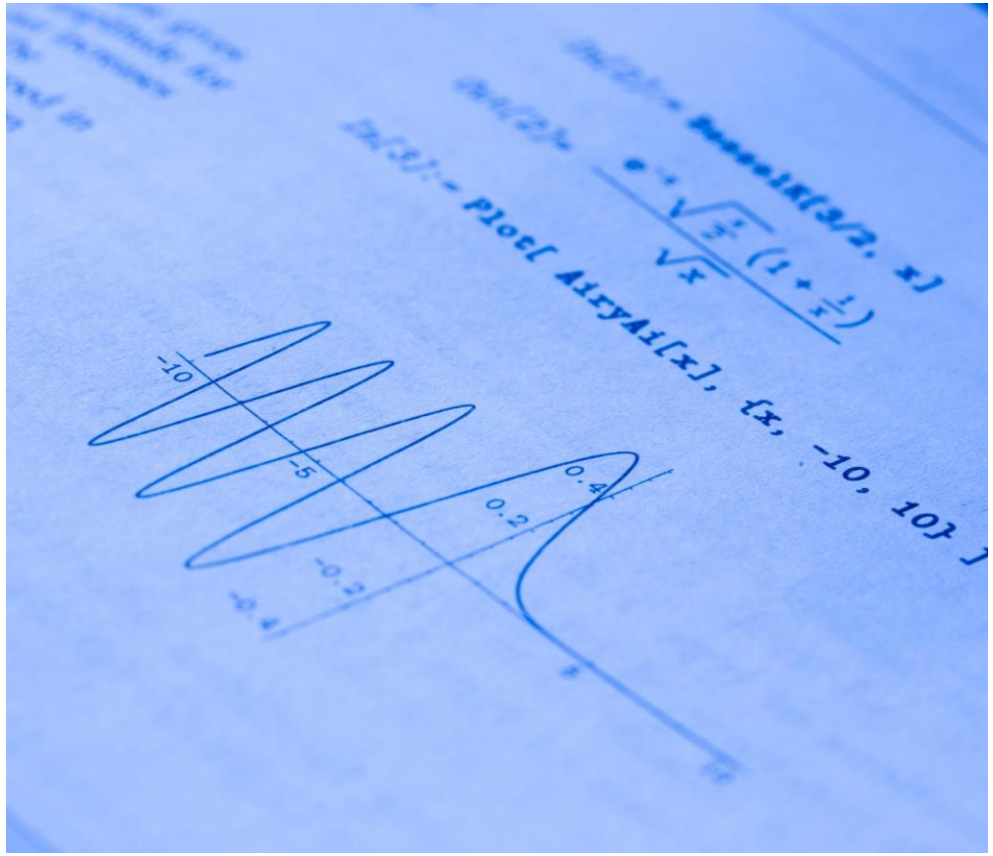
LINEAR ALGEBRA APPLICATIONS IN PREDICTIVE, PERCEPTUAL, AND PERSONALIZED INTELLIGENCE

Mathematical foundations powering advanced intelligent systems

PREDICTIVE INTELLIGENCE: BOSTON HOUSING DATASET

PURPOSE



Linear Regression and Prediction

Linear regression uses matrix operations to compute weights minimizing prediction error for median housing prices.

