Minimally Redundant Laplacian Eigenmaps

David Pfau, Christopher P. Burgess

DeepMind

Introduction

Manifold Learning

Nonparametric
(LLE, IsoMap, Laplacian Eigenmaps…)

- Learn by eigendecomposition
- Exactly solvable
- Data efficient
- Extend to held-out data by Nyström method [Bengio et al 2004]
- Scales O(n) in space
- Scales ~O(nlogn) in time

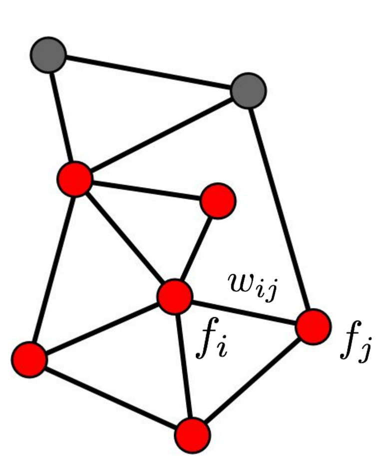
Parametric
(autoencoders, GANs…)

- Learn by gradient descent
- Not exactly solvable, but still works well in practice
- Data hungry
- Extend to held-out data trivially
- Scales O(1) in space
- Scales ~O(n) in time

One common failure mode of nonparametric manifold learning is **collapsed representations**, where the embedding does not fill the full space [Hadsell et al 2006]. We show that collapsed representations are due to the way that embedding vectors are chosen from the eigenvectors of the Gram matrix. Rather than choosing the **smallest eigenvectors**, embeddings should be chosen which have small eigenvalue **and are minimally redundant** - no dimension of the embedding should be predictable from any other.

Filtering out redundant vectors from Laplacian eigenmaps produces much more meaningful embeddings, and the remaining embedding vectors are often **automatically disentangled**, much like embeddings learned by autoencoders like **β-VAE** [Higgins et al 2017]. Thus nonparametric manifold learning can be a promising method for unsupervised disentangling **without having to learn a generative model**.

Laplacian Eigenmaps



Graph Laplacian Operator

$$(\Delta f)_i = \sum_{j \in \mathcal{N}_i} w_{ij}(f_i - f_j)$$

Matrix Form

$$L = D - W \qquad D_{ii} = \sum_{j \in \mathcal{N}_i} w_{ij}$$

Normalized Laplacian

$$L_{rw} = I - D^{-\frac{1}{2}}(I - L)D^{-\frac{1}{2}}$$

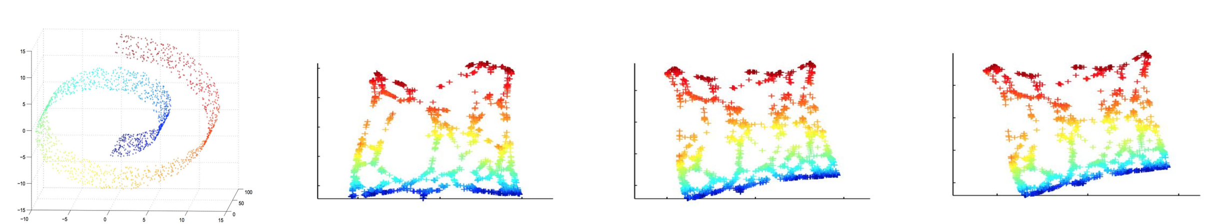
Eigenmaps Embedding Vectors

$$\Phi \Lambda \Phi^T = L$$

Graph Laplacian gives instantaneous **rate of diffusion** on a graph.

Lowest eigenvectors are slowest decaying modes of this diffusion.

