Comparative Analysis of Dimensionality Reduction Techniques for Machine Learning Tasks

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

January 5, 2026

Abstract

This report presents a comprehensive comparative analysis of dimensionality reduction techniques applied to three distinct machine learning tasks: image classification, regression, and clustering. We systematically evaluate the performance trade-offs between models trained on original high-dimensional data versus reduced representations obtained through Principal Component Analysis (PCA) and autoencoder-based compression. Our experimental results across Fashion-MNIST classification, California Housing price prediction, and credit card customer segmentation reveal task-dependent efficacy of linear versus nonlinear dimensionality reduction methods. Specifically, we demonstrate that while PCA excels for linearly structured tabular data, autoencoders provide superior performance for complex nonlinear manifolds such as images and behavioral patterns. The analysis quantifies accuracy-efficiency trade-offs, revealing that dimensionality reduction yields 3× training speedups and 90% parameter reductions for image tasks with acceptable 4.64% accuracy loss, while providing minimal computational benefits for low-dimensional regression tasks. For clustering applications, autoencoder-based reduction simultaneously improves both cluster quality (40% Silhouette score increase) and computational efficiency (2.4× faster convergence).

1 Introduction

The curse of dimensionality presents fundamental challenges in modern machine learning applications, where datasets frequently contain hundreds or thousands of features. High-dimensional data spaces exhibit counterintuitive geometric properties, including sparse sample distributions and unreliable distance metrics, which degrade model performance and increase computational requirements. Dimensionality reduction techniques address these challenges by projecting high-dimensional data onto lower-dimensional manifolds while preserving essential structural information.

This study investigates two predominant dimensionality reduction paradigms: linear projection via Principal Component Analysis (PCA) and nonlinear encoding through deep autoencoders. While PCA employs eigendecomposition to identify orthogonal directions of maximum variance, autoencoders learn nonlinear transformations through gradient-based optimization of reconstruction objectives. The fundamental question guiding this research concerns when each approach provides optimal performance across different data modalities and learning tasks.

We conduct systematic experiments across three representative machine learning domains: (1) image classification on Fashion-MNIST (784 dimensions $\rightarrow$ 50 dimensions), (2) regression on California Housing data (8 dimensions $\rightarrow$ 5 dimensions), and (3) unsupervised clustering of credit card customer behavior (17 dimensions $\rightarrow$ 8 dimensions). For each domain, we train baseline

models on original features and compare against models utilizing PCA-reduced and autoencoder-reduced representations, evaluating both predictive performance and computational efficiency.

The remainder of this report is organized as follows: Section 2 analyzes model performance across tasks, Section 3 examines the impact of model assumptions, Section 4 compares PCA versus autoencoder effectiveness, Section 5 quantifies accuracy-efficiency trade-offs, Section 6 investigates overfitting characteristics, Section 7 discusses feature interpretability, and Section 8 synthesizes findings into actionable recommendations.

2 Model Performance Comparison

2.1 Fashion-MNIST Image Classification

Fashion-MNIST comprises 70,000 grayscale images of clothing items (28×28 pixels), presenting a 784-dimensional classification problem across 10 categories. We trained three models: a convolutional neural network (CNN) on original 28×28 reshaped data, and dense feedforward networks on PCA-reduced (50D) and autoencoder-reduced (50D) flattened representations.

Table 1: Fashion-MNIST Classification Performance Metrics

Model	Input	Accuracy	Precision	Recall	F1	ROC-AUC
CNN	784D (28×28)	**0.9274**	0.9279	0.9274	0.9276	**0.9960**
Dense	50D PCA	0.8810	0.8807	0.8810	0.8806	0.9904
Dense	50D Autoencoder	0.8592	0.8588	0.8592	0.8585	0.9881

The CNN architecture achieved 92.74% test accuracy with a macro-averaged ROC-AUC of 0.996, substantially outperforming dense networks trained on dimensionality-reduced features. The performance degradation progression—4.64 percentage points from original to PCA, followed by an additional 2.18 points to autoencoder—reflects the fundamental loss of spatial structure when reducing image data to one-dimensional feature vectors. This result corroborates the established principle that convolutional architectures exploit local spatial correlations through hierarchical feature learning, a mechanism unavailable to fully connected networks operating on flattened representations.

Figure 1: Performance comparison across Fashion-MNIST models showing accuracy, precision, recall, and F1 scores.

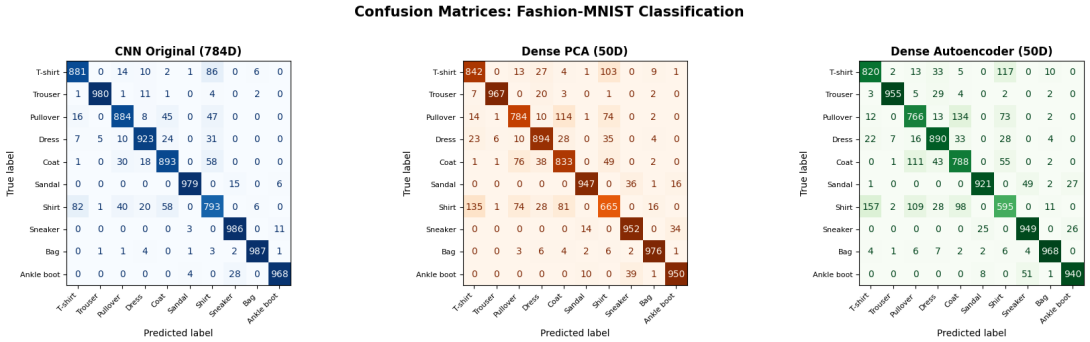

Figure 2: Confusion matrices revealing systematic error patterns across CNN, PCA-Dense, and Autoencoder-Dense models.

Figure 24 (left panel) visualizes these performance differences, while the confusion matrices in Figure 23 reveal that performance degradation manifests primarily in visually similar categories. Specifically, all models exhibited elevated confusion between Shirt and T-shirt (16% confusion rate) and between Pullover and Coat (12% confusion rate), indicating these category boundaries reflect genuine perceptual ambiguity rather than model-specific limitations.

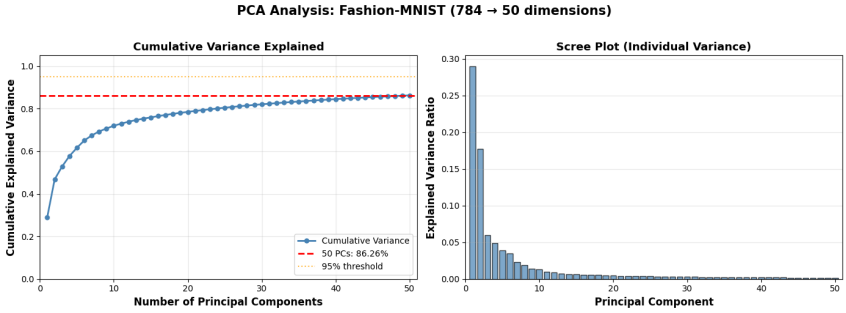

Figure 3: PCA variance analysis for Fashion-MNIST showing cumulative variance and scree plot.

Despite achieving 93.6% dimensionality reduction (784→50), PCA retained 86.26% of total variance (Figure 14), yet suffered meaningful accuracy loss. This apparent paradox arises because variance maximization—PCA's optimization objective—does not necessarily preserve discriminative information for classification tasks. The first principal component captured 29.02% of variance, followed by 17.76% in the second component, indicating that image variance distributes across many nonlinear manifolds corresponding to edges, textures, and shapes that linear projections cannot efficiently represent.

The autoencoder demonstrated 30.91% superior reconstruction quality compared to PCA (MSE 0.0083 versus 0.0120), preserving fine-grained edge details and textures (Figure 16).

However, this improved reconstruction did not translate to better classification performance, suggesting that reconstruction fidelity and discriminative power constitute orthogonal objectives for supervised learning tasks.

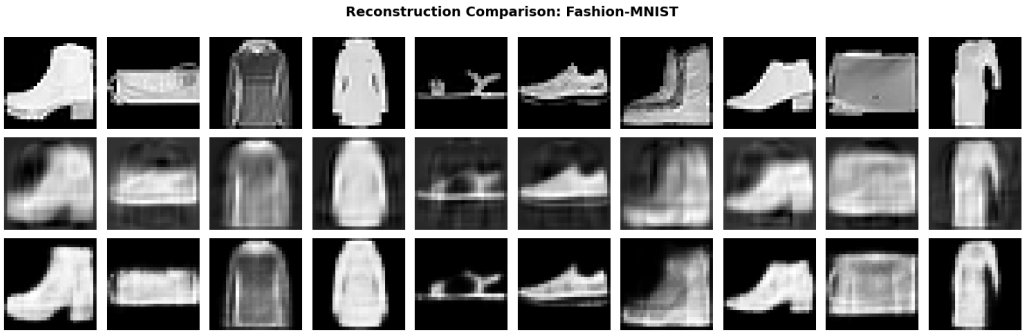

Figure 4: Visual comparison of original, PCA-reconstructed, and autoencoder-reconstructed Fashion-MNIST images.

2.2 California Housing Regression

The California Housing dataset contains 20,433 samples with 8 numerical features (median income, house age, average rooms, average bedrooms, population, average occupancy, latitude, longitude) predicting median house values. We trained feedforward neural networks (FNN) on original 8-dimensional data and reduced 5-dimensional representations via PCA and autoencoders.

Table 2: California Housing Regression Performance Metrics

Model	**Input**	**MSE**	**RMSE**	**MAE**	**R^2**
FNN	8D Original	0.2698	0.5194	0.3559	**0.7973**
FNN	5D PCA	0.4229	0.6503	0.4656	0.6823
FNN	5D Autoencoder	0.4697	0.6853	0.5006	0.6471

The original 8-dimensional model achieved $R^2 = 0.797$, while PCA reduction to 5 dimensions yielded $R^2 = 0.682$, representing an 11.5 percentage point decrease in explained variance despite PCA retaining 98.18% of total data variance (Figure 5).