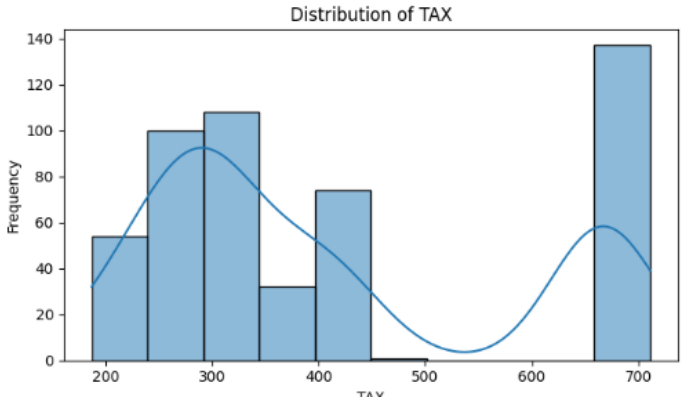
Gaussian Residual Modeling

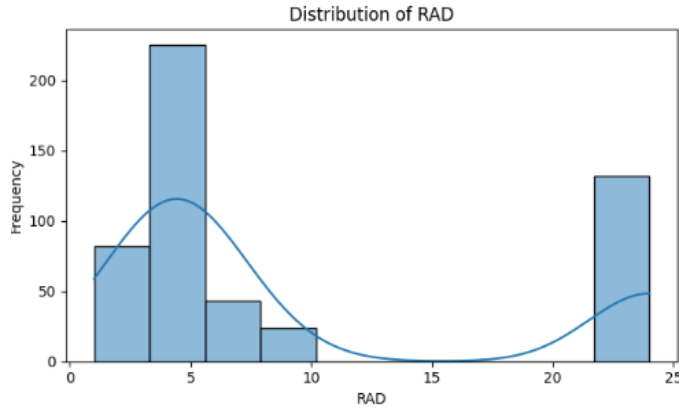
Gaussian models quantify uncertainty by treating residuals as random variables with defined mean and variance.

DATASET, PREPROCESSING, AND MODEL EVALUATION

FEATURE	DESCRIPTION
CRIM	Per capita crime rate by town
ZN	Proportion of residential land zoned for large lots
INDUS	Non-retail business acres per town
CHAS	Charles River dummy variable
NOX	Nitric oxides concentration (ppm)
RM	Average number of rooms per dwelling
AGE	% of units built prior to 1940
DIS	Weighted distances to employment centers
RAD	Accessibility index to radial highways
TAX	Property-tax rate per $10,000
PTRATIO	Pupil-teacher ratio by town
B	1000(Bk – 0.63)^2 based on Black population proportion
LSTAT	% lower-status population

