Comparative Analysis of Dimensionality Reduction Techniques for Machine Learning

Alice Gilardi, 1784496
Gianluca Vitaliano, 1741654
Sapienza, University of Rome
33515, Engineering in Computer Science and AI

January 5, 2026

1 Model Performance Comparison

We evaluate three tasks: Fashion-MNIST classification (784D→50D), California Housing regression (8D→5D), and credit card clustering (17D→8D), comparing original features against PCA and autoencoder reductions.

Table 1: Performance Summary Across Tasks

Task	Model	Metric	Value
FMNIST	CNN (784D)	Accuracy	**92.74%**
	Dense PCA (50D)	Accuracy	88.10%
	Dense AE (50D)	Accuracy	85.92%
Housing	FNN (8D)	R^2	**0.797**
	FNN PCA (5D)	R^2	0.682
	FNN AE (5D)	R^2	0.647
Credit	K-M Orig (17D)	Silhouette	0.255
	K-M PCA (8D)	Silhouette	0.277
	K-M AE (8D)	Silhouette	**0.357**

1.1 Fashion-MNIST Classification

CNN on original 784D data achieved 92.74% accuracy (ROC-AUC 0.996), outperforming dense networks on PCA-reduced features (88.10%, ROC-AUC 0.990) and autoencoder-reduced features (85.92%, ROC-AUC 0.988). The 4.64pp degradation from original to PCA, plus 2.18pp to autoencoder, reflects spatial information loss when flattening images. Despite 93.6% dimensionality reduction, PCA retained 86.26% variance but destroyed local spatial correlations critical for image understanding. **Best model**: CNN exploiting spatial structure through convolutions.

1.2 California Housing Regression

FNN on original 8D features achieved R^2=0.797 (RMSE 0.519). PCA reduction to 5D yielded R^2=0.682 (RMSE 0.650), while autoencoder achieved R^2=0.647 (RMSE 0.685). PCA outperformed autoencoder here—linear housing relationships (income, location → price) favor linear projection. PCA retained 98.18% variance yet R^2 dropped 11.5pp, indicating discarded variance contained predictive information orthogonal to maximum variance directions. **Best model**: Original FNN; dimensionality reduction unjustified for low-dimensional data.

1.3 Credit Card Clustering

K-Means on autoencoder-reduced 8D data achieved Silhouette 0.357, substantially outperforming PCA (0.277) and original 17D (0.255). Autoencoder showed lowest Davies-Bouldin index (1.108 vs 1.398 vs 1.586) and highest Calinski-Harabasz score (5233 vs 1602 vs 1295). Convergence accelerated dramatically: 9 iterations vs 25 (PCA) vs 22 (original). Nonlinear transformation captured complex spending patterns that linear PCA cannot represent. **Best model**: K-Means + Autoencoder—rare win-win with better quality AND efficiency.

1.4 Quantitative Trade-offs

2 Impact of Model Assumptions

2.1 CNN Spatial Structure Requirement

CNNs exploit local spatial coherence through translation-invariant convolutional kernels. The architecture (Conv2D→MaxPool→Dense) achieved 92.74% by learning hierarchical features (edges→textures→objects). Flattening to 50D PCA destroys spatial relationships, forcing dense networks to learn from abstract principal components rather than local patterns. Despite PCA retaining 86.26% global variance, PC1 captured only 29.02%—image variance distributes across nonlinear manifolds (shape, texture, edges) that linear projections inefficiently represent. This architectural mismatch explains the 4.64% accuracy loss. Learning curves showed CNN early stopping at epoch 45 with 2% train-val gap, indicating effective regularization through dropout (0.25, 0.5) and batch normalization despite 225,930 parameters.