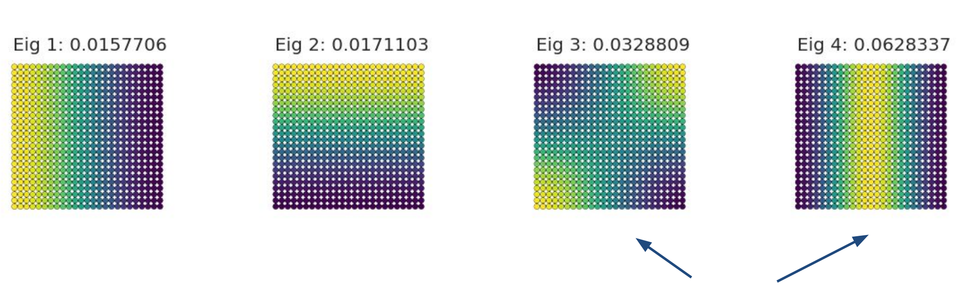
Laplacian eigenmaps constructs an embedding for points in a dataset by building nearest neighbors graph between data and using lowest eigenvectors of graph Laplacian as embedding vectors



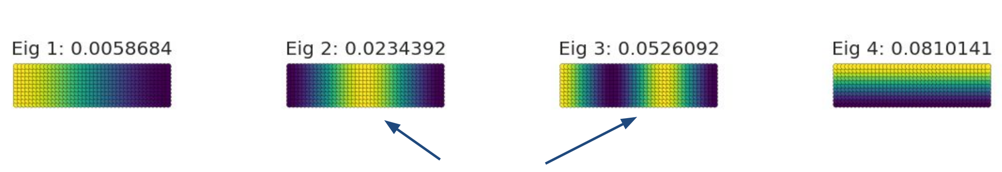
Belkin and Niyogi 2002

Minimally Redundant Embeddings

Evenly Scaled Data



Unevenly Scaled Data



Lowest eigenvectors **are not best** eigenvectors for constructing a coordinate system if the data **are not evenly scaled** in different directions

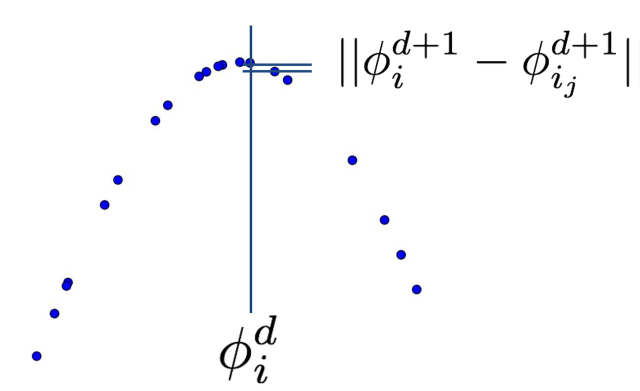
But **most real world data is not evenly scaled** in all directions!

Eigenvectors may be orthogonal but still predictable, e.g. sin(2x) and {sin(x), cos(x)}

Instead of using all lowest eigenvectors, precompute lowest eigenvectors and **filter based on predictability/redundancy**. Only unpredictable eigenvectors are added to embedding.

Math

$$\frac{1}{\sqrt{n}} \sum_i \sum_{j=1}^{k} ||\phi_i^{d+1} - \phi_{i_j}^{d+1}|| > \epsilon$$

