Unit - IV

Data Analysis

Vector Data Analysis tools - Data Analysis Tools - Network Analysis - Digital Elevation models - 3D data collection and utilisation.

Vector Data Analysis:

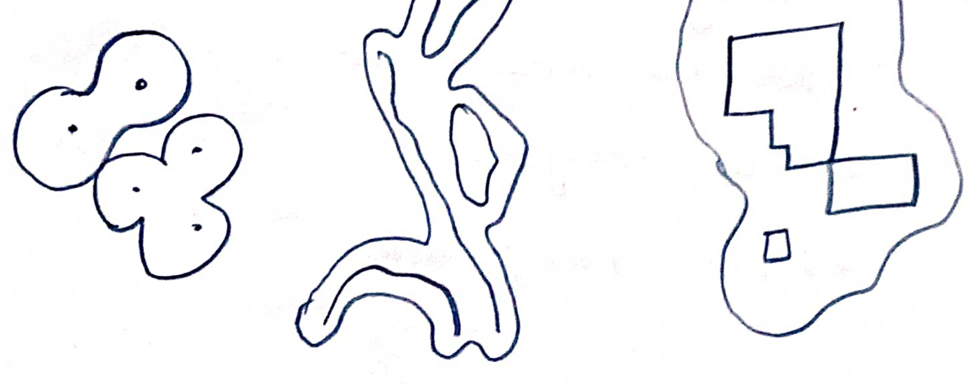
The vector data model uses points and their x, y - coordinates to construct spatial features of points, lines and polygons. These spatial features are used as inputs in vector data analysis. Therefore, the accuracy of data analysis depends on the accuracy of these features in terms of their location and shape and whether they are topological or not. Additionally, it is important to note that an analysis may apply to all, or selected, features in a layer. The following are the types of analysis used with vector data.

* Buffering
* Overlay
* Distance measurement
* Pattern Analysis
* Feature Manipulation.

Buffering:

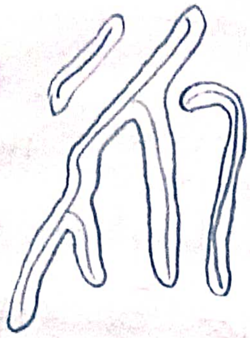
Based on the concept of proximity, buffering creates two areas; one area that is within a specified distance of select features and the other area that is beyond. The area within the specified distance is the buffer zone. A GIS typically varies the value of an attribute to seperate the buffer zone (eg. 1) from the area beyond the buffer zone (eg. 0). Besides the designation of the buffer zone, no other attribute data are added or combined.

Features for buffering may be points, lines or polygons. Buffering around points creates circular buffer zones. Buffering around lines creates a series of elongated buffer zones around each line segment. And buffering around polygons creates buffer zones that extend outward from the polygon boundaries.

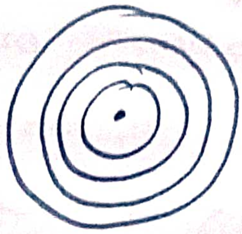


Variations in Buffering:

The buffer distance or buffer size does not have to be constant; It can vary according to the values of a given field. For example, the width of the riparian buffer can vary depending on its expanded function and the intensity of adjacent land use. A feature may have more than one buffer zone. As an example, a nuclear power plant may be buffered with distances of 5, 10, 15 and 20 miles, ~~As an example a nuclear power plant~~ the rings are not equal in area. The second ring from the plant, in fact, covers an area about three times larger than the first ring. One must consider this area difference if the buffer zones are part of an evacuation plan. Likewise, buffering around line features does not have to be on both sides of the lines, it can be on either the left side or the right side of the line feature. Likewise, buffer zones around polygons can be extended either outward or inward from the polygon boundaries. Boundaries of buffer zones may remain intact so that each buffer zone is a seperate polygon for further anlysis.



Buffering with different buffer distance



Buffering with four levels.

Regardless of its variations, buffering uses distance measurements from select features to create the buffer zones. Because buffering uses distance measurements from spatial features, the positional accuracy of spatial features in a data set also determines the accuracy of buffer zones.

Applications of Buffering:

Most applications of buffering are based on buffer zones. A buffer zone is often treated as a protection zone and is used for planning or regulatory purpose.

* Government regulations may set 2-mile buffer zones along streams to minimize sedimentation from logging operations.

* A national forest may restrict oil and gas well drilling within 500 feet of roads or highways.

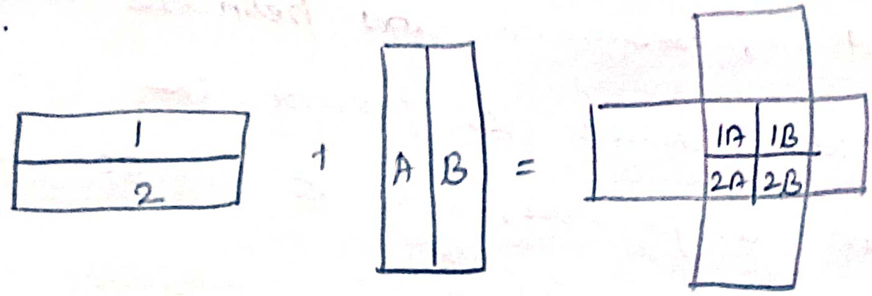
* A planning agency may set aside land along the edges of streams to reduce the effects of nutient, sediment and pesticide runoff, to maintain shade to prevent the rise of stream temperature, and to provide shelter for wildlife and aquatic life.