2.2 PCA Linearity Assumption

PCA maximizes variance through eigendecomposition: $\max_{\mathbf{w}} \mathrm{Var}(\mathbf{w}^T\mathbf{X})$ subject to $\|\mathbf{w}\| = 1$. This proves optimal for housing data where features relate linearly: PC1 (48.68% variance) loads on latitude (+0.492), population (+0.491), occupancy (+0.471), capturing urban density. PC2 (23.84%) contrasts income (-0.702) against house age (+0.702), representing socioeconomic stratification. These interpretable linear combinations enable effective regression (R^2=0.682 vs 0.647 for autoencoder). For Fashion-MNIST, nonlinear manifold structure renders PCA suboptimal—50 components needed for 86.26% variance yet only 88.10% accuracy. Credit card data exhibits nonlinear spending interactions (balance × purchase frequency × cash advances) that PCA cannot disentangle, yielding inferior clustering (Silhouette 0.277 vs 0.357 autoencoder).

2.3 K-Means Spherical Cluster Assumption

K-Means assumes spherical, convex clusters with similar variance via Euclidean distance minimization. Credit card features exhibit non-spherical distributions in 17D original space. Autoencoder's nonlinear encoding transformed irregular clusters into compact, well-separated spherical regions: 53% higher Silhouette (0.357 vs 0.255), 30% lower Davies-Bouldin (1.108 vs 1.586), 4× higher

Calinski-Harabasz (5233 vs 1295). The 2.4× convergence acceleration (9 vs 22 iterations) indicates transformed space positioned centroids near optimal configurations. 2D PCA visualization confirms enhanced separability: autoencoder clusters show minimal overlap while original data exhibits substantial boundary ambiguity.

3 Dimensionality Reduction Analysis

3.1 PCA versus Autoencoder: When Each Excels

Table 2: Reconstruction Quality (Test MSE)

Dataset	PCA	Autoencoder	Δ
Fashion-MNIST	0.0120	0.0083	+30.91%
Housing	0.0169	0.0172	-1.72%
Credit Card	0.1626	0.1144	+29.65%

PCA optimal for: (1) Linear relationships—housing R^2=0.682 vs 0.647 for autoencoder due to income/location linearity. (2) Interpretability—loading vectors enable domain validation (PC1=urban density, PC2=socioeconomic axis). (3) Computational simplicity—no training, deterministic $O(d^2n + d^3)$ eigendecomposition. (4) Guaranteed variance preservation—provably optimal linear projection.

Autoencoder optimal for: (1) Nonlinear manifolds—images/behavioral data on curved manifolds. Fashion-MNIST: 30.91% better reconstruction preserving edge details through hierarchical encoding (784→256→128→50). (2) Clustering quality—credit card: 40% Silhouette improvement, capturing interaction effects (balance×frequency×cash advances). (3) Task-specific optimization—flexible loss functions, end-to-end integration.

Skip reduction when: Original $d < 20$ (housing: 5.4% parameter reduction, 11.5% R^2 loss), maximum accuracy required, minimal computational constraints.

3.2 Variance Retention versus Reconstruction Quality

Fashion-MNIST: PCA retained 86.26% variance but achieved only 88.10% accuracy—variance maximization doesn't preserve discriminative information for supervised tasks. Autoencoder's superior reconstruction (30.91% better) failed to translate to classification advantage (85.92%), confirming reconstruction fidelity and discriminative power are orthogonal objectives. Credit card: reconstruction quality directly predicted clustering success—autoencoder's 29.65% better reconstruction correlated with 40% Silhouette improvement. Asymmetry suggests unsupervised tasks benefit from faithful representation while supervised tasks require discriminative features possibly residing in low-variance directions.

3.3 Computational Benefits versus Accuracy Loss

Fashion-MNIST: 92% parameter reduction (225,930→17,962) enables mobile deployment, 3× training speedup for 4.64% accuracy cost—acceptable for recommendation systems. Housing: negligible gains (5.4%) fail to justify 11.5% loss—avoid reduction for $d < 20$. Credit: win-win scenario—better quality (+40% Silhouette) AND 2.4× faster convergence through nonlinear feature engineering.

