1. Feature Definition and Example:

- Feature Definition: A feature is an individual measurable property or characteristic of a phenomenon being observed.

- Example: In a dataset about houses, features could include "square footage," "number of bedrooms," and "distance to the nearest school."

2. Feature Construction Circumstances:

- Circumstances for Feature Construction:

1. Insufficient Information: Limited initial features.

2. Enhancing Predictive Power: Creating new features to improve model performance.

3. Domain Knowledge: Utilizing expertise to generate meaningful features.

4. Handling Missing Data: Filling missing values based on existing features.

3. Nominal Variables Encoding:

- Encoding Nominal Variables:

- One-Hot Encoding: Creating binary columns for each category.

- Label Encoding: Assigning a unique number to each category.

4. Numeric to Categorical Conversion:

- Conversion Process:

- Binning: Grouping numeric values into ranges.

- labelling Bins: Assigning labels to the created bins.

5. Feature Selection Wrapper Approach:

- Approach Description:

- Method: Evaluating feature subsets by training models iteratively.

- Advantages:

1. Model-Dependent: Considers the impact on actual model performance.

2. Optimal Subset: Aims for the most relevant feature set.

- Disadvantages:

1. Computational Intensity: Requires multiple model fittings.

2. Overfitting Risk: Can overfit to the training data.

6. Irrelevant Feature Definition:

- Irrelevant Feature Criteria:

- Definition: A feature is considered irrelevant if it does not contribute meaningful information to the model.

- Quantification: Measured by assessing its impact on model performance.

7. Redundant Feature Identification:

- Redundant Feature Criteria:

- Definition: A feature is redundant if it provides the same information as another feature.

- Identification: Correlation analysis or assessing mutual information.

8. Distance Measurements for Feature Similarity:

- Distance Measurements:

1. Euclidean Distance: Straight-line distance between points.

2. Manhattan Distance: Sum of horizontal and vertical distances between points.

9. Euclidean vs. Manhattan Distances:

- Euclidean Distance:

- Formula: **d = [(x1 – y1 )2+ (x2 – y2)2]1/2**

- Characteristics: Shorter straight-line distance.

- Manhattan Distance:

- Formula: **d = |x1 – y1| + |x2 – y2|**

- Characteristics: Distance along grid lines.

10. Feature Transformation vs. Feature Selection:

- Feature Transformation:

- Definition: Modifying existing features into a new space.

- Example: Principal Component Analysis (PCA).

- Feature Selection:

- Definition: Choosing a subset of existing features.

- Example: Recursive Feature Elimination (RFE).

11. Brief Notes:

1. SVD (Singular Value Decomposition):

- Definition: Factorization of a matrix into singular vectors and values.

- Application: Dimensionality reduction.

2. Collection of Features Using a Hybrid Approach:

- Definition: Combining filter, wrapper, and embedded methods for comprehensive feature selection.

- Advantage: Addresses limitations of individual approaches.

3. Silhouette Width:

- Definition: Measure of how similar an object is to its own cluster compared to other clusters.

- Application: Assessing cluster quality in unsupervised learning.

4. Receiver Operating Characteristic (ROC) Curve:

- Definition: Graphical representation of a classification model's performance.

- Components: Plots true positive rate against false positive rate.