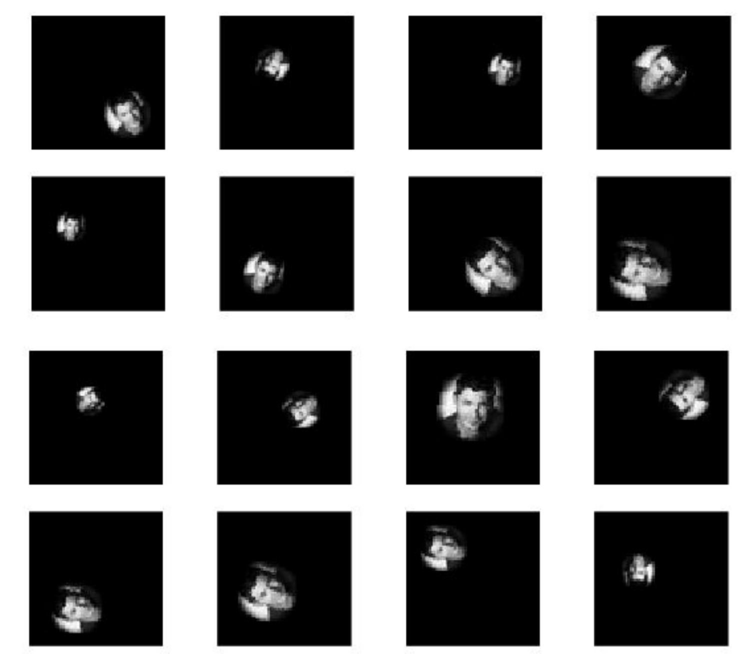


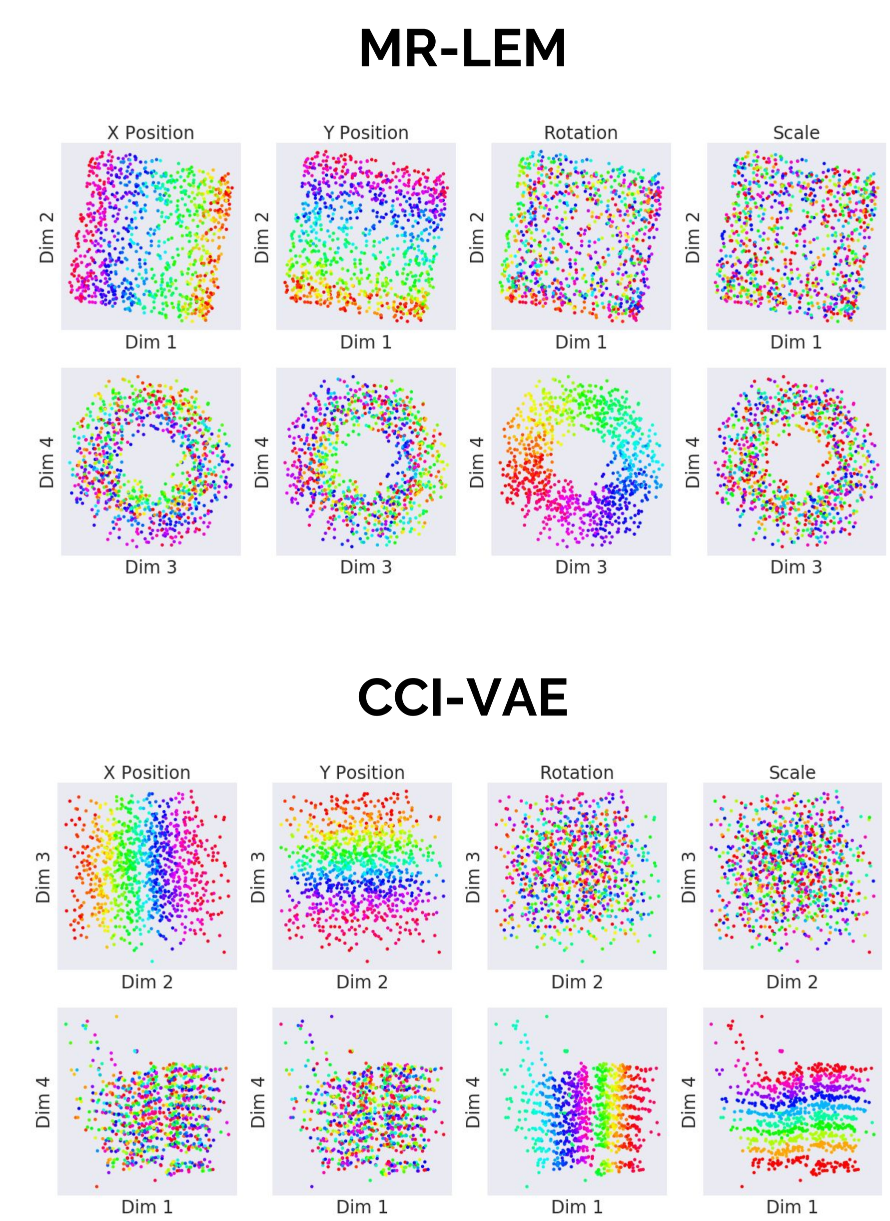
Code

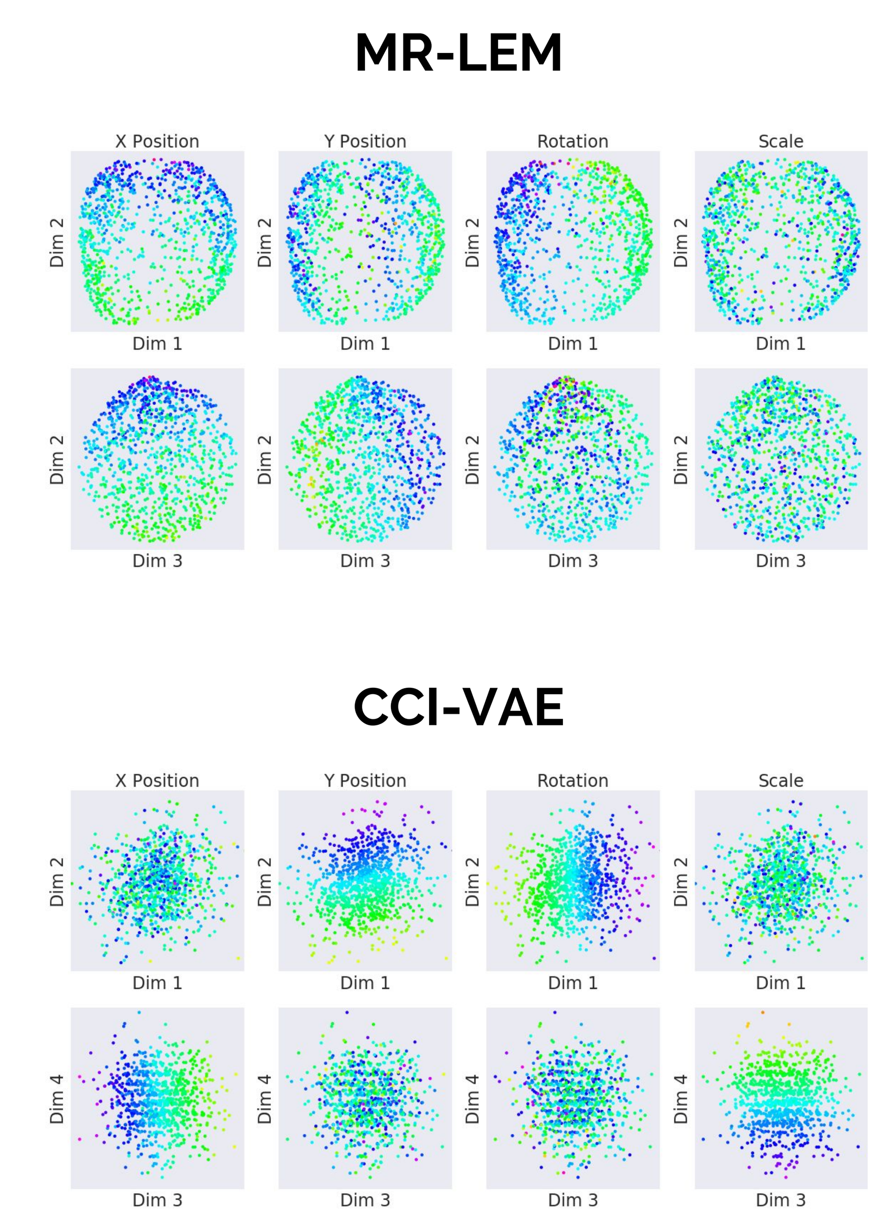
```
result = eigvec[:, :1]
m = eigvec.shape[1]
flann = pyflann.FLANN()
ind = np.arange(eigvec.shape[0])[:, None][:, (k+1)*[0]]
i = 0
while i < m-1:
  inds, dists = flann.nn(result, result, num_neighbors=k+1,
                          algorithm='kdtree')
  for j in range(i+1, m):
    i = j
    score = np.mean(np.sqrt(
        (eigvec[inds[dists != 0], i] -
         eigvec[ind[dists != 0], i])**2))
    if score > thresh:
      result = np.concatenate((result,
                               eigvec[:, i:i+1]),
                              axis=1)
      break
```

