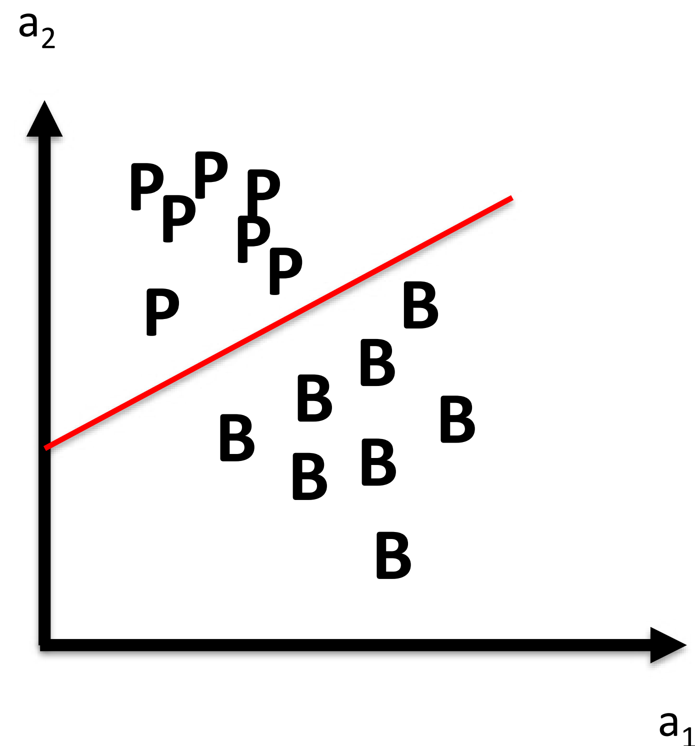
A Two Class, Wood Classifier (Pine and Birch)



IM-2

A Simple Two Class "Perceptron"





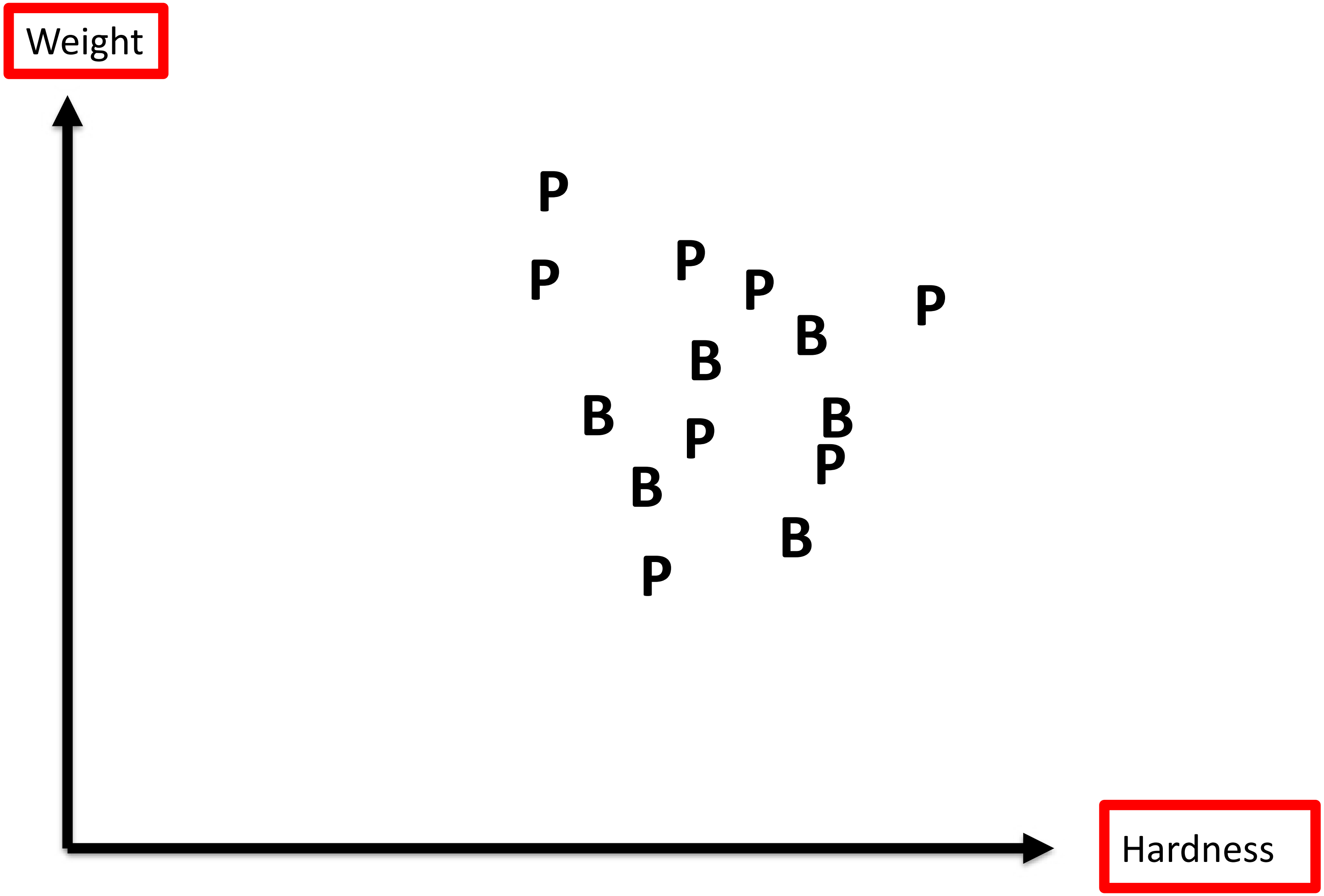
$\mathbf{w}^T=[w_1\ w_2]$

$\mathbf{a} = [a_1\ a_2]^T$

$f(\mathbf{w}, \beta) = \mathbf{w}^T\mathbf{a} + \beta$

IM-3

Feature Selection Revisited



Where Are the Class Boundaries?

A&F-8

What Model Complexity is Required?

It Depends Upon Your Application!

- Project Apollo Moon Landings
  - Relativistic mechanics not used
  - Newtonian mechanics

- GPS Computations
  - Relativity correction required

IM-4

Simplest Models: Knowledge Representation

- Uses existing knowledge to create new
  - Perspectives of the data
  - Knowledge from the data.

- Raw data is often not understandable or informative
  - additional transformation
  - New representation.

IM-5

Knowledge Representation

- General approaches:
  - Rules Based Learning
    - First-order logic
    - Decision Trees
  - Regression (Curve Fitting)
  - Descriptive Statistics
    - Average (Mean)
    - Variance
    - Type of Distribution
      - Normal (Gaussian)
        - » "Mean" is sometimes called "the norm"
      - Uniform
      - Etc

IM-6

Internal Models:
Rules Based Learning

RB-0

First Order Logic

- Logical Descriptions
  - describing data samples themselves
  - describing relationships between data samples
  - describing relationships between data and outputs

Every skier likes the snow:

∀x Skier(x) => LikesSnow(x)

All brothers are siblings:

∀x ∀y Brother(x, y) => Siblings(x, y)

http://people.westminstercollege.edu/faculty/ggagne/fall2014/301/chapters/chapter8/index.html

RB-1

Decision Trees

– Each branch is selected by the answers to a given decision

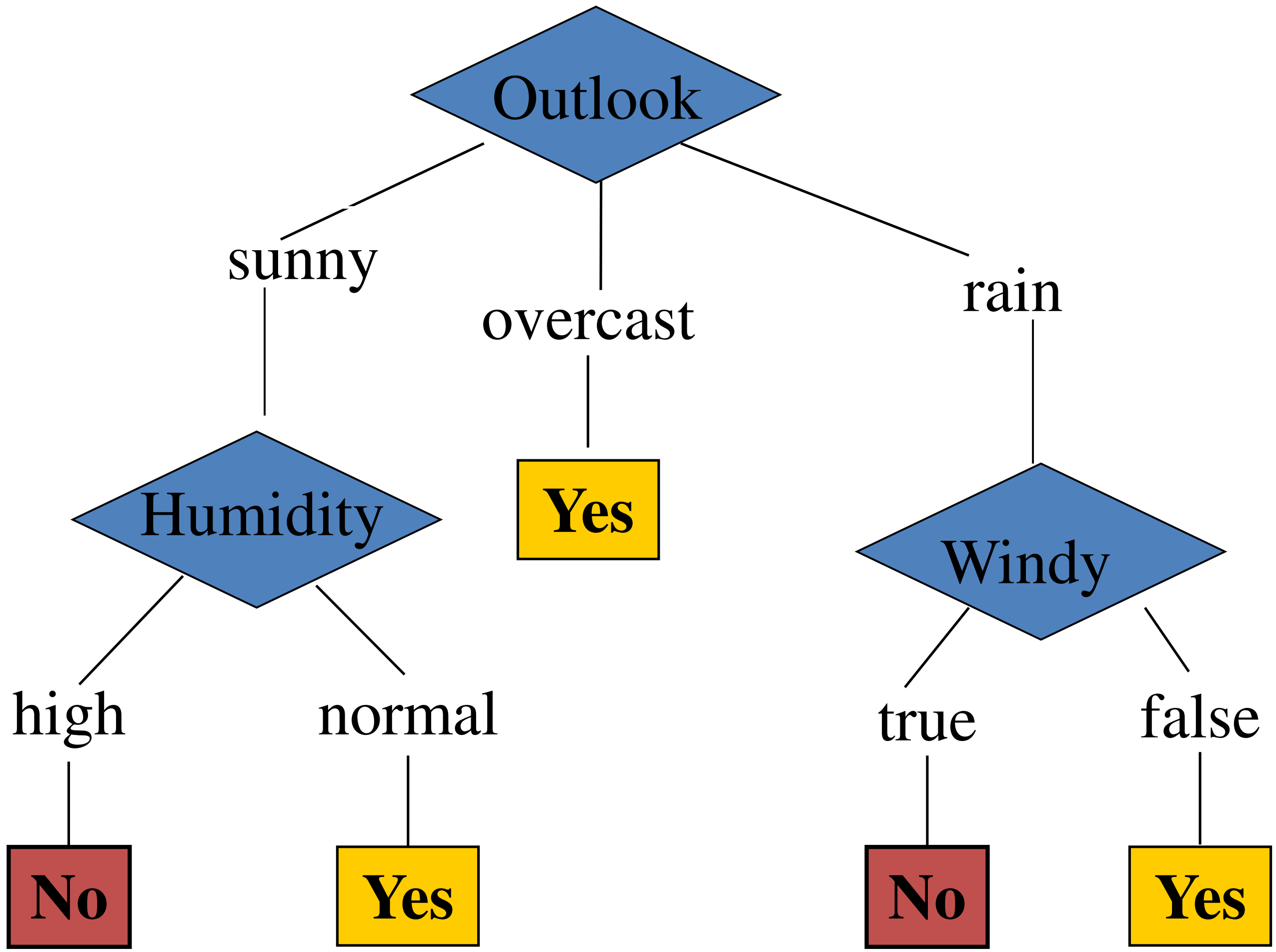
– The descent down the tree is like a series of feature space partitionings

– The series of decisions will lead from the root to a specific leaf.

  • Decision/Classification

RB-2

To ‘play frisby golf’ or not.



(Outlook==rain) and (Windy==false)

Pass it though the tree
-> Decision is yes.

RB-3

Decision Tree
Feature Space Partitioning

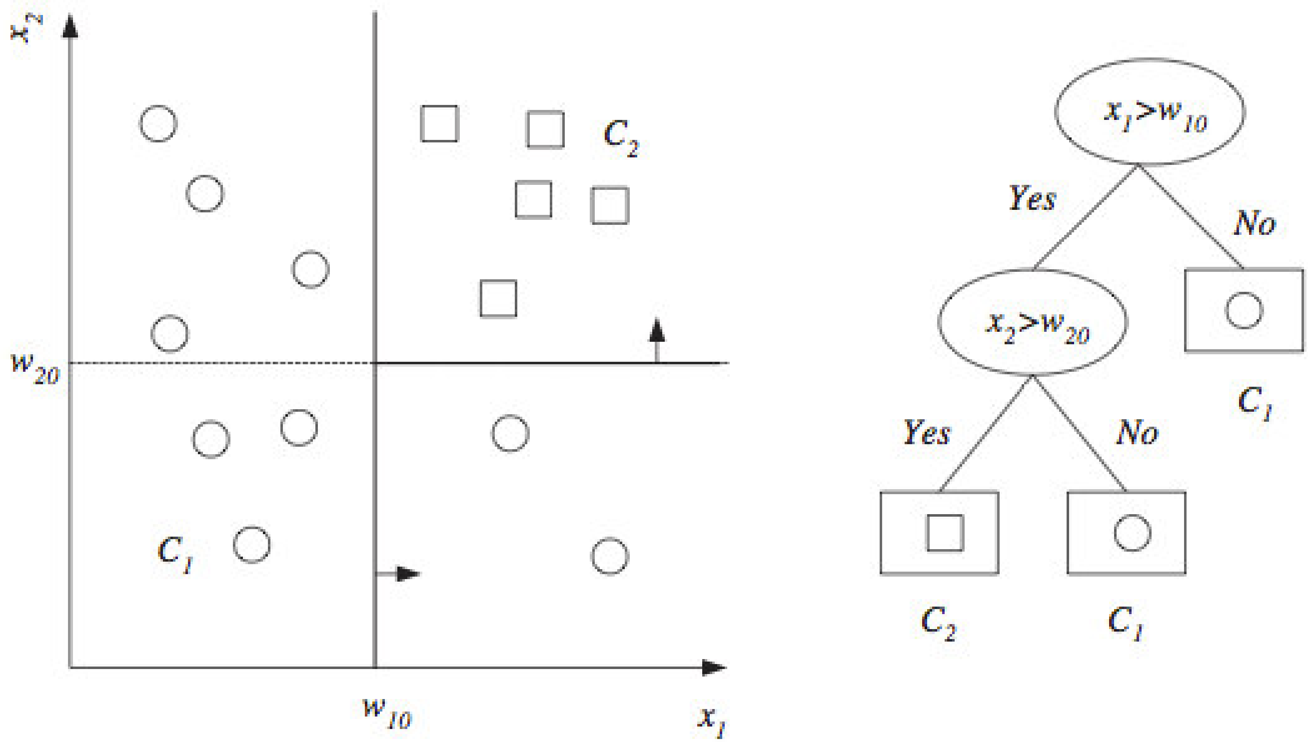


Figure 9.1 Example of a dataset and the corresponding decision tree. Oval nodes are the decision nodes and rectangles are leaf nodes. The univariate decision node splits along one axis, and successive splits are orthogonal to each other. After the first split, $\{\mathbf{x}|x_1 < w_{10}\}$ is pure and is not split further.

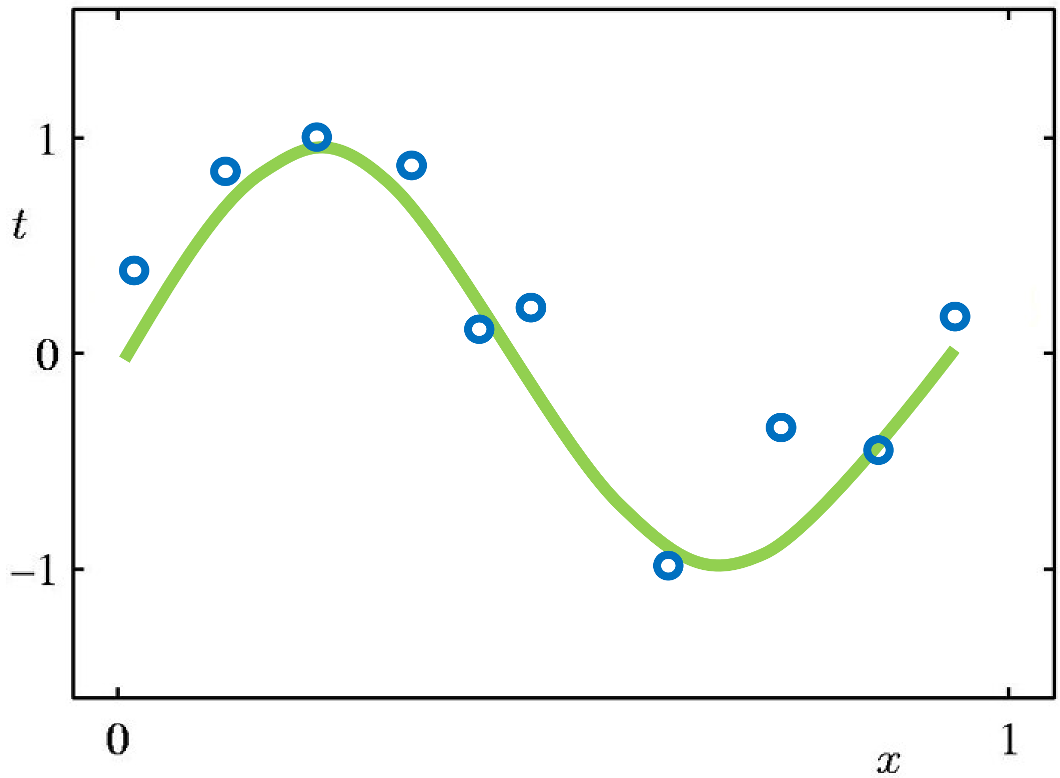
RB-4

From Alpaydin, 2010

Objective Functions

OF-0

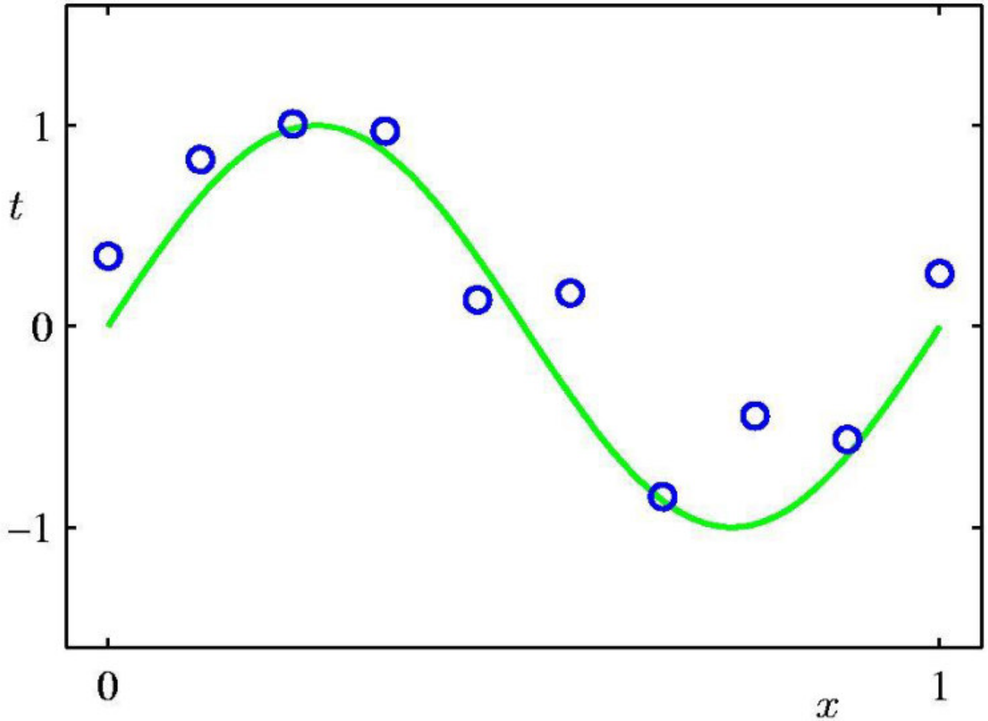
Polynomial Curve Fitting



$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

Internal Model

OF-1



$$y(x, \boxed{\mathbf{w}}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

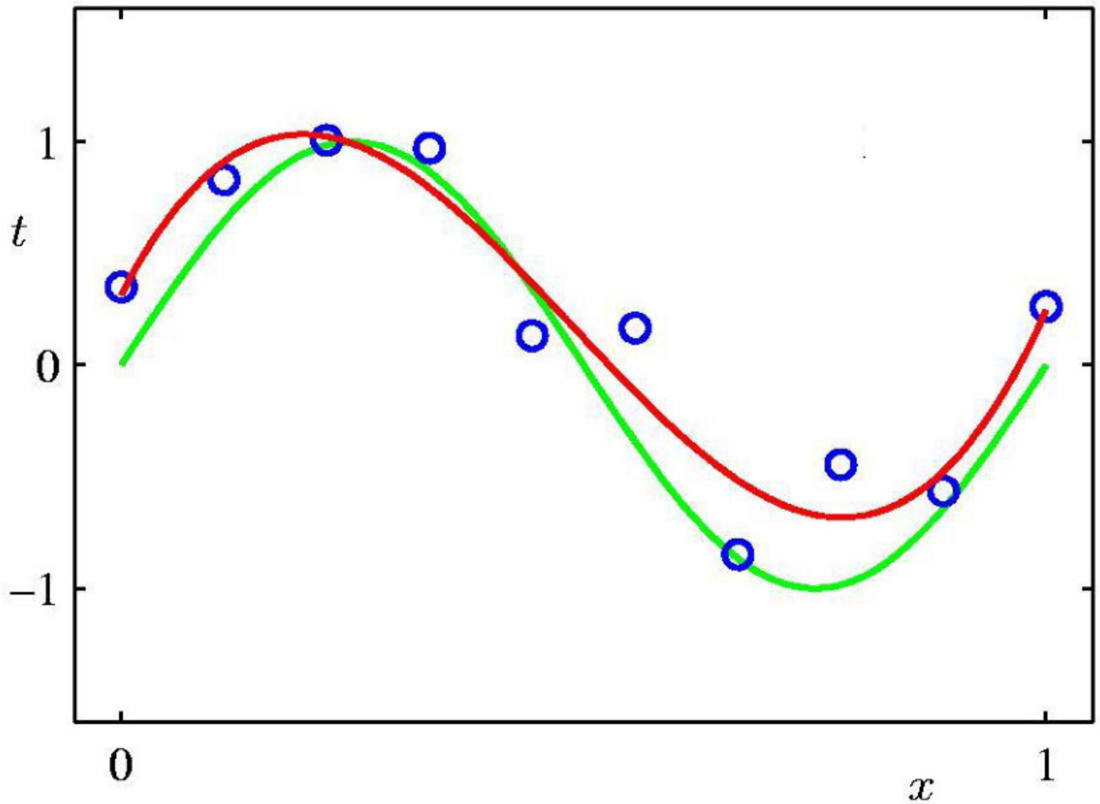
Find the weights w_j

OF-2

Polynomial Curve Fitting

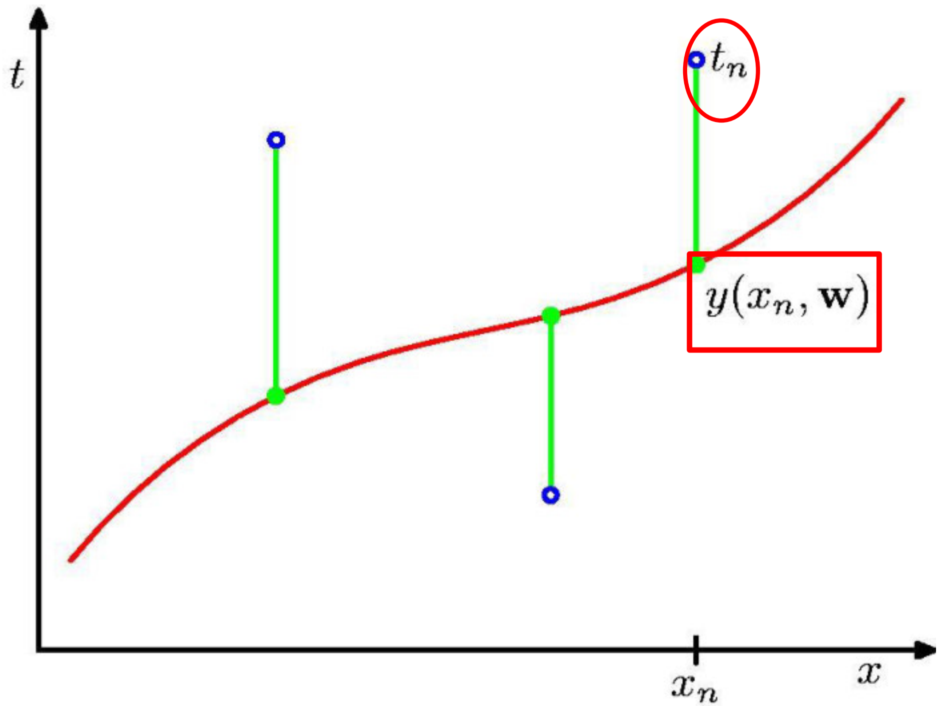
Real world system to be modelled

Regression estimated model



OF-3

Sum-of-Squares Error Function



$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{\boxed{y(x_n, \mathbf{w})} - t_n\}^2$$

