CH1

Business Pressures–Responses–Support Model

- Business pressures: result of today's competitive business climate
- Responses: to counter the pressures
- Support: to better facilitate the process

Business environment factors: markets, consumer demands, technology, and societal

Organizational Responses: Be Reactive, Anticipative, Adaptive, and Proactive Managers may take actions

Closing the Strategy Gap: closing the gap between the current performance of an organization and its desired performance, as expressed in its mission, objectives, and goals, and the strategy to achieve them.

Management: is a process by which organizational goals are achieved by using **resources**:

- Inputs: resources
- Output: attainment of goals
- Measure of success: outputs / inputs

Decision making: selecting the best solution from two or more alternatives

Mintzberg's 10 Managerial Roles:

Interpersonal	Informational	Decisional
1. Figurehead 2. Leader 3. Liaison	4. Monitor 5. Disseminator 6. Spokesperson	7. Entrepreneur 8. Disturbance handler 9. Resource allocator 10. Negotiator

Decision-Making Process (the scientific approach):

1. Define the problem (or opportunity)
2. Construct a model that describes the real-world problem.
3. Identify possible solutions to the modeled problem and evaluate the solutions.
4. Compare, choose, and recommend a potential solution to the problem.

Degree of Struct redness:

- Decisions are classified as: Highly structured, Semi-structured, Highly unstructured
- Types of Control: Strategic planning (top-level, long-range), Management control (tactical planning), Operational control

Decision Support Systems (DSS): interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems

Business Intelligence (BI): is an evolution of decision support concepts over time, BI is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.

Major objective of BI: is to enable easy access to data to provide business managers with the ability to conduct analysis

BI helps: to transform data, to information (and knowledge), to decisions, and finally to action

The Architecture of BI (major components):

- a data warehouse, with its source data
- business analytics: a collection of tools for manipulating, mining, and analyzing the data in the data warehouse
- business performance management (BPM) for monitoring and analyzing performance
- a user interface (e.g., dashboard)

Simple Taxonomy of Analytics

- Descriptive Analytics
- Predictive Analytics
- Prescriptive Analytics

CH2:

Decision Making: A process of choosing among two or more alternative courses of action for the purpose of attaining a goal(s).

Characteristics of Decision Making:

- Groupthink
- Evaluating what-if scenarios
- Experimentation with a real system!
- Changes in the decision-making environment may occur continuously
- Time pressure on the decision maker
- Analyzing a problem takes time/money

Building successful DSS requires a understanding of these concepts: Decision Making, Support, System, Models and Data

Managerial decision making: is synonymous with the entire management process.

Decision-Making Disciplines:

- Behavioral: anthropology, law, philosophy, political science, psychology, social psychology, and sociology
- Scientific: computer science, decision analysis, economics, engineering, the hard sciences

The manner by which decision makers think and react to problems

- perceive a problem
- cognitive response
- values and beliefs

Phases of Decision-Making Process:

1) Intelligence
2) Design
3) Choice
4) Implementation
5) Monitoring

Potential issues in data/information collection and estimation

1) Lack of data
2) Cost of data collection
3) Data may be insecure
4) Key data may be qualitative
5) Data change over time (time-dependence)

Problem Classification: Classification of problems according to the degree of struct redness

Problem Decomposition: Often solving the simpler subproblems may help in solving a complex problem.

Decision Making (The Design Phase) Have three Models:

- Normative models (= optimization)
- Heuristic models (= sub optimization)
- Descriptive models

Simulation: most common descriptive modeling method (mathematical depiction of systems in a computer environment)

Decision Making The Choice Phase:

- Search approaches
  - Analytic techniques (solving with a formula)
  - Algorithms (step-by-step procedures)
  - Heuristics (rule of thumb)
  - Blind search (truly random search)
- Additional activities: Sensitivity analysis, What-if analysis, Goal seeking

How Decisions are Supported:

- Support for the Intelligence Phase
  - Enabling continuous scanning of external and internal information sources to identify problems and opportunities
  - Resources/technologies: Web; ES, OLAP, data warehousing, data/text/Web mining
  - Business activity monitoring (BAM)
  - Business process management (BPM)
  - Product life-cycle management (PLM)
- Support for the Design Phase
  - Enabling generating alternative courses of action, determining the criteria for choice
  - Generating alternatives
    - **Structured/simple problems:** standard and/or special models
    - **Unstructured/complex problems**: human experts, ES, KMS, brainstorming/GSS, OLAP,
- Support for the Choice Phase
  - Enabling selection of the best alternative given a complex constraint structure
  - Use sensitivity analyses, what-if analyses, goal seeking
    - Resources (KMS, CRM, ERP, and SCM, Simulation and other descriptive models).
- Support for the Implementation Phase
  - Enabling implementation/deployment of the selected solution to the system
  - Decision communication, explanation and justification to reduce resistance to change

DSS Capabilities: it is a system intended to support managerial decisions in semi-structured and unstructured decision.