* A planning agency may create buffer zones around geographic features such as water, wetlands, critical habitats and wells to be protected and exclude these zones from landfill considerations.

Overlay:

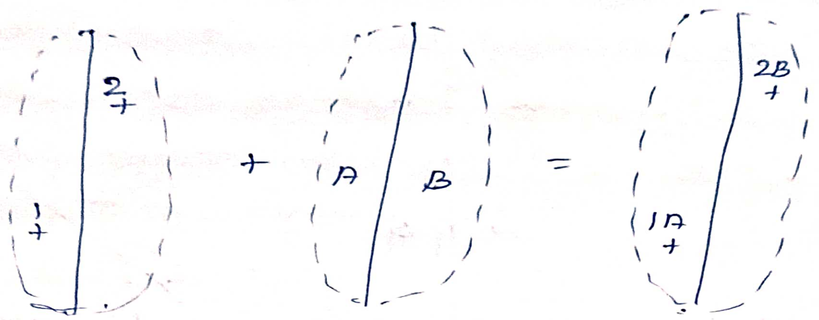
An overlay operation combines the geometrics and attributes of two feature layers to create the output. The geometry of the output represents the geometric intersection of features from the input layers. Below figure illustrates an overlay operation with two simple polygon layers.

Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbours. Feature layers to be overlaid must be spatially registered and based on the same coordinate system.



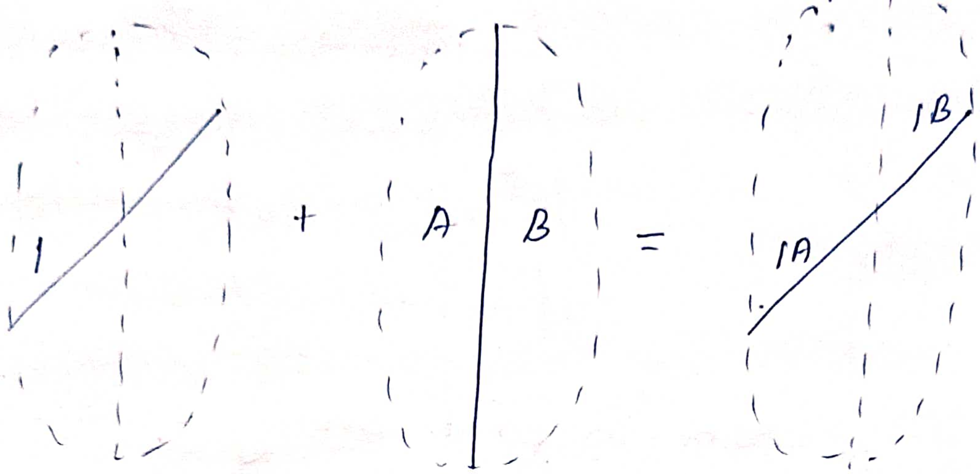
Feature Type and Overlay:

Overlay operations can take polygon, line or point layers as the inputs and create an output of a lower-dimension feature type. For eg, given the inputs of polygon and line layers, the output will be a line layer. Without going through all different combinations of feature types, this section covers three common overlay operations: point-in polygon, line-in polygon and polygon-on-polygon. To distinguish the layers in the following discussion, the layer that may be a point, line or polygon layer is called the inner layer, and the layer that is a polygon layer is called the overlay layer.

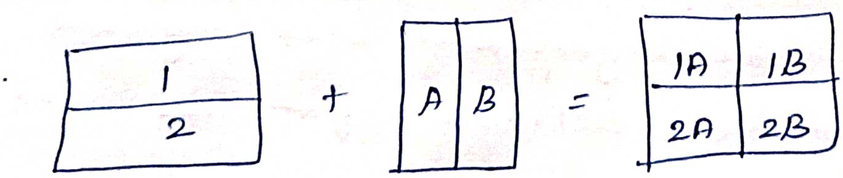


In a point-in-point overlay operation, the same point features in the input layer are included in the output but each point is assigned with attributes of the polygon within which it falls. For eg, a point-in-point overlay can find the association between wildlife locations and vegetation types.

In a line-in-polygon overlay operation, the output contains the same line features as in the input layer but each line features is dissected by the polygon boundaries on the overlay layer. Thus the output has more line segments than does the inner layer. Each line segment on the output combines attributes from the inner layer and the underlying polygon. For example, a line-in-polygon overlay can find soil data for a proposed road. The input layer includes the proposed road. The overlay layer contains soil polygons.



The most common overlay operation is polygon-on-polygon, involving two polygon layers. The output combines the polygon boundaries from the input and overlay layers to create a new set of polygons. Each new polygon carries attributes from both layers, and these attribute differ from those of adjacent polygon. For eg, a polygon-on-polygon overlay can analyze the association between elevation zones and vegetation types.