HANDLING BI-MODAL DISTRIBUTIONS

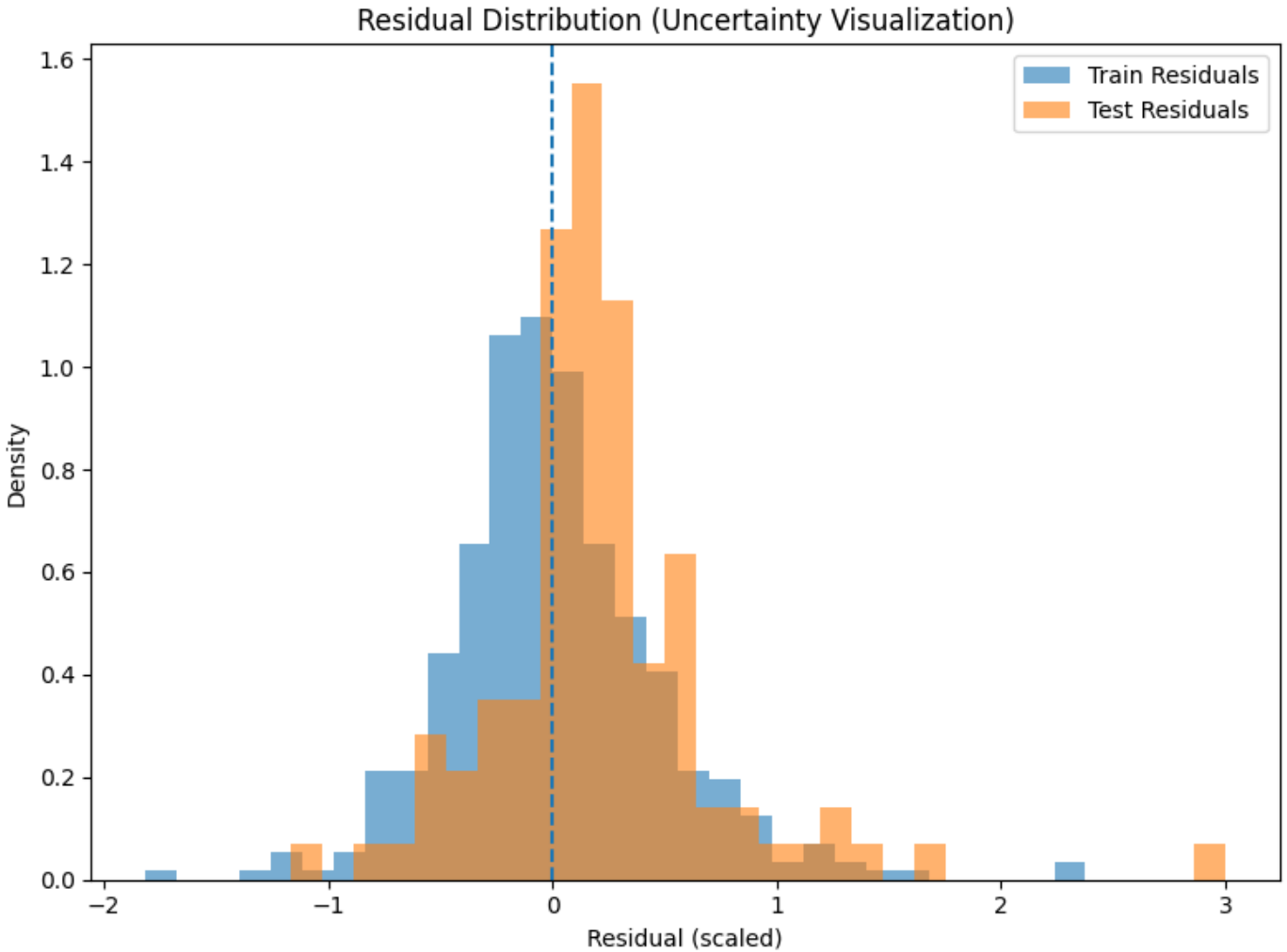




Simply Normalizing these distribution will loose the information about the two different groups in the distribution.
To handle this:
1. Create categorical feature
2. Followed by the Normalization

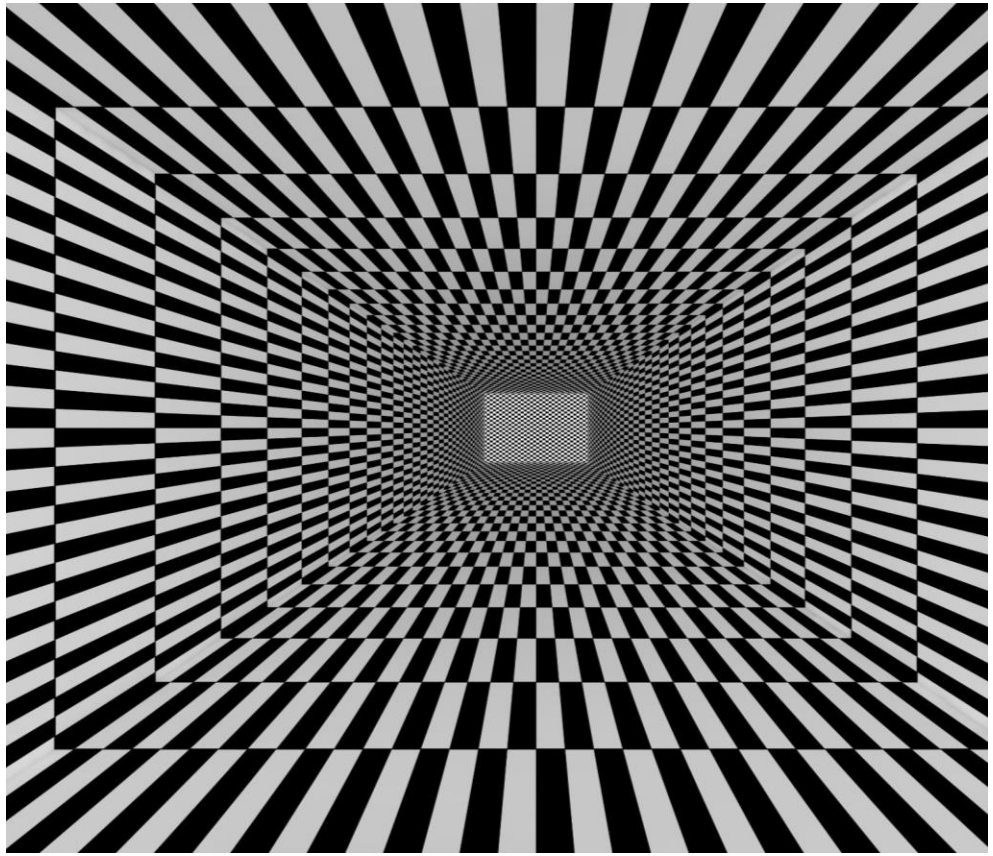
RESULTS

$\varepsilon_{train} \sim N(0.013, 0.480^2)$
$\varepsilon_{test} \sim N(0.215, 0.521^2)$



PERCEPTUAL INTELLIGENCE: DIGITS DATASET

PCA AND EIGENDIGITS



Digits Dataset and Representation

Digits dataset contains 8x8 grayscale images flattened into 64-dimensional vectors representing handwritten digits.

