Walla Walla University
SEVENTH-DAY ADVENTIST HIGHER EDUCATION

North American Simulation and Gaming Association
NASAGA

www.Mihail.Motzev.com

University of National and World Economy 1920 Sofia

Bulgarian Academic Simulation And Gaming Association

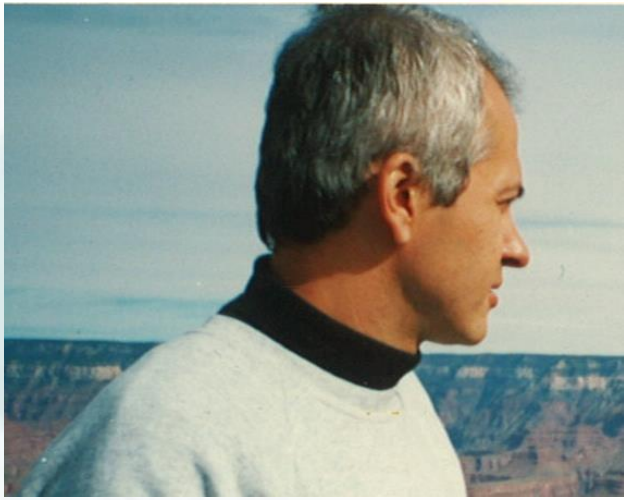
ISAGA

The International Simulation and Gaming Association

ABSEL
Association for Business Simulation and Experiential Learning

Summer School On Modelling And Complex Systems 2022:

"Statistical Learning Networks"



Mihail Motzev

Ph.D,M.Sc,P.D.D

(MRMotzev@yahoo.com)

Proud Nerd (Zubar) Generation 1

STATISTICAL LEARNING NETWORKS

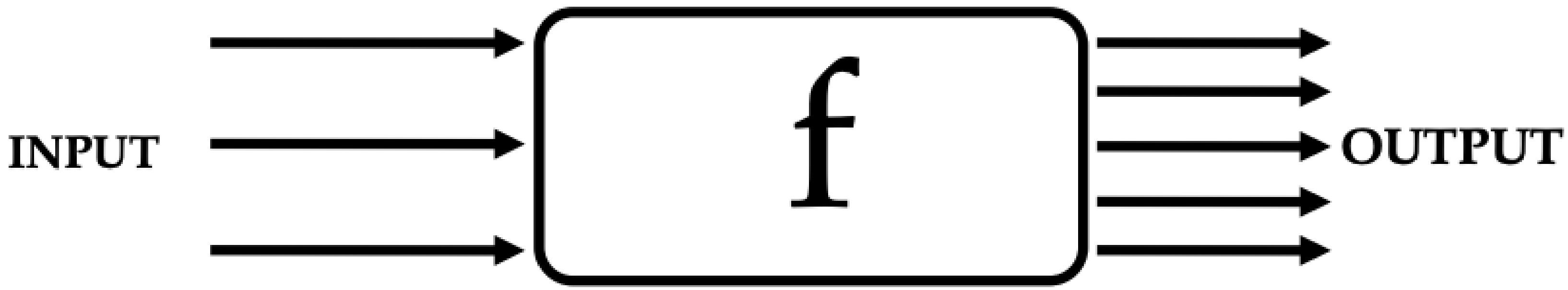
Fundamentals:

- *"Artificial Intelligence" was coined by John McCarthy (Dartmouth College - 1956) to distinguish the field from cybernetics and escape the influence of the cyberneticist Norbert Wiener.*
- *Artificial general intelligence (AGI) studies GI (the ability to take on any arbitrary problem) exclusively. Most AI research usually produced programs that can solve only one problem (narrow AI).*
- *"Statistical learning" techniques such as HMM and neural networks gain higher levels of accuracy in many practical domains such as data mining, without necessarily acquiring a semantic understanding of the datasets.*

Fundamentals:

- ***Artificial general intelligence*** *(AGI, strong AI, full AI etc.) is the hypothetical ability of an intelligent agent to understand or learn any intellectual task that a human being can.*
- ***Narrow AI*** *(weak AI) is limited to the use of software to study or accomplish specific pre-learned problem solving or reasoning tasks (expert systems).*
- *In the 1990s and early 21st century, mainstream AI achieved great commercial success and academic respectability by focusing on specific sub-problems where they can produce verifiable results and commercial applications, such as* ***artificial neural networks*** *and* ***statistical machine learning****.*

Statistical Learning Theory: supervised learning



Given a set of l examples (data)

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$$

Question: find function f such that

$$f(x) = \hat{y}$$

is a **good predictor** of y for a **future** input x (fitting the data is **not** enough!)

A framework for machine learning drawing from the fields of statistics and functional analysis.

Learning Models & Approaches

- ***Supervised learning*** **- is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.**
- ***Unsupervised learning*** **- looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision, also known as self-organization.**
- ***Semi-supervised learning*** **- an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training.**

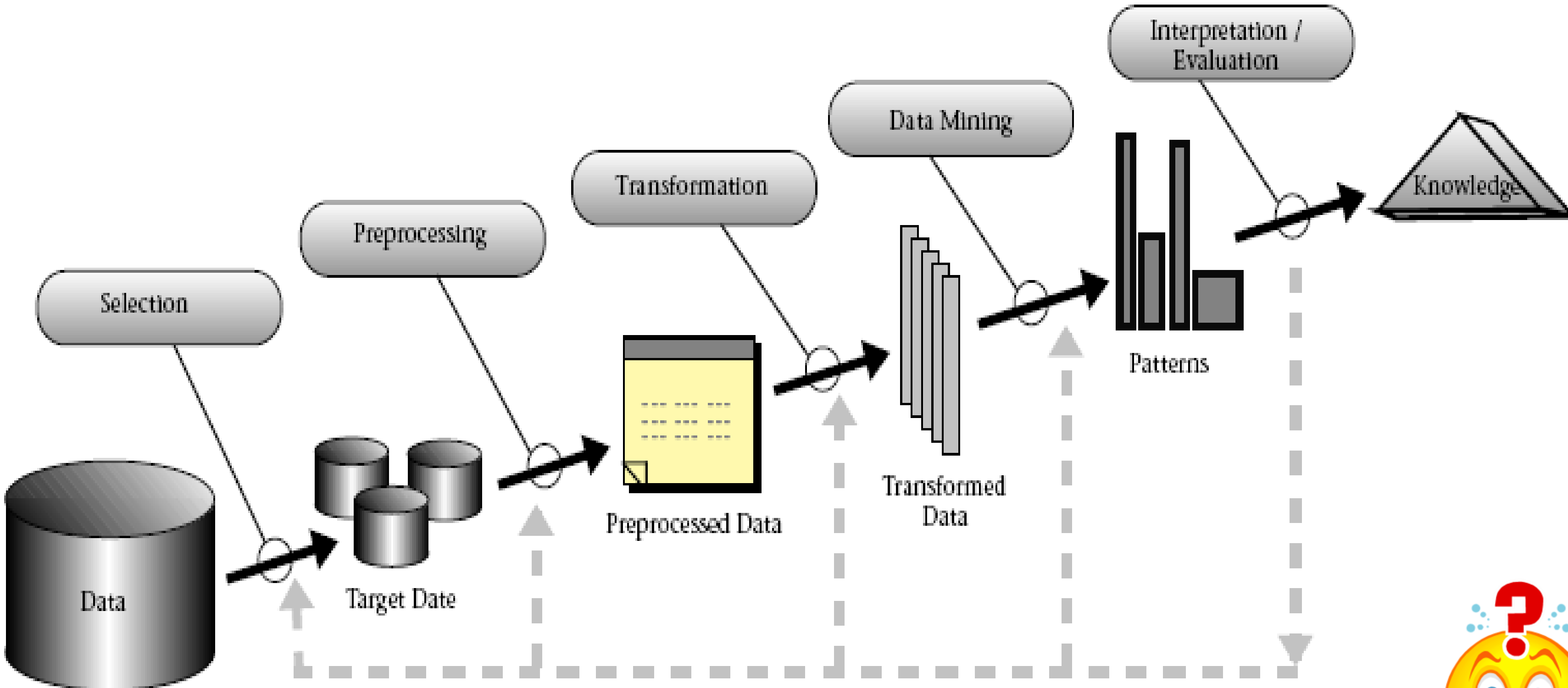
STATISTICAL LEARNING NETWORKS

Fundamentals:

- ***Network*** *– a function (model) represented by the composition of many basic functions (models).*
- ***Basic function*** *– element, unit, building block, network node, artificial neuron, partial model.*
- ***A Learning Network*** *estimates its function from representative observations of the relevant variables.*
- *From a data mining perspective, ANNs are just another way of fitting a model to observed historical data in order to be able to make classifications or predictions.*