This is an “Objective Function”

It measures how well our internal model accounts for the data

OF-4

Objective Functions

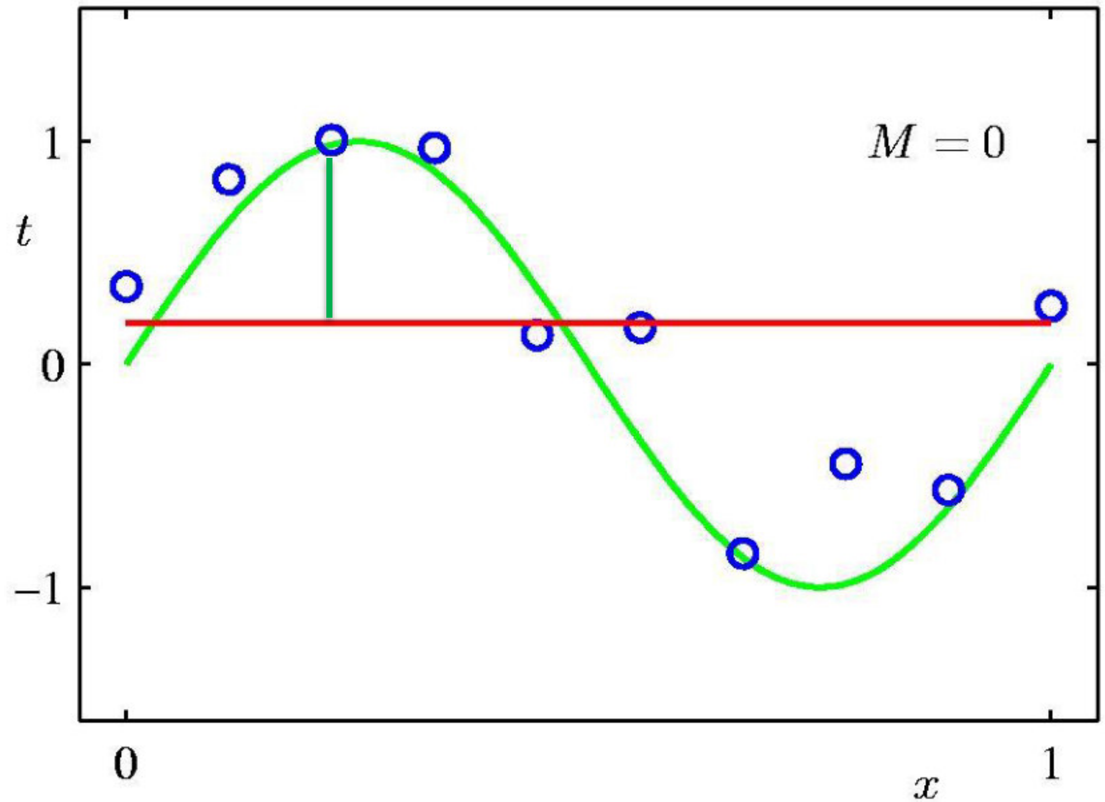
- Measures a figure of merit to be ***optimized*** during the learning process
  - Sum of Squares (for the regression example)
  - Mean Square Error (MSE)
    - Average of sum of squares
  - Least Mean Squares (LMS)
  - Statistical Measurements
    - Variance
    - Kurtosis
  - Information Theoretical Metrics
    - Mutual Information
    - Information Entropy
      - Negentropy

(Internal) Model Complexity

MC-0

0^{th} Order Polynomial $y(x, \mathbf{w}) = w_0$

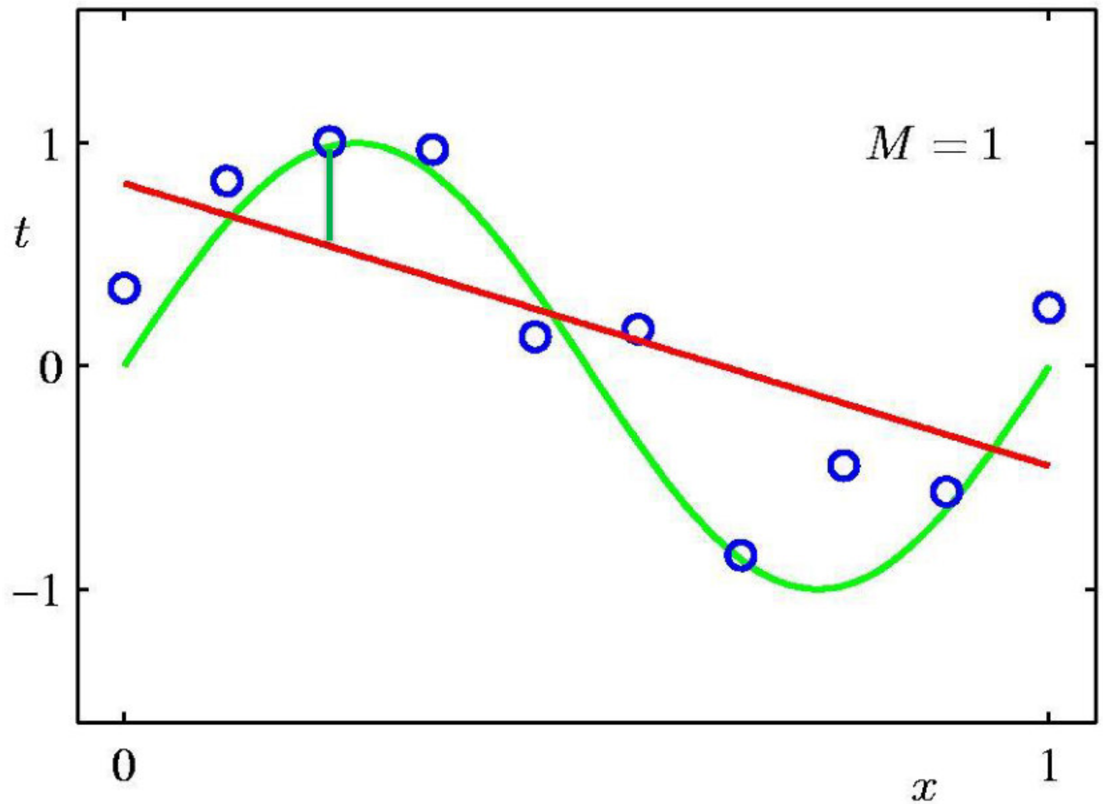
Regression estimated model —



$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

MC-2

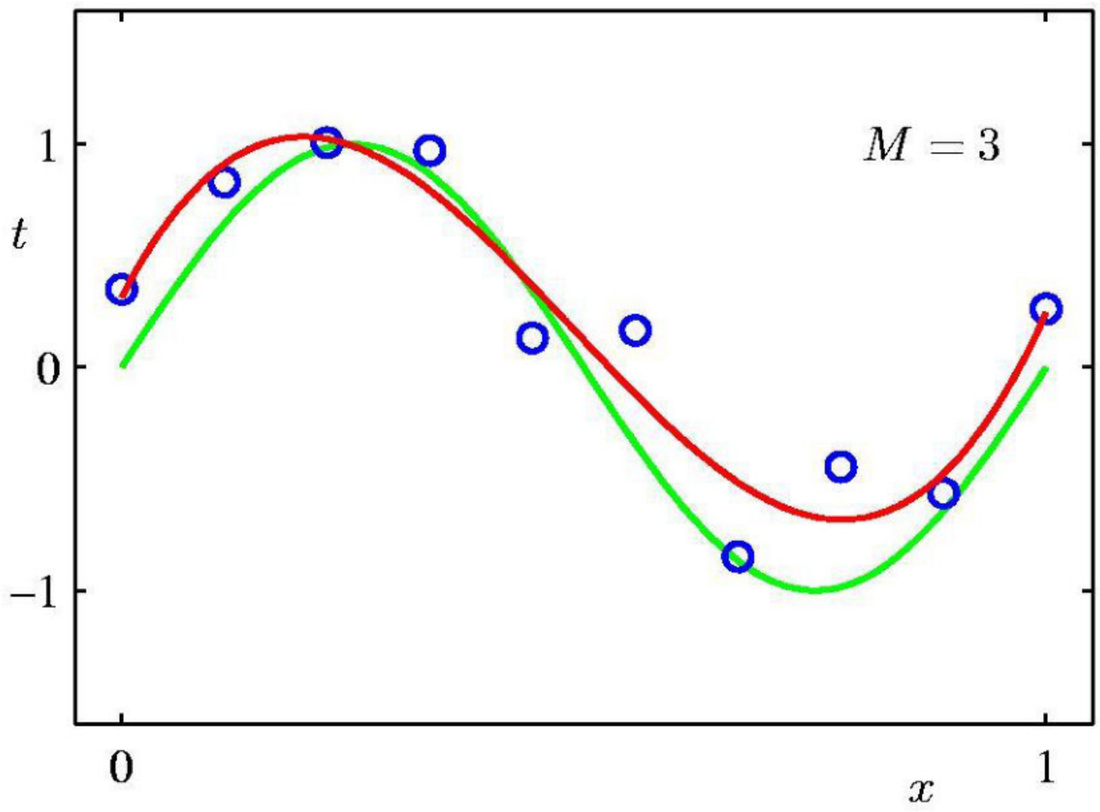
1st Order Polynomial $y(x, \mathbf{w}) = w_0 + w_1 x$



MC-2

3^{rd} Order

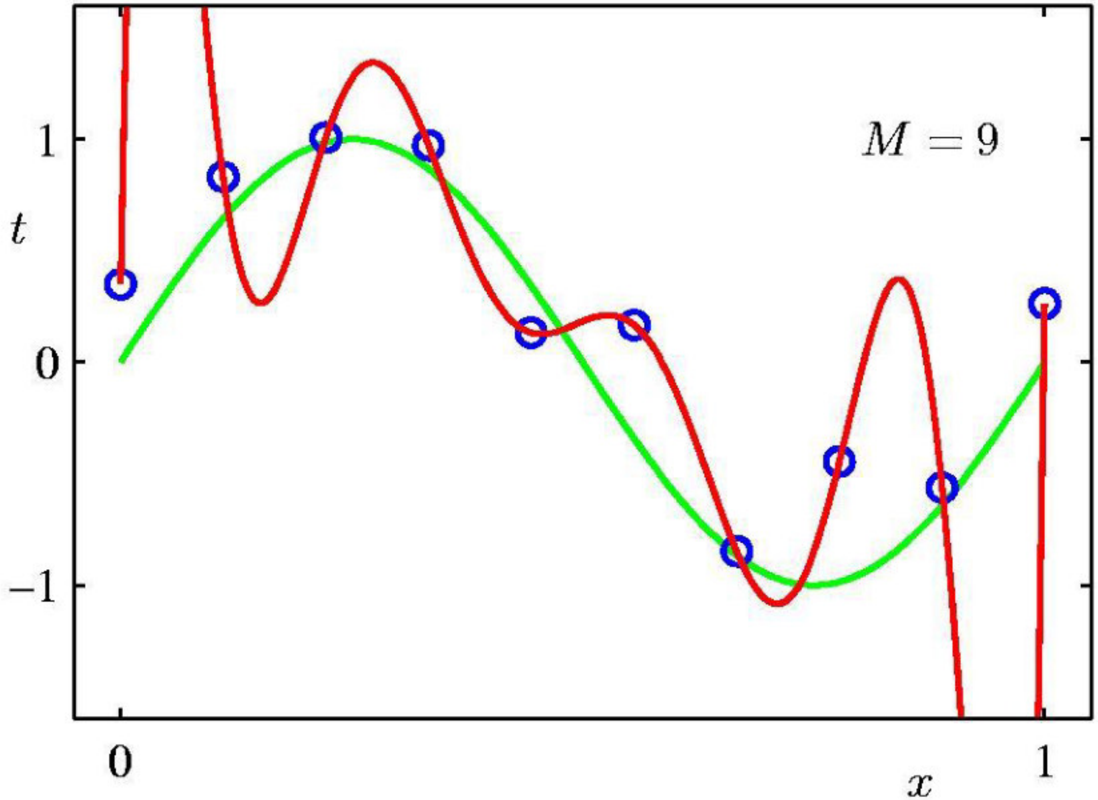
$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + w_3 x^3$$



MC-3

9th Order

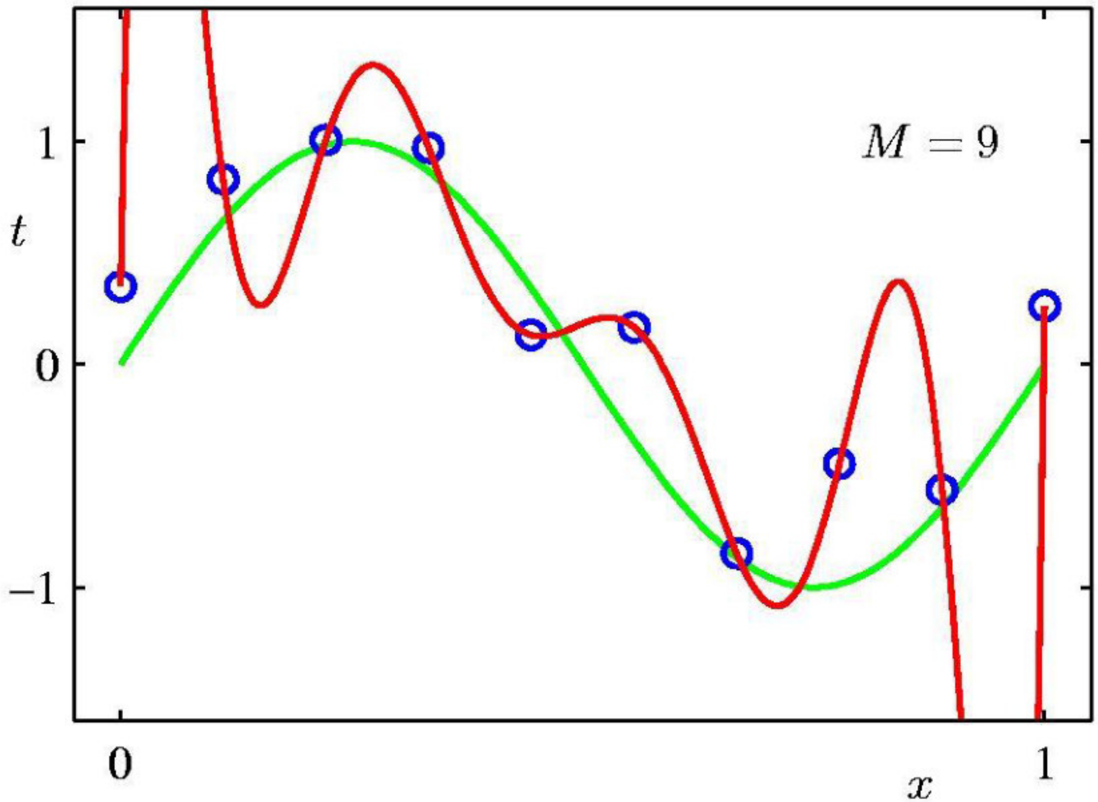
$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M$$



What Happened?!

Model Complexity

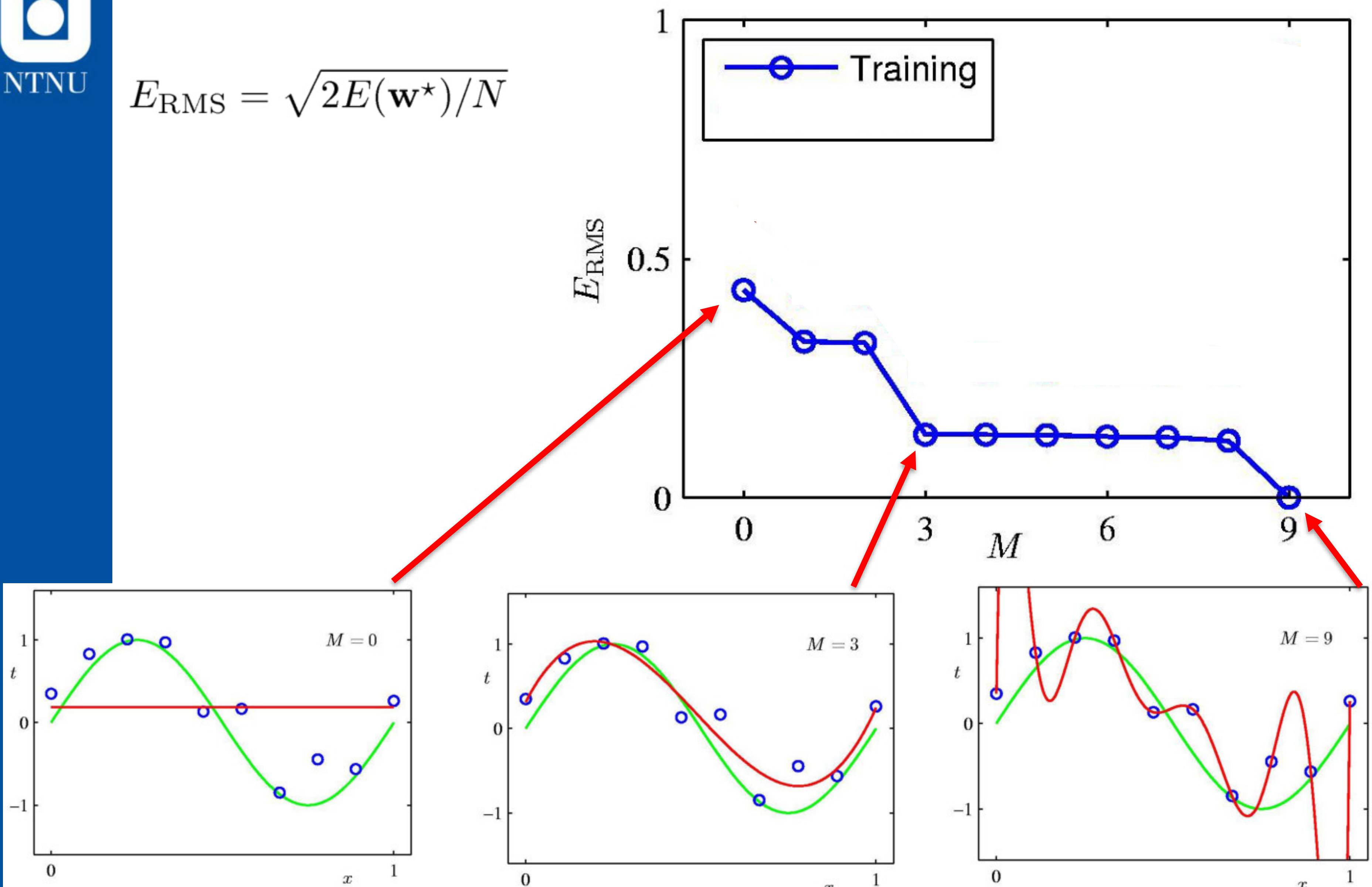
- Curse of Dimensionality (Too Much Complexity)
- Overfitting