PCA Process

PCA involves centering data, computing covariance matrix, and eigen decomposition to identify maximum variance directions.

Eigentdigits and Dimensionality Reduction

Eigentdigits capture dominant visual patterns and enable compression by projecting data onto top eigenvectors.

Reconstruction and Fidelity

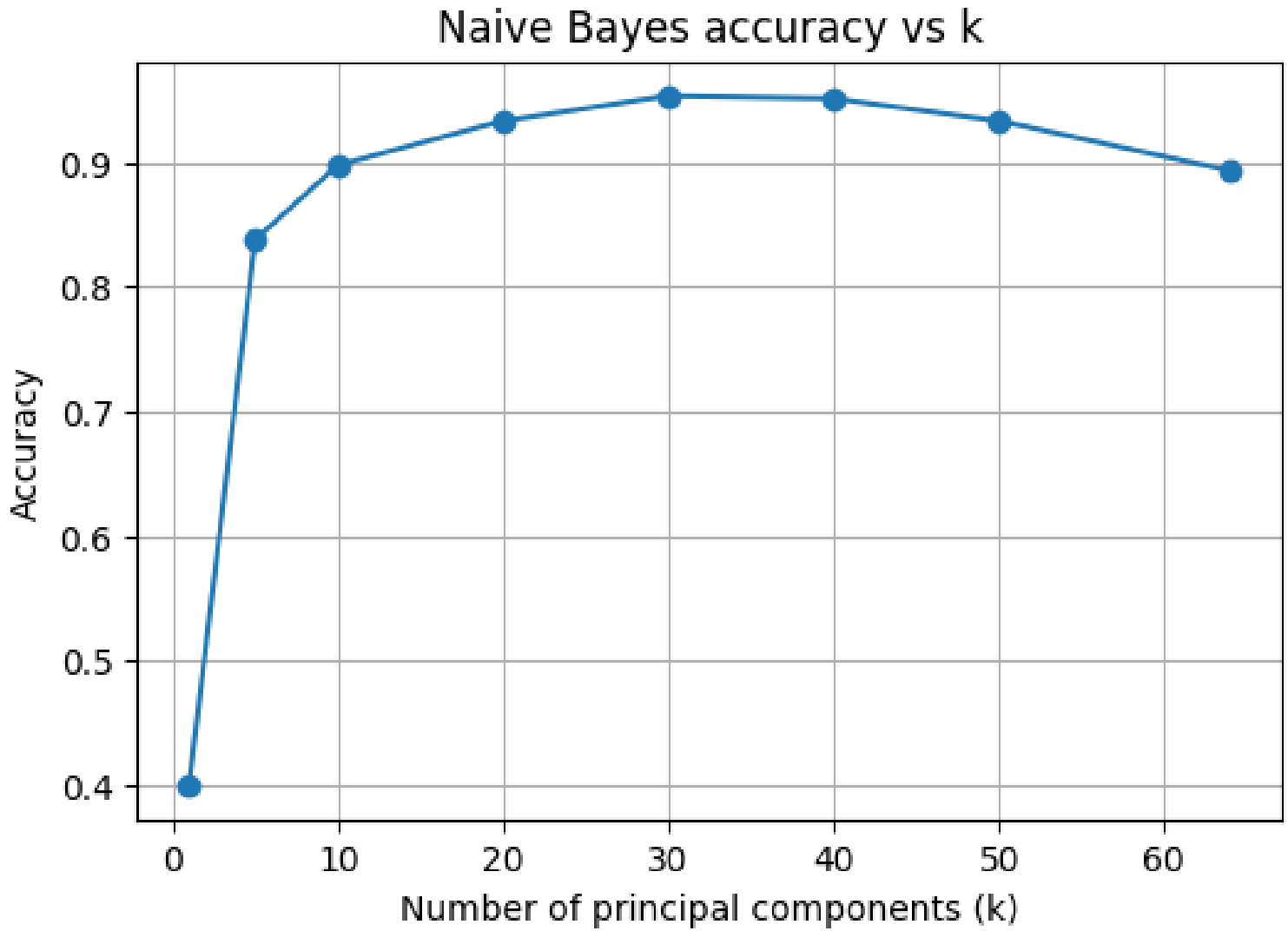
Increasing number of eigenvectors restores finer digit details, balancing compression and image quality.