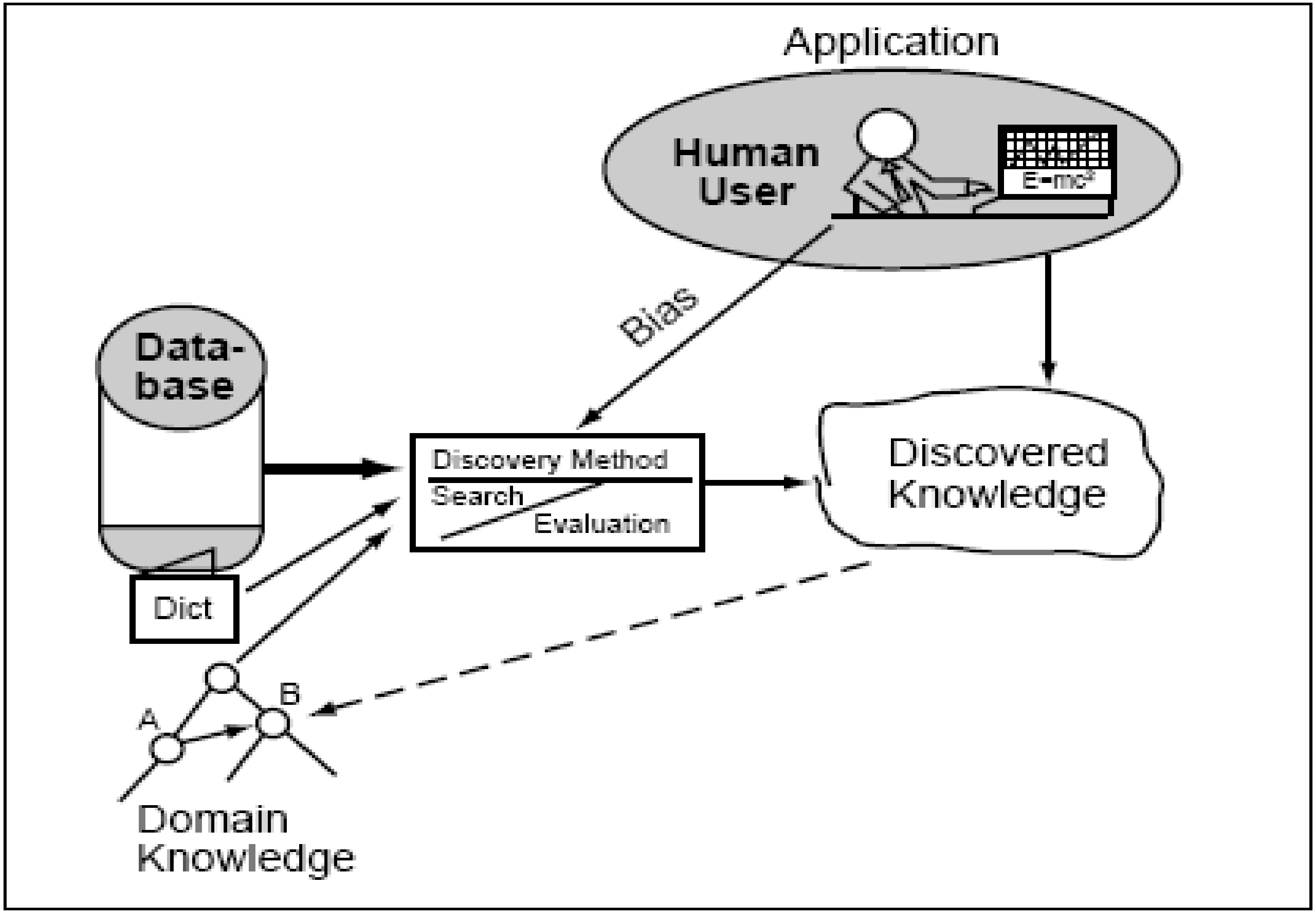
Knowledge Discovery in Databases – "Identification of underlying patterns, categories, and behaviors in large data sets, using techniques such as *neural networks* and *data mining*"

An overview of KDD process



Knowledge Discovery from Data

A Framework for Knowledge Discovery in Databases



Data Mining and Predictive analytics

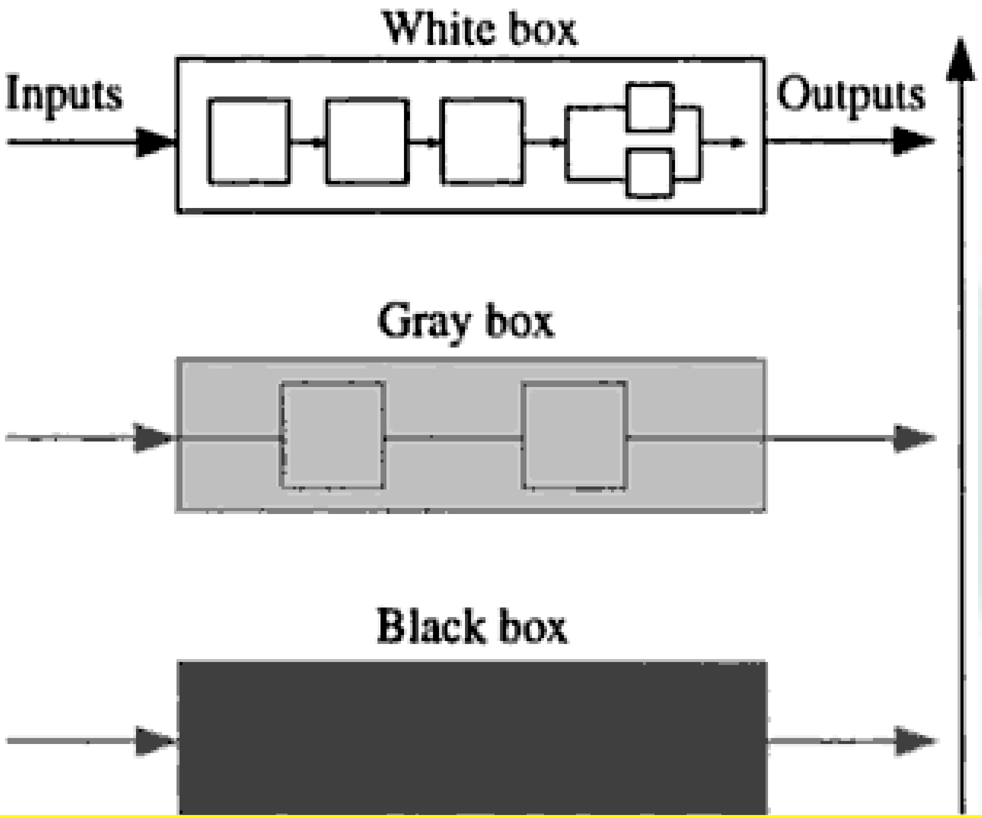
- ***Data mining*** is the process of exploration and analysis (by automatic or semi- automatic means) of large quantities of data in order to discover meaningful patterns and rules.

- ***Predictive analytics*** encompasses a variety of techniques from *statistics*, *machine learning* and *data mining* that analyze current and historical facts to make predictions about future or otherwise unknown events - technically, predictive analytics is an area of data mining that deals with extracting information from data and using it to predict trends and behavior patterns.

Data mining activities:

- ***Classification***: learning a function that maps (classifies) a data item into one of several predefined classes;
- ***Estimation (regression)***: learning a function that maps a data item into a real-valued prediction variable, building a model;
- ***Prediction (predictive modeling):*** building a model which can be used to make reliable forecasts;
- ***Affinity grouping or association rules***: finding a model that describes significant dependencies between variables;
- ***Clustering***: identifying a finite set of categories or clusters to describe the data;
- ***Description and visualization (summarization)***: finding a compact description for a subset of data.

Model Identification



Increasing internal knowledge

Data Mining

Computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems

DIRECTED DATA MINING

The goal is to use the available data to build a model that describes one particular variable of interest in terms of the rest of the available data. A top-down approach, used when we know what we are looking for. It often takes the form of predictive modeling. The model is considered as a ***black box***.

Data mining activities:

- ***Classification***: learning a function that maps (classifies) a data item into one of several predefined classes;
- ***Estimation (regression)***: learning a function that maps a data item into a real-valued prediction variable, building a model;
- ***Prediction (predictive modeling):*** building a model which can be used to make reliable forecasts.

DIRECTED DATA MINING

- ***A top-down approach*** **– often takes the form of *predictive modeling* where we know exactly what we want to predict. In this case the model is considered as a *black box*, i.e. it is not important what the model is doing, we just want the most accurate result possible.**

UNDIRECTED DATA MINING

A bottom-up approach that finds patterns in the data and leaves it up to the user to determine whether or not these patterns are important, i.e. it is about discovering new patterns inside the data. The goal is to establish some relationship among all the variables (represented with ***semitransparent boxes***).

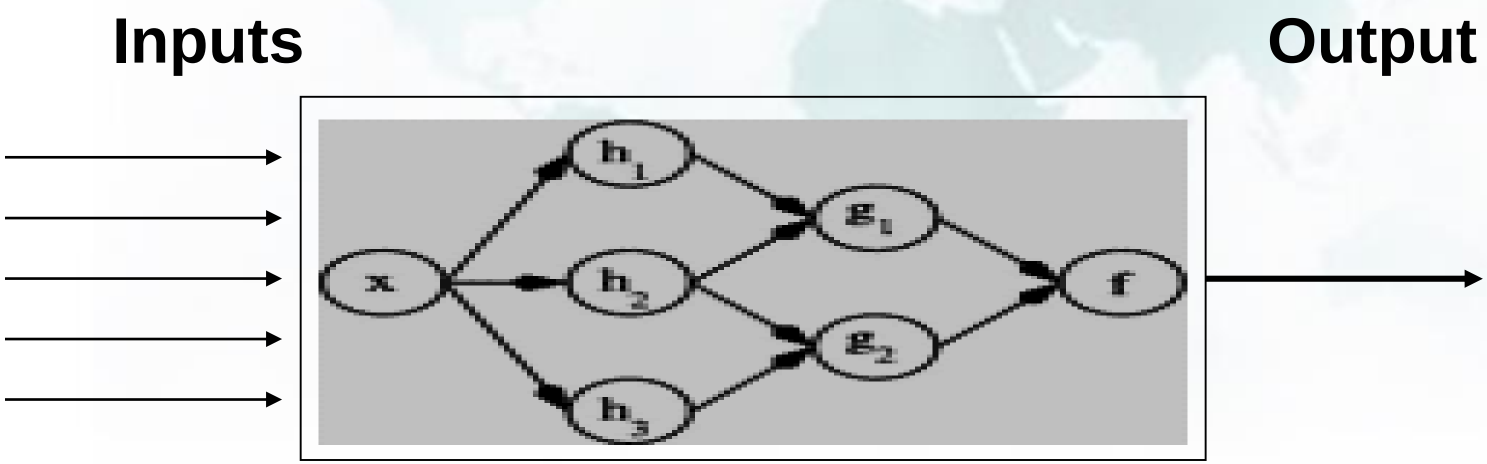
Data mining activities:

- ***Affinity grouping or association rules*****:** finding a model that describes significant dependencies between variables;
- ***Clustering*****:** identifying a finite set of categories or clusters to describe the data;
- ***Description and visualization (summarization)*****:** finding a compact description for a subset of data.