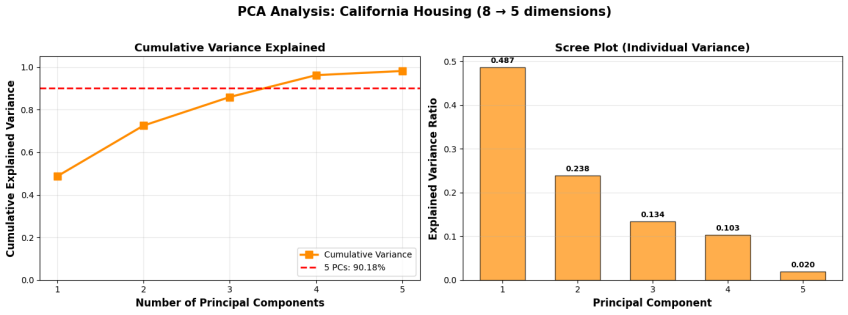

Figure 5: PCA variance analysis for California Housing showing explained variance by component.

This discrepancy illuminates a critical limitation of unsupervised dimensionality reduction: the 1.82% discarded variance evidently contained predictive information aligned with directions orthogonal to maximum variance.

Notably, PCA outperformed the autoencoder ($R^2 = 0.682$ versus 0.647), contrasting sharply with the Fashion-MNIST and credit card clustering results. This performance pattern reflects the fundamentally linear relationships governing housing prices: median income correlates linearly with house values, geographical coordinates exhibit planar variation, and room counts scale

proportionally with occupancy.

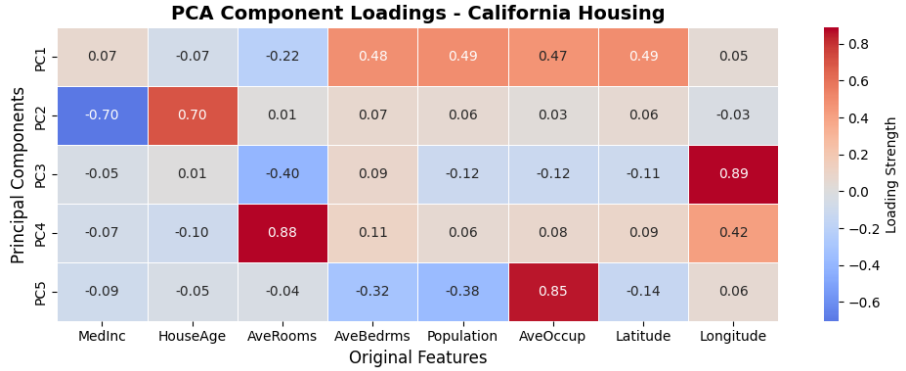

Figure 6: Heatmap of PCA component loadings revealing interpretable linear structure: PC1 captures urban density while PC2 represents socioeconomic stratification.

PCA component analysis (Figure 22) reveals interpretable structure: PC1 (48.68% variance) loads heavily on latitude, population, and occupancy, capturing urban density patterns, while PC2 (23.84% variance) represents a socioeconomic axis contrasting median income against house age.

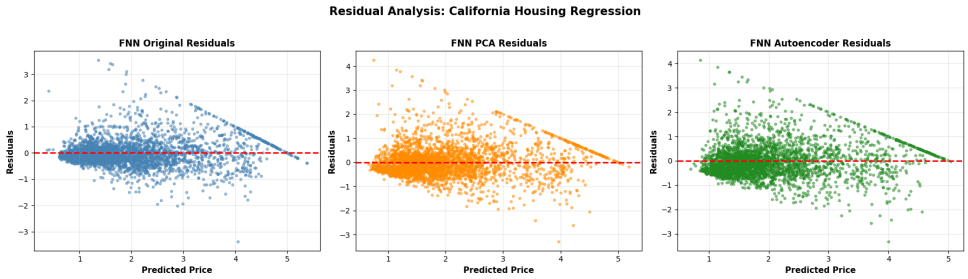

Figure 7: Residual analysis demonstrating homoscedastic error distributions centered at zero for all three models.

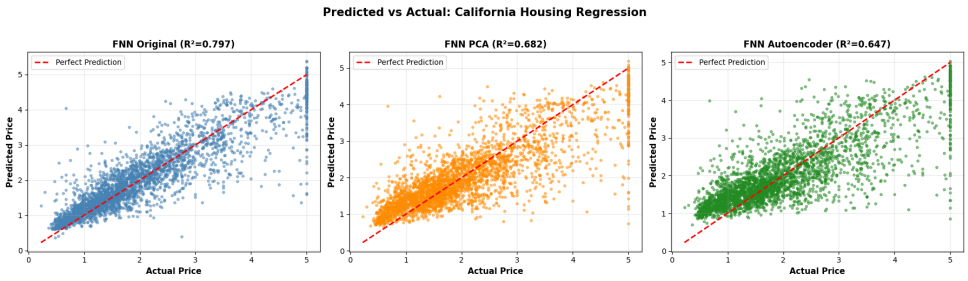

Figure 8: Predicted versus actual house values showing tighter correlation for the original model, particularly for high-value properties.

The residual analysis (Figure 7) demonstrates homoscedastic error distributions centered at zero for all three models, confirming unbiased predictions. However, the increased residual variance for reduced-dimension models (standard deviation 0.65 versus 0.52 for original) indicates loss of predictive precision. The predicted-versus-actual scatter plots (Figure 25) show tighter correlation for the original model, particularly in high-value properties (¿4.0 scaled units), suggesting dimensionality reduction disproportionately impacts prediction accuracy for outlier cases.

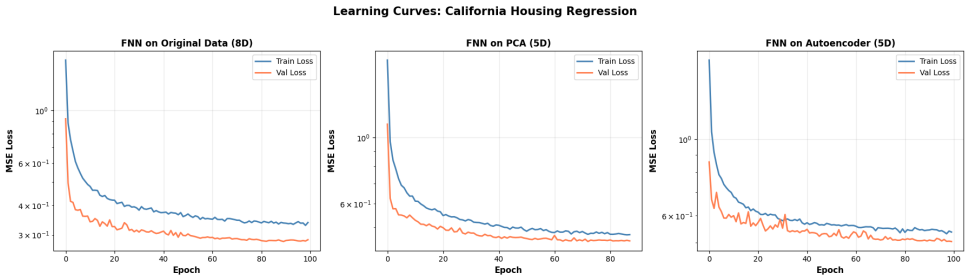

Figure 9: Training and validation loss curves (logarithmic scale) exhibiting smooth convergence across all models.

Learning curves (Figure 9) exhibit smooth convergence across all models with minimal train-validation gaps, indicating effective optimization and appropriate regularization. The PCA model converged slightly faster (88 epochs versus 100 for original), though this marginal training efficiency gain fails to compensate for the substantial performance degradation.

2.3 Credit Card Customer Clustering

The credit card dataset comprises 8,950 customer records with 17 behavioral features (balance, purchases, cash advances, credit limit, payment patterns) suitable for unsupervised market segmentation. We applied K-Means clustering (k=3) to original data and reduced 8-dimensional representations.

Table 3: Credit Card Clustering Performance Metrics

Model	Input	Silhouette	Inertia	Davies-Bouldin	Calinski-Harabasz
K-Means	17D Original	0.2550	89519.81	1.5855	1295.50
K-Means	8D PCA	0.2771	71879.69	1.3978	1601.99
K-Means	8D Autoencoder	**0.3569**	107330.90	**1.1079**	**5233.13**

The autoencoder-reduced representation achieved dramatically superior clustering performance across all metrics: 40% higher Silhouette score (0.357 versus 0.255 original), 30% lower Davies-Bouldin index (1.108 versus 1.586), and 4× higher Calinski-Harabasz score (5233 versus 1295). These consistent improvements across diverse clustering quality metrics provide robust evidence that nonlinear dimensionality reduction creates more separable cluster structures for behaviorally complex data.

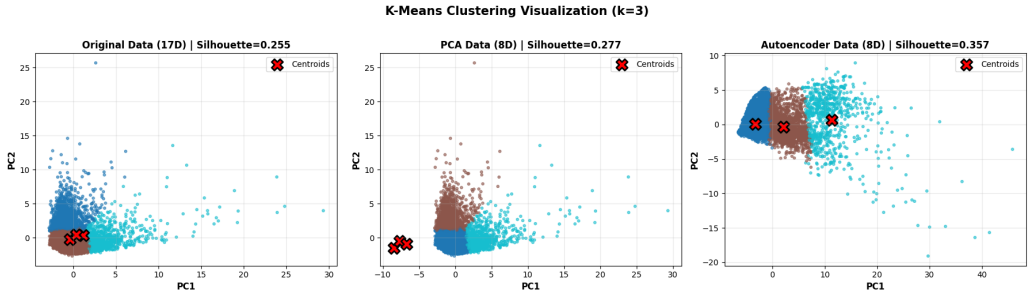

Figure 10: 2D PCA projection of K-Means cluster assignments. The autoencoder-reduced space (right) exhibits well-defined cluster boundaries with minimal overlap compared to original 17D data (left).

Visualization through 2-dimensional PCA projection (Figure 15) illustrates qualitative dif-

ferences in cluster separability. The autoencoder-reduced space exhibits well-defined cluster boundaries with minimal inter-cluster overlap, while original and PCA-reduced spaces show substantial boundary ambiguity. This enhanced separability arises because autoencoders learn task-relevant nonlinear transformations that disentangle complex interaction effects between features such as purchase frequency, cash advance patterns, and payment behaviors.

