A lossy compression account of pattern separation

Dale Zhou[1,2,3], Sharon Noh[2], Nora Harhen[3], Nidhi Banavar[3], Brock Kirwan[4], Michael Yassa[1,3], and Aaron Bornstein[2,5]

[1]University of California, Irvine, Neurobiology and Behavior, 519 Biological Sciences Quad, Irvine, 92697, United States

[2]University of California, Irvine, Center for the Neurobiology of Learning and Memory, Qureshey Research Laboratory, Irvine, 92697, United States

[3]University of California, Irvine, Department of Cognitive Sciences, Social Science Lab 334, Irvine, 92697, United States

[4]University of Pennsylvania,School of Arts & Sciences, Philadelphia, 19104, United States

[5]To whom correspondence should be addressed: aaron.bornstein@uci.edu

October 12, 2025

To demonstrate the intuition behind how distortion can support discrimination in practice, we first apply distortion and dimensionality metrics to a popular machine learning dataset. Does pattern separating handwritten numbers in an image dataset into digits benefit from dimensionality reduction or expansion [1]? To obtain encodings, images were decorrelated into orthogonalized components using principal components analysis. After dimensionality reduction, reconstructions of the original inputs were generated by decoding from 2 to 784 (the number of pixels) principal components and define a distortion metric as the mean squared error **(Figure 1A)**. Visualizing all of the images in a 2-dimensional space highlights how distinct digits can become more separated (using a non-linear dimensionality reduction method for illustration) **(Figure 1B)**. When fewer dimensions are used to encode X_1, the reconstruction X_2 has a larger error $d(X_1, X_2)$ **(Figure 1C)**. Expansion can reconstruct all the variance of the source but the addition of principal components has diminishing returns for information about fine details.

With these representations, we can assess the quality of pattern separation under differing levels of compression or expansion. The quality of separation is measured by the distance between clusters, the quality of clustering, and the classification accuracy. The Euclidean distance across encoding dimensions represents the separability of images in each class of digits. The Euclidean distance increases with the raw dimensionality (lighter blues), seemingly supporting the dimensionality expansion hypothesis **(Figure 1D)**. However, in high-dimensional statistical learning, the dimensionality d is commonly normalized by $\sqrt{d}$ to help counter the "curse of dimensionality." The curse refers to how increasing dimensionality tends to indiscriminately expand the distances between all points, making variations in distance lower, less meaningful, and less separable in practice. Indeed, after normalizing the dimensionality, Euclidean distance decreases with increasing dimensionality. Consistent with the compression hypothesis, Euclidean distance increases with distortion incurred by reducing the number of normalized dimensions.

Next, we evaluate clustering and classification of compressed versus expanded representations. Clustering quality is measured by the silhouette score, the average intra-cluster distance to the nearest-cluster distance. A score of 1 indicates good separation, 0 indicates some overlap, and -1 indicates erroneous overlap. Classification accuracy is measured by k-nearest neighbor classification using $k = 200$, where an image is classified based on the majority class among its k-closest neighbors in the space of encoding dimensions. Consistent with the compression hypothesis, the greatest silhouette scores and classification accuracies are achievable using more compact lower-dimensional representations that incur greater distortion **(Figure 1E-F)**. Taken together, these results characterize how lossy compression can support learning separable representations of similar inputs, particularly when the inputs have lower intrinsic dimensionality [2].

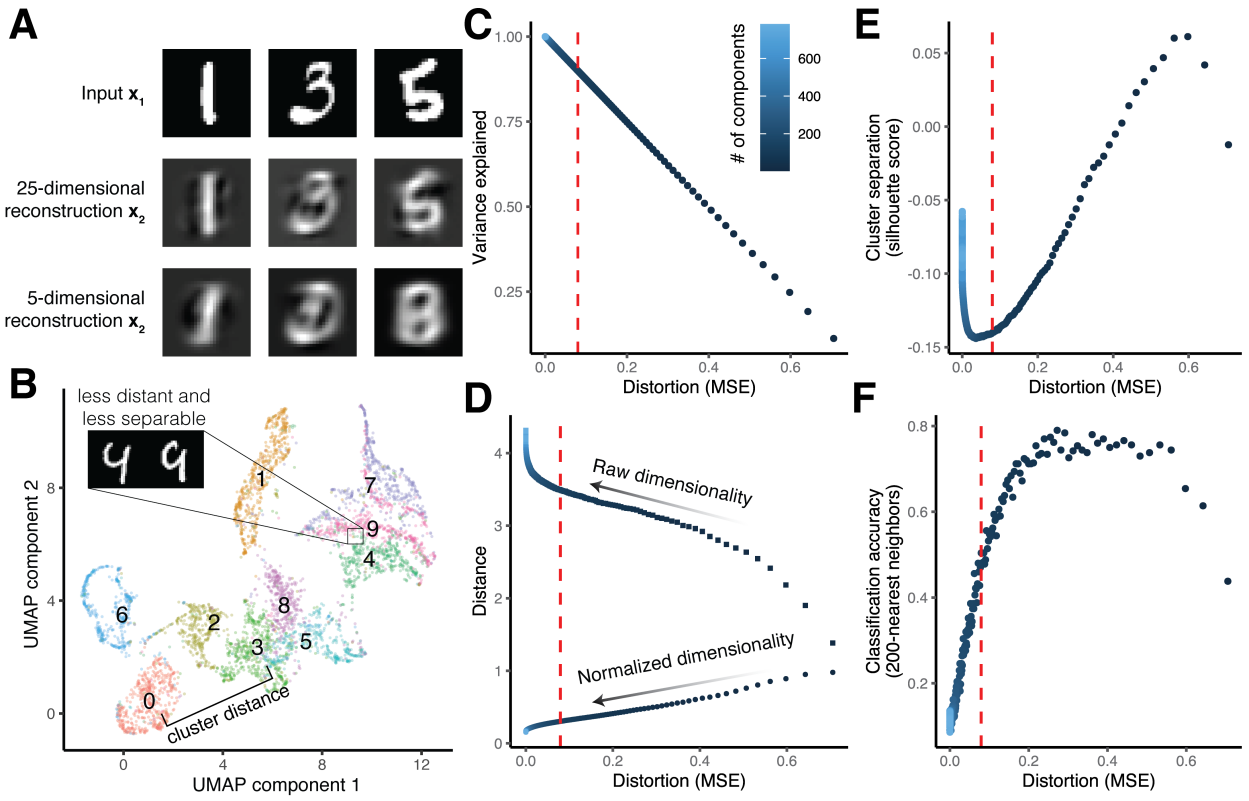

Figure 1: **(A)** Principal components analysis was used to generate a reconstruction of the images from the main 25 or 5 dimensions of variation, resulting in distortion. **(B)** A 2-dimensional embedding highlights how distant data points may be more separable and nearby data points across clusters may be more confusable. **(C)** Dimensionality reduction incurs greater distortion. The red dotted line indicates the number of components explaining over 90% of variation. **(D)** The distances between clusters increase with raw dimensiona lity but decrease with normalized dimensionality. However, cluster distances increase with normalized dimensionality reduction and increased distortion, whereas distances decrease raw dimensionality reduction and increased distortion. **(E)** Cluster separation increases with distortion. **(F)** The classification accuracy increases with distortion.

References

[1] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[2] S. Recanatesi, M. Farrell, M. Advani, T. Moore, G. Lajoie, and E. Shea-Brown, "Dimensionality compression and expansion in deep neural networks," *arXiv preprint arXiv:1906.00443*, 2019.