STATISTICAL LEARNING NETWORKS

UNDIRECTED DATA MINING

- ***A bottom-up approach*** **that finds patterns in the data which provide insights. This form of data mining is represented with** ***semitransparent boxes*** **and unlike directed** ***DM*****, here users want to know what is going on, how the model is coming up with an answer.**



Data Mining Process

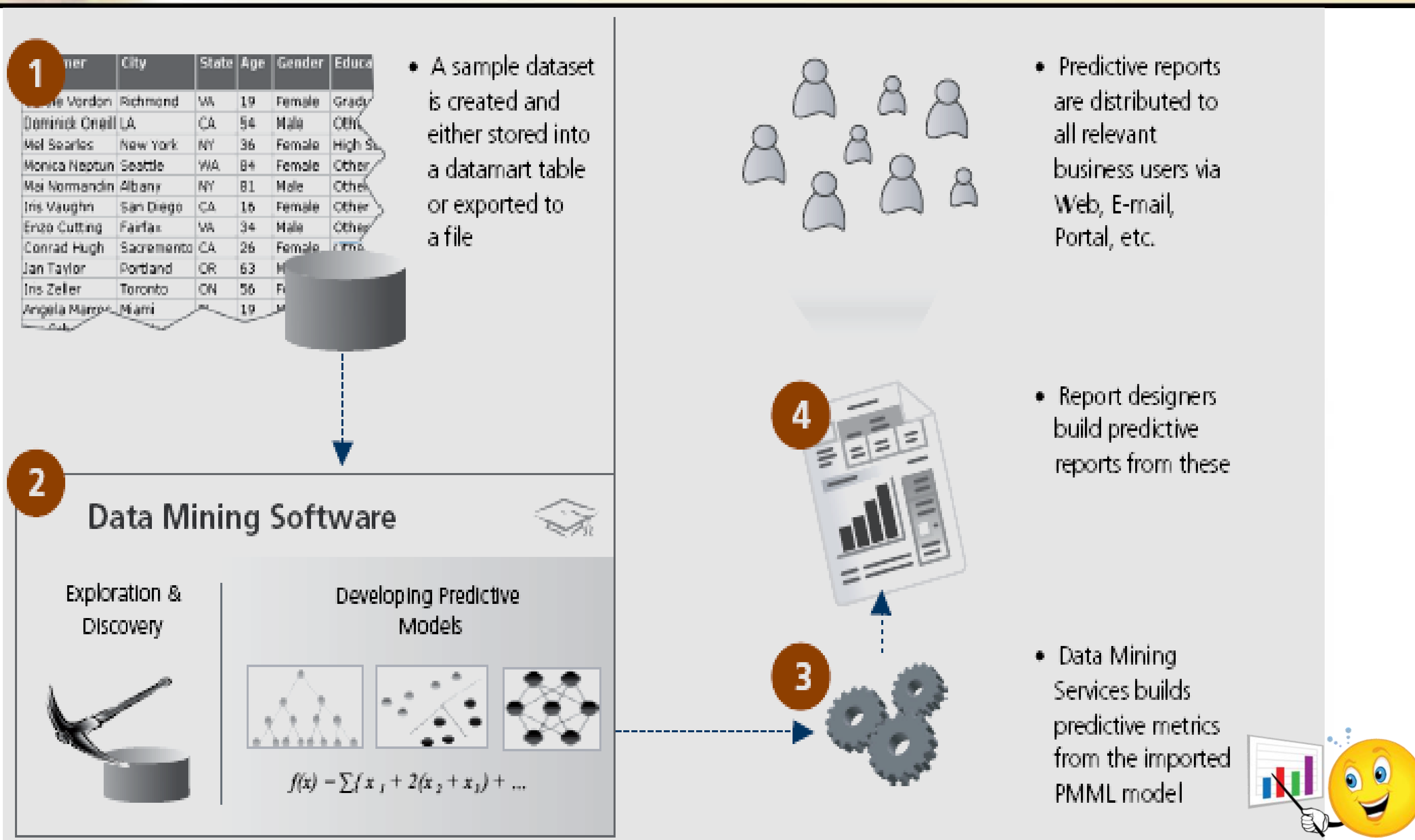
1. **Create a predictive model from a data sample**
2. **Train the model against datasets with known results**
3. **Apply the model against a new dataset with an unknown outcome (*cross-validation*)**

Notes: SAS Institute Inc. developed a five-step data mining cycle process known as ***SEMMA***: Sample, explore, modify, model, and assess.

IBM Corp. has a slightly different interpretation of the data mining process and other companies may have their own view as well.

STATISTICAL LEARNING NETWORKS

Data Mining Workflow in MicroStrategy platform



Data Mining Process - the Three Pillars of Data Mining

Three main components in Data Mining process:

1. *Data -* The power of data mining is leveraging the data that a company collects to make better informed business decisions.
2. *Modeling Skills -* The set of *modeling skills* needed to build predictive models in data mining in general is the same as in business forecasting process and which is working well for both directed and undirected data mining.
3. *Data Mining Techniques – clustering, decision trees and neural networks.*

Data mining tasks:

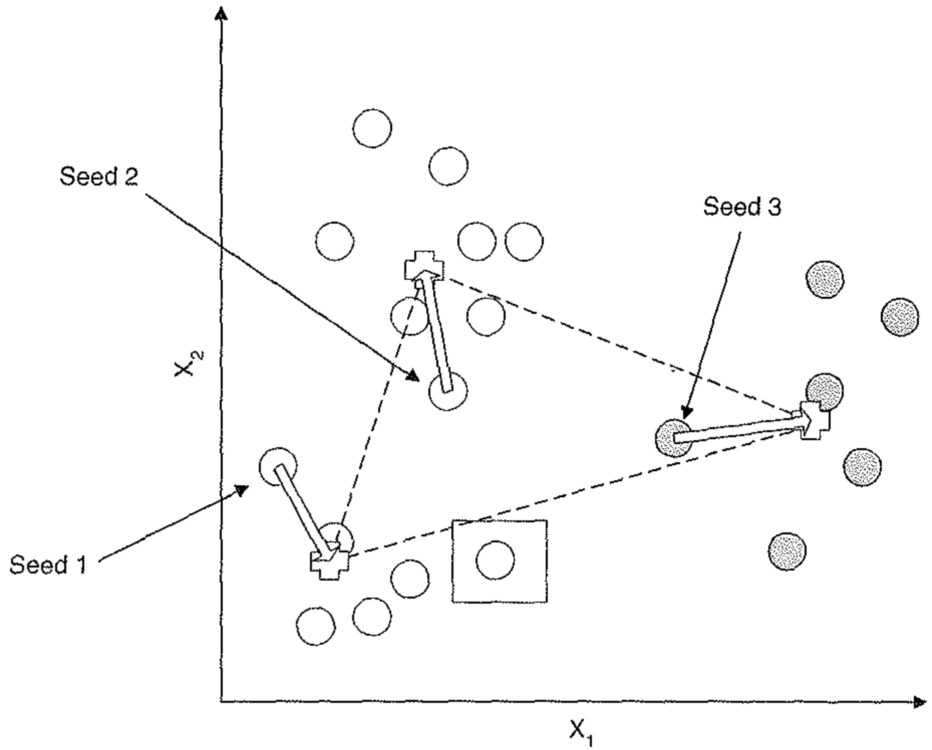
- ***classification*:** learning a function that maps (classifies) a data item into one of several predefined classes;
- ***regression*:** learning a function that maps a data item into a real-valued prediction variable;
- ***clustering*:** identifying a finite set of categories or clusters to describe the data;
- ***summarization*:** finding a compact description for a subset of data;
- ***dependency modeling*:** finding a model that describes significant dependencies between variables;
- ***change and deviation detection*:** discovering the most significant changes in the data from previously measured or normative values.

Data Mining Techniques

- ***Automatic Cluster Detection*** **- use cluster detection when we suspect that there are natural groupings that may represent groups of customers or products that have a lot in common with each other.**
- ***Decision Trees (Classification & Regression)*** **- a good choice when the data mining task is classification of records or prediction of outcomes. We should use decision trees when the goal is to assign each record to one of a few broad categories.**
- ***Artificial Neural Networks (the most widely known and the least understood of the major data mining techniques)*** **- a good choice for most classification and prediction tasks when the results of the model are more important than understanding how the model works. *ANNs* represent complex mathematical equations, with lots of summations, exponential functions, and many parameters.**

Automatic Cluster Detection

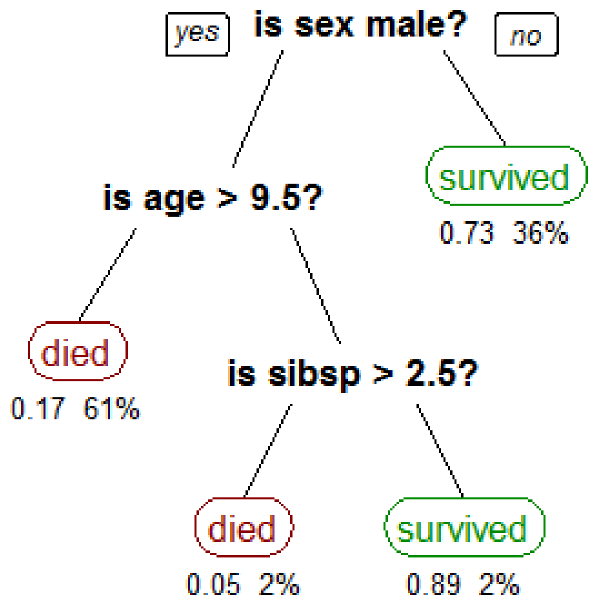
Data Mining Techniques



Grouping a set of objects in such a way that objects in the same group (cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters)