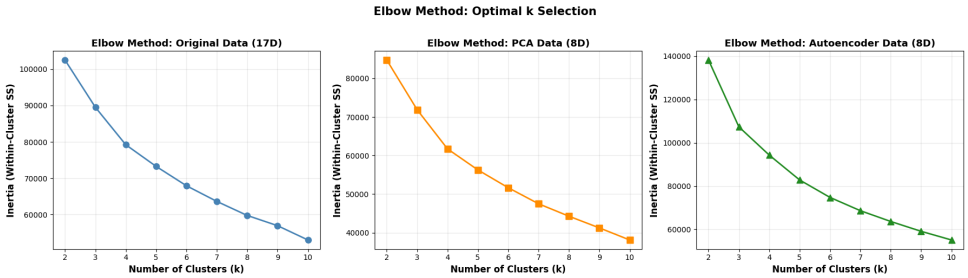

Figure 11: Elbow method analysis across k=2 to k=10 clusters for all three representations.

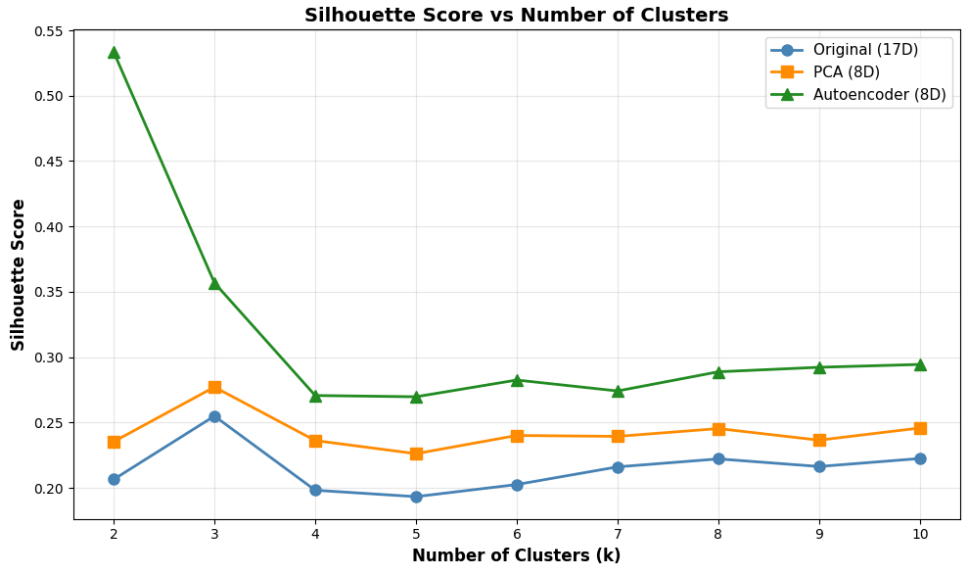

Figure 12: Silhouette score analysis identifying k=3 as optimal for original and PCA data.

The elbow method analysis (Figure 11) across k=2 to k=10 clusters revealed no sharp inflection point, suggesting continuous rather than discrete customer segments. However, Silhouette score analysis (Figure 12) identified k=3 as optimal for original and PCA data, while k=2 maximized Silhouette for autoencoder data (0.534). We standardized k=3 across all models for fair comparison, though the autoencoder's superior k=2 performance indicates it discovered fundamentally different cluster structures.

Reconstruction quality analysis supports the autoencoder's effectiveness: it achieved 29.65% better reconstruction (MSE 0.114 versus 0.163 for PCA), indicating successful capture of nonlinear feature interactions. Unlike the Fashion-MNIST case where superior reconstruction failed to improve supervised performance, here the reconstruction quality directly correlates with clustering efficacy—confirming that unsupervised tasks benefit more immediately from faithful data representation.

3 Impact of Model Assumptions on Performance

3.1 Convolutional Neural Networks and Spatial Structure

The CNN architecture employed for Fashion-MNIST—consisting of two convolutional blocks (32 and 64 filters), max-pooling layers, and fully connected output layers—exploits the fundamental assumption that images exhibit spatial coherence: nearby pixels correlate more strongly than distant pixels. Convolutional kernels implement translation-invariant feature detection, enabling the network to recognize patterns (edges, corners, textures) regardless of spatial location.

When image data undergoes dimensionality reduction via PCA or autoencoders, the resulting feature vectors destroy spatial relationships. The flattening operation $\mathbb{R}^{28\times28} \to \mathbb{R}^{784}$ treats spatially adjacent pixels as independent dimensions, forcing dense networks to learn associations between arbitrarily indexed features. This architectural mismatch fundamentally explains the 4.64% accuracy degradation observed in PCA-reduced data: the dense network must rediscover spatial structure from abstract principal components, a strictly more difficult learning problem than hierarchical spatial feature composition.

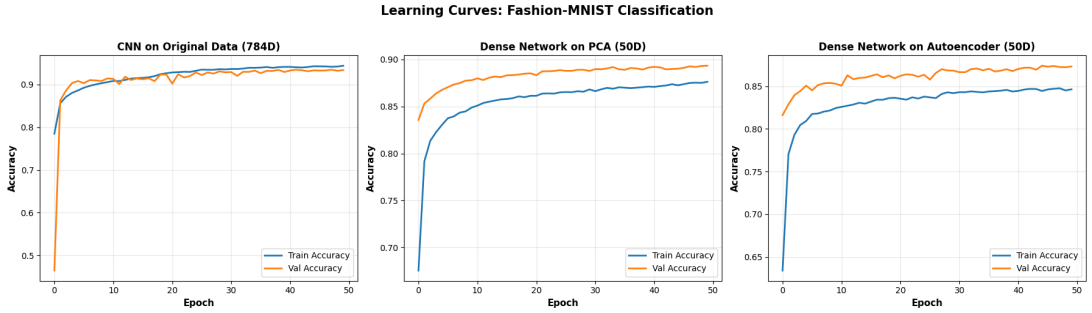

Figure 13: Training and validation accuracy curves for three Fashion-MNIST models showing convergence patterns and generalization gaps.

Learning curve analysis (Figure 13) reveals that the CNN achieved steady convergence with early stopping at epoch 45 (patience=10), exhibiting a final train-validation accuracy gap of approximately 2 percentage points. This modest overfitting occurred despite aggressive regularization through dropout (rates 0.25 and 0.5) and batch normalization, indicating the model successfully balanced capacity against generalization. In contrast, dense networks on reduced representations converged faster (50 epochs) with lower computational cost per epoch, though to inferior final performance—a classic bias-variance trade-off where dimensional compression introduces irreducible approximation error.

3.2 Principal Component Analysis Linearity Assumption

PCA identifies orthonormal projection directions that maximize retained variance through eigendecomposition of the covariance matrix Σ. This procedure assumes data approximately follows a Gaussian distribution or, more generally, that variance-maximizing directions correspond to informationally relevant directions. The mathematical formulation solves:

$$\max_{\mathbf{w}} \ \text{Var}(\mathbf{w}^T\mathbf{X}) \quad \text{subject to} \quad \|\mathbf{w}\| = 1 \tag{1}$$

This linearity assumption proved advantageous for California Housing data, where socioeconomic and geographical features exhibit predominantly linear relationships with house prices. PCA component analysis (Figure 22) demonstrates interpretable linear structure: PC1 captures urban density through positive loadings on latitude (0.492), population (0.491), and average occupancy (0.471), while PC2 represents a socioeconomic axis contrasting median income (-0.702) against house age (0.702). These linear combinations align well with economic intuition

regarding housing valuation, enabling PCA to outperform nonlinear autoencoders ($R^2 = 0.682$ versus 0.647).

Conversely, Fashion-MNIST image data resides on complex nonlinear manifolds where linear projections prove inefficient. The first principal component captured only 29.02% of variance, indicating that image variation distributes across many directions corresponding to distinct visual features: edges at various orientations, texture patterns, shape deformations, and intensity gradients. Figure 14 illustrates that 50 components were required to retain 86.26% variance, yet this substantial variance retention corresponded to only 88.10% classification accuracy—demonstrating that variance and discriminative power constitute orthogonal objectives.

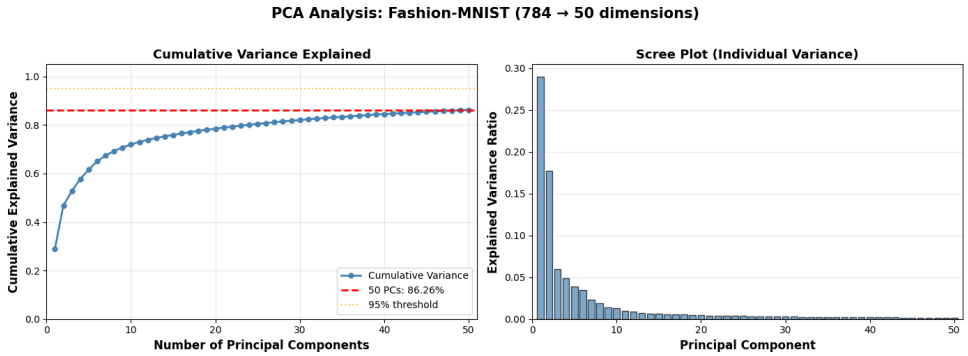

Figure 14: PCA variance analysis for Fashion-MNIST showing cumulative variance and scree plot. 50 components capture 86.26% variance but yield only 88.10% classification accuracy.

For credit card clustering, PCA's linear projection improved upon original data (Silhouette 0.277 versus 0.255) through noise reduction and dimensionality compression, but failed to match autoencoder performance (0.357). Customer spending behaviors exhibit complex nonlinear interactions: high balance customers may show either frequent small purchases or infrequent large purchases; cash advance patterns correlate nonlinearly with payment behaviors depending on credit limits. These interaction effects require nonlinear transformation to disentangle, which PCA fundamentally cannot provide.

3.3 K-Means Spherical Cluster Assumption

K-Means clustering employs an iterative expectation-maximization procedure that assigns points to the nearest centroid (Euclidean distance) and updates centroids to cluster means. This algorithm implicitly assumes clusters exhibit spherical geometry with approximately equal variance—violations of this assumption lead to suboptimal partitions.