DSS Classifications (AIS SIGDSS Classification):

1. Communication-driven and group DSS
2. Data-driven DSS
3. Document-driven DSS
4. Knowledge-driven DSS
5. Model-driven DSS

Components of DSS:

1. Data Management Subsystem
   - Includes the database that contains the data
   - Database management system (DBMS)
   - Can be connected to a data warehouse
2. Model Management Subsystem
   - Model base management system (MBMS)
3. User Interface Subsystem
4. Knowledgebase Management Subsystem
   - Organizational knowledge base

DSS Components

- Data Management Subsystem: DSS database, DBMS, Data directory, Query facility.
- Model Management Subsystem: Model base, MBMS, Modeling language, Model directory, Model execution, integration, and command processor
- User Interface Subsystem:
  - **Interface:** Application and User Interface
  - **DSS User Interface:** Portal, Graphical icons, Color coding
  - **Interfacing with PDAs**, cell phones, etc.

CH3:

What is a Data Warehouse?

- A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format
- The data warehouse is a collection of integrated, subject-oriented databases designed to support DSS functions, where each unit of data is non-volatile and relevant to some moment in time

Characteristics of DWs:

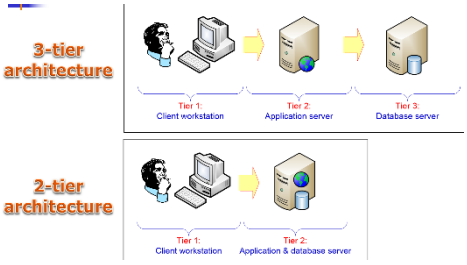
- Subject oriented
- Integrated
- Time-variant (time series)
- Nonvolatile
- Summarized
- Not normalized
- Metadata
- Web based, relational/multi-dimensional
- Client/server, real-time/right-time/active

DW Components (types):

- Data Mart: A departmental small-scale "DW" that stores only limited/relevant data
  - Dependent data mart: A subset that is created directly from a data warehouse
  - Independent data mart: A small data warehouse designed for a strategic business unit or a department
- Operational data stores (ODS): A type of database often used as an interim area for a data warehouse
- Oper marts: an operational data mart
- Enterprise data warehouse (EDW): A data warehouse for the enterprise.
- Metadata: Data about data.

Issues to consider when deciding which architecture to use:

- Which database management system (DBMS) should be used?
- Will parallel processing and/or partitioning be used?
- Will data migration tools be used to load the data warehouse?
- What tools will be used to support data retrieval and analysis?



DW Architecture:

- Three-tier architecture
  1. Data acquisition software (back-end)
  2. The data warehouse that contains the data & software
  3. Client (front-end) software that allows users to access and analyze data from the warehouse
- Two-tier architecture:
  - First two tiers in three-tier architecture is combined into one

Alternative DW Architectures:

- Independent Data Marts Architecture
- Data Mart Bus Architecture with Linked Dimensional Datamarts
- Hub and Spoke Architecture (Corporate Information Factory)
- Centralized Data Warehouse Architecture
- Federated Architecture

Ten factors that potentially affect the architecture selection decision:	
1. Information interdependence between organizational units 2. Upper management's information needs 3. Urgency of need for a data warehouse 4. Nature of end-user tasks 5. Constraints on resources	6. Strategic view of the data warehouse prior to implementation 7. Compatibility with existing systems 8. Perceived ability of the in-house IT staff 9. Technical issues 10. Social/political factors

Data Integration and the Extraction, Transformation, and Load Process:

- ETL = Extract Transform Load
- Data integration: Integration that comprises three major processes: data access, data federation, and change capture.
- Enterprise application integration (EAI): A technology that provides a vehicle for pushing data from source systems into a data warehouse
- Enterprise information integration (EII): An evolving tool space that promises real-time data integration from a variety of sources, such as relational or multidimensional databases, Web services