Decision Trees

Data Mining Techniques

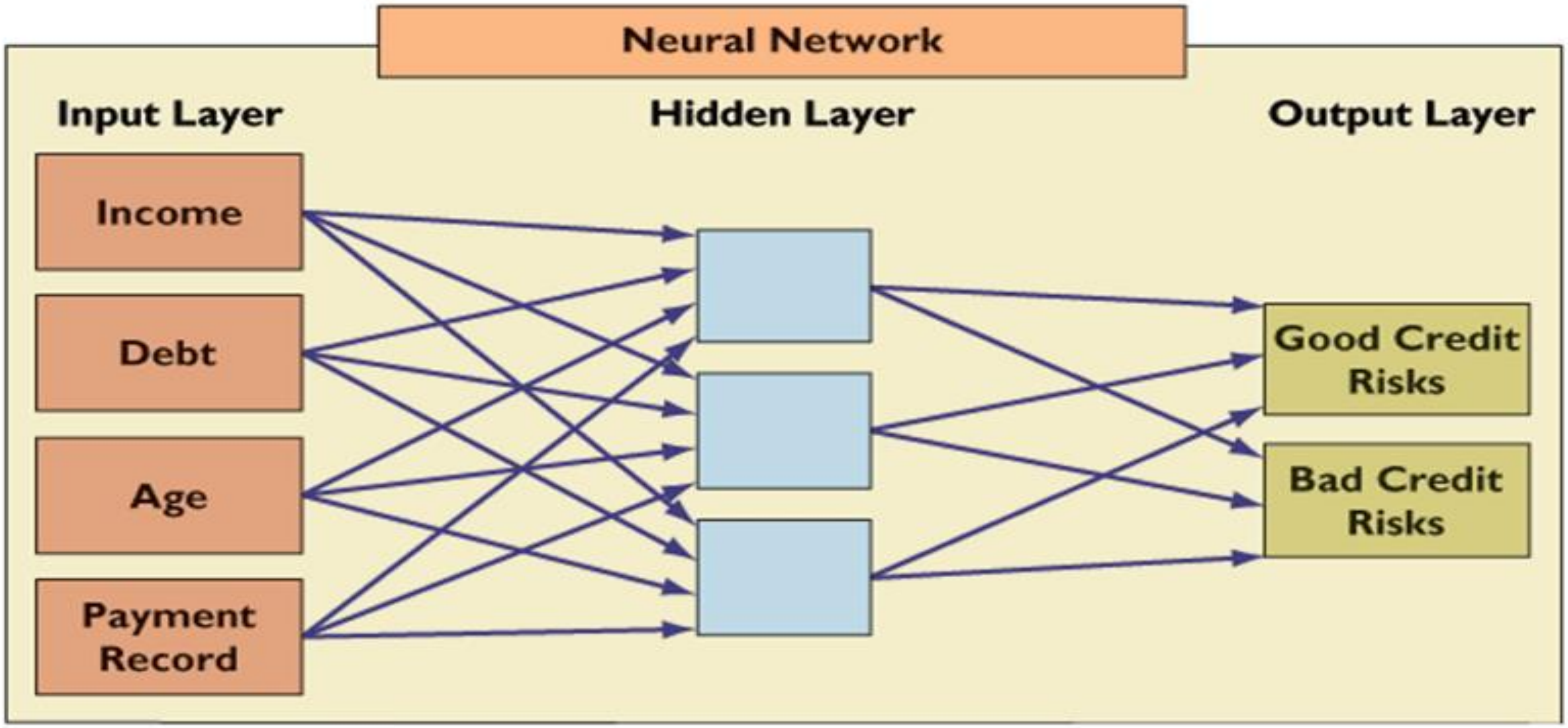


A tree showing survival of passengers on the Titanic ("sibsp" is the number of spouses or siblings aboard). The figures under the leaves show the probability of survival and the percentage of observations in the leaf

Artificial Neural Networks

Data Mining Techniques

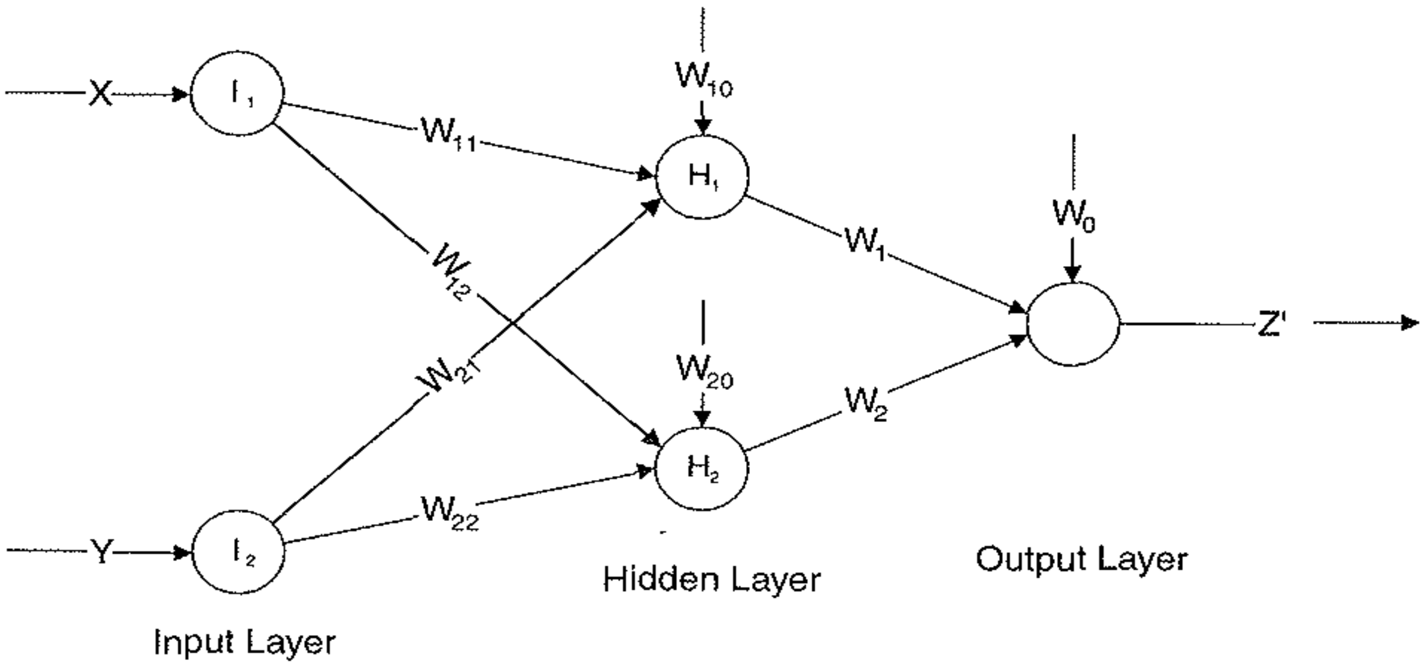
ANN – artificial systems which emulate the processing patterns of the biological brain to discover patterns and relationships in massive amounts of data *("Perceptron" - Ph. Rozenblat)*



Source: Herb Edelstein, "Technology How-To: Mining Data Warehouses," *InformationWeek*, January 8, 1996.
Copyright © 1996 CMP Media, Inc., 600 Community Drive, Manhasset, NY 11030. Reprinted with permission.

Model Building

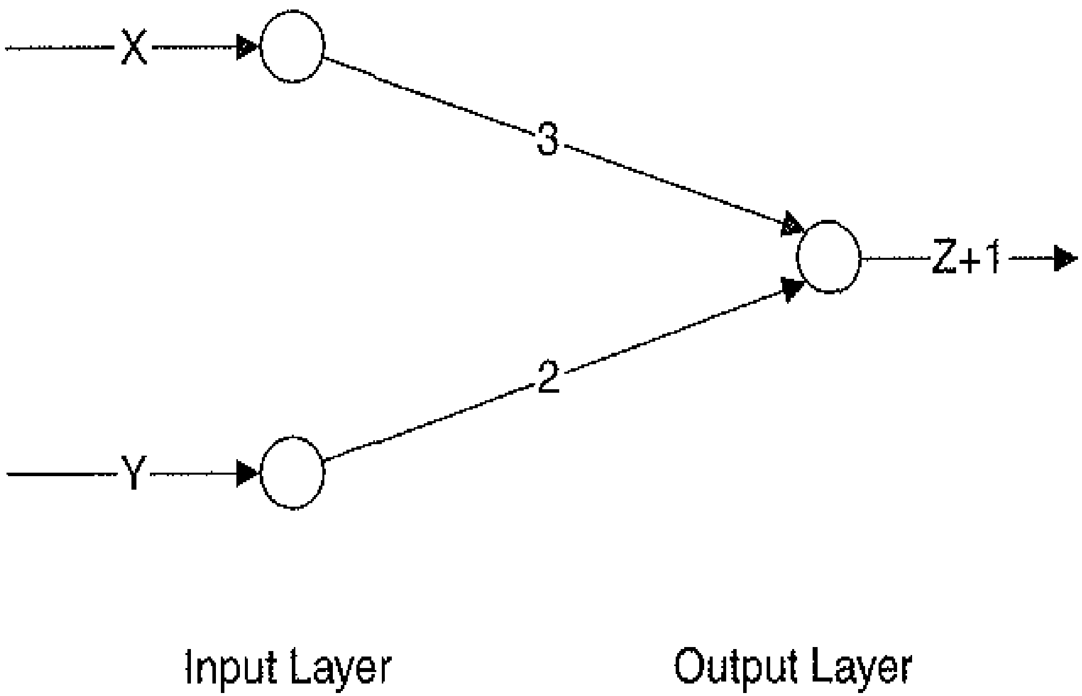
Data Mining Techniques - ANNs



A neural network with a hidden layer.

"The most widely known and the least understood of the major data mining techniques."

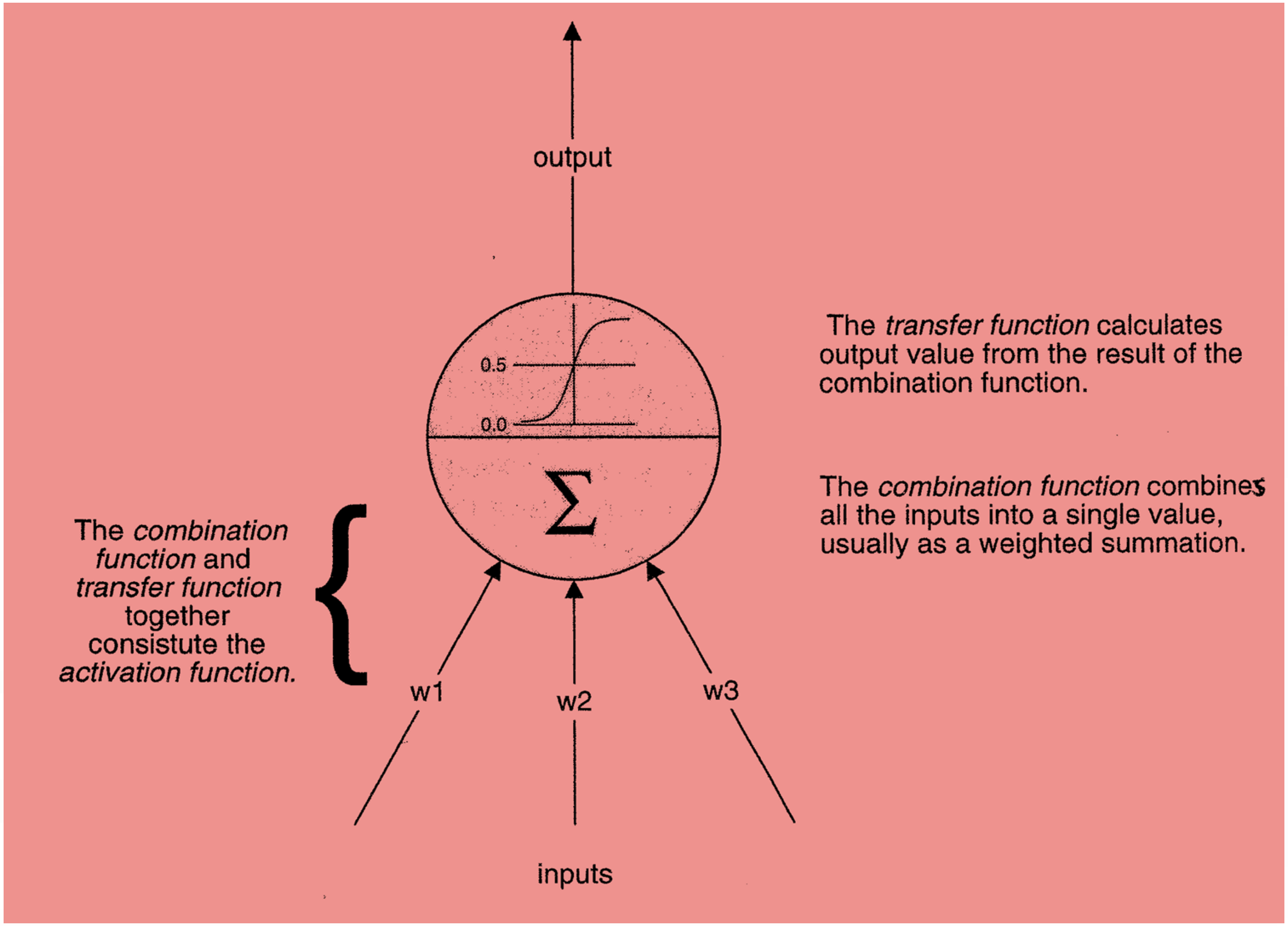
How a Neural Network Works



Neural network representation of $z=3x+2y-1$.