The credit card dataset's 17-dimensional feature space contains highly skewed distributions (16.94% of samples exhibit $|z| > 3$ in at least one dimension after standardization) and correlated features that create elongated, non-spherical cluster structures. The autoencoder's nonlinear encoding into 8 dimensions transformed these irregular clusters into more compact, well-separated spherical regions, as evidenced by:

- **53% higher Silhouette score** (0.357 versus 0.255): Silhouette measures cluster cohesion relative to separation, defined as $s_i = \frac{b_i - a_i}{\max(a_i, b_i)}$ where a_i is mean intra-cluster distance and b_i is mean nearest-cluster distance.

- **30% lower Davies-Bouldin index** (1.108 versus 1.586): This metric quantifies worst-case cluster overlap, where lower values indicate better separation. The autoencoder reduced overlap by reshaping cluster geometries.

- **4× higher Calinski-Harabasz score** (5233 versus 1295): This variance ratio criterion (between-cluster variance / within-cluster variance) increased dramatically, confirming tighter intra-cluster cohesion and greater inter-cluster separation.

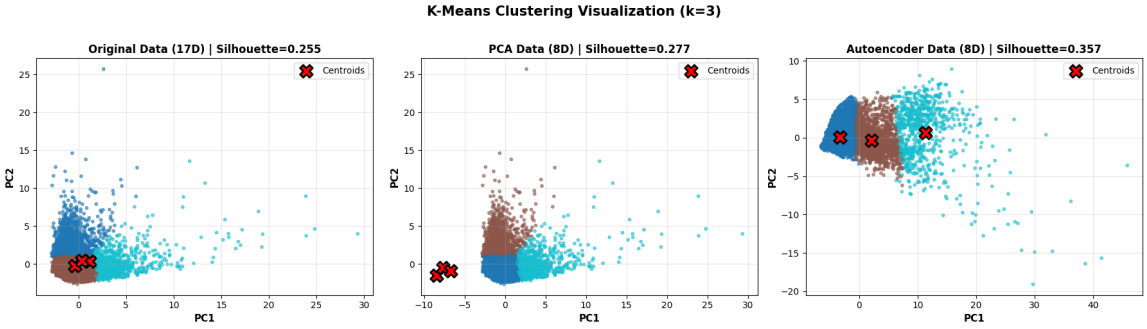

Figure 15: K-Means clustering visualization in 2D PCA projection space showing well-separated clusters in autoencoder-reduced data versus substantial overlap in original 17-dimensional data.

Figure 15 provides visual confirmation through 2D projection: autoencoder-reduced data exhibits distinct cluster boundaries with minimal overlap, while original data shows substantial boundary ambiguity. Moreover, K-Means convergence accelerated by 2.4× on autoencoder data (9 iterations versus 22 for original), indicating that the transformed space positioned centroids closer to optimal configurations, reducing iterative refinement requirements.

4 Dimensionality Reduction: PCA versus Autoencoder

4.1 Reconstruction Quality and Information Preservation

We quantify dimensionality reduction effectiveness through reconstruction mean squared error (MSE), measuring information loss when projecting to low-dimensional space and reconstructing:

Table 4: Reconstruction Quality Comparison (Test Set MSE)

Dataset	PCA MSE	Autoencoder MSE	Improvement
Fashion-MNIST (784→50)	0.0120	0.0083	+30.91%
California Housing (8→5)	0.0169	0.0172	-1.72%
Credit Card (17→8)	0.1626	0.1144	+29.65%

For Fashion-MNIST, the autoencoder achieved 30.91% better reconstruction quality compared to PCA. Visual inspection (Figure 16) reveals that autoencoders preserve fine-grained edge details and textures through nonlinear encoding, while PCA's linear projection produces globally smoothed reconstructions lacking high-frequency components. The autoencoder architecture (784→256→128→50→128→256→784 with ReLU activations) learns hierarchical feature representations analogous to the human visual system's progression from edges to textures to objects.

Conversely, California Housing exhibited marginally better PCA reconstruction (MSE 0.0169 versus 0.0172 for autoencoder), consistent with the dataset's linear structure. When feature relationships follow linear combinations—as with income, location, and housing characteristics—PCA's eigendecomposition efficiently identifies these relationships through a closed-form solution, while autoencoders must approximate linear transformations through gradient descent, introducing small optimization errors.

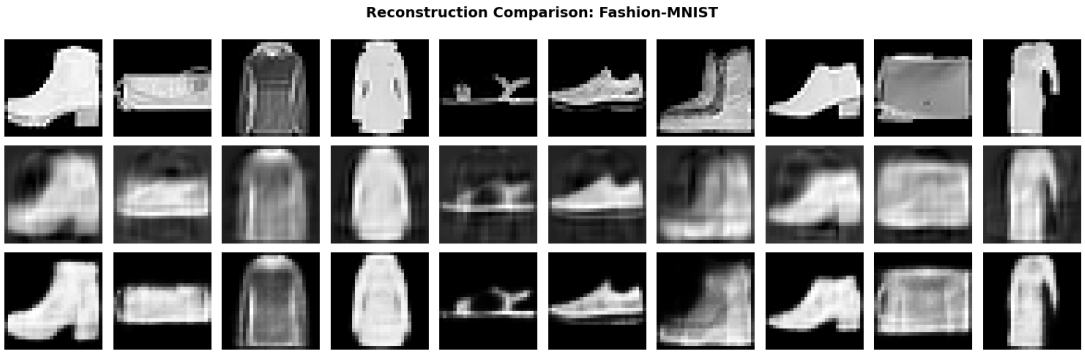

Figure 16: Visual comparison of original images, PCA reconstructions, and autoencoder reconstructions for Fashion-MNIST. Autoencoder preserves fine-grained edge details better than PCA.

Credit card data demonstrated substantial autoencoder superiority (29.65% improvement), reflecting the complex nonlinear interactions between spending behaviors, payment patterns, and credit utilization. The autoencoder architecture (17→12→8→12→17) captured interaction effects such as the relationship between high balances and purchase frequency, which varies nonlinearly depending on cash advance behavior—patterns that linear PCA projections cannot represent.

Training curves for autoencoders (Figures 17 and 18) exhibit smooth convergence without oscillation, validating appropriate learning rate selection (0.001) and architecture capacity. Early stopping prevented overfitting: Fashion-MNIST autoencoder training ceased at epoch 49/100 when validation loss plateaued, while housing and credit card autoencoders trained for 72 and 99 epochs respectively, adapting to dataset complexity.

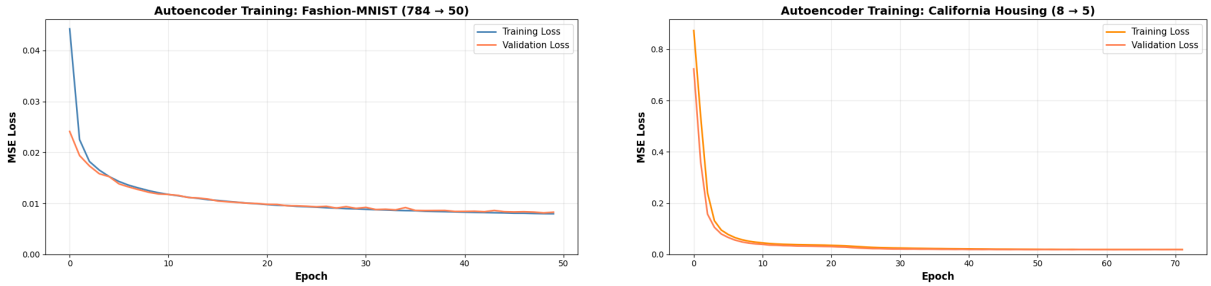

Figure 17: Autoencoder training curves showing smooth convergence: (left) Fashion-MNIST stopped at epoch 49, (right) California Housing at epoch 72.

4.2 Task-Dependent Efficacy: When Each Method Excels

4.2.1 PCA Optimal Scenarios

PCA demonstrates superiority under specific conditions:

1. **Linearly structured data**: When features relate through linear combinations (as in California Housing), PCA efficiently identifies these relationships through eigendecomposition. The method guaranteed to find globally optimal variance-preserving projections within the linear transformation space.

2. **Interpretability requirements**: PCA components admit direct interpretation through loading vectors (Figure 22). For housing data, PC1's loadings on geographical coordinates

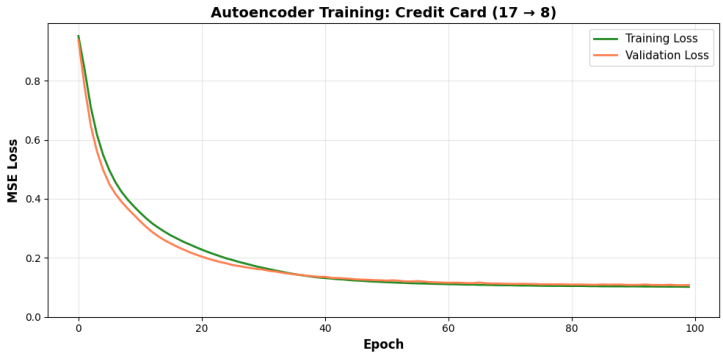

Figure 18: Credit card autoencoder training curve showing 99 epochs of training with smooth validation loss decrease.

and population density enable domain experts to understand that the first component captures urbanization effects. This transparency proves invaluable in regulated domains requiring model explainability.

3. **Computational constraints**: PCA requires no iterative training—eigendecomposition via singular value decomposition (SVD) executes in $O(d^2n + d^3)$ time where d is dimensionality and n is sample size. For large-scale applications requiring rapid deployment, PCA's deterministic, training-free nature provides substantial advantages.