DIGIT RECONSTRUCTION

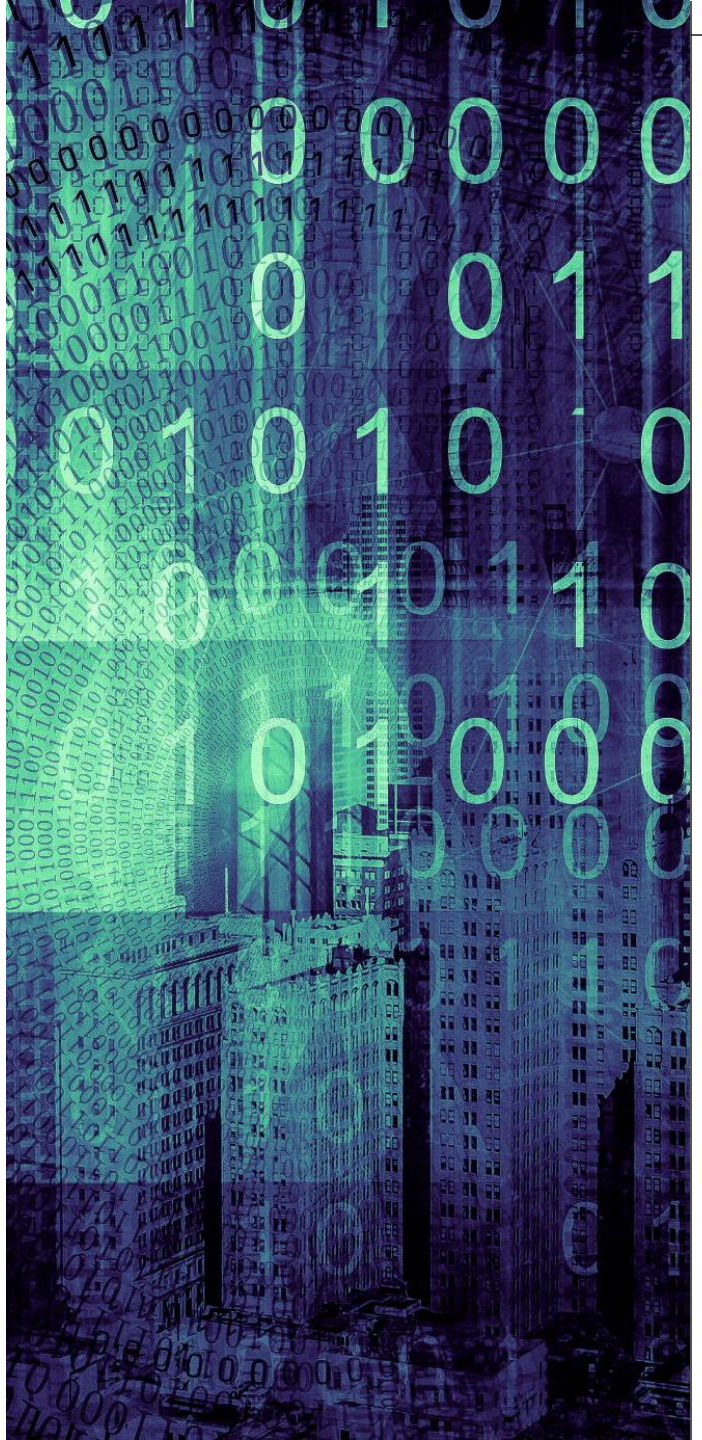
- **Low k** = strong compression → loss of fine details.
- **Higher k** = more variance retained → clearer digit structure.
- **k=64** ≈ near-full information → reconstruction is close to the original

CLASSIFICATION PERFORMANCE

Digits have a compact low-dimensional structure.
PCA removes redundancy and keeps the most discriminative patterns.
Naive Bayes benefits most when we keep **moderate k**, not the full 64.



