(1) What is Factor Analysis? Describe the steps

Ans. Factor Analysis is a statistical method used to simplify data by reducing huge amount of variables into small amount of underlying factors

STEPS IN FACTOR ANALYSIS:

(1) Testing Assumptions → assumptions must be tested before performing Factor Analysis.

(2) Construction of Correlation Matrix - A correlation matrix is created to examine relationships between variable

(3) Method of Factor Analysis - Selection of a specific method to extract factors and this is done using PCA, MLE.

(4) Determination of number of factors - the no. of factors is determined using Eigen Values, Scree Plot and Variance Explained.

(5) Rotation of Factors - Simplifies factor loadings to improve interpretability using Orthogonal and Oblique Rotation.

(6) Interpretation of Factors - Each of the factor is analyzed based on variables it groups together.

(2) State the properties of Eigen Vect

Ans: * Linearly independent for distinct eigen values.

* Scalar multiples of an eigen vect are also Eigen Vectors.

* Eigen Vectors can be normalized

(3) State the principle of clustering.

Ans: (i) Intra-Cluster Similarity -

Objects within the same cluster are as similar as possible

(ii) Inter-Cluster Dissimilarity -

Objects ~~outlier~~ in different clusters are as dissimilar as possible from one oth

(4) Mention the key problems of high-dimensional datasets.

Ans (i) Increased Data Sparsity.

(ii) Higher Computational Cost.

(iii) Difficulty in Distance-Based Learning

(iv) Overfitting in ML.

(5) What is Overfitting and Underfitting How can you prevent in a ML model?

Ans (i) Overfitting Model:

* Models knows too much.
* Low Bias and High Variance.

→ Techniques to reduce Overfitting Model:

* Regularization
* Cross- Validation
* Data Augmentation.

(ii) Underfitting Model:

* When model does not perform well in ~~training~~ training and testing dataset.
* High Bias and Low Variance.

→ Techniques to reduce underfitting:

* Increase Model Complexity.
* Increase number of features.
* Remove noise from data.