How a Neural Network Works

Linear Transfer Function



How a Neural Network Works

When to use Artificial Neural Networks

ANNs are a good choice for most classification and prediction tasks when the results of the model are more important than understanding how the model works. *ANN* represent complex mathematical equations, with lots of summations, exponential functions, and many parameters. The equations are the rule of the network but are useless for our understanding. Also, *ANN* does not work well when there is large number of inputs. This makes it more difficult for the network to find patterns and can result in long training phases that never converge to a good solution.

	ANNs	Statistical Learning Networks
Data analysis	universal approximator	structure identificator
Analytical model	indirect by approximation	direct
Architecture	unbounded network structure; experimental selection of adequate architecture demands time and experience	bounded network structure [1]; adaptively synthesised structure
A-priori-Information	without transformation in the world of ANNs not usable	can be used directly to select the reference functions and criteria
Self-organisation	deductive, given number of layers and number of nodes (subjective choice)	inductive, number of layers and of nodes estimated by minimum of external criterion (objective choice)
Parameter estimation	in a recursive way; demands long samples	estimation on training set by means of maximum likelihood techniques, selection on testing set (extremely short)
Feature	result depends on initial solution, time-consuming technique, necessary knowledge about the theory of neural networks	existence of a model of optimal complexity, not time-consuming technique, necessary knowledge about the task (criteria) and class of system (linear, non-linear)

General Prediction Model

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x_i x_j x_k$$

where

$X(x_1, x_2, \ldots, x_M)$ - input variables vector;

$A(a_1, a_2, \ldots, a_M)$ - vector of coefficients or weights.

$$Y = F(X, e)$$

where ***F*** can be any mathematical function describing the variable ***Y (the output)*** as a function of ***input variables X*** and the stochastic component ***e (model error).***

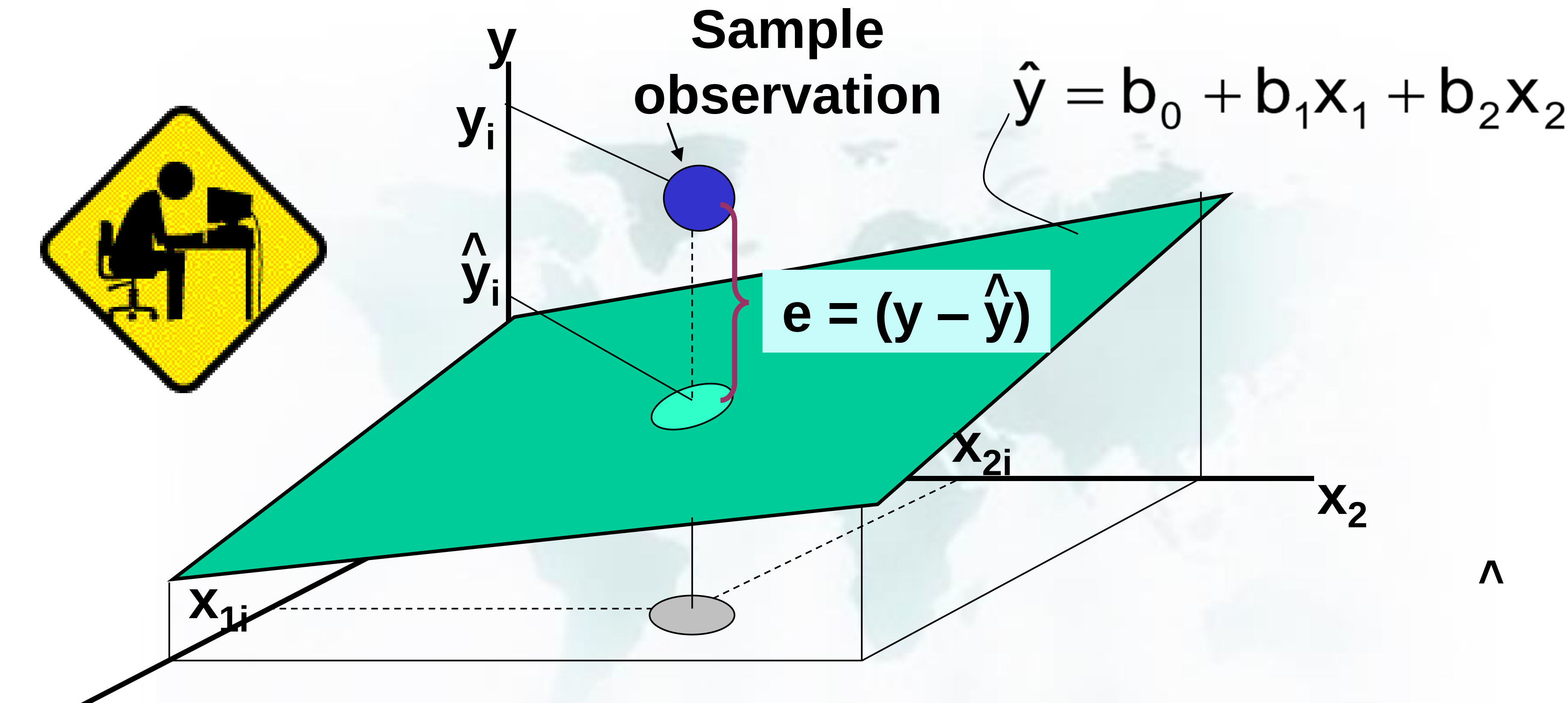
Model Building

Problems

- *Model specification;*
- *Overfitting;*
- *Autocorrelation;*
- *Multicollinearity*
- *ANNs:*
  - *number of layers;*
  - *how many input nodes;*
  - *best activation function;*
  - *ANN training;*
  - *lack of transparency (interpretation), etc.*

Model Building

Regression Analysis



Example: ***Two-Factors Regression Models***

Walla Walla University
SEVENTH-DAY ADVENTIST HIGHER EDUCATION

Model Building

Regression Models – Problems:

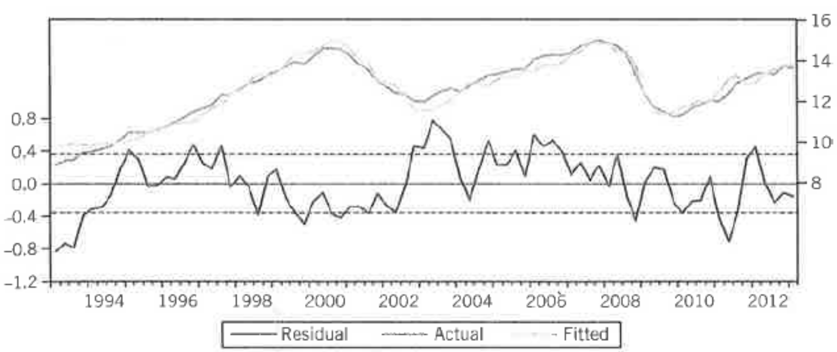
Alan Greenspan (The Map and the Territory: Risk, Human Nature, and the Future of Forecasting):

APPENDICES

Exhibit 4.7

Dependent Variable (Time Period: Q1 1993–Q1 2013, 81 obs.)		
Real Pvt Nonres Fixed Invst (SAAR, Bil.Chn.2005$) / Pvt Nonres Fixed Assets (2005 = 100)		
Independent Variable(s)	**Coefficient**	**t-Statistic***
S&P 500 (1941-43=10) / Pvt Nonres Fixed Invst Price (SA, 2005 = 100) (1 quarter ago)	0.473	19.044
Nonfarm Operating Rate (SA, % of capacity) (3 quarters ago)	0.165	6.118
Structures' share of nominal Pvt Nonres Fixed Invst	6.332	4.517
Adjusted R-sq 0.946	**Durbin-Watson** 0.585	

*t-statistic calculated using Newey-West HAC standard errors and covariance.



Source: U.S. Department of Commerce; Standard and Poor's; Federal Reserve Board; author's calculations.

Exhibit 3.3

Dependent Variable (Time Period: Jan. 1991–Dec. 2005, 180 obs.	
m/m %Δ in: CoreLogic Home Price Index (Seasonally adjusted	
Independent Variable(s)	
Freddie Mac 30yr Fixed-Rate Mortgage Rate, % p.a. (3 mo	
Adjusted R-sq 0.604	**Durbin-Watson** 0.159

*t-statistic calculated using Newey-West HAC standard errors and covarian

Exhibit 4.6

Dependent Variable (Time Period: Q1 1970–Q4 2012, 172 obs.)		
ln [Real GDP / Real GDP (4 quarters ago)]		
Independent Variable(s)	**Coefficient**	**t-Statistic***
ln [**Corp & Home Equity, Period Avg (1 quarter ago) / **Corp & Home Equity, Period Avg (5 quarters ago)]	0.127	9.691
Adjusted R-sq 0.419	**Durbin-Watson** 0.364	

*t-statistic calculated using Newey-West HAC standard errors and covariance.
**Domestic holdings of domestic corporate equities and foreign corporate equities, at market value.