MC-7

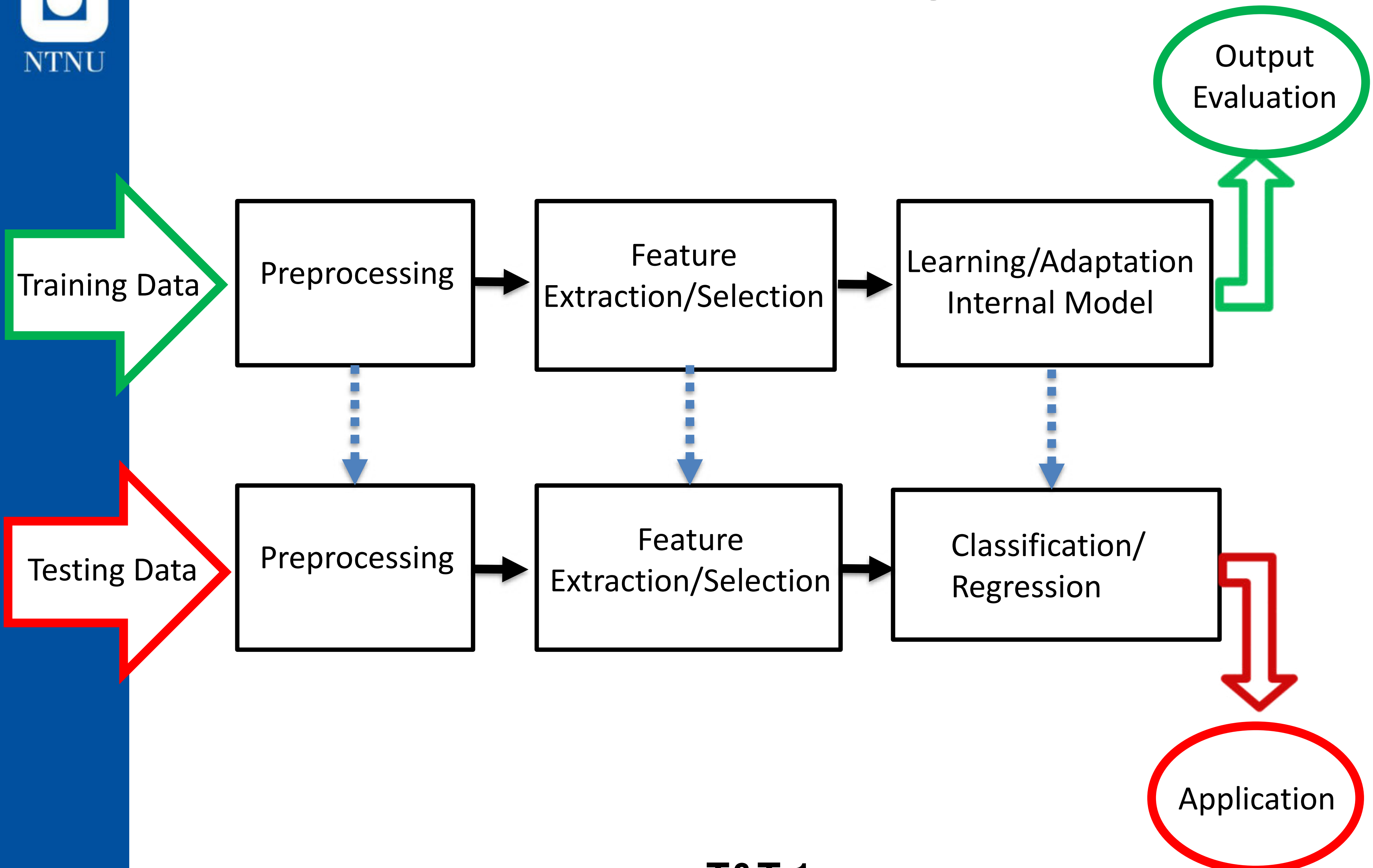
Training Performance Evaluation

$$E_{\mathrm{RMS}} = \sqrt{2E(\mathbf{w}^\star)/N}$$



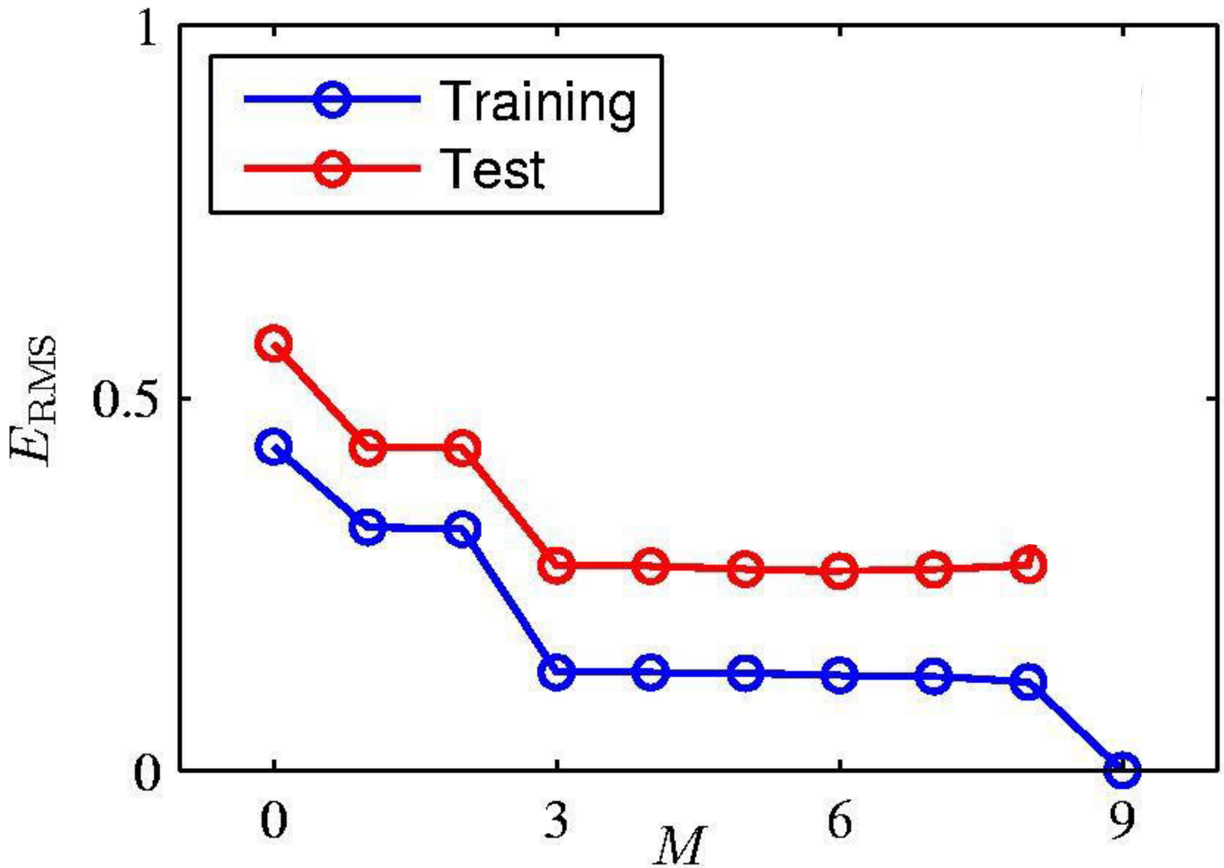
MC-8

The Machine Learning Process



T&T-1

Training Data, Testing Data & Over-fitting



MC-9

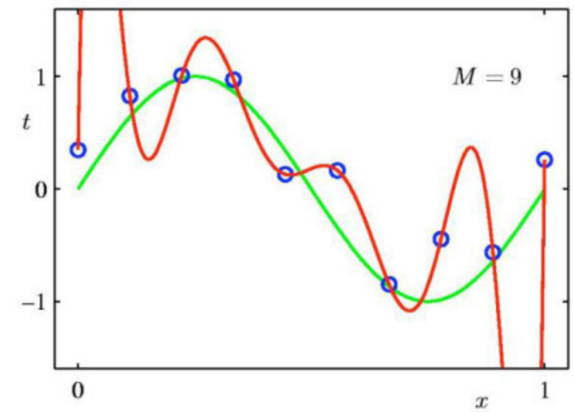
A Central Principle in ML

- ***The model complexity drives the training data requirements!***

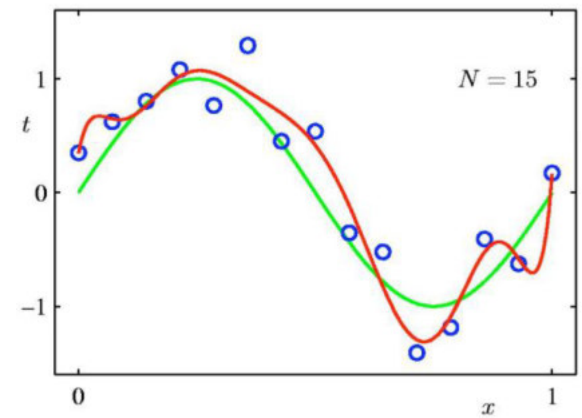
MC-10

More Data Can Fix Overfitting Problem

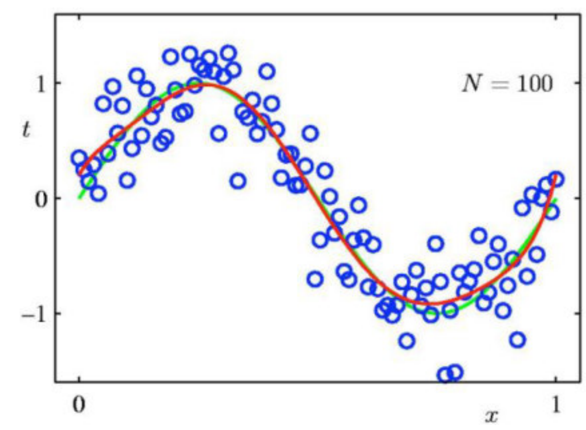
- N= 10 Data Points



- N= 15 Data Points

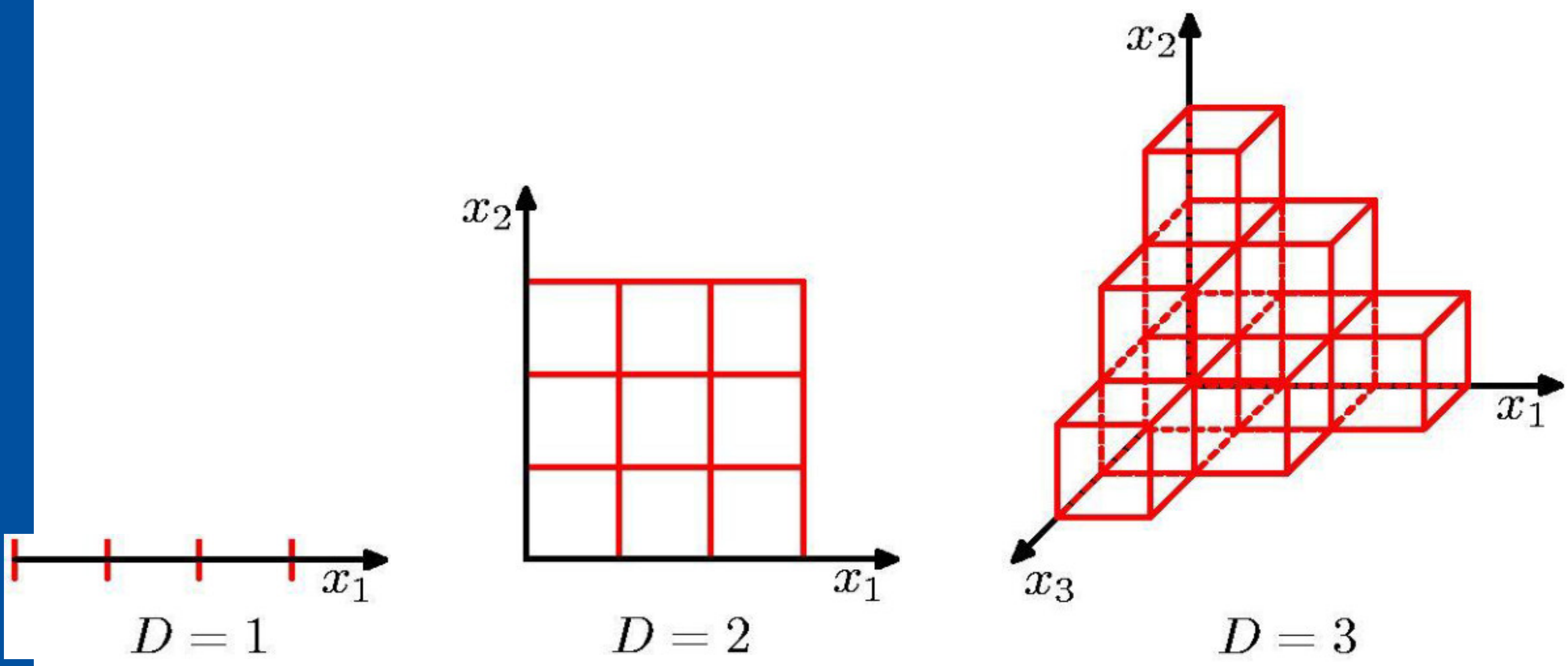


- N= 100 Data Points



MC-11

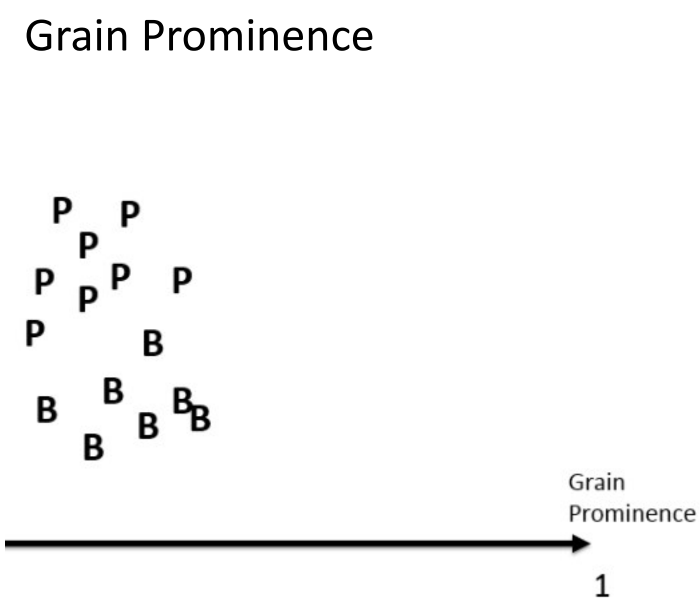
Curse of Dimensionality (Model Complexity)

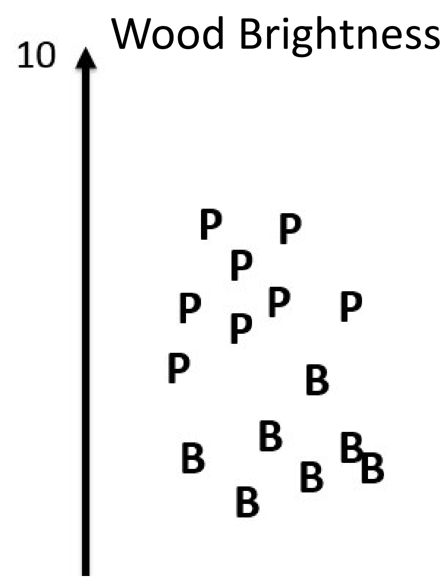


MC-12

- More complex problems, require more complex models
- More complex models, require more complex feature spaces
  - Need higher dimensionality to get good class separation

Wood classifier with 1D feature space?





MC-13

Distance Metrics

DM-0

The Distance Metric

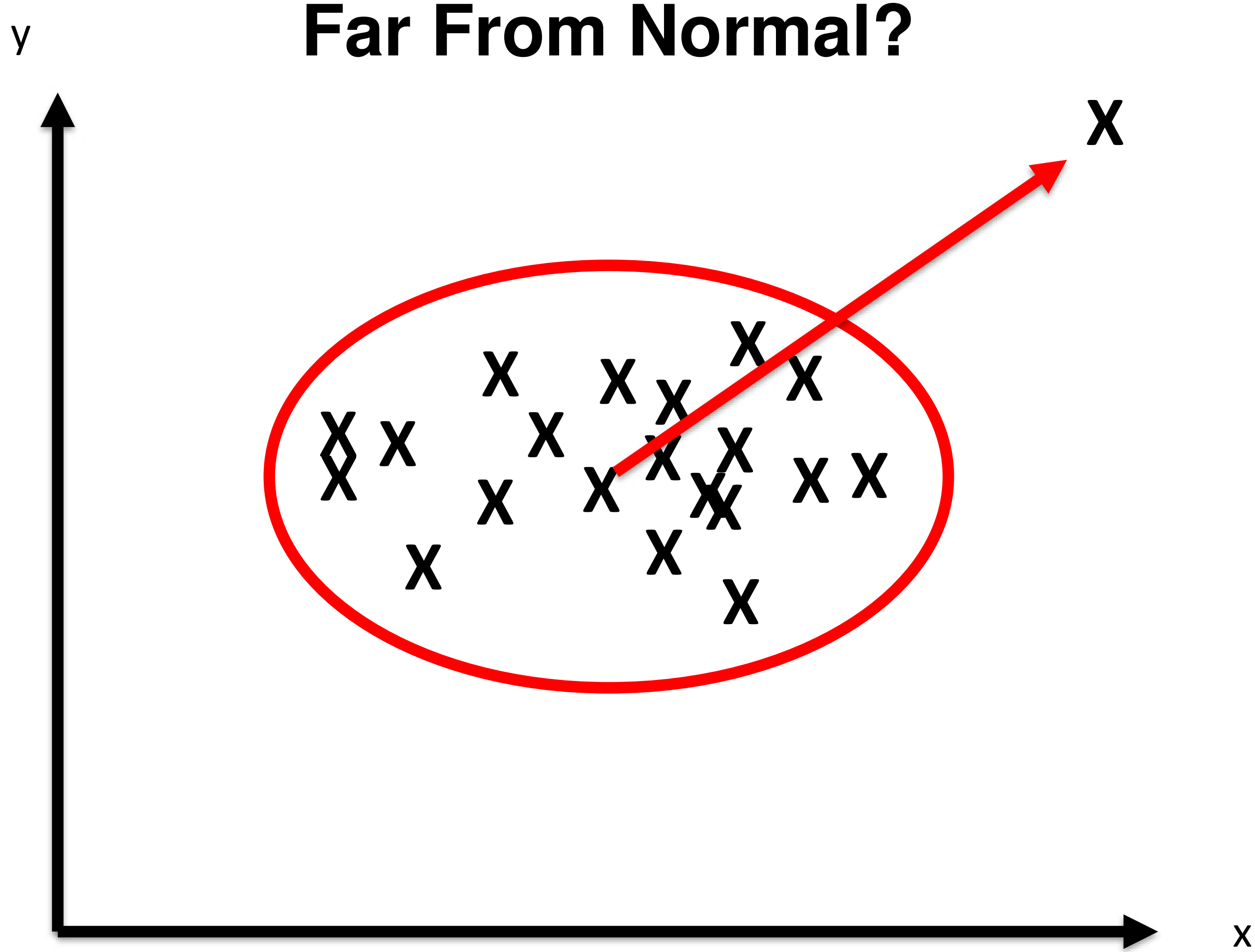
- How the similarity of two elements in a set is determined, e.g.
  - Euclidean Distance
  - Inner Product (Vector Spaces)
  - Manhattan Distance
  - Maximum Norm
  - Mahalanobis Distance
  - Hamming Distance
  - Or any metric you define over the space…

NTNU

Manhattan Distance

https://www.quora.com/What-is-the-difference-between-Manhattan-and-Euclidean-distance-measures

Far From Normal?

