

SPPU-BE-COMP-CONTENT - KSKA Git

ASSIGNMENT-3

Q1

OUTLIER ANALYSIS:

→ process of identifying data points that deviate significantly from other observations in a dataset

→ Purpose: Detect errors, anomalies, or unusual patterns.

Methods

- Statistical: z-Score, IQR method
- Distance based: K-NN, clustering
- Model based: Isolation Forest, LOF

→ used: Helps improve model accuracy & data quality.

Q2

DENDROGRAM

→ tree-like diagram used to represent hierarchical clustering results.

Features:

- Each leaf represents a data point.
- Branches show the order of cluster merging
- height of the branches indicate distance / similarity b/w clusters.

Use: Helps visualise & decide number of clusters by cutting tree at a certain level.

SPPU-BE-COMP-CONTENT - KSKA Git

Q3

SPECTRAL CLUSTERING:

→ uses graph theory and eigen values of a similarity matrix to cluster data.

Steps

- construct similarity matrix b/w points
- compute eigen values / vectors of Laplacian matrix.
- use these as features for clustering (eg. with Kmeans)

Advantage

→ Works well for nonlinearly separable & complex shaped data.

Q4

ARTIFICIAL NEURAL NETWORK (ANN)

→ computational model inspired by human brain made of interconnected nodes (neurons)

Structure:

- Input layer: receives data
- Hidden layer: Extract features & patterns.
- output layer: produces results.

Uses: Image recognition, NLP, forecasting, etc

Learning: Adjusts weights using algorithms like back-propagation.

SPPU-BE-COMP-CONTENT - KSKA Git

Q5

SIGMOID & LOGISTIC FUNCTION:

- smooth S-shaped functⁿ used for mapping any real value into a range (0,1)

$$g(x) = \frac{1}{1 + e^{-x}}$$

- use: used as activatⁿ function in logistic regression and neural networks.
- Interpretation: converts linear outputs into probabilities.

Q6

RADIAL BASIS FUNCTION:

- real valued function whose output depends only on distance from a center point

→ Formula: $\phi(x) = e^{-\frac{\|x - c\|^2}{2\sigma^2}}$

→ Use:

- in RBF Networks & SVM Kernels.
- Good for interpolation & pattern recognition.

Key idea: Points closer to center have higher influence (higher values)