Data warehouse development approaches:

- Inmon Model: EDW approach (top-down)
- Kimball Model: Data mart approach (bottom-up)

Representation of Data in DW:

- Dimensional Modeling: A retrieval-based system that supports high-volume query access
- Star schema: The most commonly used and the simplest style of dimensional modeling
  - Contain a fact table surrounded by and connected to several dimension tables
- Snowflakes schema: An extension of star schema where the diagram resembles a snowflake in shape

Multidimensionality: The ability to organize, present, and analyze data by several dimensions, such as sales by region, by product, by salesperson, and by time (four dimensions)

Multidimensional presentation:

- Dimensions: products, salespeople, market segments, business units, geographical locations, distribution channels,
- Measures: money, sales volume, head count, inventory profit, actual versus forecast
- Time: daily, weekly, monthly, quarterly, or yearly

OLAP (online transaction processing) vs. OLTP (Online analytical processing):

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data)	Data warehouse or data mart (a nonnormalized data)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

OLAP Operations:

- Slice - a subset of a multidimensional array
- Dice - a slice on more than two dimensions
- Drill Down/Up - navigating among levels of data ranging from the most summarized (up) to the most detailed (down)
- Roll Up - computing all the data relationships for one or more dimensions
- Pivot - used to change the dimensional orientation of a report or an ad hoc query-page display

Variations of OLAP:

- Multidimensional OLAP (MOLAP): OLAP implemented via a specialized multidimensional database (or data store) that summarizes transactions into multidimensional views ahead of time
- Relational OLAP (ROLAP): The implementation of an OLAP database on top of an existing relational database

DW Implementation Issues:

- Identification of data sources and governance
- Data quality planning, data model design
- Establishment of service-level agreements
- Data transport, data conversion
- End-user support
- Political issues

Successful DW Implementation Things to Avoid:

- Starting with the wrong sponsorship chain
- Setting expectations that you cannot meet
- Engaging in politically naive behavior
- Loading the data warehouse with information just because it is available

Failure Factors in DW Projects:

- Lack of executive sponsorship
- Unclear business objectives
- Cultural issues being ignored
- Unrealistic expectations

CH4:

What is a Business Report?: A written document that contains information regarding business matters.

- Purpose: to improve managerial decisions
- Source: data from inside and outside the organization (via the use of ETL)
- Format: text + tables + graphs/charts
- Distribution: in-print, email, portal/intranet

Key to Any Successful Report: Clarity, Brevity, Completeness, Correctness.

Report types (in terms of content and format)

- Informal: a single letter or a memo
- Formal: 10-100 pages; cover + summary + text
- Short report: periodic, informative, investigative

Types of Business Reports:

- Metric Management Reports: Help manage business performance through metrics (SLAs for externals; KPIs for internals). Can be used as part of Six Sigma and/or TQM
- Dashboard-Type Reports: Graphical presentation of several performance indicators in a single page using dials/gauges
- Balanced Scorecard-Type Reports: Include financial, customer, business process, and learning & growth indicators

Components of Business Reporting Systems:

- OLTP (online transaction processing)
- Data supply (volume, variety, velocity, …)
- ETL
- Data storage
- Business logic
- Publication medium
- Assurance

Data and Information Visualization: he uses of visual representations to explore, make sense of, and communicate data.

Types of Charts and Graphs:

Basic Charts and Graphs	Specialized Charts and Graphs
Line Chart	Histogram
Bar Chart	Gantt Chart
Pie Chart	PERT Chart
Scatter Plot	Geographic Map
Bubble Chart	Bullet Graph

Visual Analytics:

- Information visualization: Descriptive, backward focused "what happened" "what is happening"
- Predictive analytics: Predictive, future focused "what will happen" "why will it happen"

Performance Dashboards: Performance dashboards are commonly used in BPM software suites and BI platforms

Dashboard design: The fundamental challenge of dashboard design is to display all the required information on a single screen, clearly and without distraction

Three layers of information in Dashboard design: Monitoring, Analysis, Management

Best Practices in Dashboard Design:

- Benchmark KPIs with Industry Standards
- Wrap the Metrics with Contextual Metadata
- Present Information in Three Different Levels
- Pick the Right Visual Constructs
- Provide for Guided Analytics

What to look for in a dashboard:

- Use of visual components to highlight data and exceptions that require action.
- Transparent to the user, meaning that they require minimal training and are extremely easy to use
- Combine data from a variety of systems into a single, summarized, unified view of the business
- Enable drill-down or drill-through to underlying data sources or reports
- Present a dynamic, real-world view with timely data
- Require little coding to implement/deploy/maintain

Business Performance Management (BPM): A real-time system that alerts managers to potential opportunities, impending problems and threats, and then empowers them to react through models and collaboration.