PERSONALIZED INTELLIGENCE: MOVIELENS DATASET



MATRIX FACTORIZATION AND PREDICTION

Handling Sparse Data

Sparse user-movie rating matrix is filtered and missing values imputed using user-specific mean ratings for better analysis.

Singular Value Decomposition

Truncated SVD decomposes the rating matrix into latent factors capturing user preferences and movie attributes.

Predicting Missing Ratings

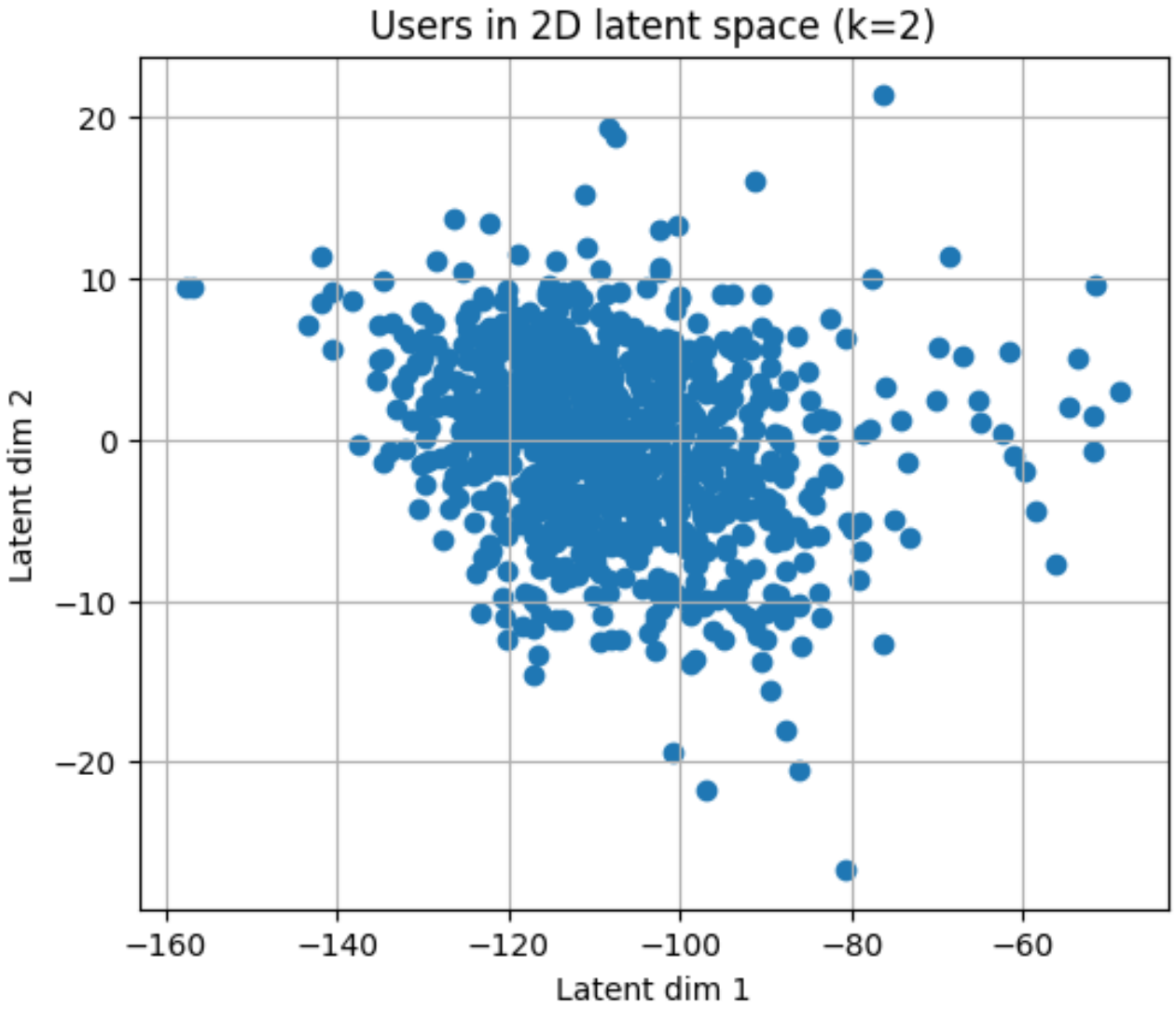
Reconstruction of the matrix from latent factors enables prediction of unknown ratings for personalized recommendations.