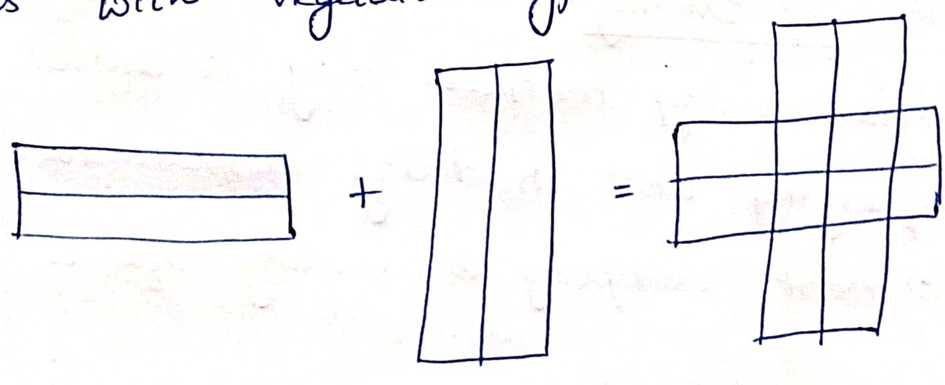
Overlay Methods:

The overlay methods are based on the Boolean connectors AND, OR and XOR. Intersect uses the AND connector. Union uses the OR connector. Symmetrical difference or difference uses XOR connector. Identity or Minus uses the following expression:

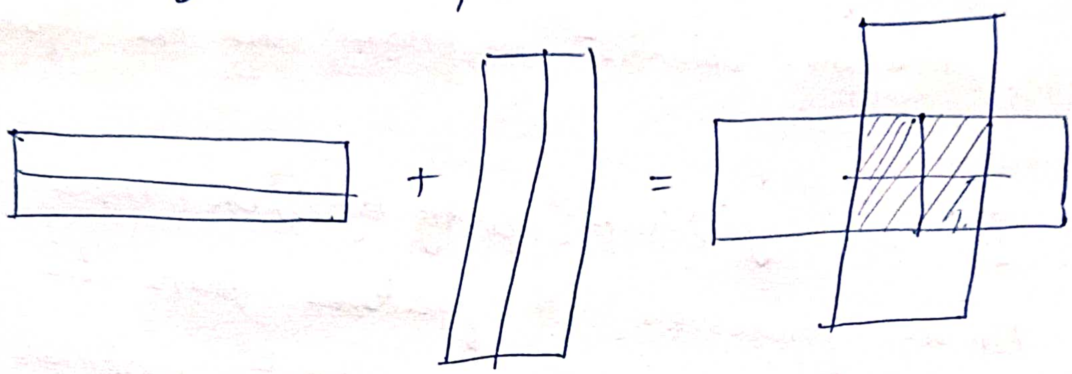
[(input layer) AND (identity layer)] OR (input layer)

Union preserves all features from the inputs. The area extend of the output combines the area extents of both input layers.

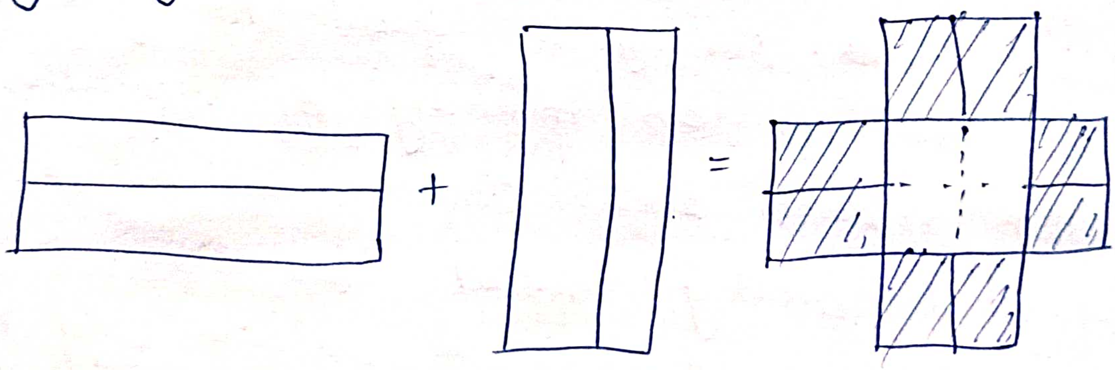
Intersect preserves only those features that falls within the area extent common to the inputs. For eg, a forest management plan may call for an inventory of vegetation types within riparian zones. Intersect will be a more efficient overlay method than Union in this case because the output contains only riparian zones with vegetation types.



Symmetrical Difference preserves features that fall within the area extent that is common to only one of the inputs. In other words, symmetrical difference is opposite to intersect in terms of the output's area extent.



Identity preserves only features that fall within the area extent of the layer defined as the input layer. The other layer is called the identity layer.



Applications of Overlay:

The overlay methods play a central role in many querying and modeling applications. Suppose an investment company is looking for a land parcel that is zoned commercial, not subject to flooding, and not more than 1 mile from a heavy-duty road. The company can first create a 1-mile road buffer and overlay the buffer zone layer with the zoning and floodplain layers. A subsequent query of the overlay output can select land parcels that satisfy the company's selection criteria.

A more specific application of overlay is to help solve the aereal interpolation problem. Areal interpolation involves transfering known data from one set of polygons to another. For eg, census tracts may represent source polygons with known populations in each tract from the U.S. Census Bureau, and school districts may represent target polygons with unknown population data. Using overlay the population of the school districts can be calculated using the population given in the census tract.

Data Analysis Tools:

i) **Distance Measurement:**

Distance measurement refers to measuring straight line distances between features. Measurements can be made from points in a layer to points in another layer, or from each point in a layer to its nearest point or line in another layer. In both cases, distance measures are stored in a field. Distance measures can be used directly for data analysis.