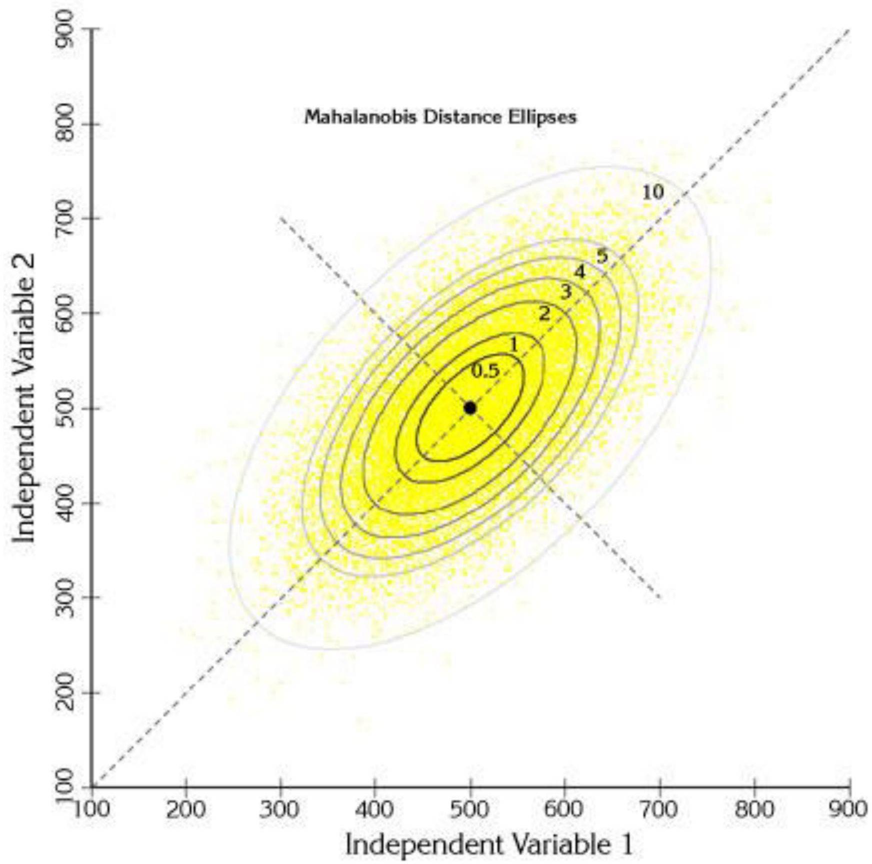


Center = Mean

Spread = Variance

DM-3

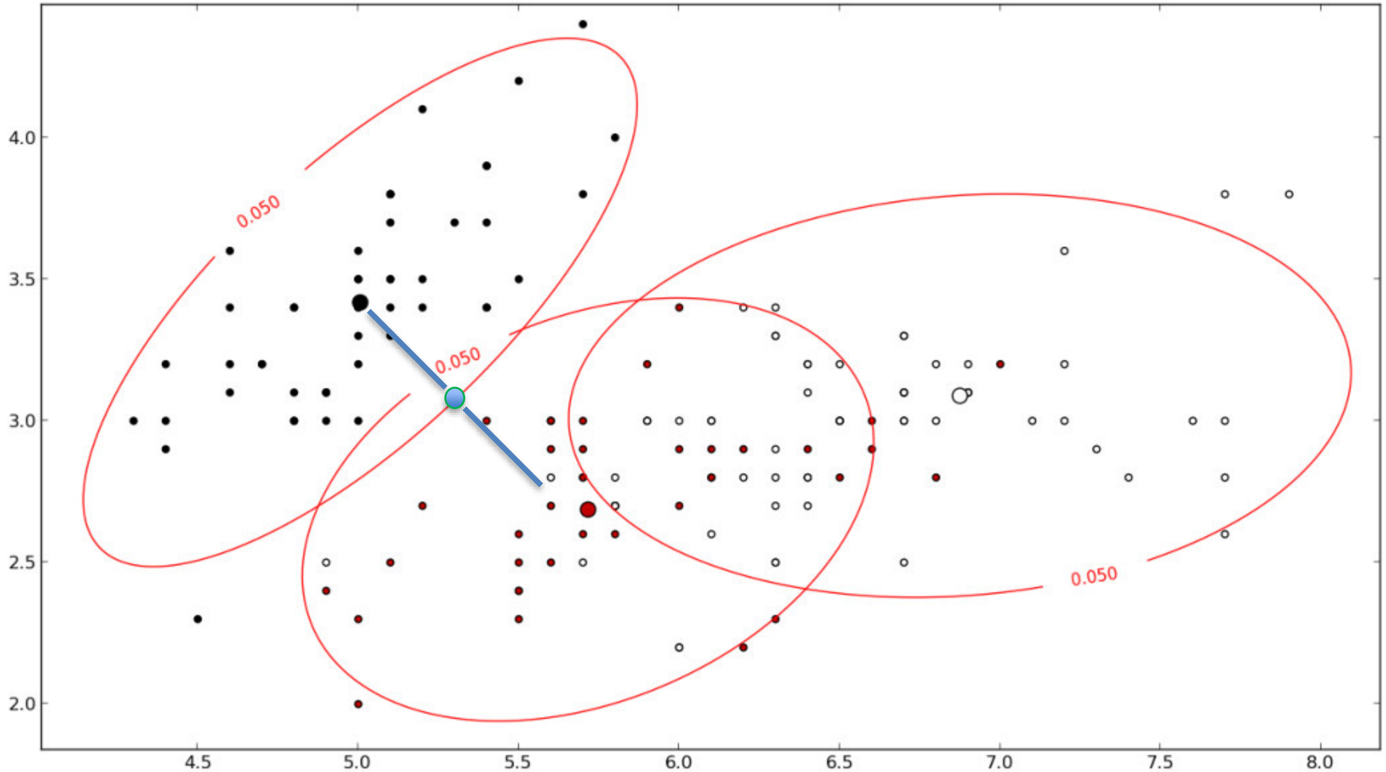
Mahalanobis Distance



http://www.jennessent.com/arcview/mahalanobis_description.htm

DM-4

Mahalanobis Distance



http://stats.stackexchange.com/questions/62092/bottom-to-top-explanation-of-the-mahalanobis-distance

Unsupervised Learning

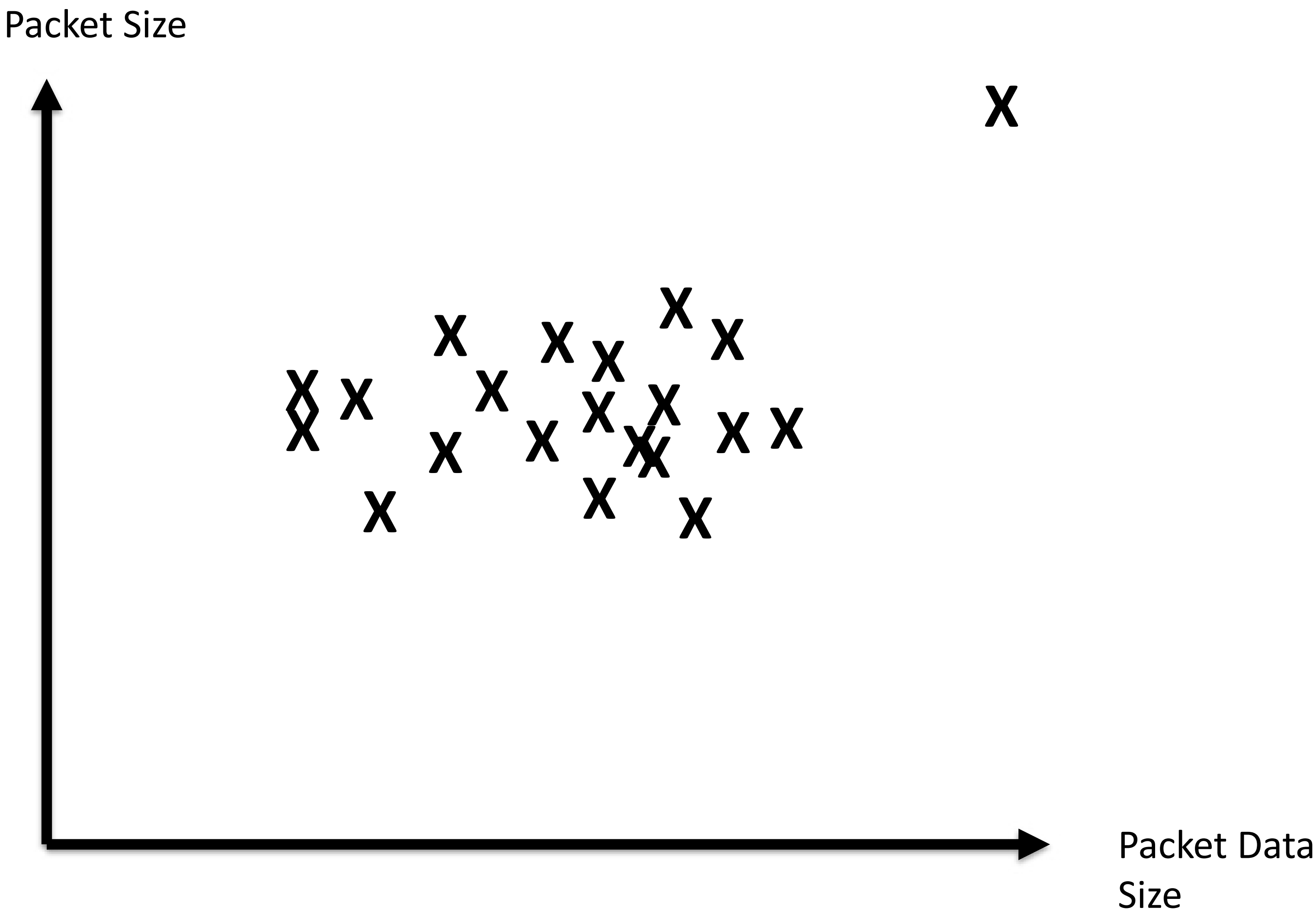
U-0

Clustering

- Partitional

- Hierarchical

Anomaly Detection with Unlabelled Data



U-C-1

Recap of Wood Classification

- 2 Optical Attributes or Features
  - Brightness
  - Grain prominence

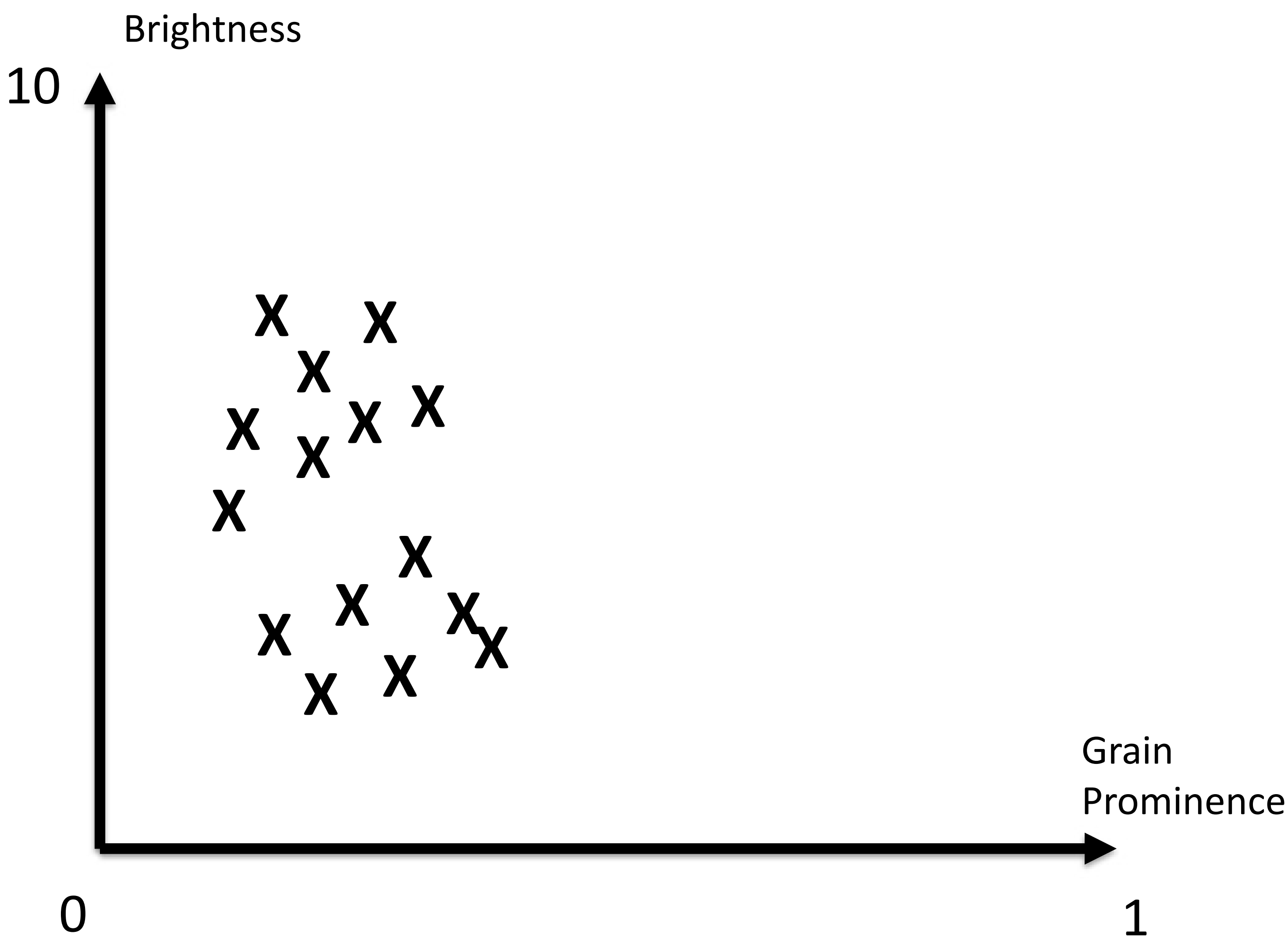
- Yielded a 2-Dimensional **Feature Space**

- We had **SUPERVISED** learning:
  - We started with known pieces of wood
  - Gave each plotted training example its class **LABEL**

- We chose our features well, we saw good ***clustering/separation*** of the different classes in the features space.

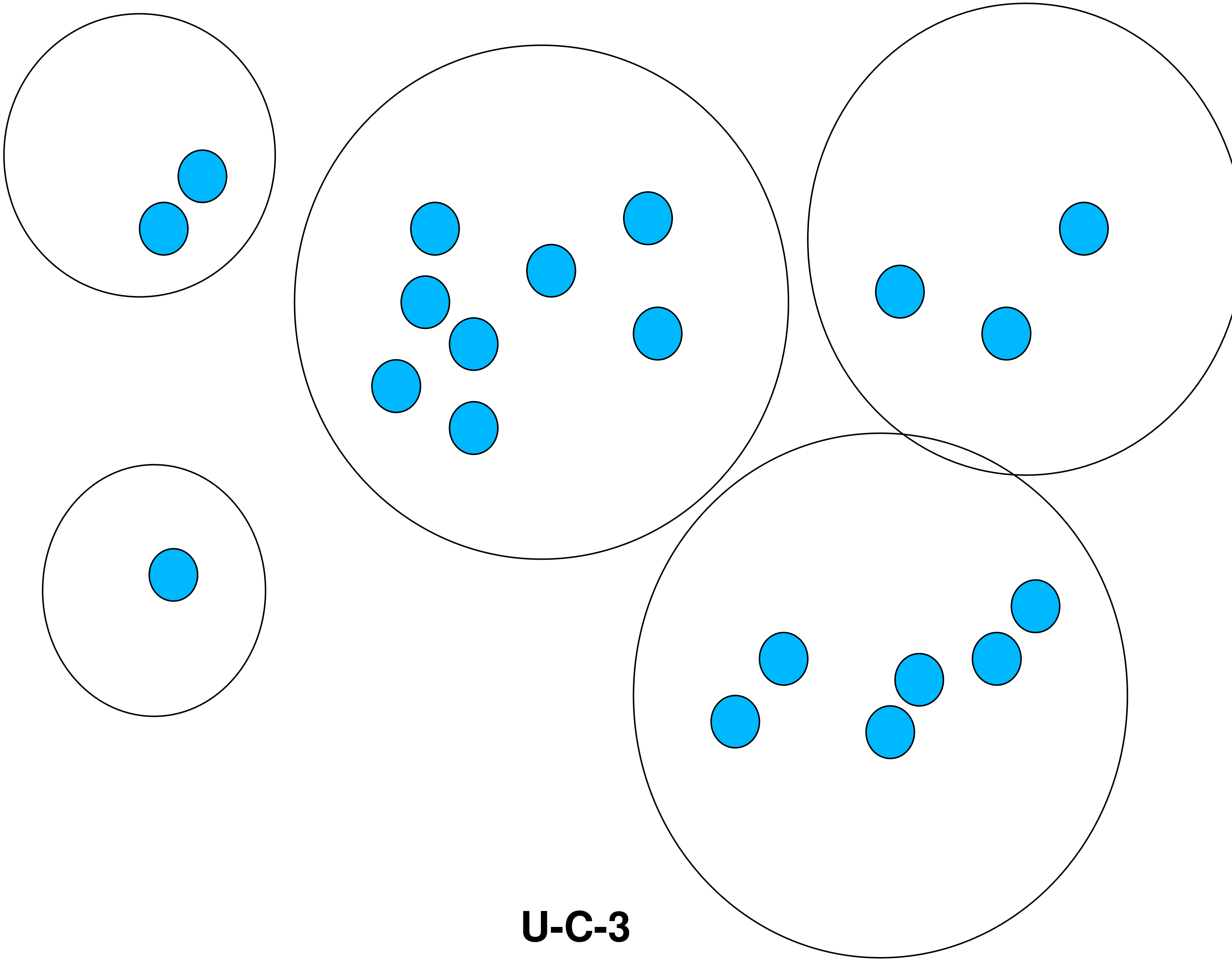
U-C-2

Unlabelled Data



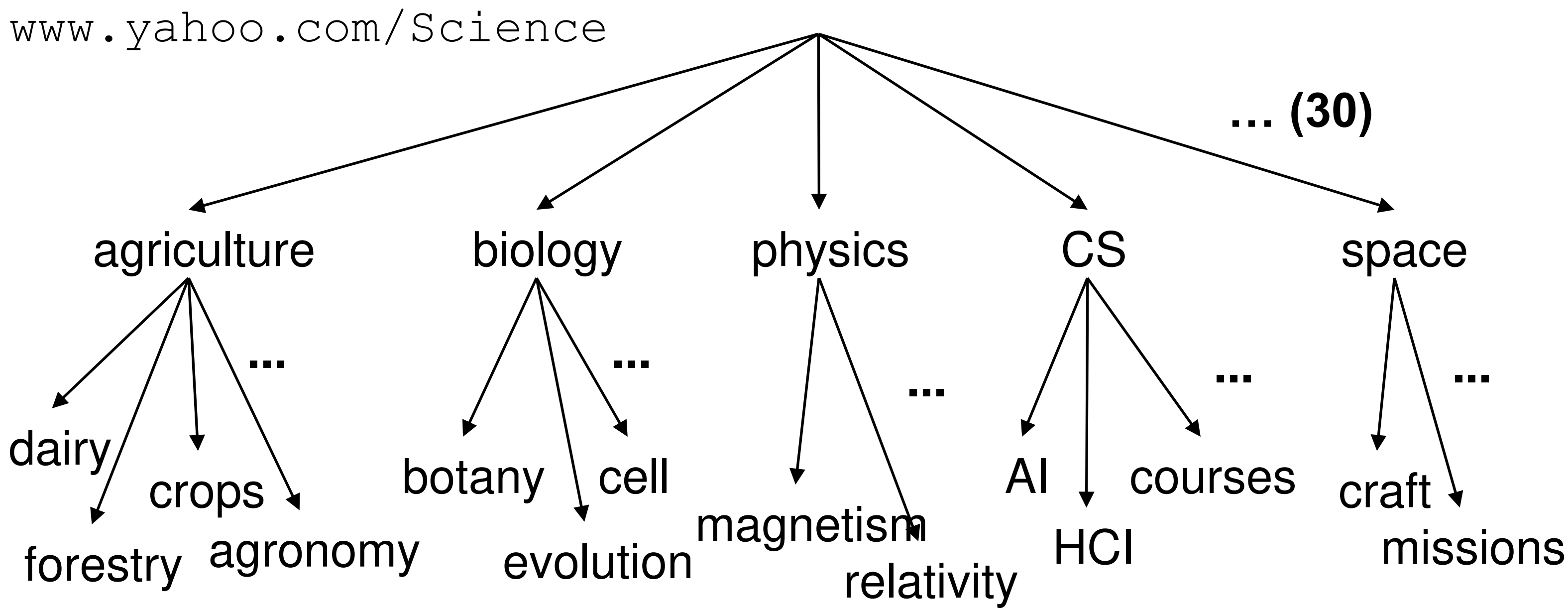
U-C-3

Partitional Clustering



U-C-3

Hierarchical Clustering: Corpus browsing



Essentials of Clustering

- Similarities
  - Natural Associations
  - Proximate*
- Differences
  - Distant*

*Implies a distance metric

U-C-3

Essentials of Clustering

- What is a “Good” Cluster?
  - Members are very “similar” to each other
    - Within Cluster Divergence Metric σ_i
      - Variance also works
    - Relative Cluster Sizes versus Data Spread

Partitional Clustering Methods

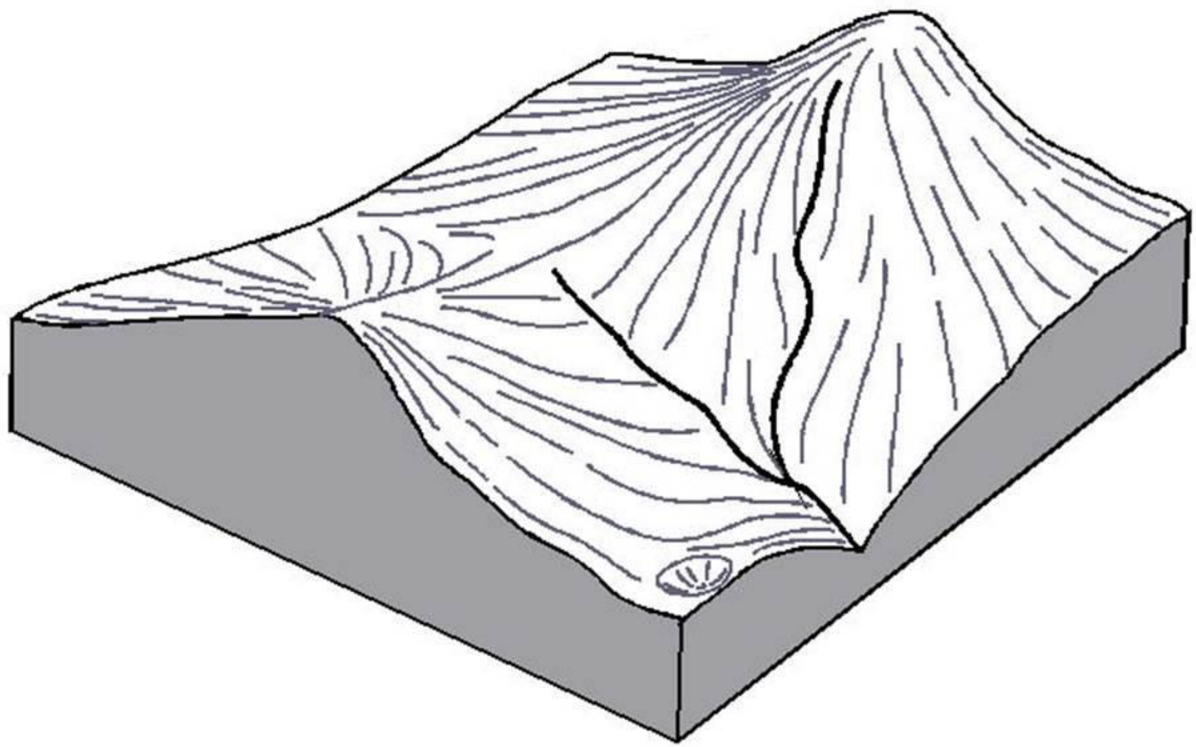
- K-Means Clustering
- Gaussian Mixture Models
- Canopy Clustering
- Vector Quantization

Unsupervised Learning/Clustering

Self Organizing Maps (SOM)