Table 3: Computational Trade-offs

Dataset	Params	Train	Infer	Cost
FMNIST	-92.0%	3.0×	2.86×	-4.64% acc
Housing	-5.4%	1.42×	1.08×	-11.5% R^2
Credit	-52.9% mem	1.33×	1.66×	+40% Sil

4 Overfitting and Regularization

4.1 Train-Validation Gaps

Fashion-MNIST CNN: 94.27% train vs 93.37% validation (0.90pp gap) despite 225,930 parameters on 51,000 samples (4.43 ratio). Minimal overfitting attributable to dropout (0.25, 0.5), batch normalization, early stopping at epoch 45 (patience=10), and learning rate reduction. Reduced-dimension models showed smaller gaps through information bottleneck regularization—compression forces networks to discard irrelevant details.

Housing: All models exhibited 3-5% relative gaps (original 4.3%, PCA 5.3%, autoencoder 3.4%). Learning curves showed smooth exponential decay without oscillation, confirming stable optimization. Residuals centered at zero with homoscedastic distributions, indicating unbiased predictions.

Credit card: K-Means deterministic, validated through multiple initializations (n_init=10) and consistent train-test Silhouette scores (±2% relative). Elbow/Silhouette analysis (k=2-10) showed smooth monotonic trends without noise overfitting.

4.2 Regularization Mechanisms

Dropout randomly zeros activations (probability p), forcing redundant representations. Batch normalization reduces internal covariate shift, providing mild regularization through minibatch noise. Early stopping prevented excessive training: CNN stopped epoch 45/50, housing models 73-100/100. Learning rate reduction (ReduceLROnPlateau, factor=0.5, patience=5) enabled fine-grained optimization. Information bottleneck in reduced models acts as implicit regularization—50D encoding for Fashion-MNIST retains only essential discriminative features.

5 Feature Importance & Interpretability

5.1 PCA Component Interpretation

Housing PC1 (48.68% variance): Latitude +0.492, Population +0.491, AveOccup +0.471 → urban density. Northern California cities (high latitude) exhibit high population density and occupancy. PC2 (23.84%): MedInc -0.702, HouseAge +0.702 → socioeconomic axis. Negative PC2 = high-income newer housing (gentrified), positive = older lower-income neighborhoods. Loading heatmap confirms components capture coherent feature groups, enabling domain expert validation. This transparency proves critical for regulated applications requiring explainability—autoencoders lack direct interpretation despite superior reconstruction.

Fashion-MNIST confusion matrices reveal systematic errors: Shirt↔T-shirt (16% confusion), Pullover↔Coat (12%). Shirt achieved lowest F1 (0.78 CNN, 0.69 PCA, 0.63 autoencoder), indicating genuine perceptual ambiguity in collar styles. Best classes: Trouser/Bag (F1 0.99), Ankle boot (0.97)—distinctive shapes create separable features. Error consistency across models suggests inherent category ambiguity rather than dimensionality reduction artifacts.

6 Computational Efficiency

Fashion-MNIST: 92% parameter reduction (2.64 MB→0.26 MB) enables edge deployment. 3× training speedup (300s→100s) through reduced gradient computation. 2.86× inference acceleration (0.38ms→0.13ms per sample) critical for real-time applications.

Housing: Minimal benefits—5.4% parameter reduction, 1.42× training speedup insufficient for 11.5% R^2 loss. Original model preferred unless strict memory constraints.

Credit: 52.9% memory reduction (0.40 KB→0.19 KB centroids), 2.4× convergence acceleration (22→9 iterations). Autoencoder's nonlinear transformation positioned centroids near optimal configurations, minimizing iterative refinement. Near-linear K-Means complexity ($O(nkdi)$) explains 1.33× training speedup from 17→8D and 22→9 iterations.

7 Cross-Task Insights

7.1 Dimensionality Curse Manifestation

The curse of dimensionality manifests task-dependently. Fashion-MNIST (784D) exhibits sparse sample distribution where nearest neighbors in pixel space are often visually dissimilar. CNNs mitigate this through inductive bias (local connectivity, weight sharing), achieving 92.74% despite high dimensionality. Housing (8D) shows no curse—Euclidean distances remain meaningful, explaining simple FNN success (R^2=0.797). Credit card (17D) demonstrates moderate curse: K-Means improves with 8D autoencoder reduction (Silhouette 0.255→0.357), confirming dimensionality reduction beneficial when $20 < d < 100$.