BPM: refers to the business processes, methodologies, metrics, and technologies used by enterprises to measure, monitor, and manage business performance.

BPM Also called:

- corporate performance management (CPM)
- enterprise performance management (EPM)
- strategic enterprise management (SEM)

BPM encompasses three key components

- A set of integrated, closed-loop management and analytic processes supported by technology …
- Tools for businesses to define strategic goals and then measure/manage performance against them
- Methods and tools for monitoring key performance indicators (KPIs), linked to organizational strategy

A Closed-Loop Process Steps:

1. Strategize
2. Plan
3. Monitor/analyze
4. Act/adjust

Common tasks for the strategic planning process:

1. Conduct a current situation analysis
2. Determine the planning horizon
3. Conduct an environment scan
4. Identify critical success factors
5. Complete a gap analysis
6. Create a strategic vision
7. Develop a business strategy
8. Identify strategic objectives and goals

Operational plan: plan that translates an organization's strategic objectives and goals into a set of well-defined tactics and initiatives, resources requirements, and expected results for some future time period (usually a year). can be:

- Tactic-centric (operationally focused)
- Budget-centric plan (financially focused)

Performance measurement system: A system that assists managers in tracking the implementations of business strategy by comparing actual results against strategic goals and objectives.

Key performance indicator (KPI): A KPI represents a strategic objective and metrics that measure performance against a goal

Distinguishing features of KPIs:

- Strategy
- Targets
- Ranges
- Encodings
- Time frames
- Benchmarks

Key performance indicator (KBI):

- Outcome KPIs: lagging indicators e.g., revenues.
- Driver KPIs: leading indicators e.g., sales leads

Operational areas covered by driver KPIs:

- Customer performance
- Service performance
- Sales operations
- Sales plan/forecast

Performance Measurement System:

- Balanced Scorecard (BSC): A performance measurement and management methodology that helps translate an organization's financial, customer, internal process, and learning and growth objectives and targets into a set of actionable initiatives
- Six Sigma: A performance management methodology aimed at reducing the number of defects in a business process to as close to zero defects per million opportunities (DPMO) as possible
- The DMAIC performance model: A closed-loop business improvement model that encompasses the steps of defining, measuring, analyzing, improving, and controlling a process

Comparison of Balanced Scorecard and Six Sigma:

Balanced Scorecard	Six Sigma
• Strategic management system • Relates to the longer-term view of the business • Designed to develop balanced set of measures • Identifies measurements around vision and values • Critical management processes are to clarify vision/strategy Balances customer and internal operations without a clearly defined leadership role • Emphasizes targets for each measurement • Emphasizes learning of executives based on the feedback • Focuses on growth • Heavy on strategic content • Management system consisting of measures	• Performance measurement system • Provides snapshot of business's performance and identifies measures that drive performance toward profitability • Designed to identify a set of measurements that impact profitability • Establishes accountability for leadership for wellness and profitability • Includes all business processes-management and operational • Balances management and employees' roles; balances costs and revenue of heavy processes • Emphasizes aggressive rate of improvement for each measurement, irrespective of target • Emphasizes learning and innovation at all levels based on the process feedback; enlists all employees' participation • Focuses on maximizing profitability

CH5:

Definition of Data Mining: The nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases.

- Keywords in this definition: Process, nontrivial, valid, novel, potentially useful, understandable.
- Other names for Data Mining: knowledge extraction, pattern analysis, knowledge discovery, information harvesting, pattern searching, data dredging

Data Mining Characteristics/Objectives:

- Source of data for DM is often a consolidated data warehouse (not always!).
- DM environment is usually a client-server or a Web-based information systems architecture.
- Data is the most critical ingredient for DM which may include soft/unstructured data.
- The miner is often an end user
- Striking it rich requires creative thinking

Data: a collection of facts usually obtained as the result of experiences, observations, or experiments.