U-C-7

SOMs
Topology Preserving Projections



U-C-8

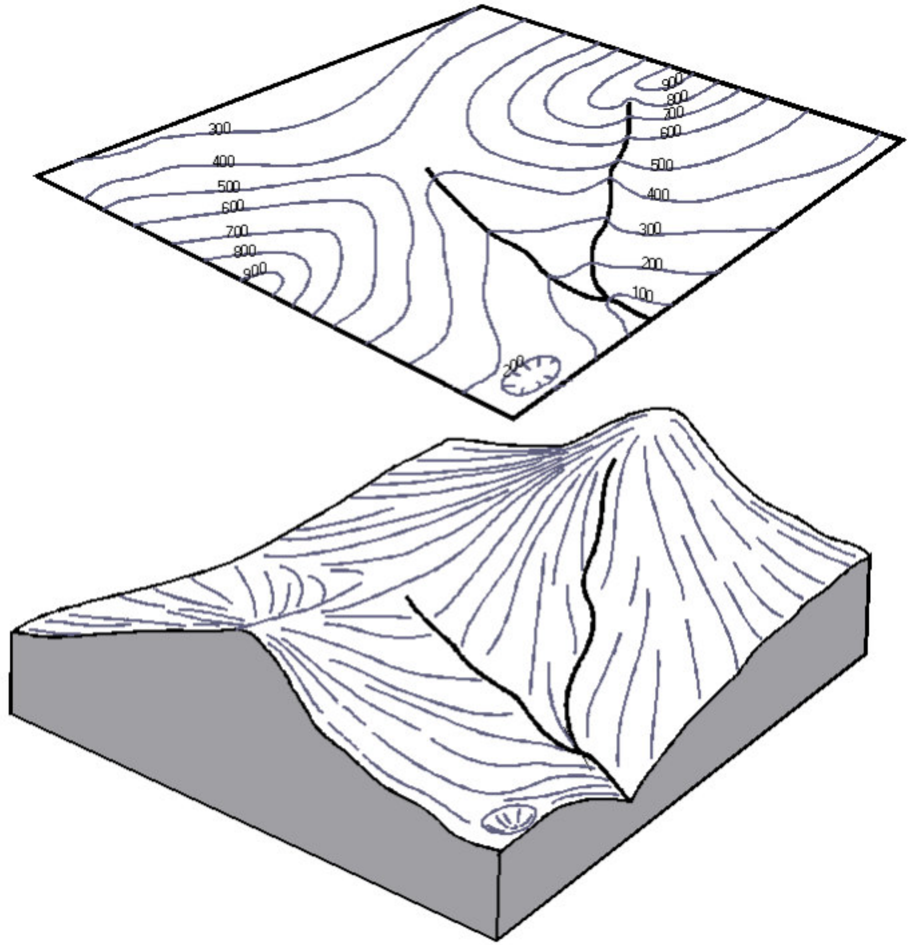
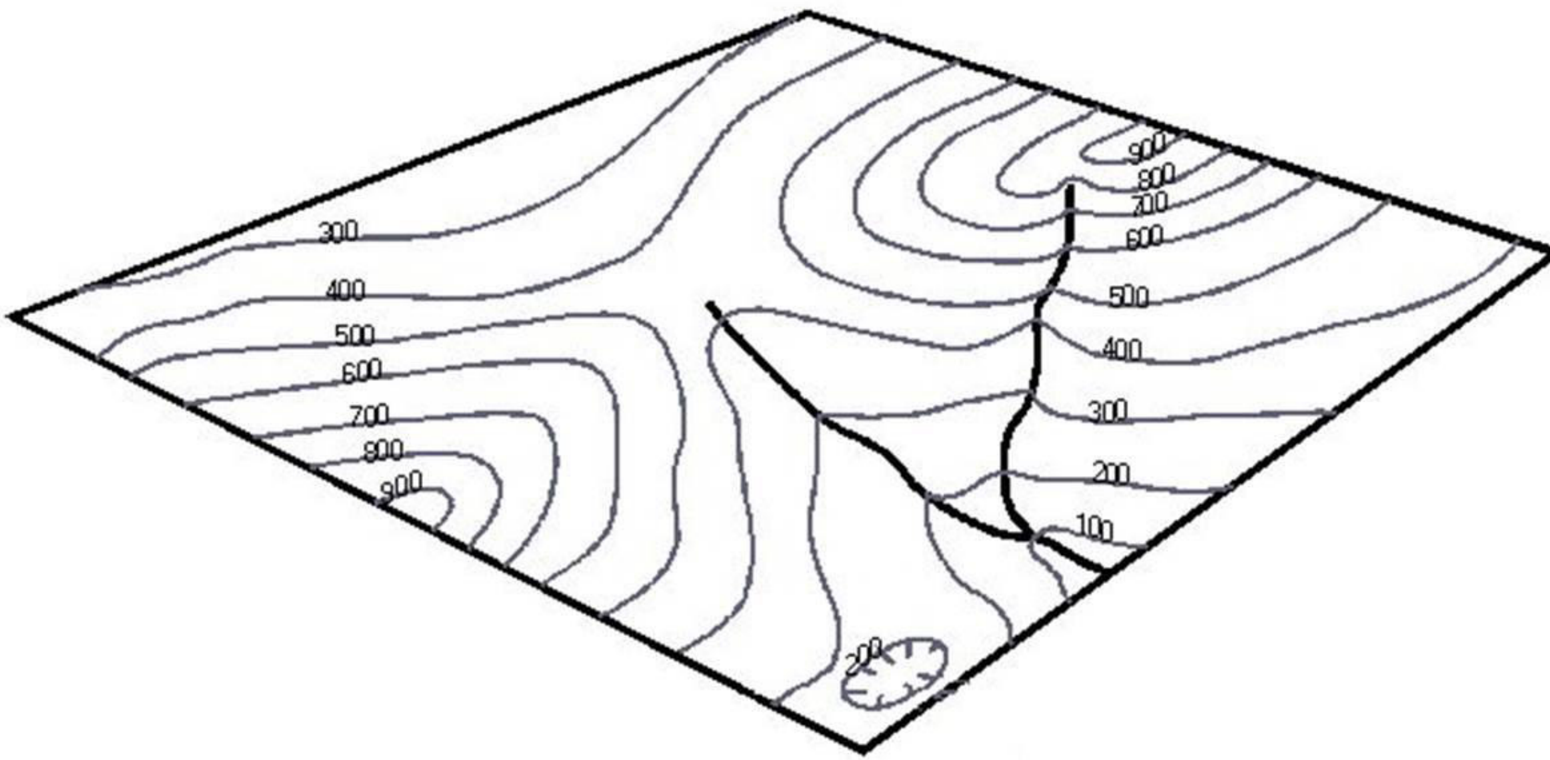


Figure 2. The relationship between a topographic map (top) and the corresponding land surface (bottom).

http://www.cita.utoronto.ca/~murray/GLG130/Exercises/F2.gif

U-C-9

Topology Preserving Projections



http://www.cita.utoronto.ca/~murray/GLG130/Exercises/F2.gif

U-C-10

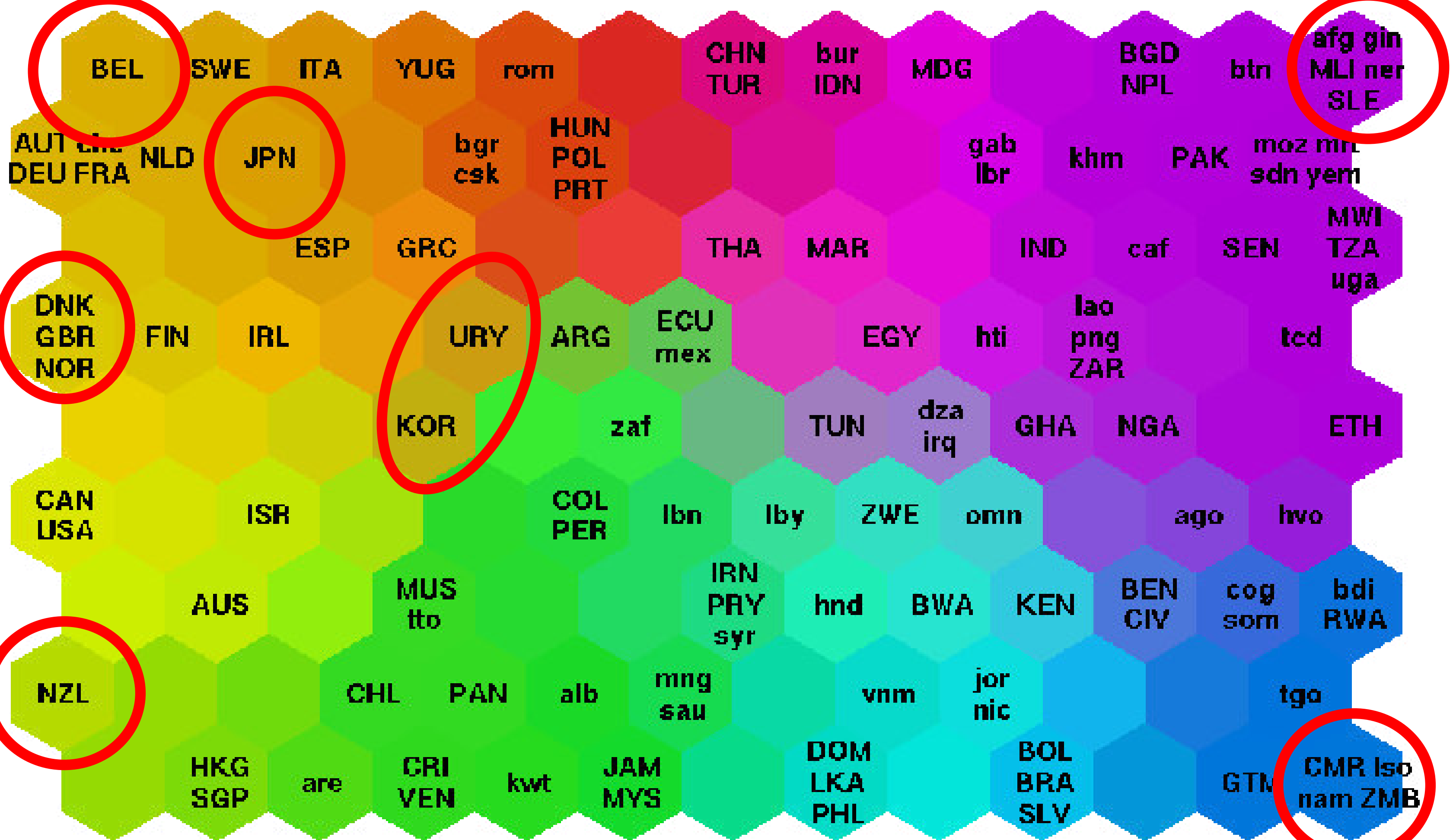
Topology Preserving Projections

- How will the distance metric handle polymorphous data?
  - Units of time (different units of time?)
    - Sprint performance data: years of age and seconds to finish
  - Units of space
    - (meters, lightyears)
    - Surface area
    - Volumetric
  - Units of mass (grams, kilograms, tonnes)
  - Units of $$$
    - NOK
    - USD

U-C-11

Proximity By Colour and Location

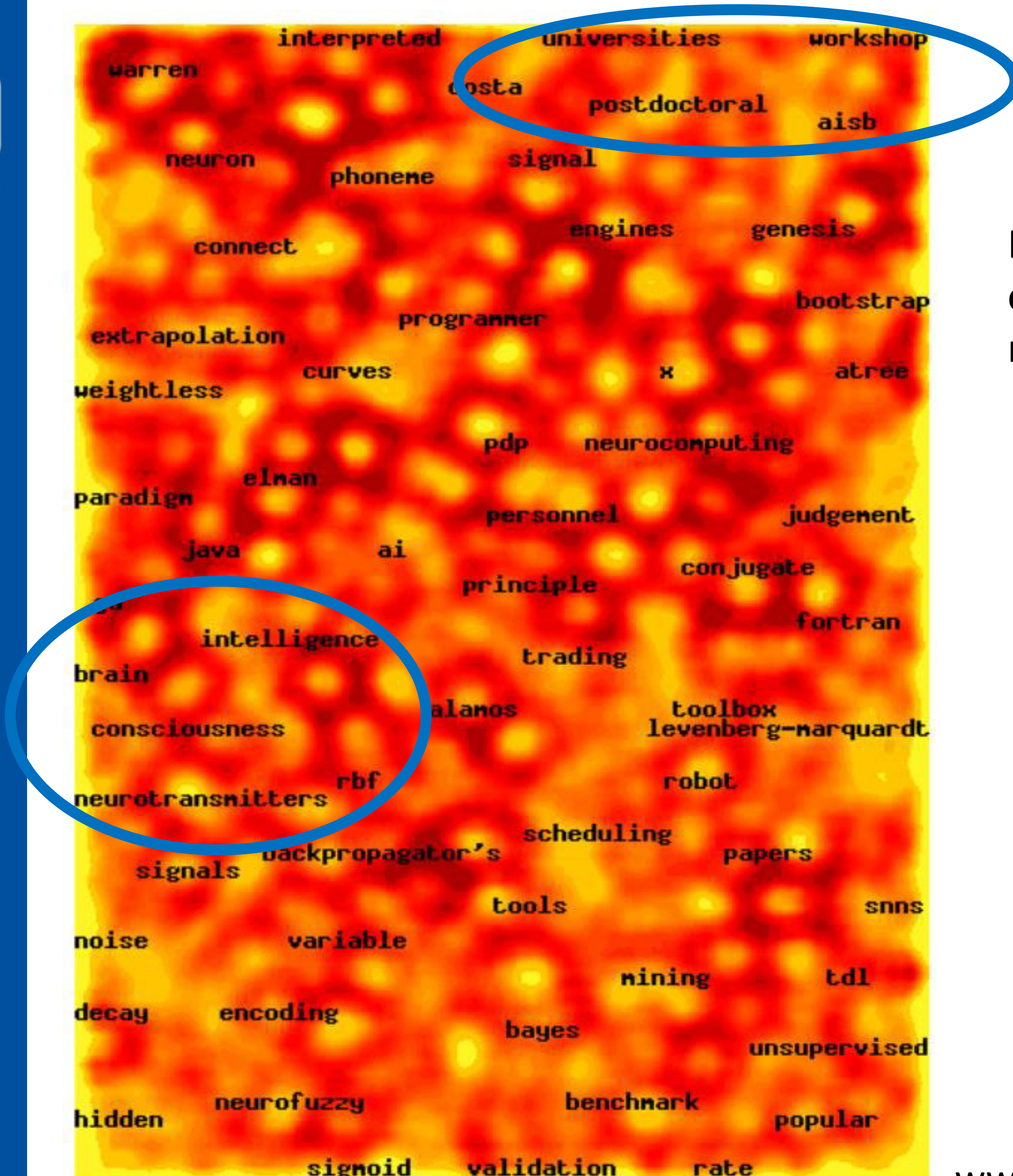
Poverty Map of the World (1997)



http://www.cis.hut.fi/research/som-research/worldmap.html

U-C-12

NTNU



Map of Labels in Titles From comp.ai.neural-nets-news newsgroup

www.cs.hmc.edu/courses/2003/fall/cs152/slides/som.pdf

U-C-13

Learning As Search

LAS-0

- Exhaustive search
  - DFS
  - BFS

- Gradient search
  - Can Get Stuck in Local Optimal Solution

- Simulated annealing
  - Avoids Local Optima

- Genetic algorithms

LAS-1

Exact vs Approximate Search

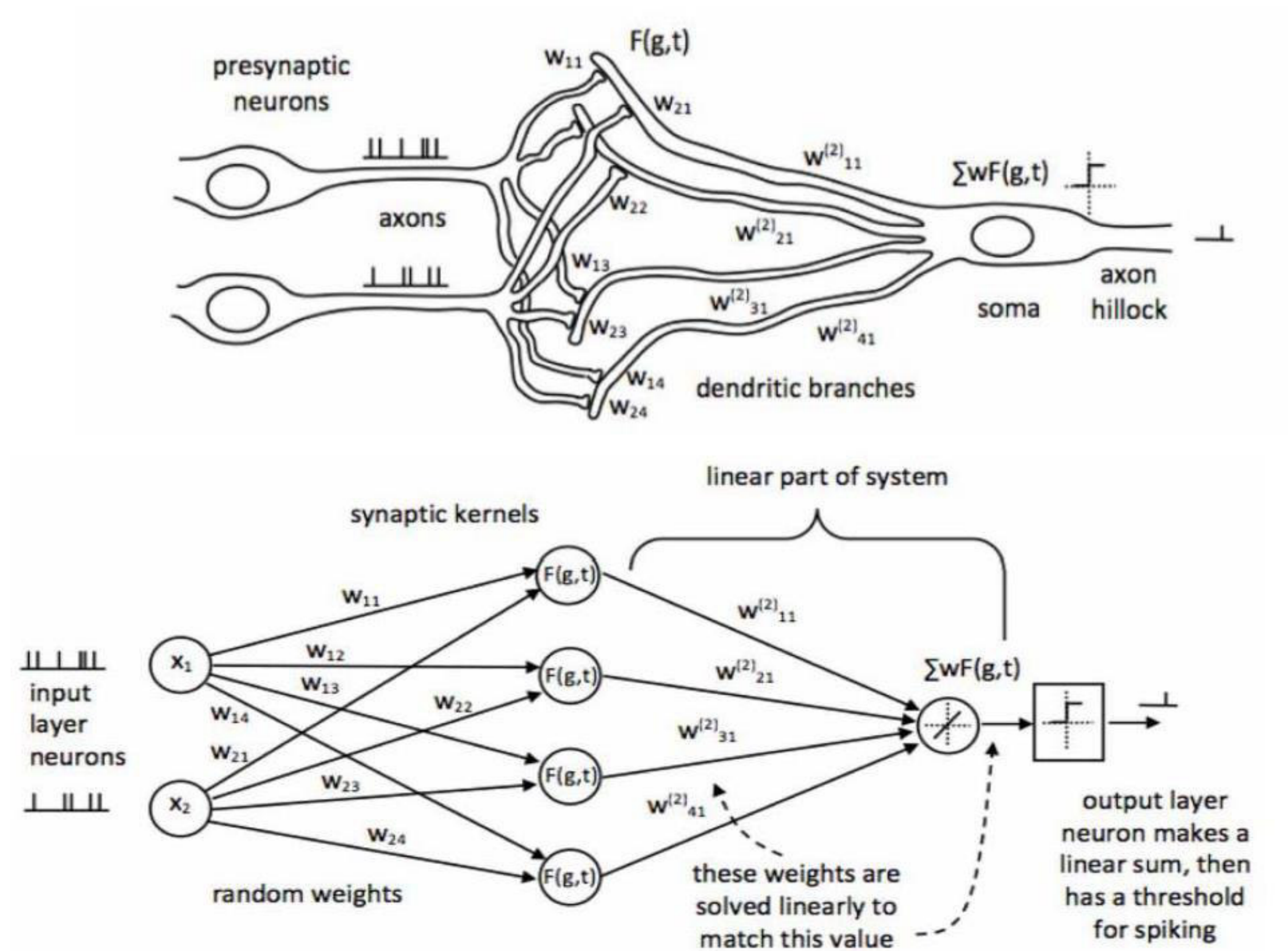
- Exact:
  - Hashing techniques
  - String matching (“Murder”)

- Approximate:
  - Approximate Hashing
  - Partial strings
  - Elastic Search
    - “murder”
    - “merder”

Artificial Neural Networks (ANN)

ANN-0

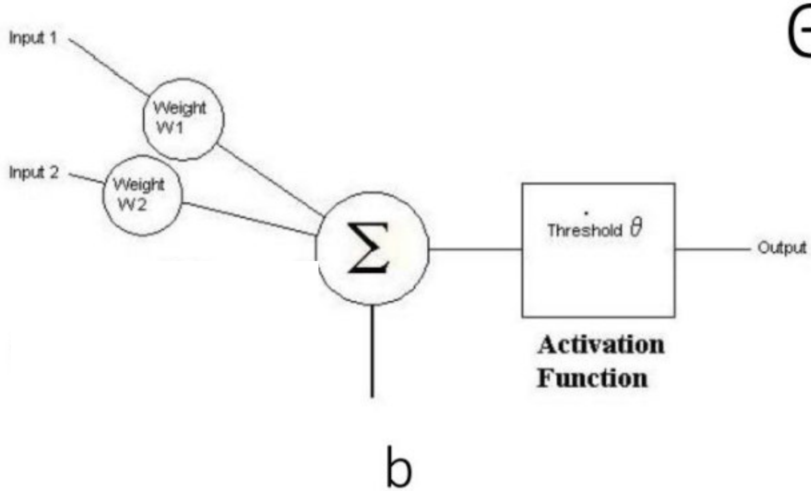
Inspired by Natural Neural Nets

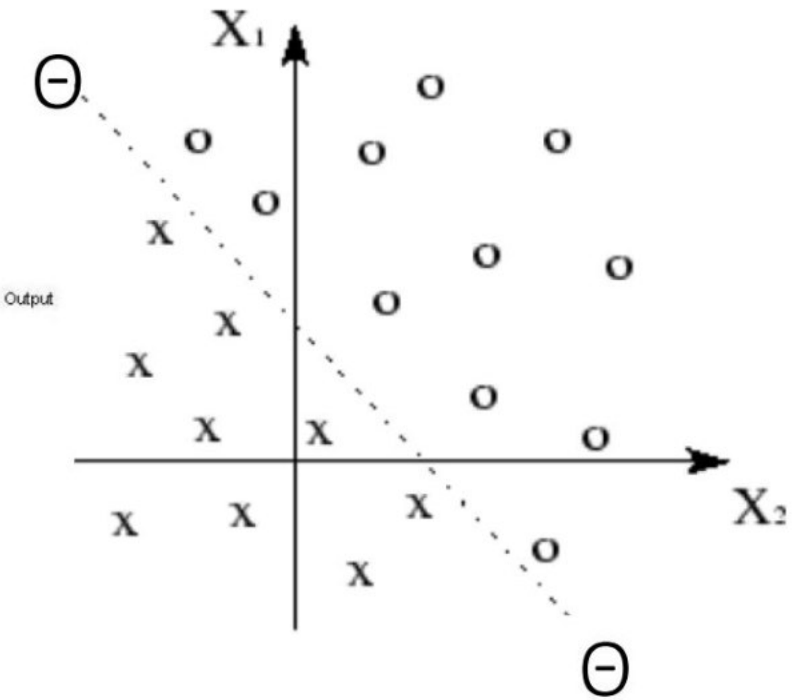


Tapson, Jonathan, et al. "Synthesis of neural networks for spatio-temporal spike pattern recognition and processing."