7.2 Bias-Variance Trade-off

Dimensionality reduction introduces bias (information loss) while reducing variance (fewer parameters, implicit regularization). Fashion-MNIST: PCA's 4.64% accuracy bias enabled 92% parameter reduction with smaller train-val gap (1.5% vs 2% CNN). Housing: 11.5% R^2 bias outweighs variance reduction—optimal complexity achieved at 8D. Credit: autoencoder's nonlinear bias proved beneficial, creating separable feature space reducing both approximation error and estimation variance (2.4× faster convergence).

7.3 Theoretical Foundation

PCA provably finds optimal k-dimensional linear subspace minimizing reconstruction: $\min_{\mathbf{W}_k} \|\mathbf{X} - \mathbf{X}\mathbf{W}_k\mathbf{W}_k^T\|_F^2$ where $\mathbf{W}_k$ contains top k eigenvectors. This optimality explains housing success—linear ground truth aligns with PCA objective. Autoencoders solve $\min_\theta \mathbb{E}[\|\mathbf{x} - \text{dec}(\text{enc}(\mathbf{x}))\|^2]$ via gradient descent, enabling nonlinear transformations. Universal approximation theorem guarantees sufficient capacity for any continuous function, explaining superior Fashion-MNIST/credit card reconstruction. Gradient descent finds local minima—housing's 1.72% worse reconstruction reflects optimization challenges for linear relationships better solved by closed-form PCA.

7.4 Deployment Recommendations

Fashion-MNIST: Cloud inference uses CNN (max accuracy 92.74%). Mobile apps deploy PCA model (0.26 MB, 88.10% acceptable). Edge devices leverage aggressive quantization of PCA model (<100KB). Housing: original model universally preferred—8D negligible burden. Credit: autoencoder preprocessing recommended universally—training cost amortized, 1.66× inference speedup compounds over millions of segmentations.

7.5 Domain Generalization

Results extend to: (1) Computer vision—CNNs dominate, reduction only for extreme constraints. (2) Tabular $d < 50$—tree methods often superior; reduction rarely beneficial. (3) High-dimensional sparse (text, genomics)—specialized methods (TF-IDF, pathway analysis) outperform generic PCA/autoencoders. (4) Time series—RNNs/Transformers exploit temporal structure; reduction destroys dependencies. (5) Unsupervised—autoencoders consistently valuable with complex interactions.

8 Conclusions

Optimal dimensionality reduction depends critically on data structure. PCA excels for linear data (housing R^2=0.682) with interpretability (PC loadings), while autoencoders dominate nonlinear manifolds (images 30.91% better reconstruction, credit 40% Silhouette improvement). Fashion-MNIST: CNN spatial structure yields 92.74% despite 15.7× more dimensions than PCA—architecture matters more than dimensionality. Housing: reduction counterproductive for $d < 20$ (5.4% gains, 11.5% loss). Credit: win-win autoencoder scenario—better quality AND efficiency through disentangling behavioral interactions.

Key insights: (1) Variance preservation $\neq$ task performance—PCA's 98.18% housing variance retention yielded only 68.23% R^2. (2) Reconstruction quality predicts unsupervised success but not supervised performance—orthogonal objectives. (3) Benefits scale nonlinearly with dimensionality ratio—15.7× reduction yields 3× speedup, but 1.6× reduction yields 1.42× speedup. (4) Information bottleneck provides implicit regularization—compressed models show smaller train-val gaps.

Decision framework: Use PCA for linear data with interpretability needs and $d > 50$. Deploy autoencoders for nonlinear manifolds requiring reconstruction quality or clustering. Skip reduction for $d < 20$, maximum accuracy requirements, or when computational constraints minimal. Fashion-MNIST: PCA acceptable for resource-constrained deployment (88.10%, 92% fewer parameters). Housing: always use original. Credit: always use autoencoder (superior across all metrics).