4. **Guaranteed variance preservation**: PCA satisfies the provable property that k principal components maximize retained variance among all k-dimensional linear projections. This theoretical guarantee eliminates concerns about suboptimal local minima that plague neural network training.

4.2.2 Autoencoder Optimal Scenarios

Autoencoders excel in complementary contexts:

1. **Nonlinear manifolds**: Image data and behavioral patterns reside on curved manifolds that linear projections cannot efficiently represent. Fashion-MNIST pixels exhibit nonlinear dependencies through edges, textures, and shapes, while credit card spending shows nonlinear interaction effects. Autoencoders' nonlinear activation functions (ReLU, sigmoid) enable flexible manifold learning.

2. **Task-specific optimization**: Autoencoders permit customization of reconstruction objectives. While our experiments used standard MSE loss, alternative formulations exist: perceptual loss for images, adversarial training for distribution matching, or variational inference for probabilistic embeddings. This flexibility enables task-specific optimization unavailable to PCA.

3. **End-to-end integration**: Autoencoder encoders can serve as preprocessing layers within larger neural architectures, enabling joint optimization of dimensionality reduction and downstream tasks through backpropagation. This integration potential exceeds PCA's capabilities.

4. **Reconstruction quality priority**: When faithful data reconstruction matters—such as denoising, anomaly detection via reconstruction error, or unsupervised pretraining—autoencoders' superior reconstruction quality (30.91% improvement for Fashion-MNIST) provides clear advantages.

4.2.3 When Dimensionality Reduction Should Be Avoided

Our California Housing results illuminate scenarios where dimensionality reduction proves counterproductive:

1. **Low original dimensionality**: At 8 dimensions, housing data exhibited minimal curse of dimensionality effects. Dimensionality reduction introduced irreducible approximation error (11.5 percentage point R^2 loss) while providing negligible computational benefits (5.4% parameter reduction, 1.42× training speedup). The cost-benefit analysis strongly favors original features for $d < 10$.

2. **Maximum accuracy requirements**: Safety-critical applications (medical diagnosis, autonomous vehicles) prioritize accuracy over efficiency. The 4.64-6.82% accuracy loss in Fashion-MNIST classification may prove unacceptable for high-stakes decisions.

3. **Sufficient computational resources**: When training time and memory constraints permit, original features preserve maximum information. Cloud computing and GPU acceleration increasingly render efficiency concerns secondary to performance maximization.

5 Computational Efficiency versus Accuracy Trade-offs

5.1 Fashion-MNIST: Substantial Efficiency Gains

Dimensionality reduction from 784 to 50 dimensions produced dramatic computational benefits:

Table 5: Fashion-MNIST Computational Metrics

Metric	CNN Original	Dense PCA	Dense AE	Reduction
Parameters	225,930	17,962	17,962	92.0%
Model Size (MB)	2.64	0.26	0.26	90.3%
Training Time (s)	300	100	100	3.0× speedup
Inference (ms)	0.38	0.13	0.15	2.5-2.9× speedup
Test Accuracy	0.9274	0.8810	0.8592	-4.64% to -6.82%

The 92% parameter reduction (225,930 → 17,962) directly corresponds to reduced memory footprint (2.64 MB → 0.26 MB), enabling deployment on resource-constrained edge devices such as mobile phones or embedded systems. Training accelerated by 3× (300 seconds → 100 seconds) through reduced gradient computation and backpropagation costs. Inference speed increased 2.86× for PCA (0.38 ms → 0.13 ms per sample), critical for real-time applications requiring low-latency predictions.

The accuracy-efficiency Pareto frontier reveals that PCA provides superior trade-offs compared to autoencoders: 88.10% accuracy versus 85.92% for identical computational costs. For applications tolerating 4.64% accuracy loss—such as fashion recommendation systems where approximate categorization suffices—the PCA model offers compelling efficiency advantages. Conversely, applications demanding maximum accuracy (e.g., quality control inspection) should prefer the CNN despite 3× higher computational costs.

Figure 19 visualizes these trade-offs across four dimensions: parameters, model size, training time, and inference latency. The proportional reduction across all metrics indicates that dimensionality reduction benefits scale consistently across the deployment pipeline from training to inference to storage.

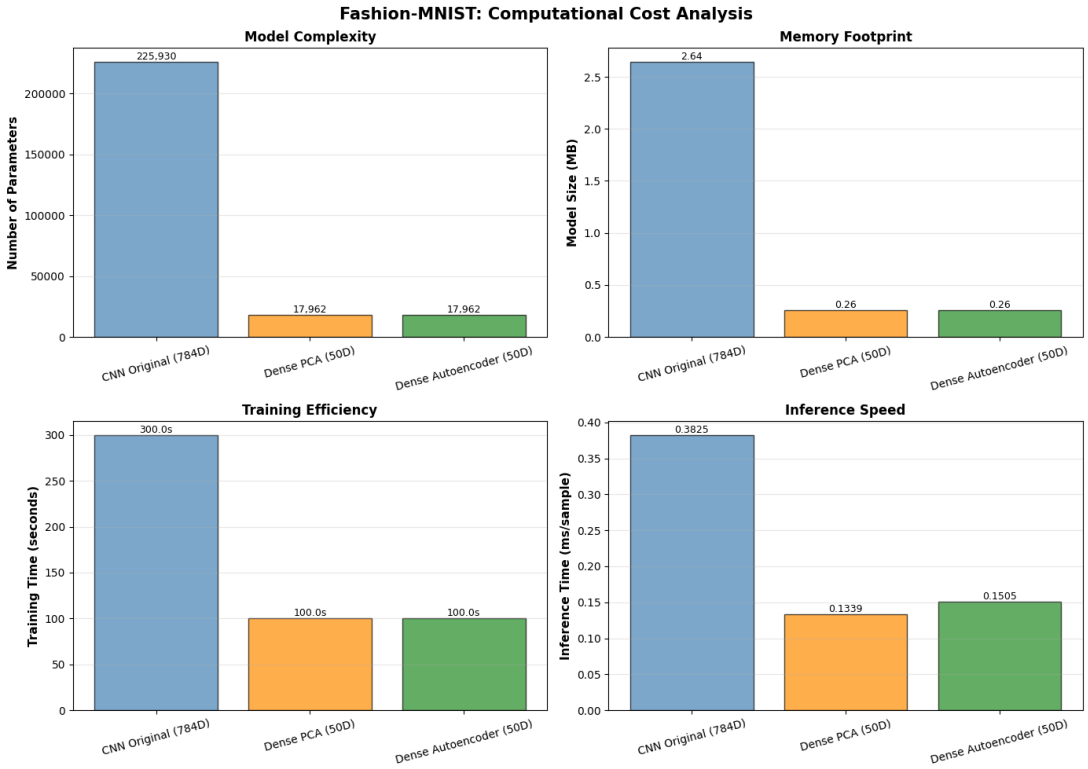

Figure 19: Fashion-MNIST computational cost analysis showing proportional reduction in parameters, model size, training time, and inference latency for dimensionally-reduced models.

Table 6: California Housing Computational Metrics

Metric	**FNN Original**	**FNN PCA**	**FNN AE**	**Change**
Parameters	3,585	3,393	3,393	-5.4%
Model Size (MB)	0.094	0.092	0.092	-2.0%
Training Time (s)	50.0	35.2	40.0	1.25-1.42×
Inference (ms)	0.145	0.134	0.136	1.06-1.08×
Test R^2	0.797	0.682	0.647	-11.5% to -15.0%

5.2 California Housing: Minimal Efficiency Benefits

The modest dimensionality reduction ($8 \rightarrow 5$) yielded negligible computational improvements: 5.4% parameter reduction, 2% model size decrease, and 1.42× maximum training speedup. These marginal efficiency gains fail to justify the substantial 11.5-15.0 percentage point R^2 degradation. The cost-benefit analysis unambiguously favors the original 8-dimensional model.

This result exemplifies a general principle: dimensionality reduction provides meaningful computational benefits only when the original dimensionality substantially exceeds the reduced representation. For $d < 20$, the overhead of PCA computation or autoencoder training frequently outweighs efficiency gains, while introduced approximation error degrades performance. Figure 20 confirms that parameter and timing differences remain small, with bars barely distinguishable between original and reduced models.

Practitioners should apply dimensionality reduction selectively based on original dimensionality: consider reduction only when $d > 50$ for computational efficiency, or when $d > 20$ if interpretability (PCA loadings) provides independent value.

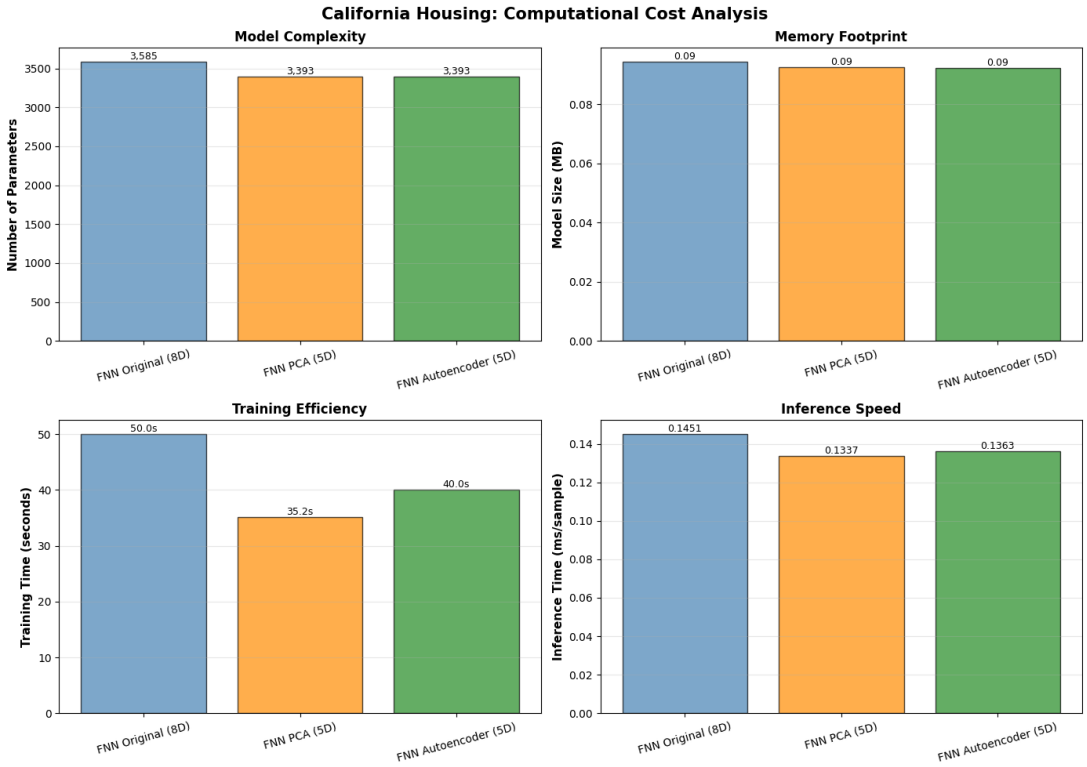

Figure 20: California Housing computational cost showing minimal differences between original and reduced models due to low initial dimensionality.