ANN-1

Perceptron (1950s)

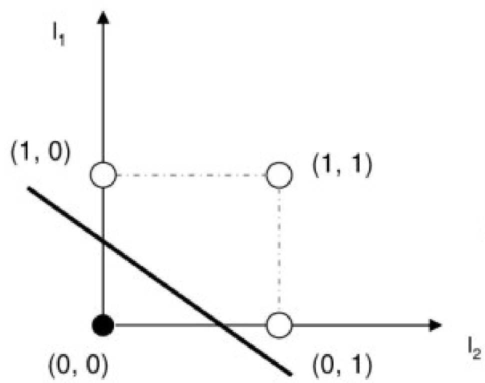




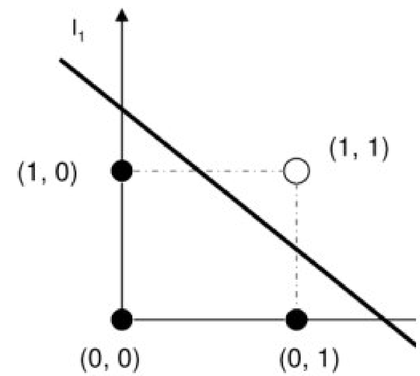
Perceptron Can Learn Simple Boolean Logic

OR & AND Decision Boundaries

OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1



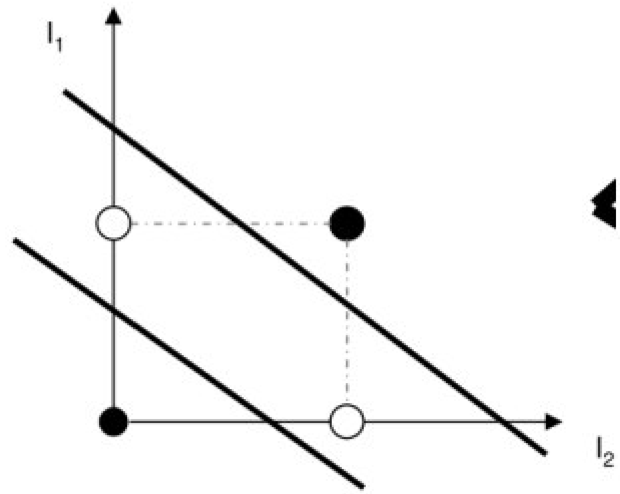
AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1



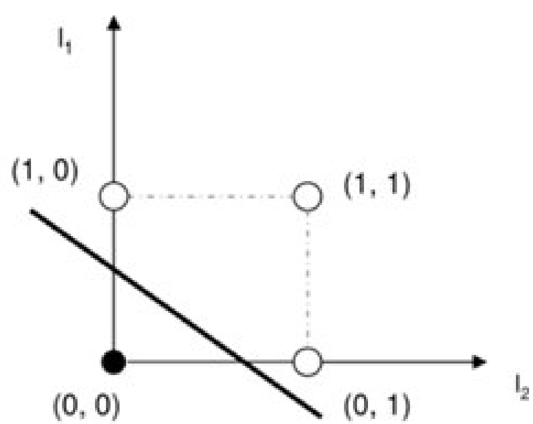
Single Boundary, Linearly Separable

Perceptron Cannot Learn XOR

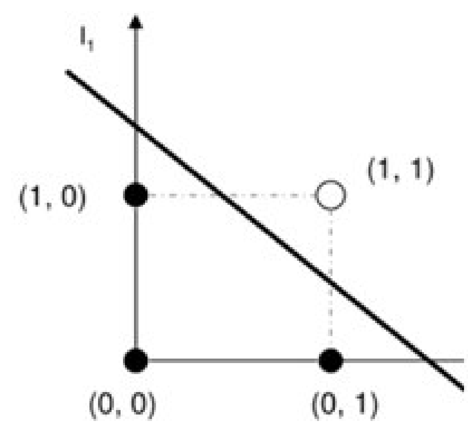
XOR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	0



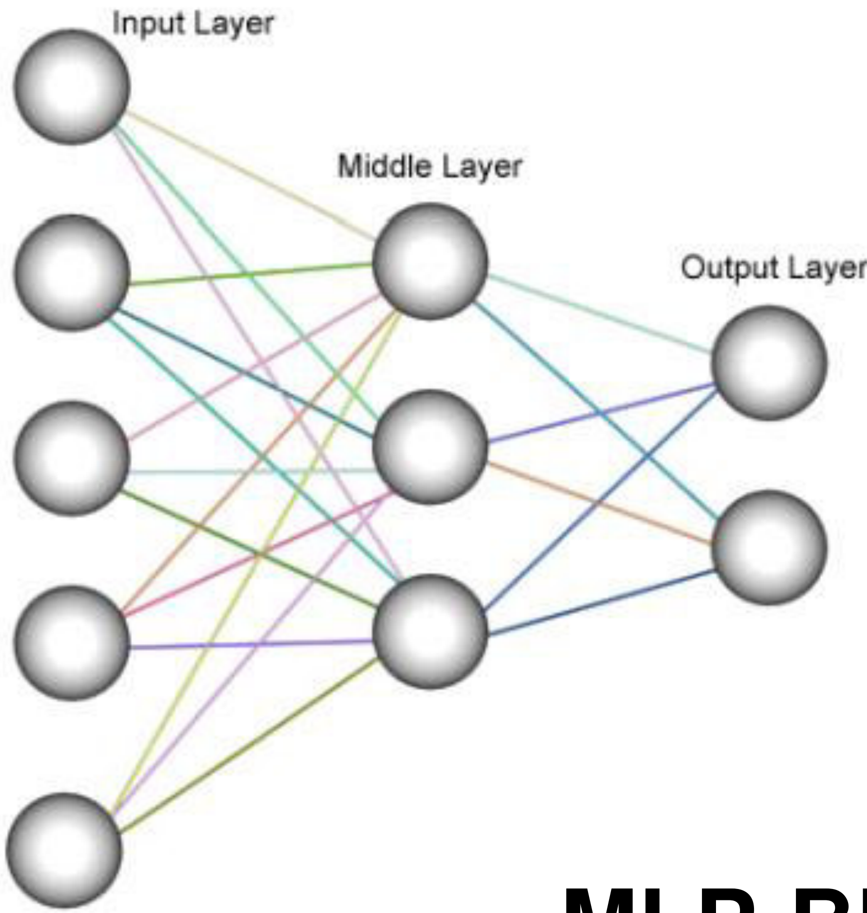
OR		
I_1	I_2	out
0	0	0
0	1	1
1	0	1
1	1	1

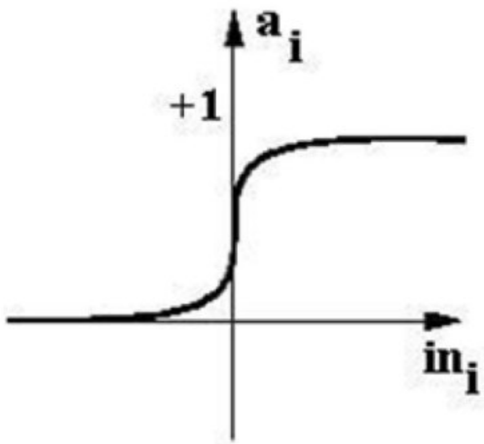


AND		
I_1	I_2	out
0	0	0
0	1	0
1	0	0
1	1	1



Multi-Layer Perceptron
Error Back-Propagation Network

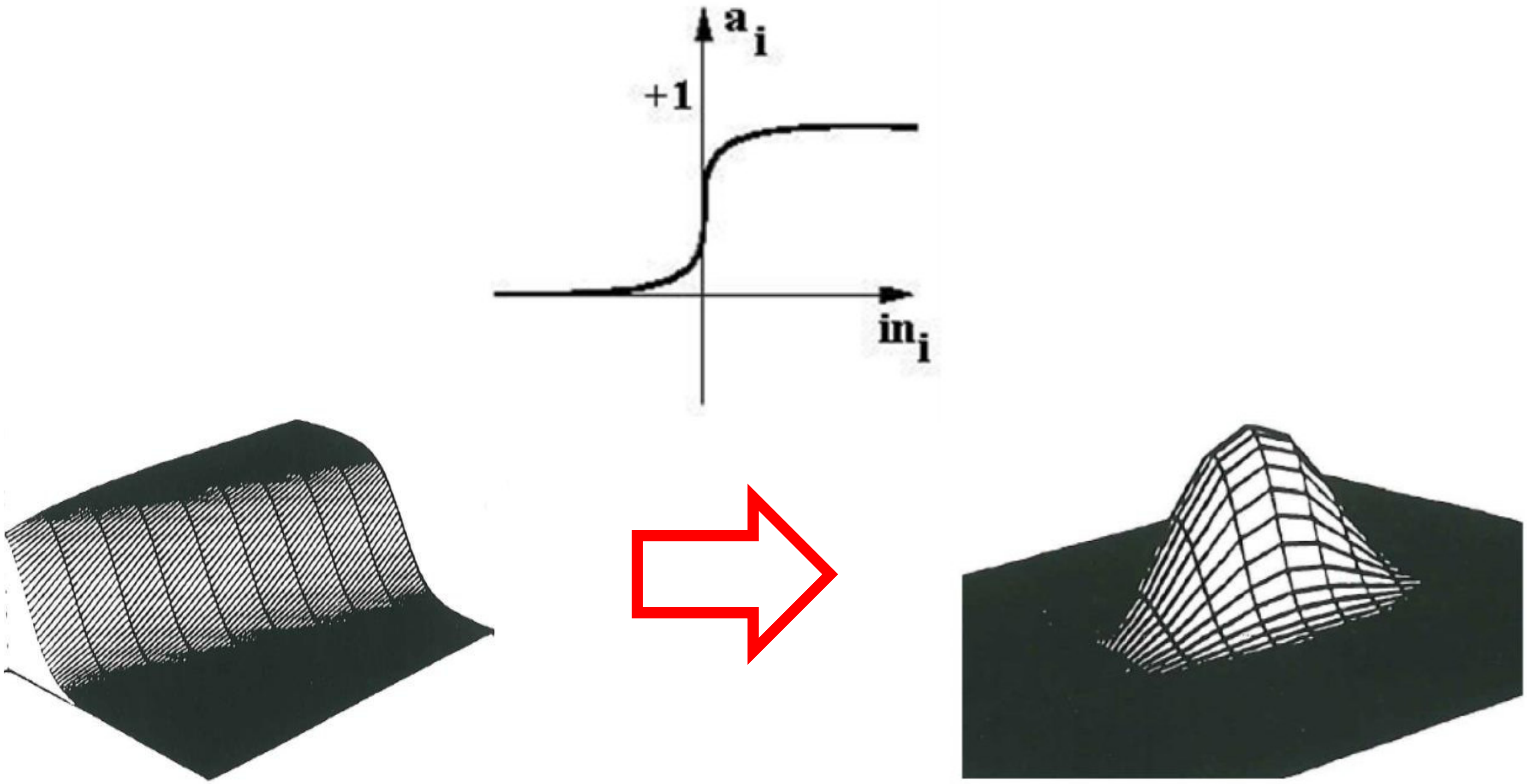




MLP-BP

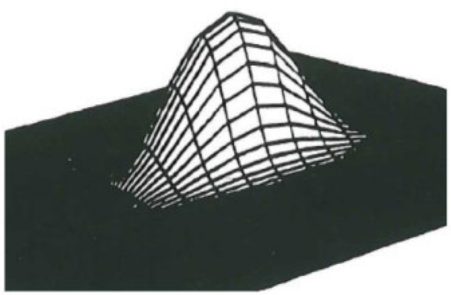
ANN-5

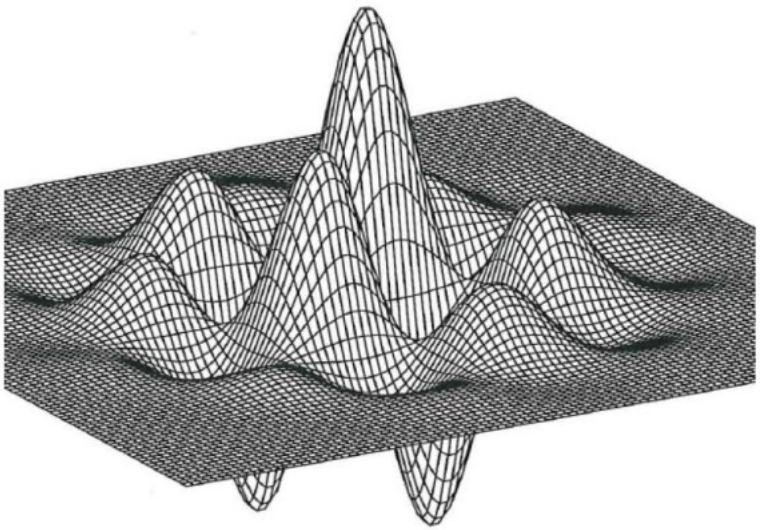
MLP-BP Internal Model Building Block



5 MLP-BP Neurons

MLP-BP "Universal Voxel"





ANN-8

NeuroFuzzy Methods

NF-0

Neuro Fuzzy Overview

- Neuro-Fuzzy (NF) is a hybrid intelligence / soft computing
  - (*Soft?)
- A combination of Artificial Neural NetworkS (ANN) and Fuzzy Logic (FL)
- Opposite of fuzzy logic is
  - Crisp
  - Sharp
- ANN are black box statistics, modelled to simulate the activity of biological neurons
- FL extracts human-explainable linguistic fuzzy rules
- Applications in Decision Support Systems and Expert Systems

NF-1

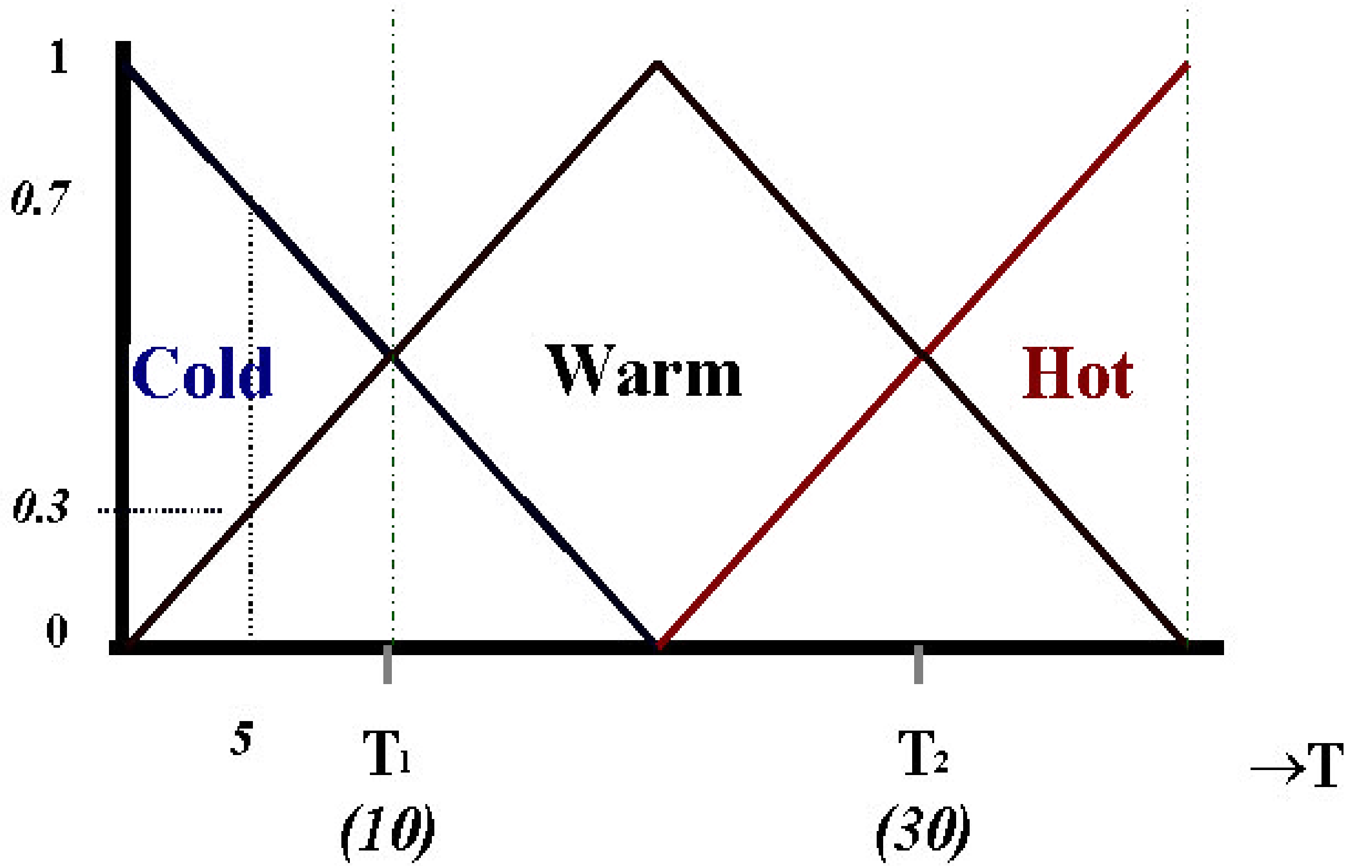
Fuzzy Basics

- FL uses **linguistic variables** that can contains several **linguistic terms**
  - Temperature (linguistic variable)
    - Hot (linguistic terms)
    - Warm
    - Cold

  - Consistency (linguistic variable)
    - Watery (linguistic terms)
    - Gooey
    - Soft
    - Firm
    - Hard
    - Crunchy
    - Crispy

NF-2

Triangular Fuzzy Membership Functions



http://sci2s.ugr.es/keel/links.php

Fuzzy Inference

- Sharp antecedent: “If the tomato is red, then it is sweet”

  - Fuzzy antecedent:
    - “If the piece of wood is more or less dark ($\mu_{dark} = 0.7$)”

  - Fuzzy consequent(s):
    - “The piece of is more of less pine ($\mu_{pine} = 0.64$)”
    - “The piece of is more of less birch ($\mu_{birch} = 0.36$)”

http://ispac.diet.uniroma1.it/scarpiniti/files/NNs/Less9.pdf

Combining ANN/FL

- ANN black box approach requires sufficient data to find the structure (generalization learning)
  - NO PRIORS required
  - But cannot extract linguistically meaningful rules from trained ANN

- Fuzzy rules require prior knowledge
  - Based on linguistically meaningful rules

http://www.scholarpedia.org/article/Fuzzy_neural_network

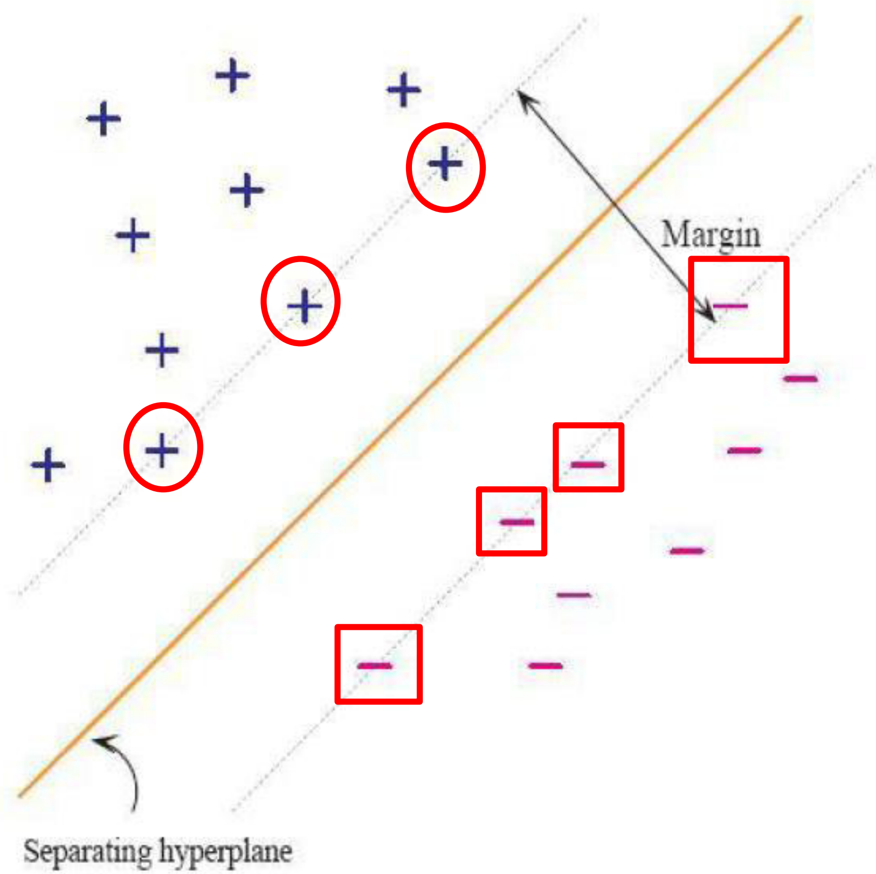
NF-5

Combining ANN/FL

- Combining the two gives us higher level of system intelligence
  - Intelligence(?)
- Can handle the usual ML tasks
  - (regression, classification, etc)