Poster

Transformed Faces Results





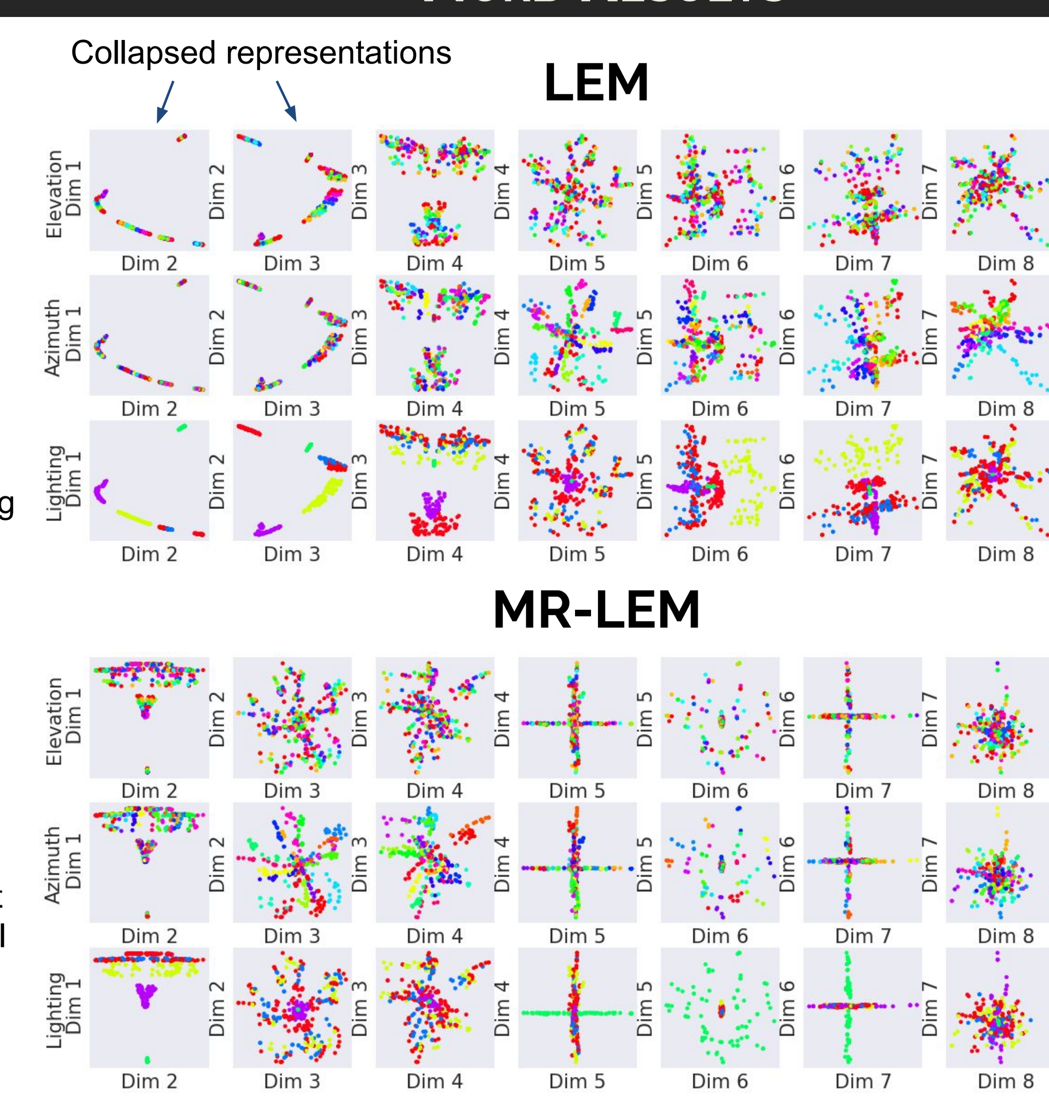


Above: samples from dataset of a face image translated, rotated and scaled.