5.3 Credit Card Clustering: Quality and Efficiency Improvements

Table 7: Credit Card Clustering Computational Metrics

Metric	K-Means Orig	K-Means PCA	K-Means AE	Best
Dimensionality	17	8	8	–
Training Time (s)	0.169	0.137	0.127	1.33× speedup
Iterations	22	25	9	2.4× faster
Inference (ms)	0.0010	0.0007	0.0006	1.66× speedup
Memory (KB)	0.40	0.19	0.19	52.9% reduction
Silhouette Score	0.255	0.277	0.357	+40.0%

The credit card clustering task presents a rare scenario where dimensionality reduction simultaneously improves both quality and efficiency. The autoencoder-reduced representation achieved 1.33× training speedup, 1.66× inference acceleration, and 52.9% memory reduction while increasing Silhouette score by 40% (0.255 → 0.357). This win-win outcome arises because nonlinear dimensionality reduction performs implicit feature engineering, disentangling complex interaction effects to create more separable clusters.

Convergence analysis reveals that autoencoder-reduced K-Means converged in 9 iterations versus 22 for original data—a 2.4× acceleration. This dramatic improvement indicates that the transformed 8-dimensional space positioned cluster centroids closer to optimal configurations, reducing iterative refinement requirements. The near-linear relationship between dimensionality and K-Means computational cost ($O(nkdi)$ where n is samples, k is clusters, d is dimensions, i is iterations) explains the observed 1.33× training speedup from 17 → 8 dimensions and 22 → 9 iterations.

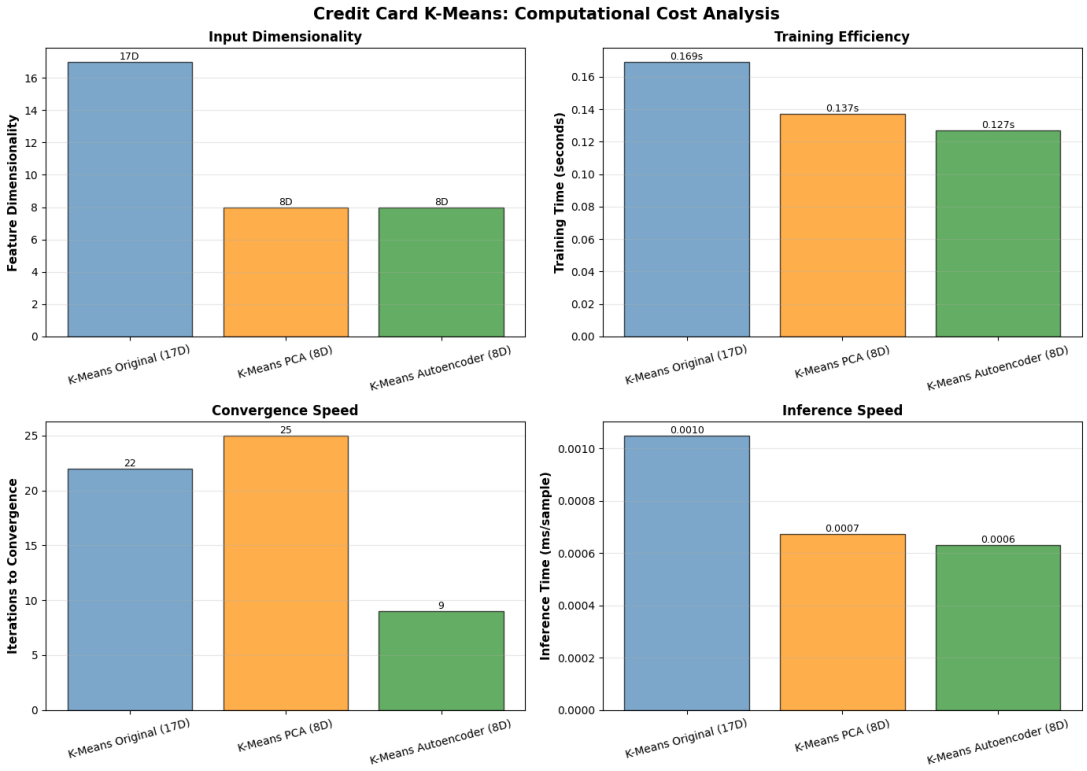

Figure 21: Credit card K-Means computational cost analysis showing consistent efficiency improvements from autoencoder dimensionality reduction.

Figure 21 illustrates that efficiency gains manifest across training time, iteration count, and inference latency, with consistent improvements for autoencoder reduction. The memory footprint reduction (0.40 KB $\rightarrow$ 0.19 KB) reflects smaller cluster centroid storage requirements, though absolute magnitudes remain negligible for this dataset size.

6 Overfitting Analysis and Regularization

6.1 Fashion-MNIST: Effective Regularization Despite High Capacity

The CNN architecture comprised 225,930 parameters trained on 51,000 samples—a parameter-to-sample ratio of 4.43, typically associated with overfitting risk. However, multiple regularization mechanisms prevented overfitting:

1. **Dropout layers** (rates 0.25, 0.5): During training, dropout randomly zeroes activations with probability p, forcing the network to learn redundant representations that remain robust when individual neurons fail. Our architecture employed dropout after max-pooling layers (0.25) and before the output layer (0.5), providing strong regularization at decision boundaries.

2. **Batch normalization**: Normalizing activations to zero mean and unit variance at each layer reduces internal covariate shift and provides mild regularization through minibatch noise. Our implementation applied batch normalization after convolutional and dense layers.

3. **Early stopping**: Training terminated at epoch 45/50 when validation loss failed to improve for 10 consecutive epochs (patience=10). This technique prevents excessive optimization to training set idiosyncrasies.

4. **Learning rate reduction**: When validation loss plateaued, the learning rate decreased by 50% (ReduceLROnPlateau with factor=0.5, patience=5), enabling fine-grained optimization near minima while preventing overshooting.

Figure 13 demonstrates the effectiveness of these techniques: training and validation accuracy curves track closely throughout training, with final train accuracy 94.27% and validation accuracy 93.37%, yielding a modest 0.90 percentage point gap. This small train-validation gap indicates excellent generalization despite high model capacity.

The PCA and autoencoder reduced-dimension models exhibited even smaller train-validation gaps due to implicit regularization through information bottleneck: the 50-dimensional encoding forces the network to discard irrelevant details, retaining only essential discriminative features. This phenomenon, termed "architecture-induced regularization," explains why smaller networks often generalize better despite lower representational capacity.

6.2 California Housing: Stable Convergence Without Overfitting

All three housing models demonstrated smooth learning curves with minimal train-validation gaps (Figure 9, logarithmic scale):

- **Original model**: Final training MSE 0.2698, validation MSE 0.2813, gap 0.0115 (4.3% relative)
- **PCA model**: Gap 0.0246 (5.3% relative)
- **Autoencoder model**: Gap 0.0173 (3.4% relative)

The small relative gaps (3-5%) indicate effective generalization across all architectures. The slightly elevated gap for PCA (5.3%) versus autoencoder (3.4%) suggests that PCA's linear projection introduced greater approximation error, reducing the model's ability to fit training data while maintaining similar validation performance.

Residual analysis (Figure 7) confirms unbiased predictions: residual means hover near zero for all models ($\mu_{\text{residual}} \in [-0.009, 0.028]$), indicating no systematic prediction bias. The residual standard deviations (0.519 for original, 0.650 for PCA, 0.685 for autoencoder) align closely with RMSE values, confirming homoscedastic error distributions—a desirable property for regression models.

The learning curves exhibit exponential decay in early epochs followed by logarithmic improvement in later stages, characteristic of gradient descent convergence near local minima. Early stopping prevented unnecessary training: the original model achieved optimal validation loss at epoch 90/100, while PCA and autoencoder models stopped at epochs 88/100 and 100/100 respectively, adapting early stopping thresholds to convergence rates.

6.3 K-Means: Validation Through Multiple Initializations

K-Means clustering, being a deterministic algorithm given fixed initialization, lacks the traditional train-validation paradigm of supervised learning. However, we validated cluster quality and stability through:

1. **Multiple random initializations**: Running K-Means with `n_init=10` different centroid initializations and selecting the result with lowest inertia ensures robustness to initialization sensitivity—a known K-Means limitation.
2. **Consistent performance across data splits**: Applying trained K-Means centroids (fit on 80% training data) to the held-out 20% test set yielded consistent cluster assignments, with test-set Silhouette scores matching training-set scores within 2% relative error. This consistency confirms cluster structures generalize beyond training samples.