Data: lowest level of abstraction (from which information and knowledge are derived).

What Does DM Do: DM extract patterns from data.

Pattern: A mathematical (numeric and/or symbolic) relationship among data items

Types of patterns

• Association	• Cluster (segmentation)
• Prediction	• Sequential (or time series) relationships

Data Mining Applications:

- Customer Relationship Management
- Banking & Other Financial
- Retailing and Logistics
- Manufacturing and Maintenance
- Insurance

Data Mining Process:

- A manifestation of best practices
- A systematic way to conduct DM projects
- Different groups has different versions
- Most common standard processes:
  - CRISP-DM (Cross-Industry Standard Process for Data Mining)
  - SEMMA (Sample, Explore, Modify, Model, and Assess)
  - KDD (Knowledge Discovery in Databases)

Data Mining Process: CRISP-DM:

- Step 1: Business Understanding
- Step 2: Data Understanding
- Step 3: Data Preparation (!)
- Step 4: Model Building
- Step 5: Testing and Evaluation
- Step 6: Deployment

Data Mining Methods: Classification:

- Most frequently used DM method
- Part of the machine-learning family
- Employ supervised learning
- Learn from past data, classify new data
- The output variable is categorical (nominal or ordinal) in nature

Assessment Methods for Classification:

- Predictive accuracy: Hit rate
- Speed: Model building; predicting
- Robustness
- Scalability
- Interpretability: Transparency, explain ability

Classification Techniques:

- Decision tree analysis
- Statistical analysis
- Neural networks
- Support vector machines
- Case-based reasoning
- Bayesian classifiers
- Genetic algorithms

Decision Trees algorithms:

- Splitting criteria - (Which variable, what value, etc.)
- Stopping criteria - (When to stop building the tree)
- Pruning (generalization method) - (Pre-pruning versus post-pruning)

Alternative splitting criteria:

- Gini index determines the purity of a specific class as a result of a decision to branch along a particular attribute/value
- Information gain uses entropy to measure the extent of uncertainty or randomness of a particular attribute/value split
- Chi-square statistics

Cluster Analysis for Data Mining:

- Used for automatic identification of natural groupings of things
- Part of the machine-learning family
- Employ unsupervised learning
- Learns the clusters of things from past data, then assigns new instances
- There is not an output variable
- Also known as segmentation

Clustering results may be used to:

- Identify natural groupings of customers
- Identify rules for assigning new cases to classes for targeting/diagnostic purposes
- Provide characterization, definition, labeling of populations
- Decrease the size and complexity of problems for other data mining methods
- Identify outliers in a specific domain (e.g., rare-event detection)

Cluster Analysis methods:

- Statistical methods.
- Neural networks
- Fuzzy logic
- Genetic algorithms

Step for K-Means Clustering Algorithm:

- *k* : pre-determined number of clusters
- Algorithm (Step 0: determine value of *k*)
- Step 1: Randomly generate *k* random points as initial cluster centers.
- Step 2: Assign each point to the nearest cluster center.
- Step 3: Re-compute the new cluster centers.

Association Rule Mining:

- A very popular DM method in business
- Finds interesting relationships (affinities) between variables (items or events)
- Part of machine learning family
- Employs unsupervised learning
- There is no output variable
- Also known as market basket analysis

Common Data Mining Blunders:

1. Selecting the wrong problem for data mining
2. Ignoring what your sponsor thinks data mining is and what it really can/cannot do
3. Not leaving insufficient time for data acquisition, selection and preparation
4. Looking only at aggregated results and not at individual records/predictions
5. Being sloppy about keeping track of the data mining procedure and results

Association rules algorithms: The algorithms help identifies the frequent item sets, then converted to association rules

- Apriori
- Eclat
- FP-Growth

CH6:

Neural networks (NN): a brain metaphor for information processing

Artificial neural network (ANN)

- Many uses for ANN for - pattern recognition, forecasting, prediction, and classification
- Many application areas - finance, marketing, manufacturing, operations, information systems

Biological	Artificial
Soma	Node
Dendrites	Input Output
Axon	Weight
Synapse	Fast
Slow	Few neurons (~100s)
Many neurons (10^9)	