Middle: Comparison of Minimally Redundant Laplacian Eigenmaps (MR-LEM) and CCI-VAE [Burgess et al 2017] on data with transformations enumerated on a grid. MR-LEM **recovers the true topology** of the rotation transformation while CCI-VAE does not.

RIght: CCI-VAE and MR-LEM on data with transformation sampled from a Gaussian. MR-LEM still disentangles the transformations

NORB Results

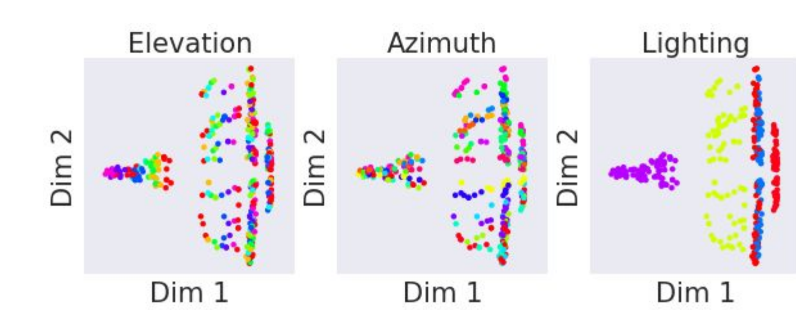


Comparison between LEM with and without filtering by redundancy on one plane object from NORB.

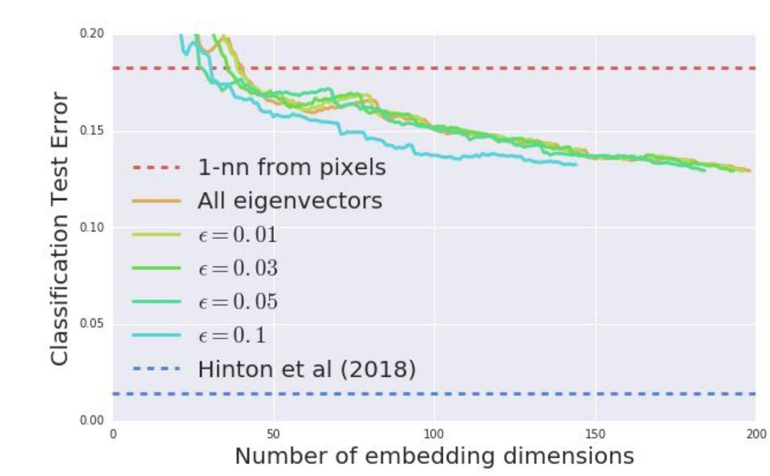
Without filtering, the **first 3 dimensions are redundant** and one lighting condition (green) is collapsed to a single point across the first >20 dimensions.

With filtering, the collapsed dimensions are removed and dimensions 5 and 6 expand out the unusual lighting condition. The angular component of dimensions 2-4 encode azimuth, while the radial components encode elevation.

Can **disentangle** lighting, azimuth and elevation using **less than 1000 samples**.



Detail of first two embedding dimensions of NORB. Dim 1 encodes lighting and elevation hierarchically.



Classification performance on full small NORB dataset. Filtering eigenvectors improves 1-NN classification. Comparison against Capsule Networks [Hinton et al 2018]

References

Belkin, M. and Niyogi, P. *Laplacian Eigenmaps*. NIPS, 2002

Bengio, Y., Paiement, J. F., Vincent, P., Delalleau, O., Le Roux, N., and Ouimet, M. *Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering*. NIPS, 2004

Hadsell, R., Chopra, S., LeCun, Y. *Dimensionality Reduction by Learning an Invariant Mapping*. CVPR, 2006

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. *β-VAE: Learning basic visual concepts with a constrained variational framework*. ICLR, 2017

Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., and Lerchner, A. *Understanding Disentangling in β-VAE*. NIPS Workshop on Learning Disentangled Representations, 2017 arXiv:1804.03599

Hinton, G. E., Sabour, S., Frosst, N., *Matrix Capsules with EM Routing*. ICLR, 2018

Paper