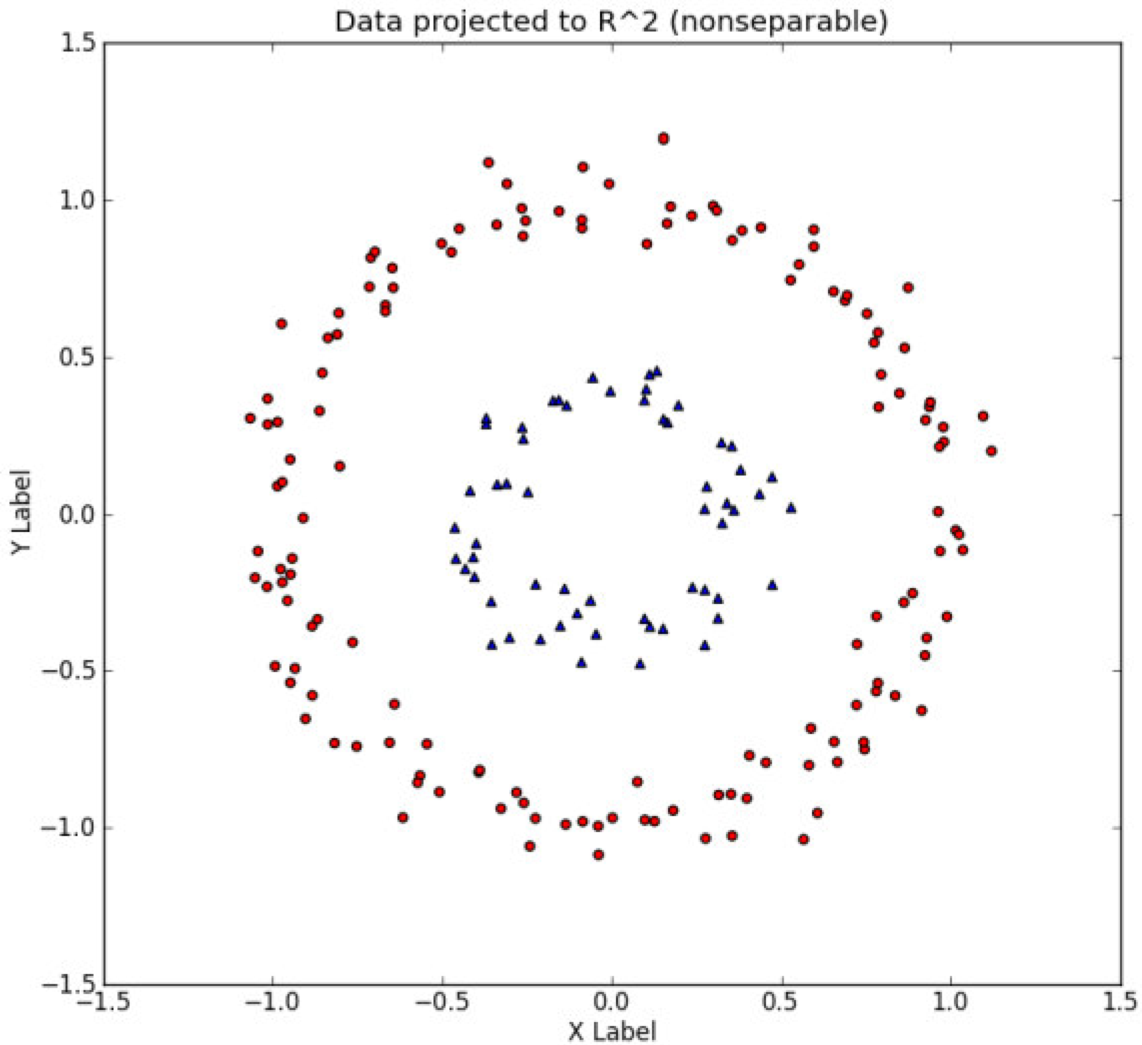
http://www.scholarpedia.org/article/Fuzzy_neural_network

Support Vector Machines



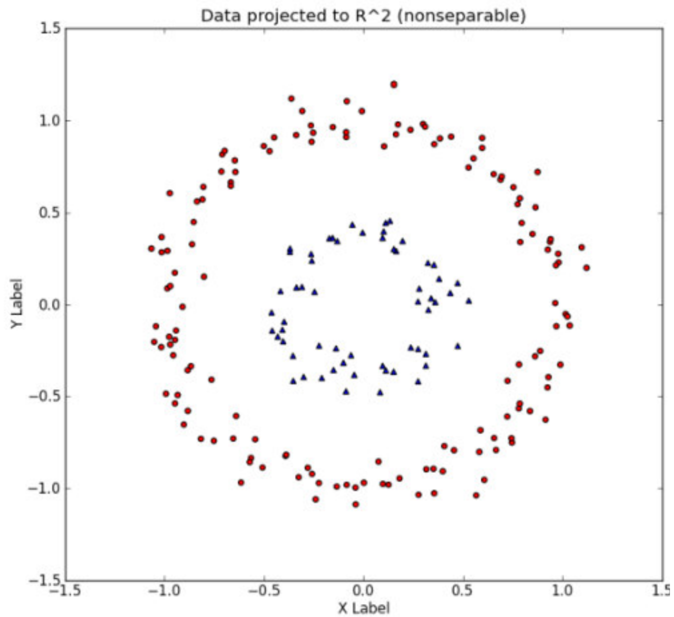
SVM-1

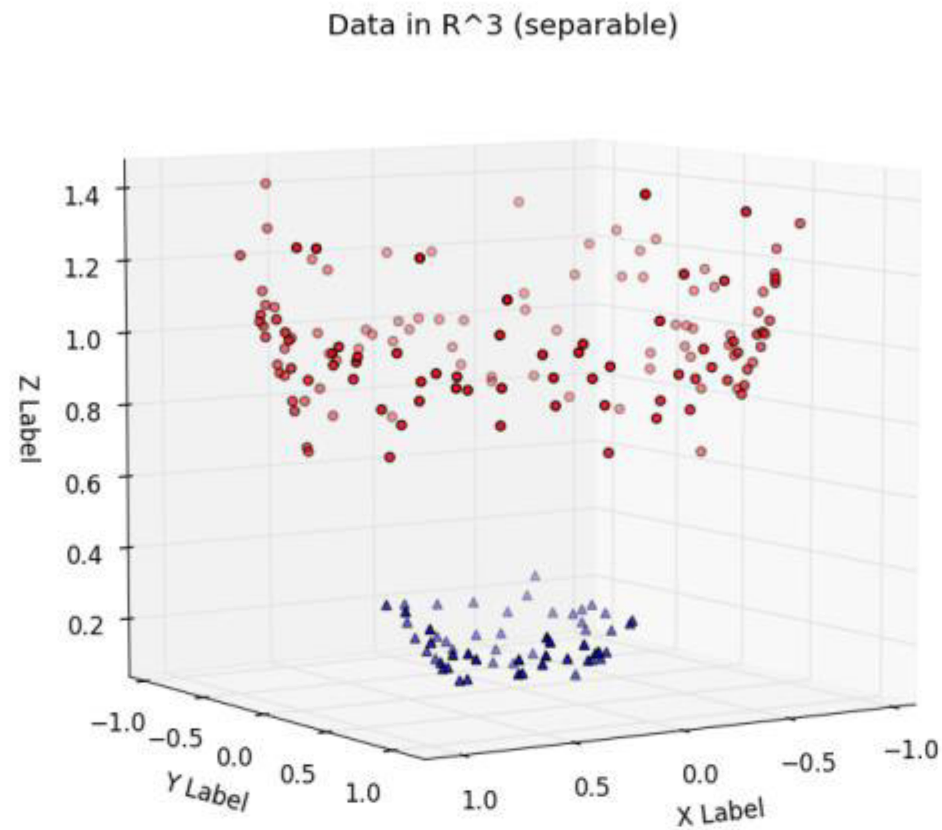
This Feature Space Isn’t Linearly Separable



SVM-2

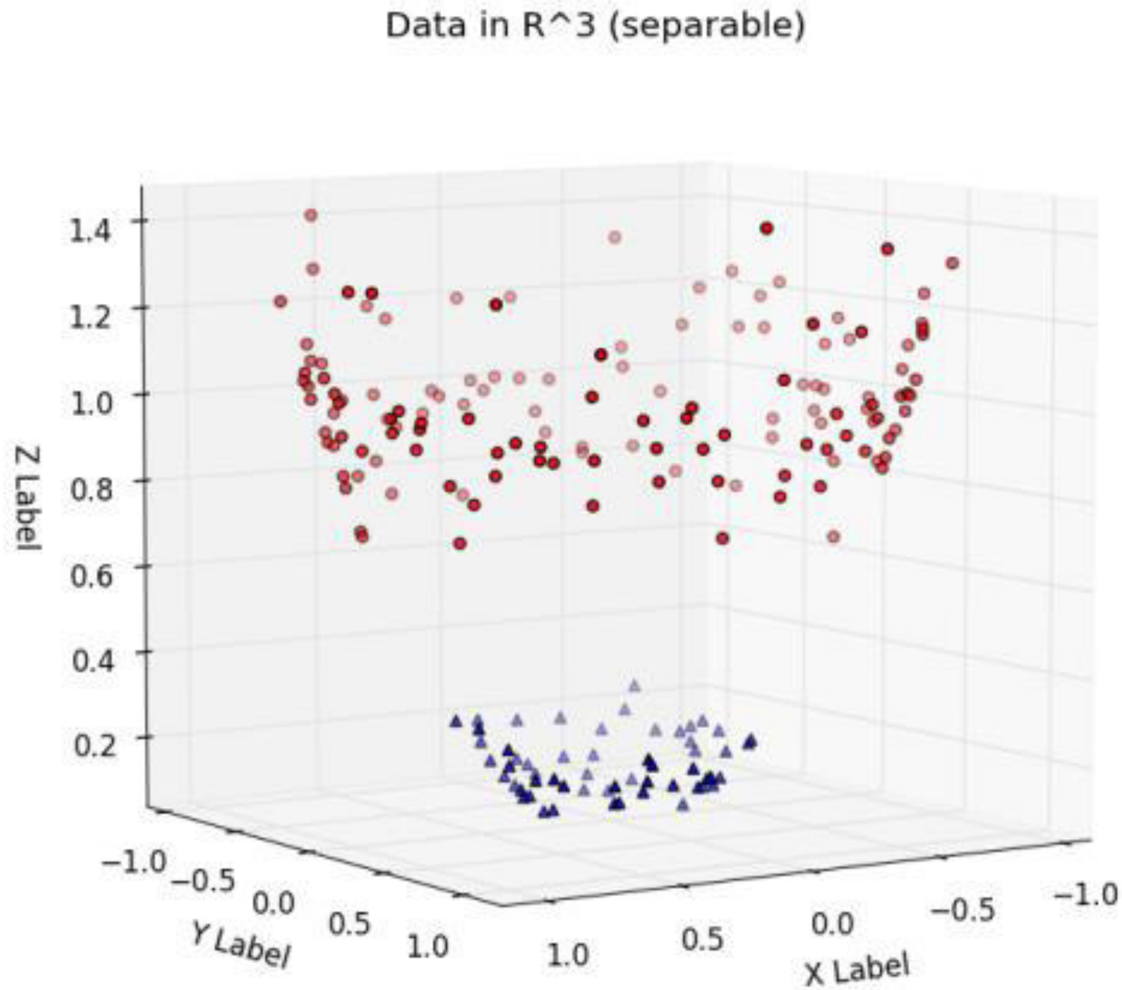
Apply the Kernel Trick!





SVM-3

Perhaps a Different Feature Space?



SVM-4

Another Type of Learning

- Supervised Learning
  - Labelled Data
- Unsupervised Learning
  - Unlabelled Data
- Reinforcement Learning
  - Situational Signals from Environment

RL-1

Reinforcement Learning

- The learner/agent is not told which actions to take
- Correct ***action models*** are reinforced with a reward signal
- May also be a penalty signal
  - Eg: actions that use battery power
- Learner/agent must discover which actions yield the most reward
- learner/agent interacts with environment and uses **trial and error**

RL-2

Exploration and Exploitation

- To obtain a reward, a reinforcement learning agent must prefer actions that it has tried in the past and found to be effective in producing reward.
  - But to discover such actions, it has to try actions that it has not selected before.
  - The agent has to **exploit what it already knows** in order to obtain reward
  - But it also has to **explore what it doesn't know** order to make better action selections in the future.
  - RL systems can learn to forgo an immediate reward in favour of maximizing total reward over long term.

    Exploitation versus exploration

Ensemble Approaches

- Basic idea:

  Build different “experts”, and let them vote

Why do they work?

- Suppose there are 25 base classifiers
- Each classifier has error rate, ε = 0.35 (35%)
- Assume independence among classifiers
- Probability that the ensemble classifier makes a wrong prediction
  - (13 out of 25 get it wrong):

$$\sum_{i=13}^{25} \binom{25}{i} \varepsilon^{\mathrm{i}} (1-\varepsilon)^{25-i} = 0.06 = 6\%$$

Where We Get All These Different Data Sets

Generating “new” datasets by “Bootstrapping”

- sample N items with replacement from the original N

x_1	x_2	x_3	x_4	x_5	y
187	80	120	30	4.5	0
160	70	119	36	5.6	0
150	80	185	60	8.8	1
192	92	140	50	6.8	1
168	110	155	45	7.8	1

N = 4

187	80	120	30	4.5	0
150	80	185	60	8.8	1
150	80	185	60	8.8	1
168	110	155	45	7.8	1
168	110	155	45	7.8	1

N = 3

160	70	119	36	5.6	0
160	70	119	36	5.6	0
150	80	185	60	8.8	1
192	92	140	50	6.8	1
168	110	155	45	7.8	1

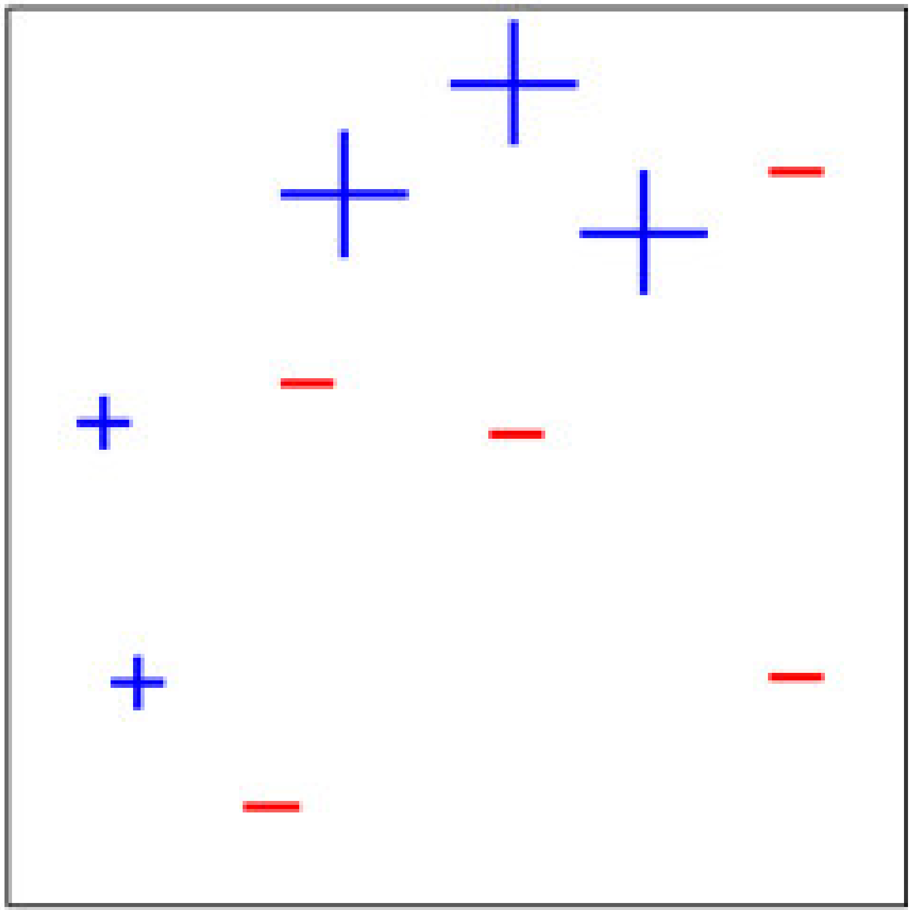
EA-3

“Bagging”

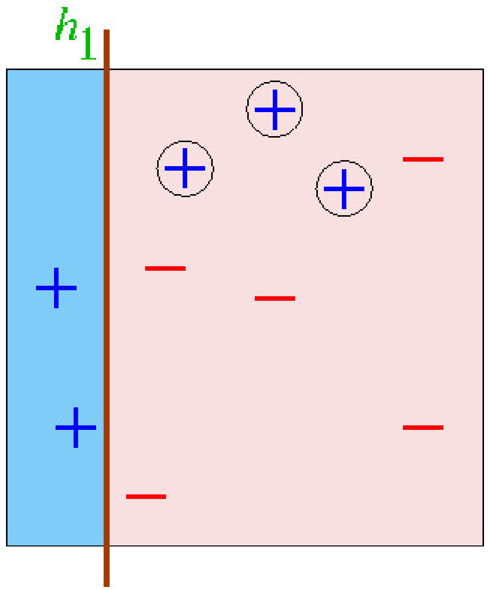
- Multiple ML/Classification Algorithms
  - Ensemble Aggregation
- Need Multiple Training/Testing Data Sets
  - Bootstrapping

Bootstrapping + Aggregating = Bagging

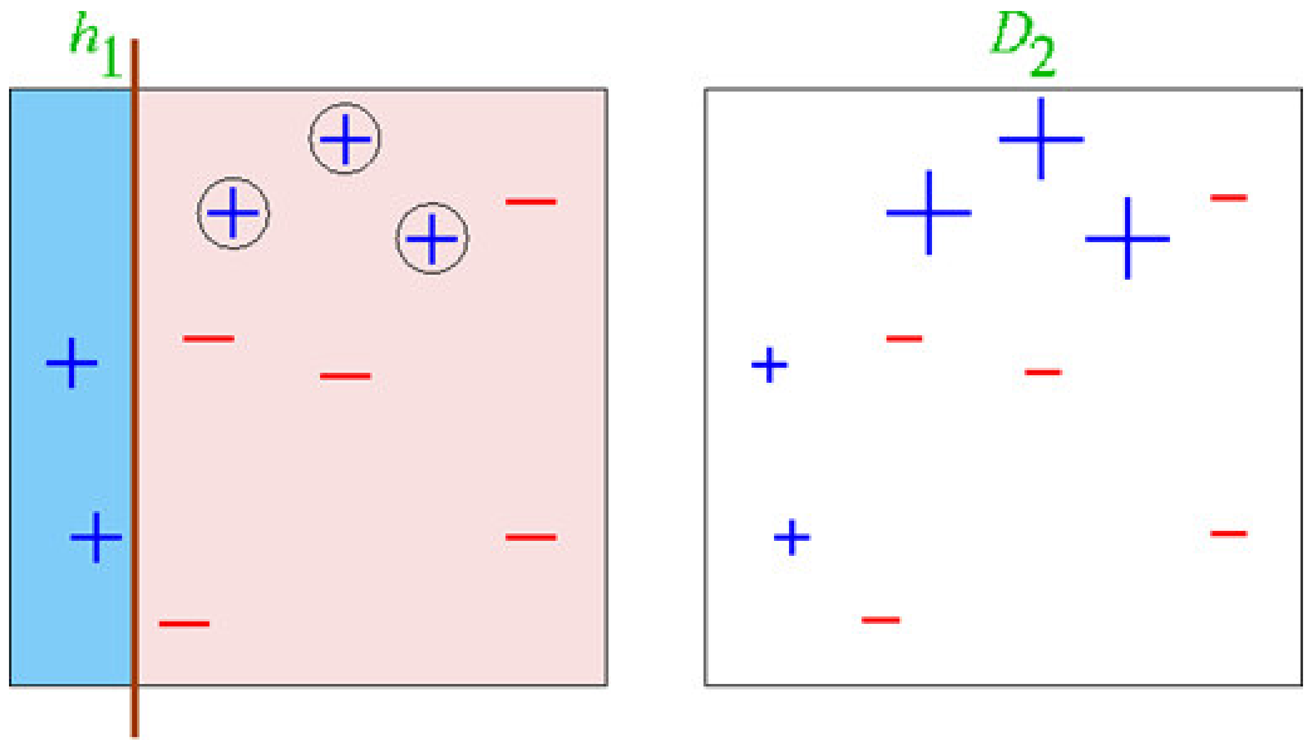
A Difficult Classification Problem



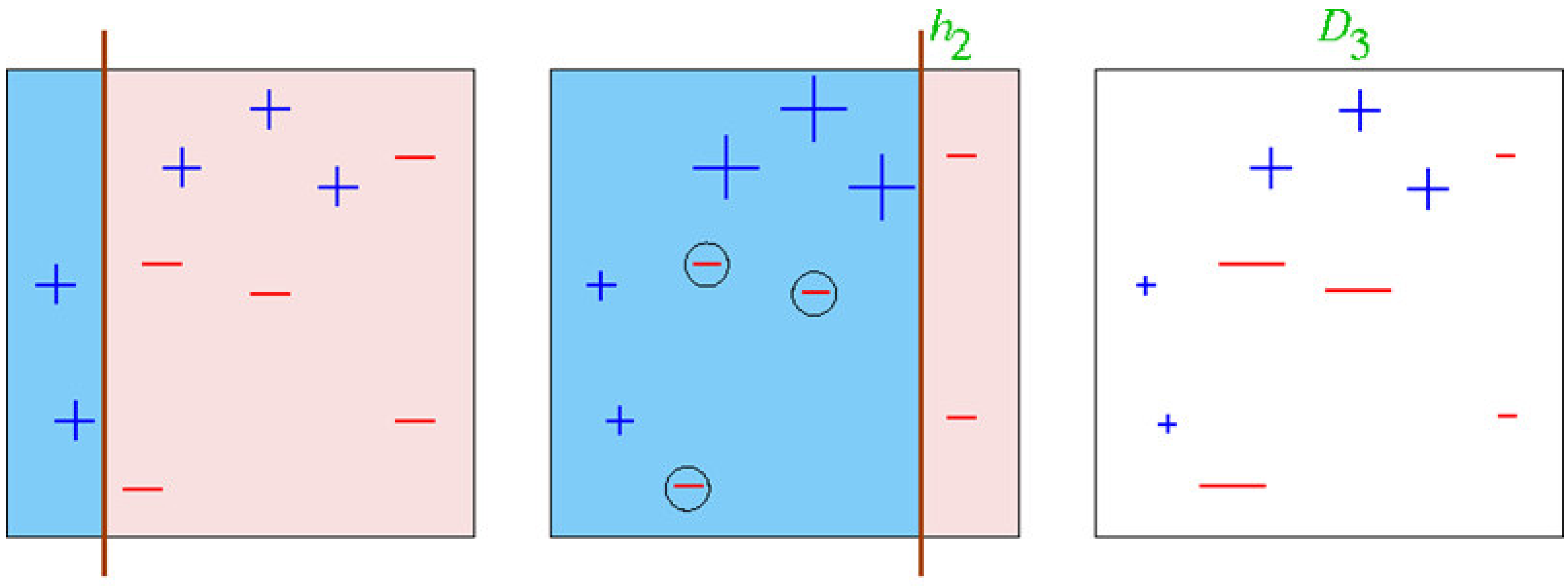
First classifier



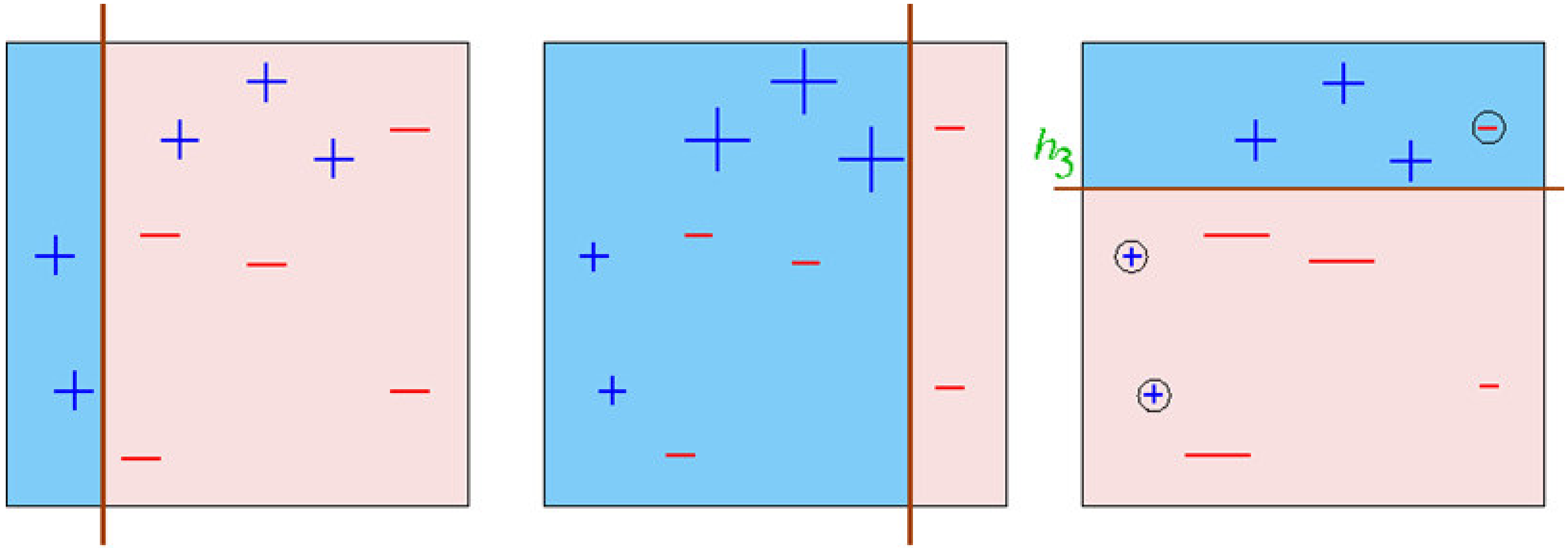
Next classifier Focuses on Data Partition D_2