Personalization with Linear Algebra

Linear algebra uncovers hidden structures in user behavior and items to produce actionable recommendation predictions.

USERS REPRESENTATION IN 2D SPACE

- Each point represents a user's embedding learned from the user–movie rating matrix.
- Coordinates come from Truncated SVD with k=2, capturing two dominant latent preference factors.
- Users close together likely share similar rating patterns.
- The dense center indicates most users have general preferences in this subset.

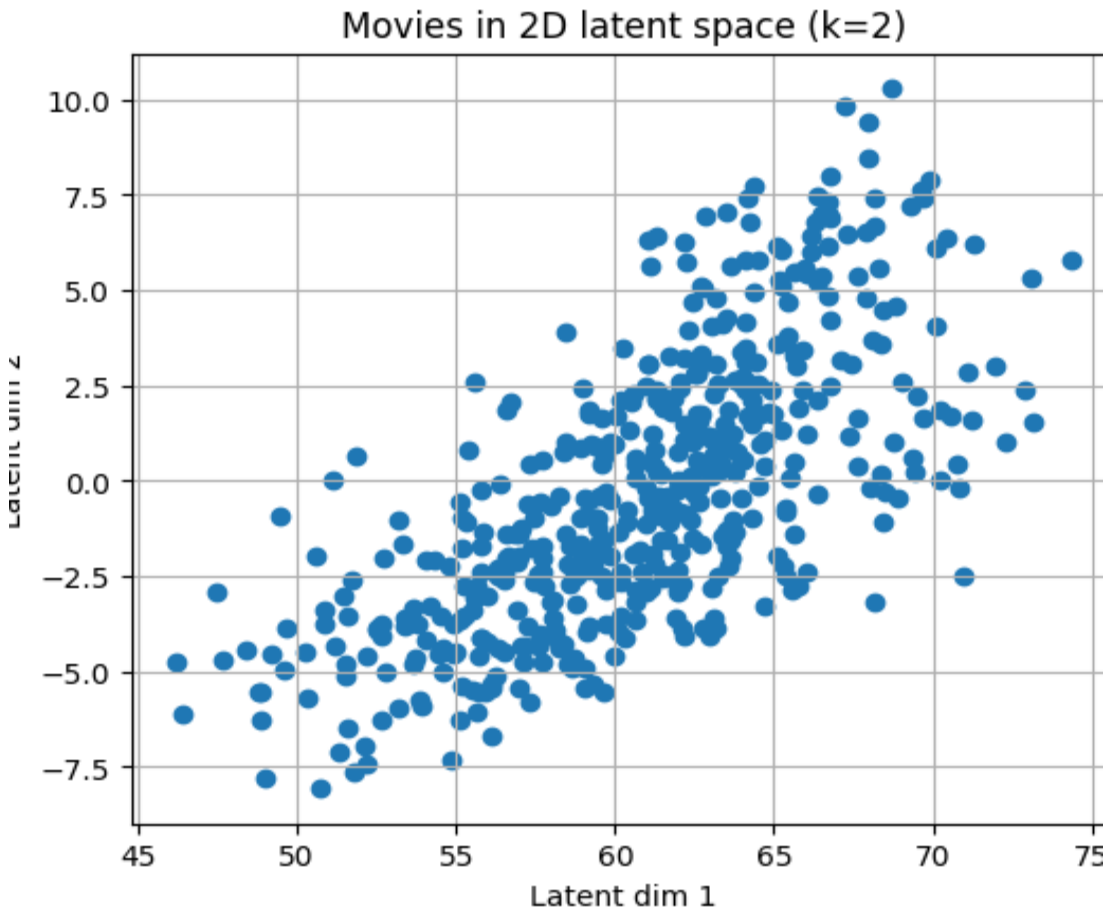


USERS REPRESENTATION IN 2D SPACE

•Each point represents a movie embedding learned from the user–movie rating matrix.

•Coordinates reflect two dominant latent “movie attribute” factors inferred from ratings.

•The band/diagonal structure suggests a strong shared latent trend in this filtered subset.



RECONSTRUCTION ERROR

- **Lower reconstruction error does not guarantee better recommendations.**
- This metric can favor overfitting to imputed entries.
- That's why we select k using **validation RMSE on held-out ratings**, not just training reconstruction error.