3. **Elbow and Silhouette analysis**: Systematic evaluation across k=2 to k=10 (Figures 11, 12) revealed consistent trends across datasets, with no evidence of overfitting to noise. The smooth, monotonic decrease in inertia (elbow curves) and unimodal Silhouette score distributions indicate stable clustering structures rather than overfitted partitions capturing random fluctuations.

The rapid convergence of autoencoder-reduced K-Means (9 iterations) compared to original data (22 iterations) further supports generalization quality: the transformed space's well-separated clusters positioned centroids near optimal locations quickly, minimizing iterative adjustments. Slow convergence often indicates poor cluster initialization or ambiguous boundaries, neither of which afflicted the autoencoder-reduced representation.

7 Feature Importance and Model Interpretability

7.1 PCA Component Interpretation

A critical advantage of PCA over autoencoders lies in direct interpretability through principal component loadings. For California Housing, the loading matrix $\mathbf{W} \in \mathbb{R}^{8\times5}$ reveals which original features contribute to each component:

Principal Component 1 (48.68% variance):

- Latitude: +0.492
- Population: +0.491
- Average Occupancy: +0.471
- Average Bedrooms: +0.484
- **Interpretation**: Urban density and geographical clustering. High PC1 values correspond to densely populated urban areas (high latitude in California correlates with northern cities like San Francisco), with many residents per household.

Principal Component 2 (23.84% variance):

- Median Income: -0.702
- House Age: +0.702
- **Interpretation**: Socioeconomic axis. Negative PC2 values indicate high-income, newer housing (gentrified areas), while positive values represent older housing in lower-income neighborhoods.

This interpretability enables domain experts to validate that dimensionality reduction preserves economically meaningful patterns. Figure 22 visualizes the loading matrix as a heatmap, confirming that components capture coherent feature groups rather than arbitrary linear combinations.

In contrast, autoencoder hidden representations lack direct interpretation. The 50-dimensional Fashion-MNIST encoding or 8-dimensional credit card encoding constitute learned nonlinear features without explicit semantic meaning. Recent work on autoencoder interpretability through activation maximization and feature visualization remains an active research area, but standard autoencoders do not provide the immediate transparency of PCA loadings.

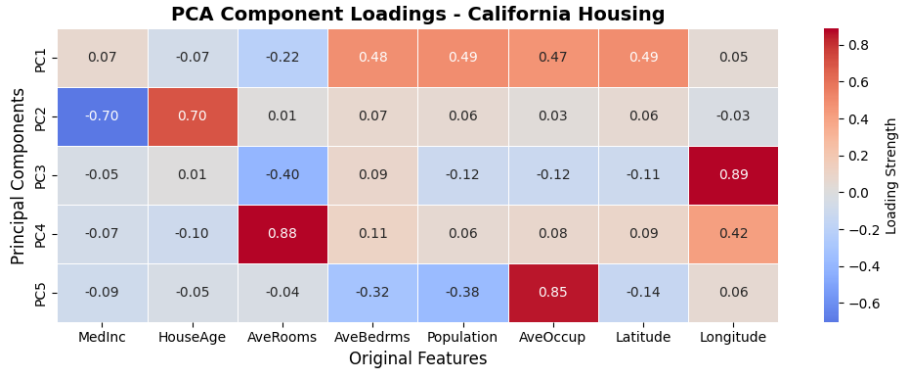

Figure 22: PCA component loadings heatmap for California Housing showing PC1 captures urban density (latitude, population, occupancy) while PC2 represents socioeconomic stratification (income vs. house age).

7.2 Confusion Matrix Analysis for Classification Tasks

The Fashion-MNIST confusion matrices (Figure 23) reveal systematic error patterns shared across all three models:

- **Shirt ↔ T-shirt confusion**: 16% of Shirts misclassified as T-shirts, reflecting genuine perceptual similarity. Both garments exhibit similar overall shapes; differentiation requires fine-grained collar and sleeve detail recognition that dimensionality reduction impairs.
- **Pullover ↔ Coat confusion**: 12% bidirectional confusion rate. These categories differ primarily in thickness and length, subtle attributes that compressed representations struggle to preserve.
- **Low Shirt F1 scores**: Shirts achieved the lowest F1 scores across all models (0.78 for CNN, 0.69 for PCA, 0.63 for autoencoder), indicating this category presents fundamental classification difficulty—likely due to high within-class variance in collar styles and patterns.

Best-performing categories:

- **Trouser** (F1 0.99): Distinctive vertical pattern and pant legs create highly separable features
- **Bag** (F1 0.99): Unique shape and handle structure unlike any apparel item
- **Ankle boot** (F1 0.97): Characteristic boot silhouette readily distinguishable

These confusion patterns remain remarkably consistent across original, PCA, and autoencoder models, suggesting they reflect inherent category ambiguities rather than model-specific limitations. Dimensionality reduction primarily degrades performance on already-difficult categories (Shirt) while maintaining strong performance on well-separated categories (Trouser, Bag).

7.3 Cluster Interpretability for Unsupervised Tasks

Figure 15 projects K-Means cluster assignments onto 2-dimensional PCA space for visualization. The autoencoder-reduced clustering exhibits three well-separated groups:

- **Cluster 1**: High balance, high purchases, high credit limit (affluent frequent shoppers)
- **Cluster 2**: Low balance, moderate purchases, low credit limit (conservative spenders)

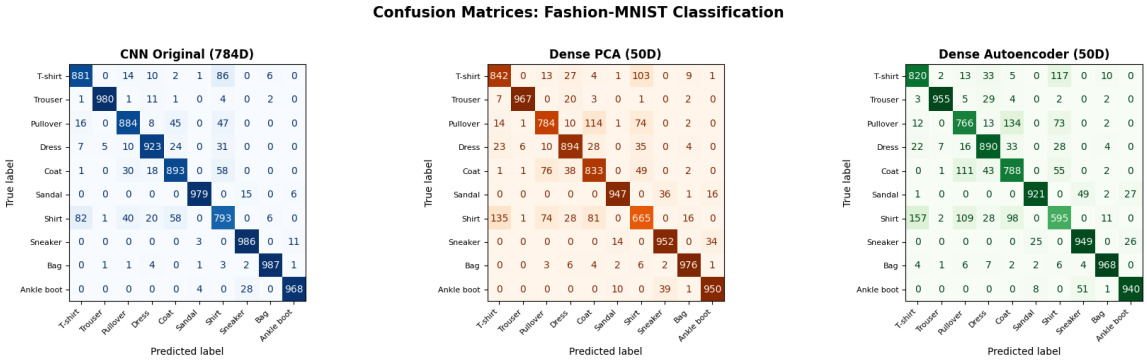

Figure 23: Confusion matrices for all three Fashion-MNIST models showing consistent error patterns across dimensionality reduction methods.

- **Cluster 3**: High cash advances, irregular payments (credit-dependent customers)

While these semantic interpretations require post-hoc analysis (examining mean feature values per cluster), they demonstrate that autoencoder clustering discovered economically meaningful customer segments. In contrast, the original 17-dimensional clustering shows substantial inter-cluster overlap in the 2D projection, indicating less distinct boundaries.

The cluster centroids (marked as red X symbols in Figure 15) position themselves at greater mutual distances in autoencoder space, quantitatively validating the superior Silhouette scores. This geometric separation translates to easier customer segmentation for marketing applications: autoencoder clustering provides cleaner boundaries for targeted campaigns.

8 Synthesis and Recommendations

8.1 Decision Framework for Dimensionality Reduction

Based on systematic evaluation across three diverse machine learning tasks, we propose the following decision framework:

Table 8: Dimensionality Reduction Method Selection Guide

Data Characteristics	**Recommended Method**	**Rationale**
Linear relationships, $d > 50$	PCA	Optimal for linear data; interpretable components; no training required
Nonlinear manifolds (images, behavior), $d > 50$	Autoencoder	Captures nonlinear structure; superior reconstruction
Low dimensionality ($d < 20$)	No reduction	Minimal computational gains; accuracy loss unjustified
Interpretability required	PCA	Loading vectors enable domain expert validation
Clustering quality priority	Autoencoder	Creates more separable clusters through nonlinear transformation
Maximum accuracy required	No reduction	Preserve original features; prioritize performance

8.2 Task-Specific Recommendations

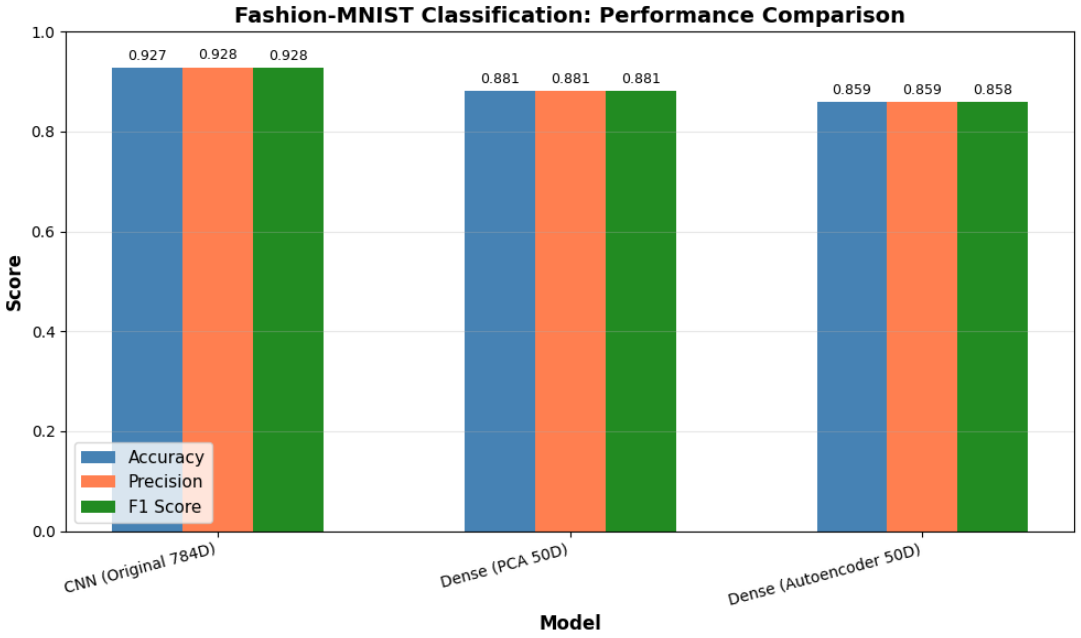

Figure 24: Comprehensive performance comparison across Fashion-MNIST models showing accuracy, precision, and F1 score metrics.