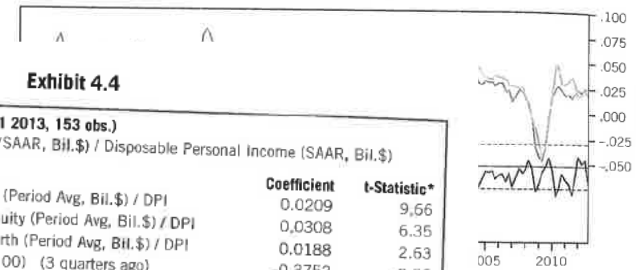
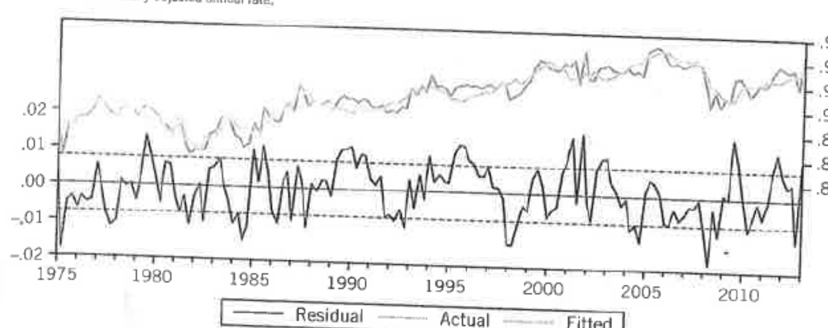


Exhibit 4.4

Dependent Variable (Time Period: Q1 1975–Q1 2013, 153 obs.)		
Personal Consumption Expenditures (***SAAR, Bil.$) / Disposable Personal Income (SAAR, Bil.$)		
Independent Variable(s)	**Coefficient**	**t-Statistic***
Household (incl. NPOs) Stock Net Worth (Period Avg, Bil.$) / DPI	0.0209	9.66
Household (incl. NPOs) Homeowners' Equity (Period Avg, Bil.$) / DPI	0.0308	6.35
Household (incl. NPOs) All Other Net Worth (Period Avg, Bil.$) / DPI	0.0188	2.63
6-Month Certificates of Deposit (% p.a./100) (3 quarters ago)	−0.3752	−9.56
[**Adjusted PI / DPI] (2 quarters ago)	0.2666	2.30
Adjusted R-sq 0.903	**Durbin-Watson** 1.089	

*t-statistic calculated using Newey-West HAC standard errors and covariance.
**Adjusted PI = (0.9*Wages and Salary Disbursements) + (1.0*Personal Current Transfer Receipts) + (0.5*All Other Personal Income).
***Seasonally adjusted annual rate.



Source: Federal Reserve Board; U.S. Department of Commerce.

Model Building

Machine Learning - Interpretations

Simple numerical example

Consider the following data set :

y	a	b	c
9	1	8	1
9	2	7	2
9	3	6	3
9	4	5	4
9	5	4	5
9	6	3	6
9	7	2	7
6	99	1	5

Model:

Y = F (a,b,c)

Solutions:

y = 9.3- 0.033a- 0.033b

y = 0.00001 + b + c

y = 9 - 0.0319a + 0.0319c

“The first general, working learning algorithm for supervised, deep, feedforward, multilayer perceptron(s) was published by Alexey Ivakhnenko and Lapa in 1967” (*Wikipedia*).

GMDH

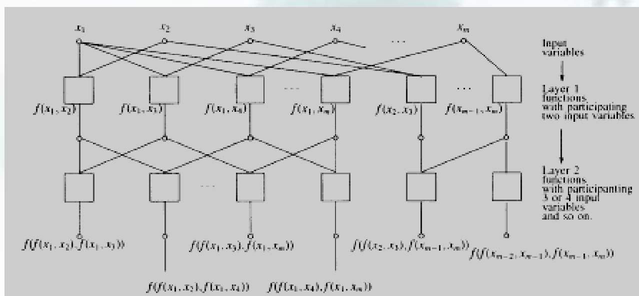
Alexey G. Ivakhnenko.

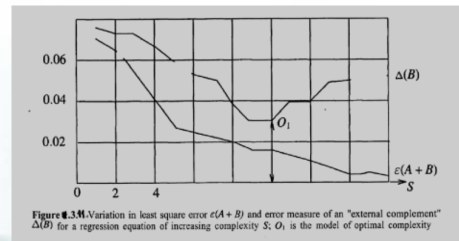
(1913-2007)

Two State Prizes of the USSR, Medal “For Labor”, Order of Friendship of Peoples …

Gödel's incompleteness theorems

Genetic selection of pairwise features





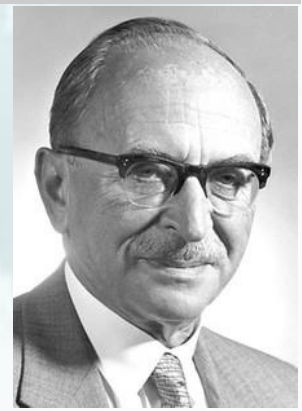
Gabor’s principle of "freedom of decisions choice"

Knowledge extraction from experimental data, Self-Organization etc…

Kurt Gödel (1906-1978)

Notable awards:

- Albert Einstein Award (1951)
- National Medal of Science (USA) in Mathematical, Statistical, and Computational Sciences (1974)

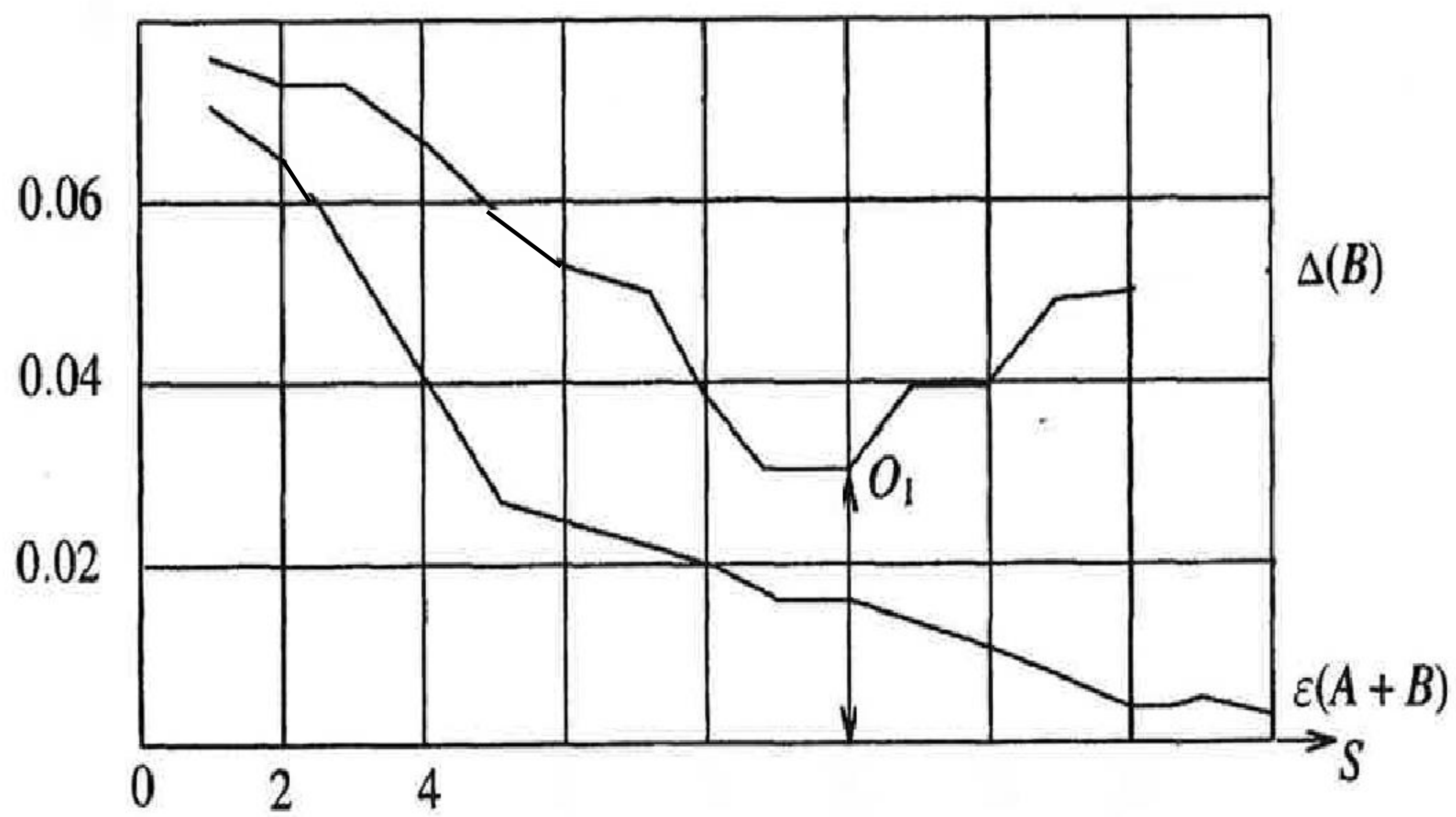


Dennis Gabor (1900-1978)

Numerous (>20) awards:

- Nobel Prize in Physics (1971)
- Honorary Doctorate, Delft University of Technology (1971)

Overfitting – Internal vs External (Cross) Validation

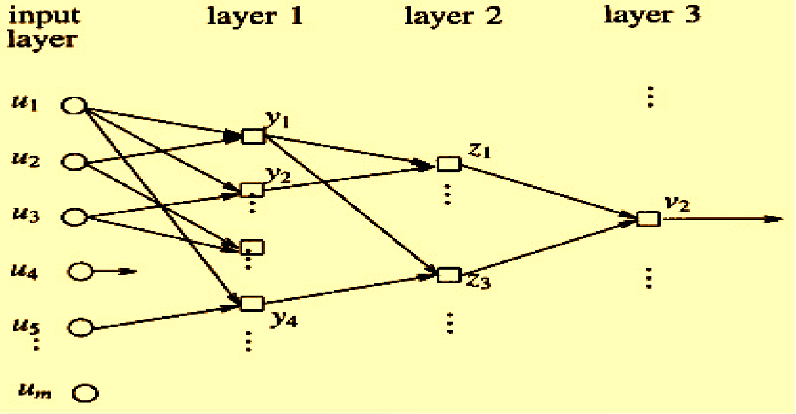


Variation in least square error $\varepsilon(A+B)$ and error measure of an "external complement" $\Delta(B)$ for a regression equation of increasing complexity S; O_1 is the model of optimal complexity