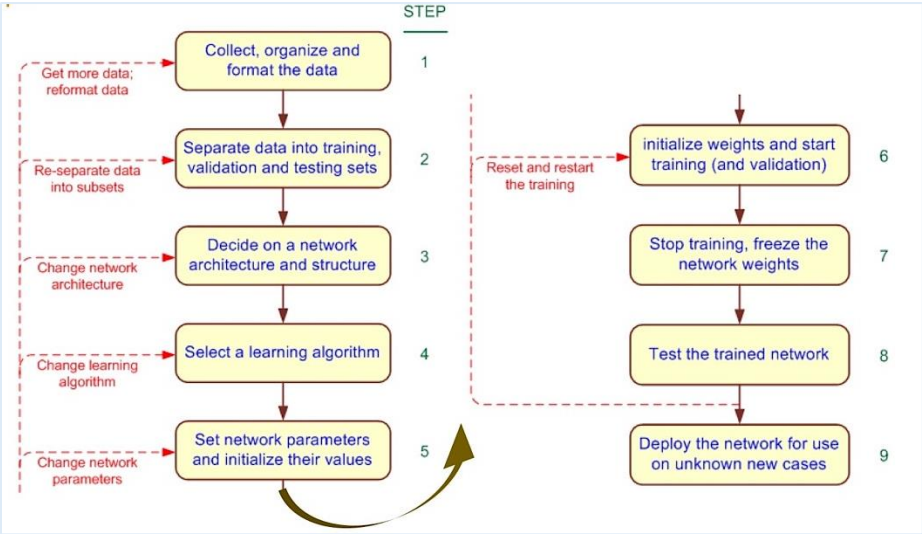
Element of ANN:

- Processing element (PE)
- Network architecture - Hidden layers & Parallel processing
- Network information processing:
  - Inputs
  - Outputs
  - Connection weights
  - Summation function

Architecture of a neural network is driven by the task it is intended to address (Classification, regression, clustering, general optimization, association)

- Feedforward, multi-layered perceptron with backpropagation learning algorithm
- Recurrent, self-organizing feature maps, Hopfield networks

Process of ANN:



AN Learning Process - A Supervised Learning Process

1. Compute temporary outputs.
2. Compare outputs with desired targets.
3. Adjust the weights and repeat the process.

Support Vector Machines (SVM):

- SVM are among the most popular machine-learning techniques.
- SVM belong to the family of generalized linear models… (capable of representing non-linear relationships in a linear fashion).
- SVM achieve a classification or regression decision based on the value of the linear combination of input features.
- Because of their architectural similarities, SVM are also closely associated with ANN.

Goal of SVM: to generate mathematical functions that map input variables to desired outputs for classification or regression type prediction problems.

A hyperplane is a geometric concept used to describe the separation surface between different classes of things.

A kernel function in SVM uses the kernel trick (a method for using a linear classifier algorithm to solve a nonlinear problem).

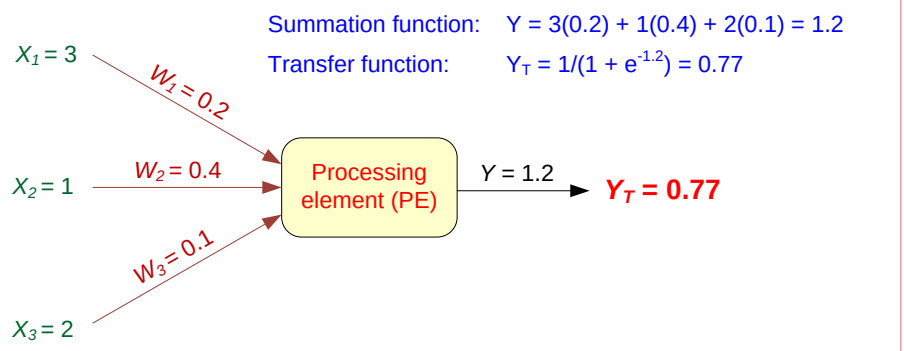
Cross Validation: is often used to determine the best value for k and the distance measure

How Does an SVM Work?

- Following a machine-learning process, an SVM learns from the historic cases.

SVM Applications: SVMs are the most widely used kernel-learning algorithms for wide range of classification and regression problems

(العملي مهم جدا)



$$d(i, j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \ldots + |x_{ip} - x_{jp}|^q)}$$

If $q = 1$, then d is called Manhattan distance

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \ldots + |x_{ip} - x_{jp}|$$

If $q = 2$, then d is called Euclidean distance

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \ldots + |x_{ip} - x_{jp}|^2)}$$