Pattern Analysis:

Pattern analysis is the study of the spatial arrangements of points or polygon features in two dimentional space. Pattern analysis uses distance measurements as inputs and statistics for describing the distribution pattern. At the general level, a pattern analysis can reveal if a point distribution pattern is random, dispersed or clustered.

A classic technique for point pattern analysis, nearest neighbour analysis uses the distance between each point and its closest neighboring point in a layer to determine if the point pattern is random, regular, or clustered. The nearest neighbour statistic is the ratio (R) of the observed average distance between nearest neighbours (dobs) to the expected average for a hypothetical random distribution (dexp).

$$R = \frac{d_{obs}}{d_{exp}}$$

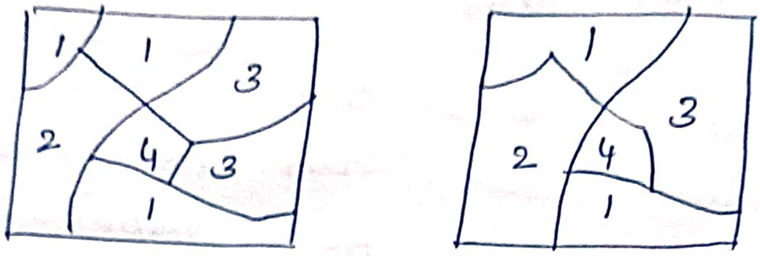
The ratio R is less than 1 if the point pattern is more clustered than random and greater than 1 if the point pattern is more dispersed than random.

__Feature Manipulation:__

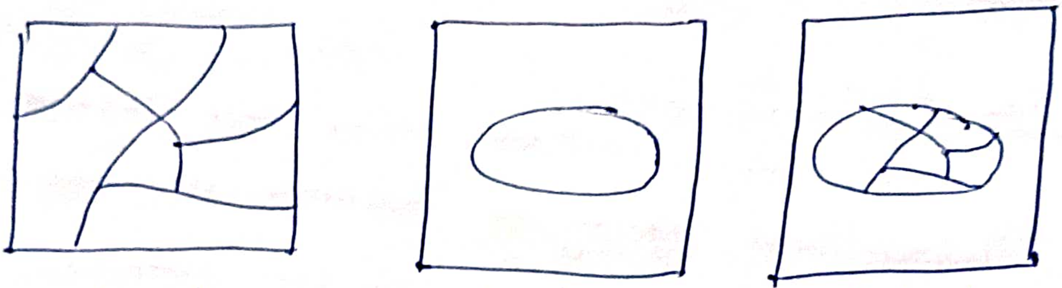
Tools are available in a GIS package for manipulating and managing features in one or more feature layers. When a tool involves two layers, the layers must be based on the same co-ordinate system.

Like overlay, these features tools are often needed for data preprocessing and data analysis, however unlike overlay, these tools do not combine geometries and attributes from input layers into a single layer.

Dissolve aggregates features in a feature layer that have the same attribute value or values. For example, we can aggregate roads by highway number or counties by state. An important application of Dissolve is to simplify a classified polygon layer.

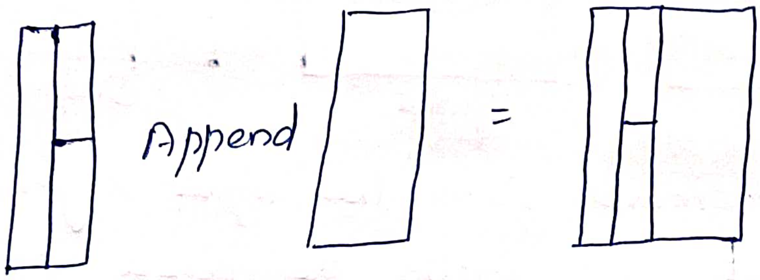


(a) (b)


Input layer Clip Layer Output.

Clip creates a new layer that includes only those features of the input layer, including their attributes that fall within the area extent of the clip layer. Clip is a useful tool, for example for cutting a map acquired elsewhere to fit a study area.

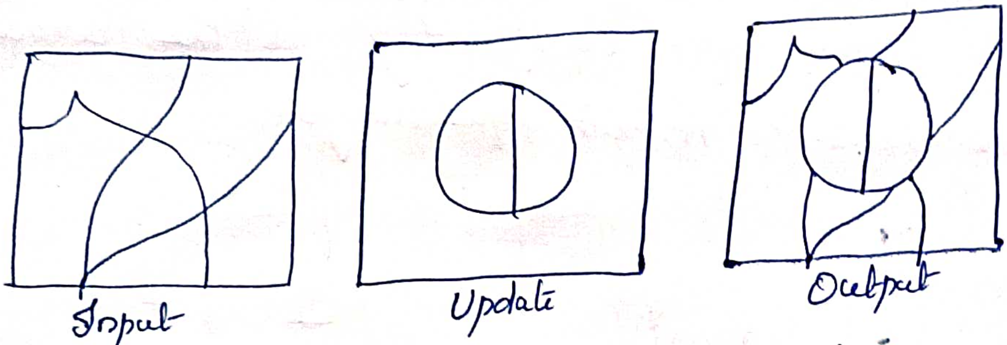
Append creates a new layer by piecing together two or more layers, which represent the same feature and have the same attributes. For eg, append can put together a layer of four input layers, each corresponding to the area extent of a USGS-7.5 minute quadrangle.