Statistical Learning Networks of Active Neurons

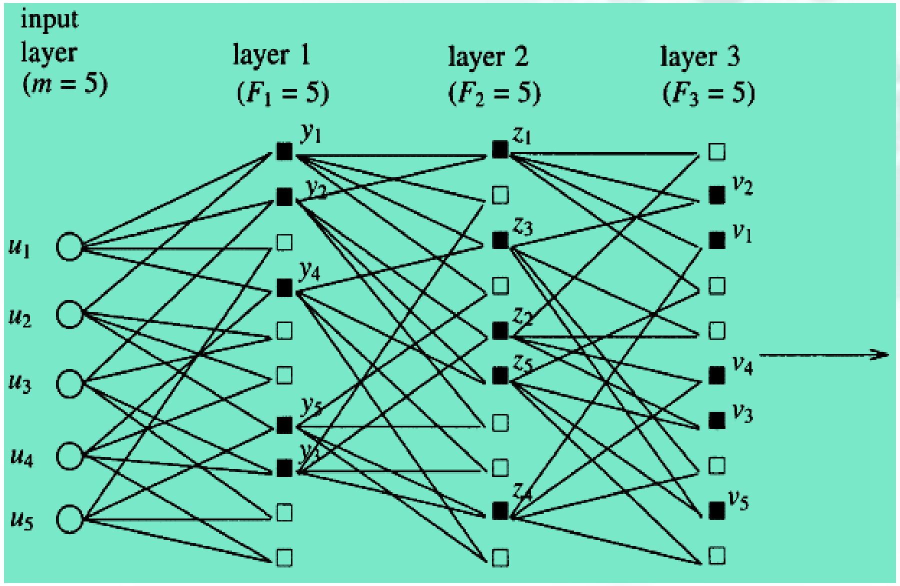
Multilayered Net of Active Neurons (MLNAN)

In this approach, neither the number of neurons and the number of layers in the network, nor the actual behavior of each created neuron is predefined. The modeling process is self-organizing because all of them (the number of neurons, the number of layers, and the actual behavior of each created neuron) are adjusting during the process of self-organization.



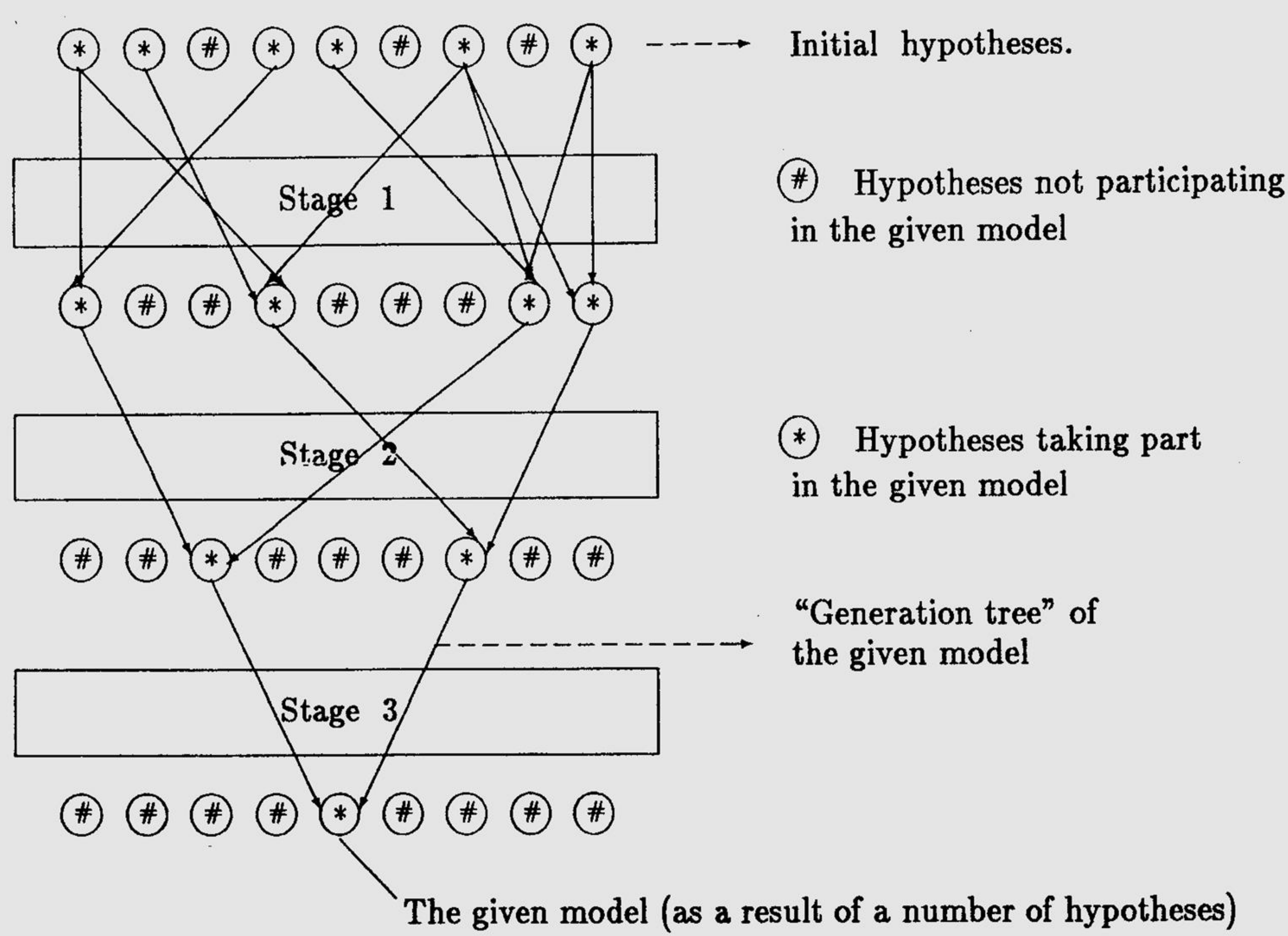
Multilayer network structure representing the output flow to unit 2 of layer 3

This method grows a tree-like network out of data of input and output variables in a pair-wise combination and competitive selection from a simple single unit to a desired final solution that does not have a predefined model. The basic idea is that first the elements on a lower level are estimated and the corresponding intermediate outputs are computed and then the parameters of the elements of the next level are estimated.

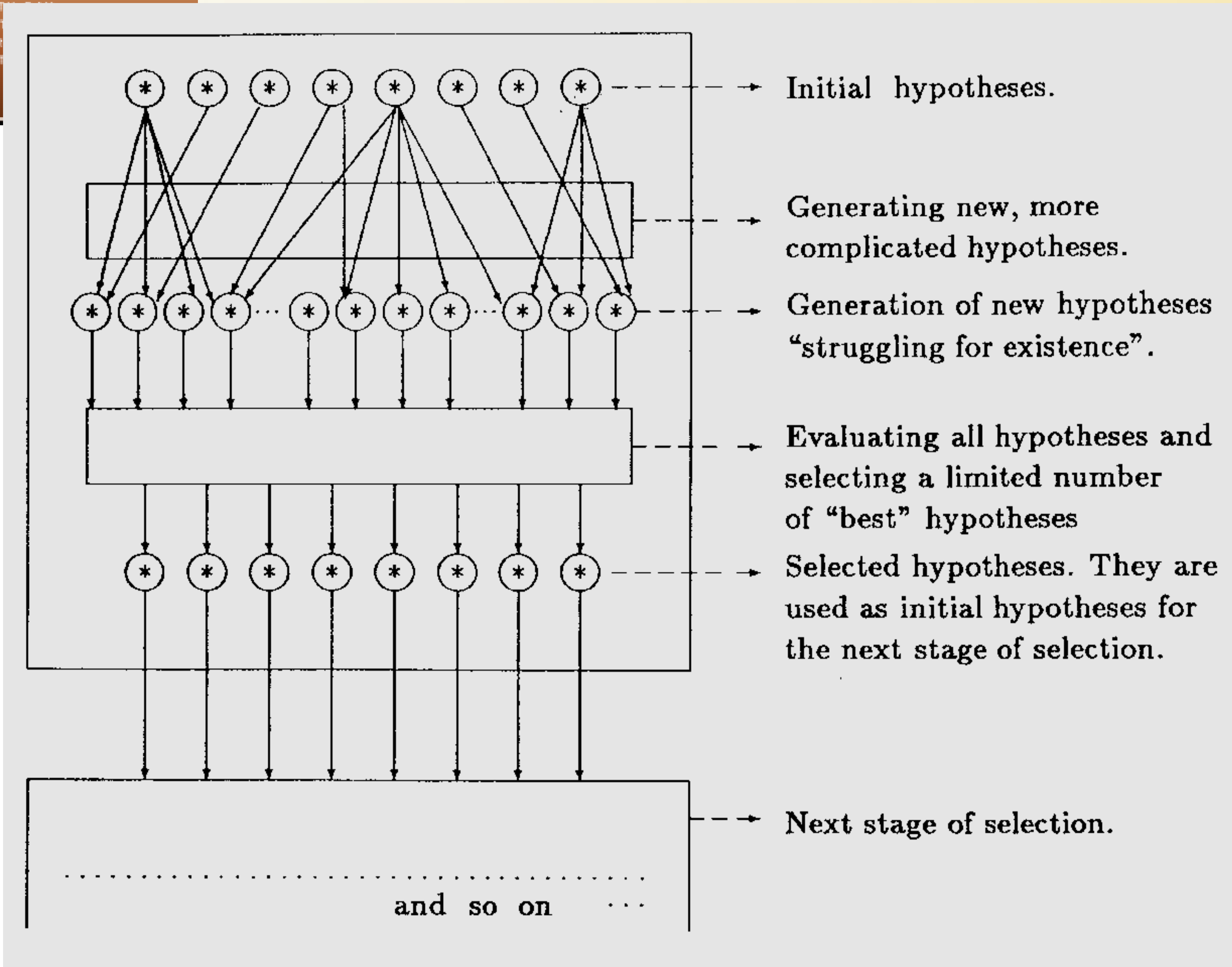


Multilayer network structure with five input arguments and selected nodes:

Multi-Stage Selection Algorithm

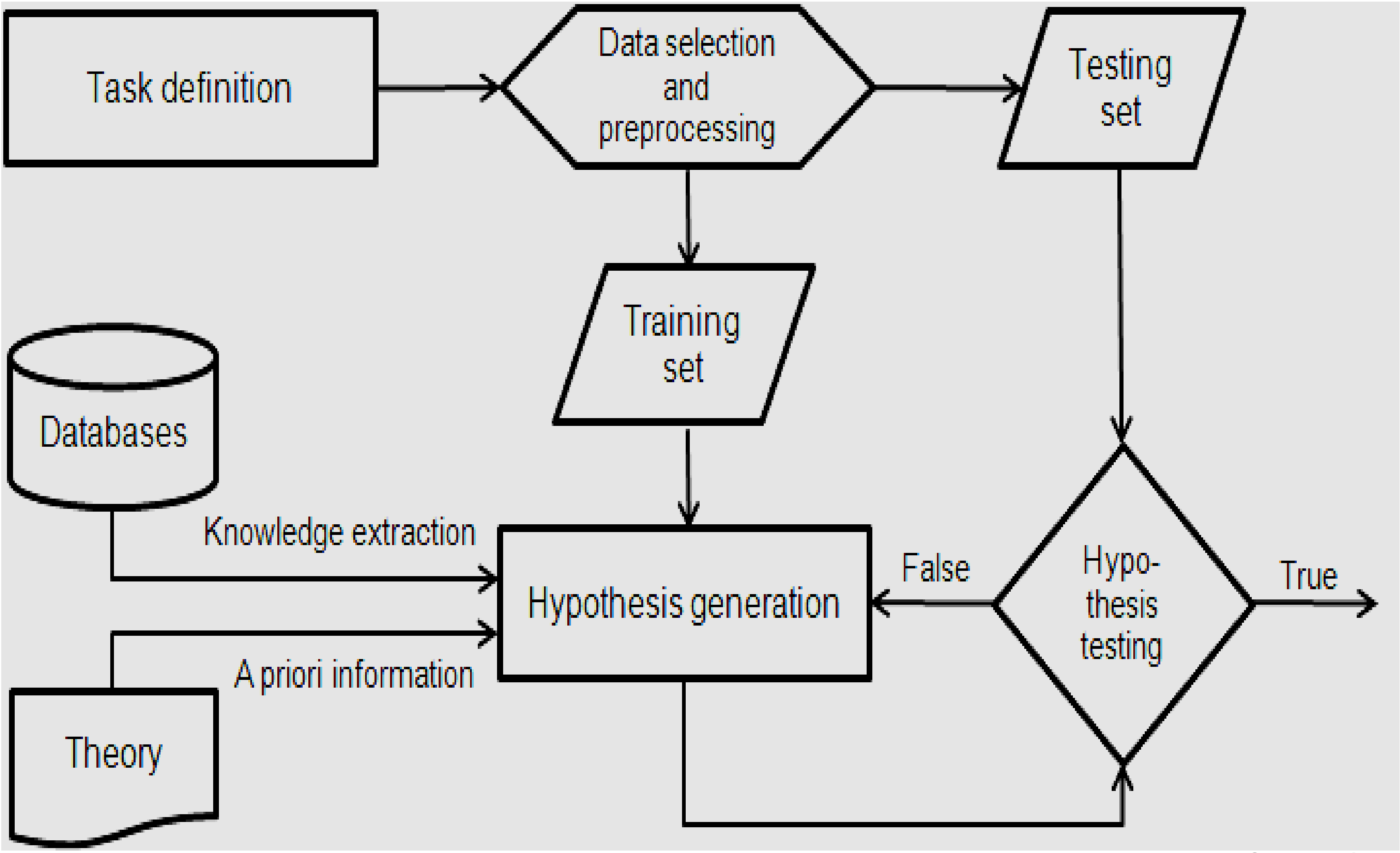


Pair-Wise Selection Using External Criteria



Model Selection

Cross Validation and a-priori information



Overfitting and Model Selection

- ***Cross Validation*** **-** also called ***rotation estimation*** or ***out-of-sample testing***, is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set.
- Involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the ***training set***), and validating the analysis on the other subset (called the ***validation set*** or ***testing set***).
- Two types of cross-validation can be distinguished: ***exhaustive*** and ***non-exhaustive cross-validation.***

Cross Validation

- ***Exhaustive cross-validation*** **- learn and test on all possible ways to divide the original sample into a training and a validation set.**

- *Leave-p-out cross-validation* *- involves using p observations as the validation set and the remaining observations as the training set. This is repeated on all ways to cut the original sample on a validation set of p observations and a training set.*

- ***Leave-one-out cross-validation*** - a particular case of ***leave-p-out cross-validation*** with p = 1.

Cross Validation

- ***Leave-one-out cross-validation:***

1. Select (it could be random) observation i for the testing set and use the remaining observations in the training set. Compute the error on the test observation.

2. Repeat the above step for i = 1, 2, … N-1, where N is the total number of observations.

3. Compute the forecast accuracy measures based on all errors obtained.

A total of 8 models will be trained and tested:

Cross Validation

- ***Non-exhaustive cross-validation* -** do not compute all ways of splitting the original sample. Those methods are approximations of ***leave-p-out cross-validation.***

- *k-fold cross-validation* - the sample is randomly partitioned into k equal sized subsamples. When k = n (the number of observations), *k-fold cross-validation* is equivalent to *leave-one-out cross-validation*.

- ***holdout method* -** randomly assign data points to two sets A and B (training set and test set).

- ***repeated random sub-sampling validation*** or ***Monte Carlo cross-validation*** creates multiple random splits of the dataset into training and validation data

Cross Validation

- ***Nested cross-validation* -** cross-validation is used simultaneously for selection of the best set of hyperparameters and for error estimation.

- *k*l-fold cross-validation* - contains an outer loop of k folds and an inner loop of l folds. One by one, a set is selected as (outer) test set and the k - 1 other sets are combined into the corresponding outer training set.

- ***k-fold cross-validation with validation and test set* -** k*l-fold cross-validation when l = k - 1. One by one, a set is selected as a test set. Then, one by one, one of the remaining sets is used as a validation set and the other k - 2 sets are used as training sets until all possible combinations have been evaluated.

Cross Validation with Time Series data

- ***Rolling forecasting origin*** **-** since it is not possible to get a reliable forecast based on a very small training set, the earliest observations ***n*** are not considered as testing sets.

1. We select the observation at time (n+i) for the testing set and use the observations at times t= {1, 2, … (n+i-1)} to estimate the forecasting model. Then we compute the error on the forecast for the time (n+i).
2. The above step should be done for all i= {1, 2, … (T-n)}, where T is the total number of observations and the forecast error should be measured on each (n+i) period accordingly.
3. In the end, we compute the forecast accuracy measures based on all errors obtained.

Model Selection

Prediction (simulation) error:

$$e_t = y_t - F_t$$

where $\boldsymbol{e_t}$ is the error at period ***t* (*t*={1, 2, 3...N});**

N is the prediction interval (or the size of the dataset);

$\boldsymbol{y_t}$ is the actual value at period ***t*** and

$\boldsymbol{F_t}$ is the forecast for period ***t*.**

Mean Forecast Error (forecast bias):

$$\mathrm{MFE} = \frac{1}{\mathrm{N}}\sum_{t=1}^{N} e_t$$

Two common Measures of Fit

- Measures of fit are used to gauge how well the forecasts match the actual values

MSE (mean squared error)

- – Average squared difference between y_t and F_t

MAD (mean absolute deviation)

- – Average absolute value of difference between y_t and F_t
- – Less sensitive to extreme values

Model Selection

- Mean Absolute Deviation (MAD)
  - Average absolute error – most useful to measure the forecast error in the same units as the original series.

$$\text{MAD} = \frac{\sum \left|\text{Actual} - \text{Forecast}\right|}{n} = \frac{\sum \left|e(t)\right|}{n}$$

Model Selection

- Mean Squared Error (MSE)
  - Average of squared error – provides a penalty for large forecasting errors (it squares each)

$$MSE = \frac{\sum (\text{Actual} - \text{forecast})^2}{n - 1}$$

MSE vs. MAD

Mean Squared Error

$$MSE = \frac{\sum (y_t - F_t)^2}{n-1}$$

Mean Absolute Deviation

$$\mathrm{MAD} = \frac{\sum | y_t - F_t |}{n}$$

where:

y_t = Actual value at time t

F_t = Predicted value at time t

n = Number of time periods

- MSE
  - Squares errors
  - More weight to large errors

MAD

- Easy to compute
- Weights errors linearly

Model Selection

- Mean Percentage Error (MPE)
  - Average percentage error – useful when it is necessary to determine whether a forecasting method is biased. If the forecast is unbiased MPE will produce a % that is close to 0. Large –% means overestimating. Large +% - the method is consistently underestimating.

$$\text{MPE} = \frac{\sum(\text{Actual} - \text{Forecast}) / \text{Actual}}{n} \times 100$$

Walla Walla University
SEVENTH-DAY ADVENTIST HIGHER EDUCATION

Model Selection

- Coefficient of variation of the Root Mean Squared Error, CV(RMSE): The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power and CV(RMSE) helps to compare forecasting errors of different models.

$$\mathrm{CV(RMSE)} = \mathrm{RMSE}/\bar{y} \qquad RMSE = \sqrt{\mathrm{MSE}}$$

Model Selection

Measures of Trueness (Systematic error, Statistical Bias):

- **Mean Percentage Error (MPE)**

$$\text{MPE}\ (\%) = \frac{1}{\text{N}} \sum_{t=1}^{N} (e_t \,/\, y_t)\text{x}100$$

- **Root Mean Squared Error (RMSE)**

$$RMSE = \sqrt{\text{MSE}} \qquad \text{MSE} = \sum (e_t)^2 / (n-1)$$

Walla Walla University
SEVENTH-DAY ADVENTIST HIGHER EDUCATION

Model Selection

- Mean Absolute Percent Error (MAPE) - Puts errors in perspective:
  - Average absolute percent error – useful when the size of the forecast variable is important in evaluating. It provides an indication of how large the forecast errors are in comparison to the actual values of the series. It is also useful to compare the accuracy of different techniques on same/different series.

$$\text{MAPE} = \frac{\Sigma(\left|\text{Actual} - \text{forecast}\right|) / \text{Actual}*100}{n}$$

Model Selection

Measures of Precision (Random Error):

- **Mean Absolute Percentage Error (MAPE)**

$$\text{MAPE}\ (\%) = \frac{1}{\text{N}}\sum_{t=1}^{N}\left(|e_t| / y_t\right)\text{x}100$$

- **Coefficient of Variation of the RMSE, CV(RMSE)**

$$\text{CV(RMSE)} = \text{RMSE}/\overline{\text{y}}$$

Model Selection – a Multiple Criteria Approach

Thank You!

Questions?

Thank You!

and I’ll

See You again…