1.(a)

Linear Methods: Principal Component Analysis (PCA)

Non-linear Methods: t-Distributed Stochastic Neighbor Embedding (t-SNE)

1.(b)

The "curse of dimensionality" refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings. Issues include an exponential increase in volume as dimensions increase, leading to data becoming sparse. This sparsity is problematic for any method that relies on statistical significance. To mitigate these issues, dimensionality reduction techniques are used to project the data into a lower-dimensional space where the data points are easier to manage and analyze, while preserving as much of the significant variability as possible.

1.(c)

t-SNE Limitations:

(1)Sensitive to parameter settings (e.g., perplexity size).

(2)Not guaranteed to be consistent across different runs.

(3)Primarily suited for visualization rather than higher-level analyses due to its focus on local structure.

PCA Limitations:

(1)Assumes that the principal components are linear combinations of the original features.

(2)Can be significantly affected by outliers in the data.

(3)Captures maximal variance under the assumption of linearity, which may not handle complex polynomial relationships between features.

2(a)

文字图案

中度可信度描述已自动生成

2(b)

Number of samples (M): 1797

Number of features (N): 64

2(c)

(iv)

Columns of U: Each column represents an eigenvector, or principal component, which is a direction of maximum variance in the data.

Rows of U: Each row shows how much each original feature contributes to a particular principal component.

(vii)

图表, 散点图

描述已自动生成

3

Sigmas for each data point:

[1.14873428 1.36361809 1.14721298 1.20844536 1.10727882 1.26853676 1.28945465 1.20274048 1.44690933 1.38643761 1.2586483 1.42256851 1.13922615 1.38453598 1.26054993 1.15405884 1.21300926 0.9897583 1.19285202 1.24001236]