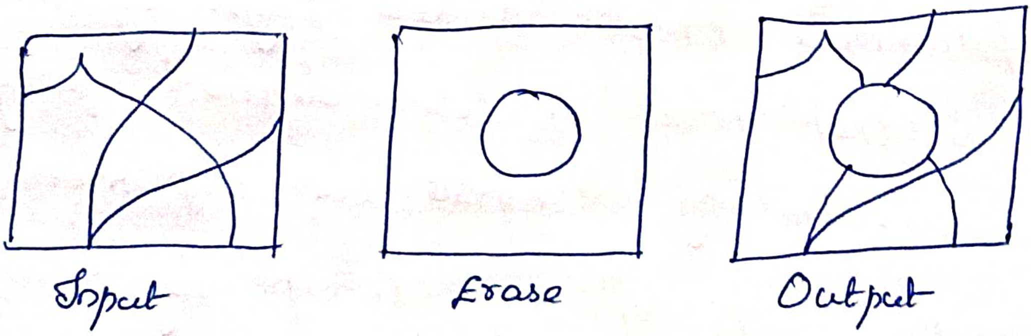
Select creates a new layer that contains features selected from a user-defined query expression. For eg, we can create a layer showing high-canopy closure by selecting stands that have 60 to 80 percent closure from a stand layer.

Eliminate creates a new layer by removing features that meet a user-defined query expression. For eg, eliminate can implement the minimum mapping unit concept by removing polygons that are smaller than the defined unit in a layer.

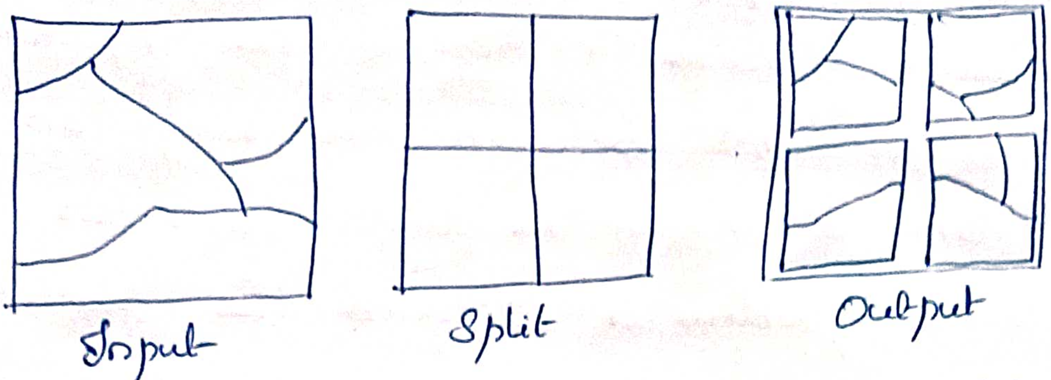
Update uses a "cut and paste" operation to replace the input layer with the update layer and its features. As the name suggests, update is useful for updating an existing layer.



Erase removes from the input layer those features that fall within the area extent of the erase layer. Suppose a suitability analysis stipulates that potential sites cannot be within 300 meters of any stream. A stream buffer layer can be used in this case as the erase layer to remove itself from further consideration.



Split divides the input layer into two or more layers. A split layer, which shows area subunits, is used as the template for dividing the input layer.



For eg, a national forest can split a stand layer by district so that each district office can have its own layers.

Network Analysis:

A network with an appropriate attribute can be used for a variety of applications. Some applications are directly accessible through GIS tools. Others, require the integration of GIS and specialized software in operations research and management science.

Shortest Path Analysis:

Shortest path analysis finds the path with the minimum cumulative impedance between nodes on a network. Because the link impedance can be measured in distance or time, a shortest path may represent the shortest route or the fastest route. Shortest path analy. typically begins with an impedance matrix in which a value represents the impedance of a direct link between two nodes on a network and an ∞ means no direct connection.

The problem is to find the shortest distances from a node to all other nodes. A variety of algorithm can be used to solve the problem, among them the most commonly used algorithm is Dijkstra's Algorithm.

Dijkstra's Algorithm:

The idea of the algorithm is very simple.

1. It maintains a list of unvisited vertices.
2. It chooses a vertex and asigns a maximum possible cost to every other vertex.
3. the cost of the source remains zero as it actually takes nothing to reach from the source vertex to itself.
4. In every subsequent step of the algorithm it tries to improve (minimize) the cost for each vertex. Here, the cost can be distance, money or time taken to reach that vertex from the source vertex. The minimization of cost is a multi-step process.

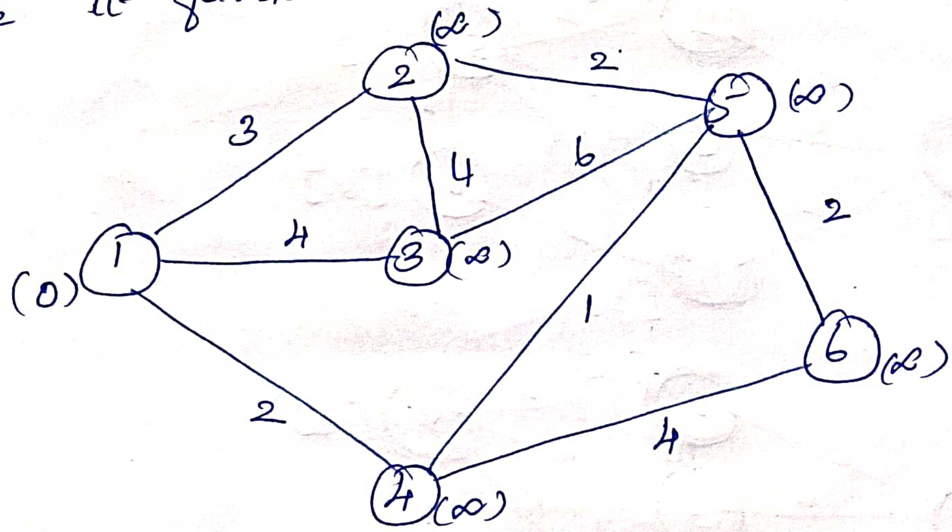
a) For each unvisited neighbour(2,3,4) of the current vertex (1) calculate the new cost from vertex (1)