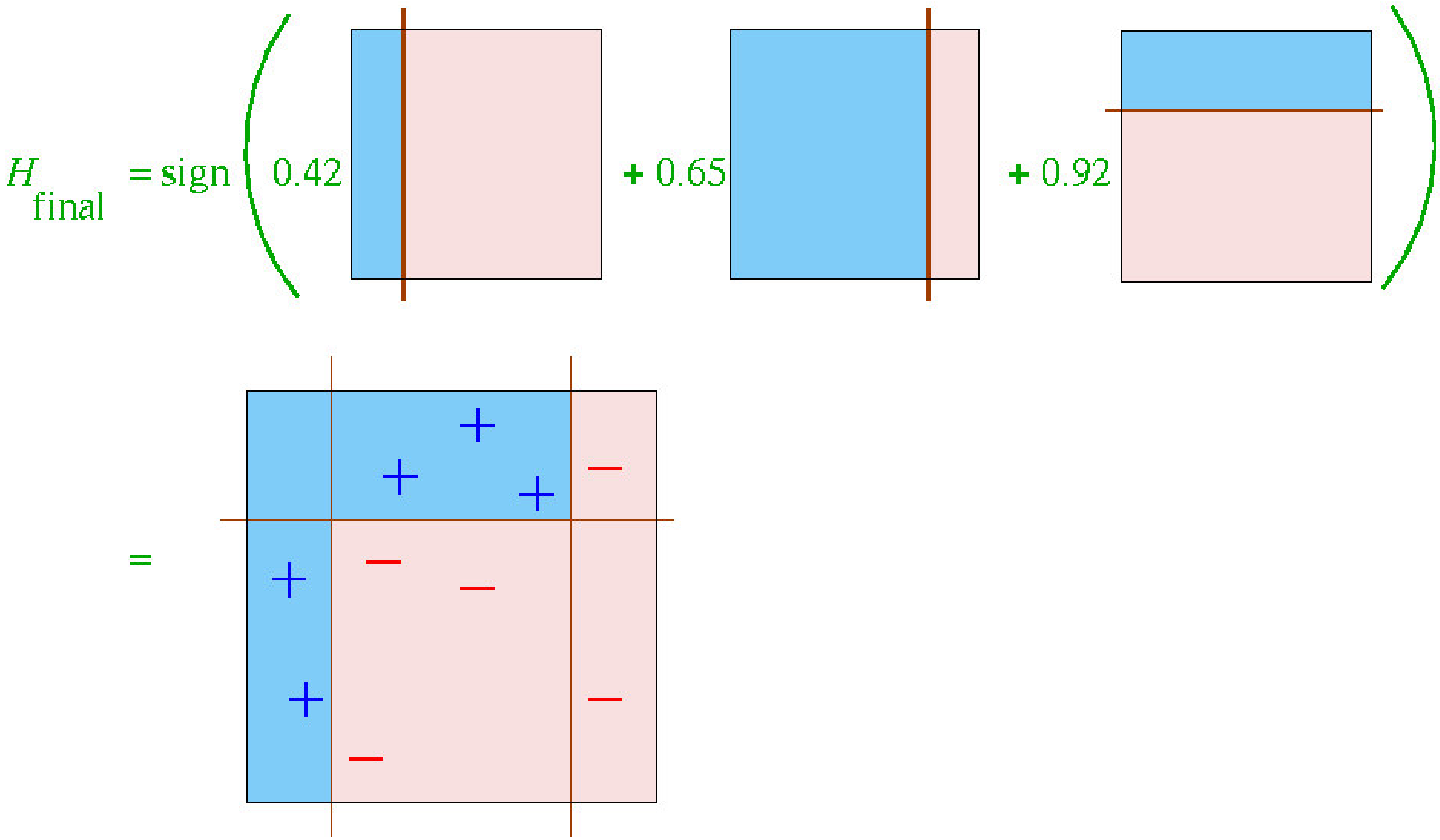
Next classifier Focuses on Data Partition D_3



Result is 3 Separate Classifiers



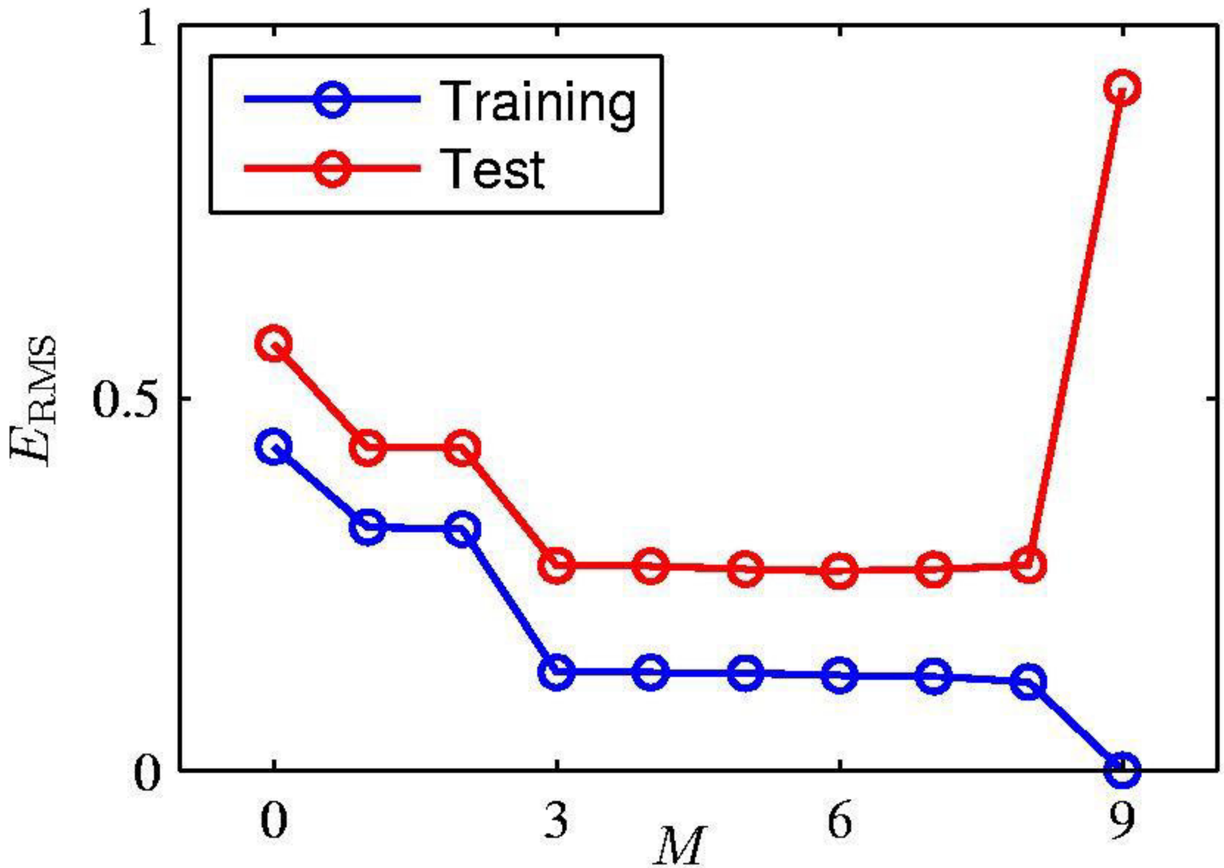
Final Classifier learned by Boosting



Performance Evaluation

PE-0

Training and Testing Performance



PE-1

Classifier Performance Evaluation: Testing Data

- Not all of the data is used to find the best fit
- Some of the data is held back, to test the fit
- A good model with sufficient data will learn to "generalize"
  - It will converge on the hidden structure in the data
  - If the data contains a good representation of the system under study (by implication, the structure in the system)

PE-2

Classifier Evaluation Metrics: Precision and Recall

- **Precision**: exactness – what % of tuples that the classifier labeled as positive are actually positive

$$precision = \frac{TP}{TP + FP}$$

- **Recall:** completeness – what % of positive tuples did the classifier label as positive?

$$recall = \frac{TP}{TP + \boxed{FN}}$$

Should have been positives

- Perfect score is 1.0
- Inverse relationship between precision & recall

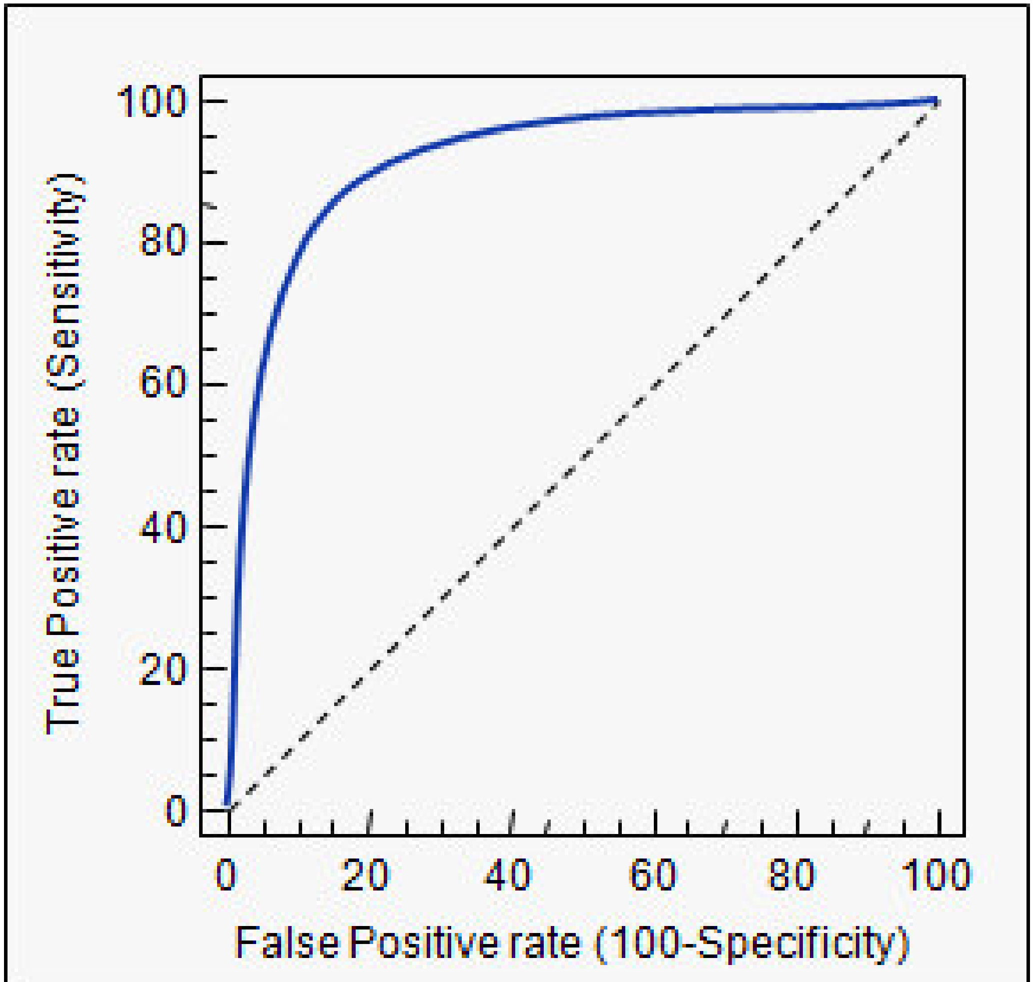
Classifier Evaluation Metrics: Confusion Matrix

Actual class\Predicted class	C_1	$\neg C_1$
C_1	**True Positives (TP)**	**False Negatives (FN)**
$\neg C_1$	**False Positives (FP)**	**True Negatives (TN)**

Actual class\Predicted class	buy_computer = yes	buy_computer = no	Total
buy_computer = yes	**6954**	**46**	7000
buy_computer = no	**412**	**2588**	3000
Total	7366	2634	10000

PE-4

ROC Curve:
Receiver Operator Characteristic
Sensitivity (TPR) Vs FPR (1-Specificity)



PE-5

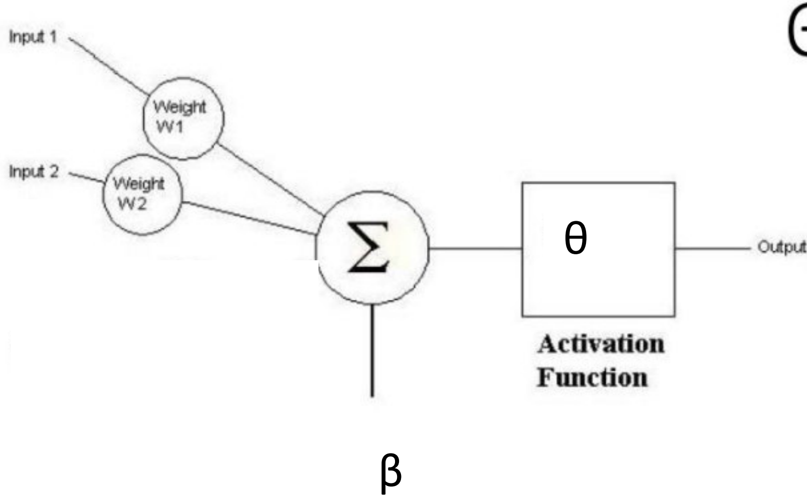
- Objective Functions
  - ML “introspection” of learning performance in training
  - Used to evaluate **training** performance

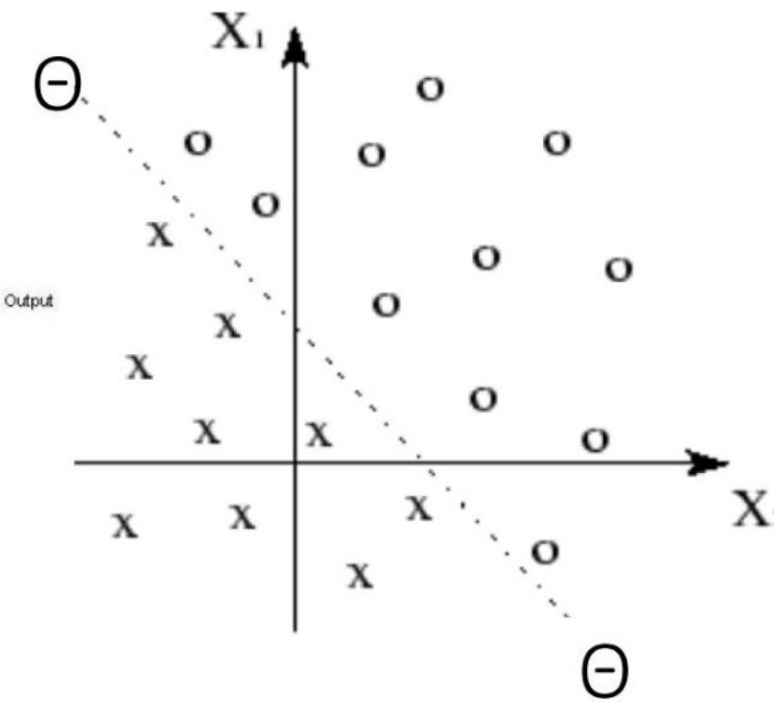
- ML Performance Evaluation
  - Used to evaluate **testing** performance
  - **BEWARE OF TRAINING BY OTHER MEANS**

Misc Advanced ML Topics

AT-0

Training By Other Means (Changing Parameter ϴ)





AT-1

Polymorphous versus Homogeneous Data

- DF Malware File Structure
  - File Size <-Bytes (integer)
  - Data Section Size <-Proportion (real)
  - Data Entropy <- Dimensionless (real)
  - API Calls <- (Strings?)
    (Hex)

Data Standardization

Z –Statistics Homogenize the Data

- All data are shifted to have zero mean
- All data are re-scaled to have unit variance

- Enables data fusion for statistical analysis
  - eg: Correlation analysis

Mean (μ_x) of Original Data

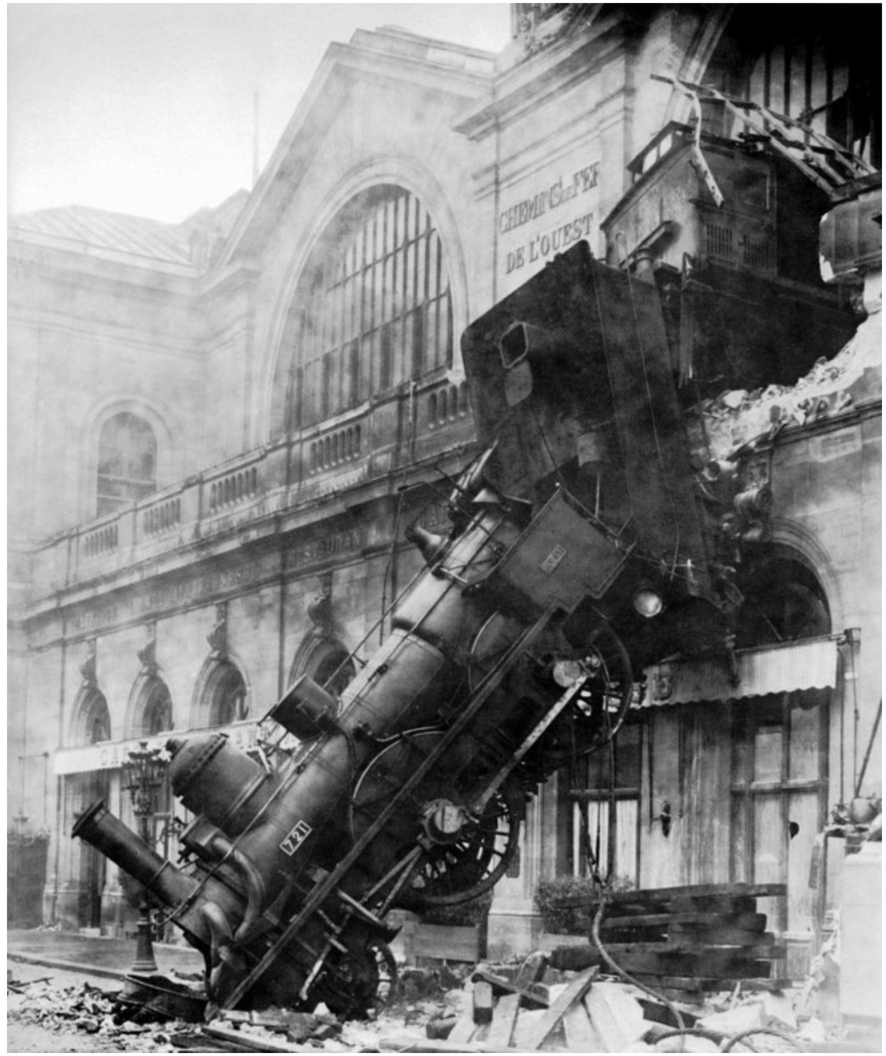
$$z = \frac{x - \mu_x}{\sigma_x}$$

Standard Deviation (σ_x) of Original Data

NB: variance = σ_x^2

AT-3

NTNU



An Ultimate Optimization Strategy, For Solving Every Problem

There is No Free Lunch!

- "No Free Lunch Theorems for Optimization" Wolpert & Macready 1997
- A good approach to solving one type of problem isn't necessarily a good approach for solving other types.
- Power lifting athletes can't run marathons.
  - Different basic body types
  - Divergent regimes of training and adaptation designed for adaptation to execute a specific task
- Marathon runners can't power lift.
  - Same reasons
- Biometric Template Attacks
  - Simplex HC for facial biometrics
  - GA for iris biometrics

Thank You!

- Questions
- Comments
- Feedback
- Improvements