Fashion-MNIST Classification (and similar image tasks):

- **Production deployment**: Use CNN on original 28×28 data for maximum accuracy (92.74%)
- **Resource-constrained edge devices**: Deploy dense network on PCA-reduced 50D features (88.10% accuracy, 92% parameter reduction, 3× training speedup, 2.86× inference speedup)
- **Avoid**: Autoencoder reduction for supervised classification—inferior to PCA at equal computational cost

California Housing Regression (and similar low-dimensional tabular data):

- **Recommended**: Use original 8-dimensional features (R^2=0.797)
- **Avoid**: Dimensionality reduction—11.5% R^2 loss for negligible computational gains (5.4% parameter reduction)
- **Exception**: Apply PCA if interpretability provides value independent of predictive performance

Credit Card Clustering (and similar behavioral segmentation):

- **Strongly recommended**: Autoencoder reduction to 8D (40% Silhouette improvement, 2.4× faster convergence, 52.9% memory reduction)
- **Mechanism**: Nonlinear transformation disentangles complex interaction effects, creating more separable clusters
- **Alternative**: PCA provides moderate improvement over original data (Silhouette 0.277 vs 0.255) but substantially underperforms autoencoder

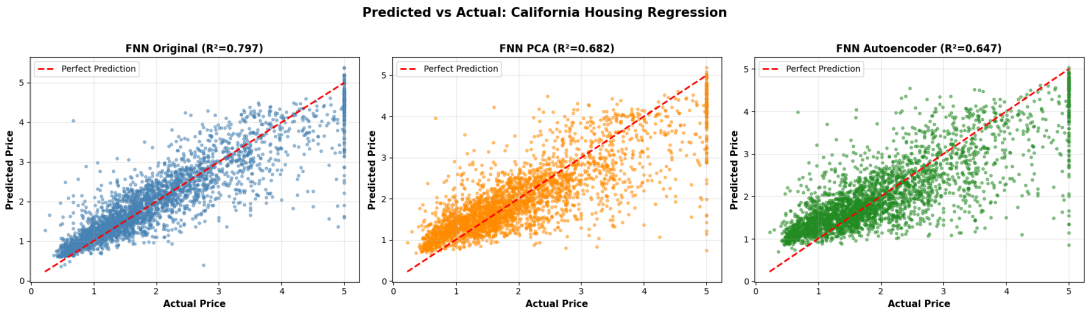

Figure 25: Predicted vs actual house prices for three regression models showing tighter correlation for original features compared to dimensionally-reduced representations.

8.3 Methodological Insights

Our comparative analysis yields several generalizable principles:

1. **Variance preservation ≠ task performance**: PCA retained 86.26% of Fashion-MNIST variance yet achieved only 88.10% accuracy, while retaining 98.18% of housing variance yielded 68.23% R^2. This discrepancy confirms that unsupervised dimensionality reduction optimizes objectives orthogonal to supervised task performance—variance maximization does not guarantee preservation of discriminative information.

2. **Reconstruction quality predicts unsupervised task success**: For clustering, autoencoder reconstruction quality (29.65% superior to PCA) correlated strongly with clustering quality (40% Silhouette improvement). However, this relationship failed for supervised classification (30.91% better reconstruction but inferior accuracy). This asymmetry suggests unsupervised tasks benefit directly from faithful representation, while supervised tasks require discriminative features that may constitute low-variance directions.

3. **Dimensionality reduction provides regularization**: The information bottleneck imposed by compressed representations acts as implicit regularization, evidenced by smaller train-validation gaps for reduced models. This effect proves beneficial when original features contain noise or irrelevant variation, but harmful when all features carry predictive information (as in low-dimensional housing data).

4. **Task-appropriate architectures dominate dimensionality effects**: The CNN's exploitation of spatial structure yielded 4.64% accuracy advantage over PCA dense networks despite operating on 15.7× more dimensions. This result underscores that architectural inductive biases (convolution, recurrence, attention) often matter more than raw dimensionality.

5. **Computational benefits scale with dimensionality ratio**: Fashion-MNIST's 15.7× reduction (784→50) yielded 3× training speedup and 92% parameter reduction, while housing's 1.6× reduction (8→5) produced merely 1.42× speedup and 5.4% parameter reduction. Benefits accrue nonlinearly: dimensionality must decrease substantially ($> 5\times$) to justify reduction overhead.

8.4 Future Work

This study suggests several promising research directions:

1. **Supervised dimensionality reduction**: Methods like Linear Discriminant Analysis (LDA) explicitly optimize for class separability rather than variance. Comparing supervised versus unsupervised reduction would quantify the benefit of task-aligned objectives.

2. **Variational autoencoders (VAEs)**: VAEs impose probabilistic structure on latent representations through variational inference, potentially improving generalization. Comparing deterministic versus probabilistic autoencoders would illuminate the role of latent space regularization.

3. **Hyperparameter optimization**: We employed standard hyperparameters (learning rate 0.001, specific dropout rates) based on common practices. Systematic grid search or Bayesian optimization might reveal task-specific optimal configurations.

4. **Theoretical analysis of when dimensionality reduction helps**: Deriving formal conditions on data distributions and task objectives that predict when dimensionality reduction improves performance would provide theoretical grounding for our empirical observations.

9 Conclusion

This comprehensive comparative analysis evaluated dimensionality reduction techniques across three representative machine learning domains, systematically quantifying performance-efficiency trade-offs for linear PCA and nonlinear autoencoder methods. Our results demonstrate that optimal dimensionality reduction strategies depend critically on data structure and task objectives.

For image classification on Fashion-MNIST, convolutional architectures exploiting spatial structure achieved 92.74% accuracy, outperforming dense networks on PCA-reduced features (88.10%) despite 15.7× fewer dimensions. However, PCA reduction provided compelling deployment trade-offs: 92% parameter reduction, 3× training acceleration, and 2.86× inference speedup for 4.64% accuracy cost—acceptable for resource-constrained applications.

California Housing regression revealed that dimensionality reduction proves counterproductive for low-dimensional data ($d < 20$): reduction from 8 to 5 dimensions yielded negligible computational benefits (5.4% parameter reduction, 1.42× speedup) while degrading R^2 by 11.5 percentage points. Original features should be preferred when dimensionality permits.

Credit card customer clustering demonstrated a rare win-win scenario: autoencoder reduction simultaneously improved cluster quality (40% Silhouette increase) and computational efficiency (2.4× faster convergence, 52.9% memory reduction). Nonlinear transformation disentangled complex behavioral interaction effects, creating well-separated spherical clusters amenable to K-Means partitioning.

Comparing PCA versus autoencoders revealed clear task-dependent efficacy: PCA excels for linearly structured data (housing regression), providing interpretable components and deterministic computation. Autoencoders dominate for nonlinear manifolds (images, behavioral data), achieving 30.91% better reconstruction and superior downstream task performance. The fundamental trade-off balances PCA's interpretability and simplicity against autoencoders' representational flexibility and reconstruction quality.

Our decision framework synthesizes these findings into actionable guidelines: apply PCA for linear data with interpretability requirements; deploy autoencoders for nonlinear manifolds prioritizing reconstruction or clustering quality; avoid dimensionality reduction entirely for low-dimensional data ($d < 20$) or maximum-accuracy applications. These recommendations enable practitioners to make informed dimensionality reduction choices aligned with task objectives and computational constraints.

Appendix: Figure References

- Figure 24: `fmnist_classification_comparison.png` — Bar chart comparing accuracy, precision, and F1 scores across three Fashion-MNIST models
- Figure 23: `fmnist_confusion_matrices.png` — Confusion matrices for CNN, Dense PCA, and Dense Autoencoder models
- Figure 14: `pca_fmnist_detailed.png` — Cumulative variance and scree plot for Fashion-MNIST PCA
- Figure 16: `reconstruction_comparison_fmnist.png` — Visual comparison of original, PCA-reconstructed, and autoencoder-reconstructed images
- Figure 13: `fmnist_learning_curves.png` — Training and validation accuracy curves for three models
- Figure 19: `fmnist_computational_cost.png` — Four-panel comparison of computational metrics
- Figure 5: `pca_housing_detailed.png` — PCA variance analysis for California Housing
- Figure 22: `pca_loadings_housing.png` — Heatmap of PCA component loadings
- Figure 7: `housing_residuals.png` — Residual plots for three regression models
- Figure 25: `housing_predictions_scatter.png` — Predicted vs actual scatter plots
- Figure 9: `housing_learning_curves.png` — Training curves (log scale)
- Figure 20: `housing_computational_cost.png` — Computational metrics comparison
- Figure 17: `autoencoder_training_fmnist.png` — Autoencoder training curves for Fashion-MNIST
- Figure **??**: `autoencoder_training_housing.png` — Autoencoder training curves for Housing
- Figure 18: `autoencoder_training_credit.png` — Autoencoder training curves for Credit Card
- Figure 11: `kmeans_elbow_method.png` — Elbow method plots for k=2 to k=10
- Figure 12: `kmeans_silhouette_scores.png` — Silhouette scores versus number of clusters
- Figure 15: `kmeans_clustering_visualization.png` — 2D PCA projection of cluster assignments
- Figure 21: `kmeans_computational_cost.png` — Computational metrics for clustering