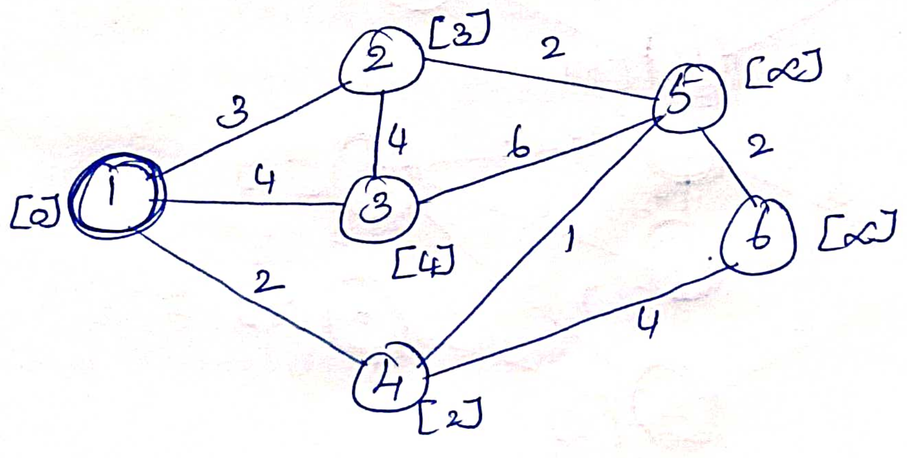
b) For eg, the new cost of vertex 2 is calculated as the minimum of the two or sum.

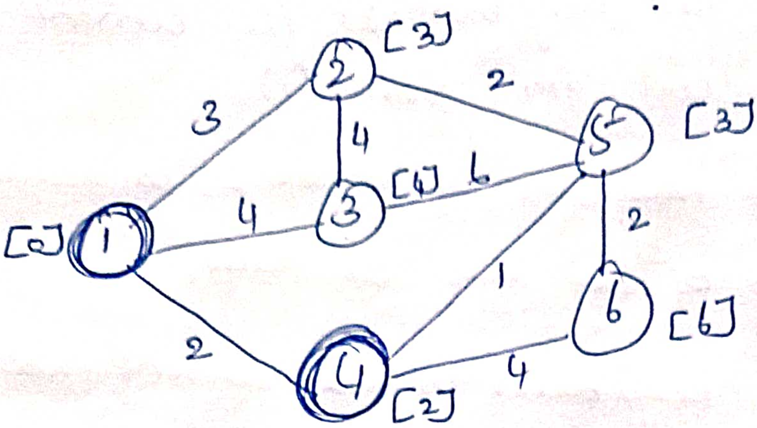
5) When all the neighbours of the current node are considered, it marks the current node as visited and is removed from the unvisited list.

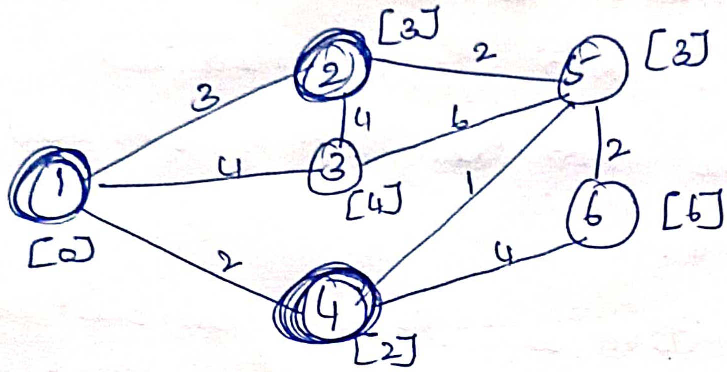
6) Select a vertex from the list of unvisited node and repeat step 4.

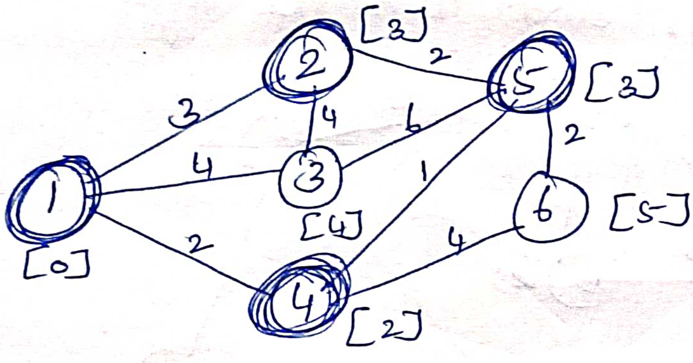
7) At the end there will be no possibilities to improve it further and then the algorithm ends.

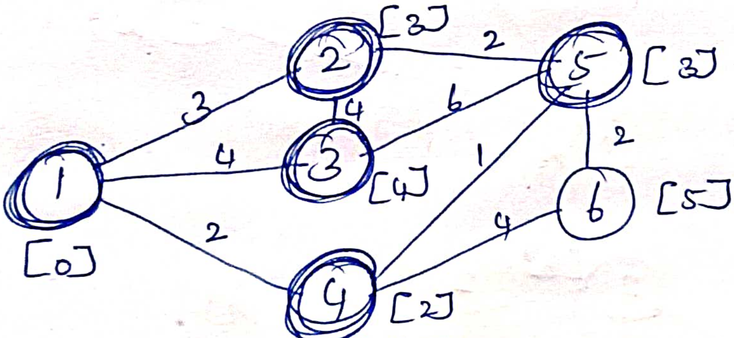


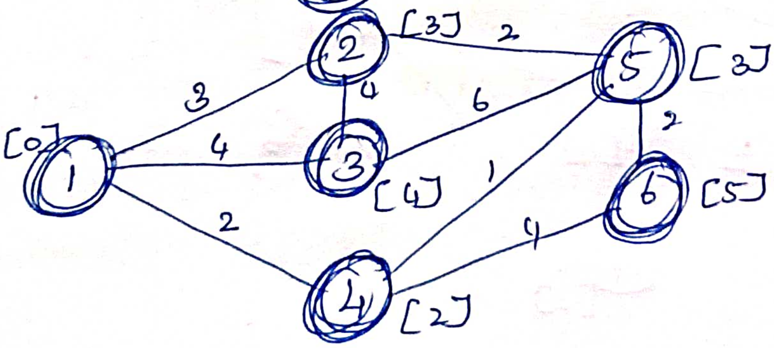



1 [0]
2 [3]
3 [4]
4 [2]
5 [3]
6 [6]
3 4 2 4 6 2 1 4 2


1 [0]
2 [3]
3 [4]
4 [2]
5 [3]
6 [6]
3 4 2 4 6 2 1 4 2


1 [0]
2 [3]
3 [4]
4 [2]
5 [3]
6 [5]
3 4 2 4 6 2 1 4 2


1 [0]
2 [3]
3 [4]
4 [2]
5 [3]
6 [5]
3 4 2 4 6 2 1 4 2


1 [0]
2 [3]
3 [4]
4 [2]
5 [3]
6 [5]
3 4 2 4 6 2 1 4 2

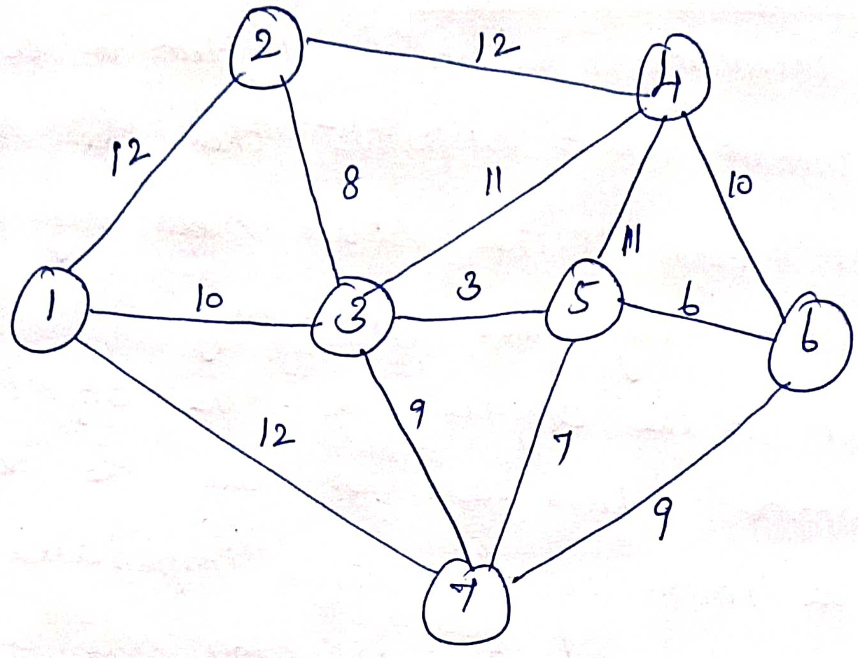
Traveling Salesman Problem:

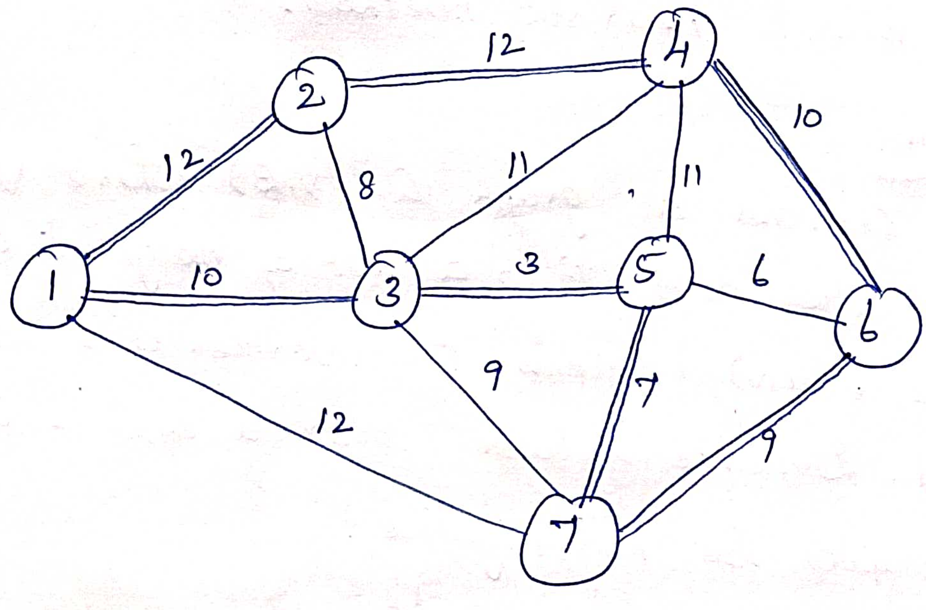
The traveling salesman problem is a routing problem, which stipulates that the salesman must visit each of the select stops only once and the salesman may start from any stop but must return to the original stop. The objective is to determine which route or tour the salesman can take to minimize the total impedance value. A common solution to the traveling salesman problem uses a heuristic method.

Algorithm:

Let us consider a graph $G=(V,E)$, where V is a set of cities and E is a set of weighted edges. An edge $e(u,v)$ represents that vertices u and v are connected. Distance between vertex u and v is $d(u,v)$ which should be non-negative.

1. Select the source node
2. Mark the current node as selected and compute the distance to all the adjacent nodes
3. Choose the adjacent node with a minimum distance.
4. Repeat the steps 2 and 3 till all the nodes are marked as selected.





Vehicle Routing Algorithms:

The vehicle routing problem is an extension of traveling salesman problem. Given a fleet of vehicles and customers, the main objective of the vehicle routing problem is to schedule vehicle routes and visits to customers in such a way that the total travel time is minimized. Additional constraints such as time windows, vehicle capacity and dynamic conditions may also exist. Because vehicle routing involves complex modeling applications, it requires the integration of GIS and special routing software in operations research and management science.

Digital Elevation Model (DEM):

Digital Elevation Model is the digital representation of the land surface elevation with respect to any reference datum. DEM is frequently used to refer to any digital representation of a topographic surface. DEM is the simplest form of digital representation of topography.

DEM's are used to determine terrain attributes such as elevation at any point, slope and aspect. Terrain features like drainage basins and channel networks can also be identified from the DEMs. DEMs are widely used in hydrologic and geologic analysis, hazard monitoring, natural resources exploration, agricultural management, etc.

Hydrologic applications of the DEM include groundwater modeling, estimation of the volume of proposed reservoirs, determining landslide probability, flood prone area mapping etc.

Three main type of structures are used

a) Regular square grids

b) Triangulated Irregular Networks

c) Contours.

A digital elevation model is a digital model or 3D representation of a terrain's surface, created from terrain elevation data.

There are three similar names as DEM, Digital Terrain Model and Digital Surface Model. DEM is a subset of DTM, which also represents other morphological elements.

Creation of DEM's.

Several methods are available to create DEM

a) Conversion of printed Contour lines

The first method is conversion of printed contour lines and use it in raster or vector form

The elevation contours are "tagged" with elevations. Any other additional elevation data are created from hydrography layer.

b) Photogrammetry:

This can be done manually or automatically.

i) Manually, an operator looks at a pair of stereophotos through a stereoplotter and must move two dots together until they appear to be one lying just at the surface of the ground.

ii) Automatically, an instrument calculates the parallax displacement of a large number of points.

3D Data Collection and Utilisation:

Similar to 2D maps, 3D GIS maps depict objects in greater detail by adding another dimension. 3D technology in GIS maps is explanatory illustrations that represent the scale of real-world objects.

3D models assist appearance, survey in a large number of different domains. For instance, 3D maps show the height of a hotel or a mountain and not just its location. The 3D tools have to be used along with 2D GIS and then imagined in a 3D setting.

There was a time when more than one software was needed in order to view objects on the streets and different places in a city. Modern GIS has changed the dynamics of Geographic and Earth Science. With the development of digital medium, modern GIS interface lets its operators not only envisage and evaluate, but also manage geographical facts and figures.

Over the years, GIS has made a significant impact in creating, mapping as an essential tool to solve problems. Conventionally, GIS information was based on a 2-D recording, which apparently limited its usage in most applications. Incorporating 3D technology in GIS customizes the whole experience, making it more personal and enabling detailed visualization.

Uses of GIS

1) City Planning:

Today, a majority cities face shortage of basic amenities like water, electricity and space to live. The problem can be attributed to improper allocation of resources. A typical 3D model would consist of building information, satellite imagery and traffic data that can be used by urban planners.

2) Building Infm Modeling:

Building Information Modeling is an important technology that depicts real-world settings of an environment. The combination of BIM and GIS provides the necessary know-how to build a robust model.

3) Coastal Modeling and Analysis:

Coastal areas are important as they connect a country with the rest of the world for trade. Coastal areas globally face major

major threats and challenges for development.
It is important for planners to recognize what all factors affect the construction and preservation of ports, fisheries and effective planning of resources on 3D GIS.

4) Disaster Response:

3D GIS can helps people and societies better handle natural disasters. In case of a calamity, detailed mapping may give a broad idea to disaster response teams by making them aware of the environment they would be dealing in.