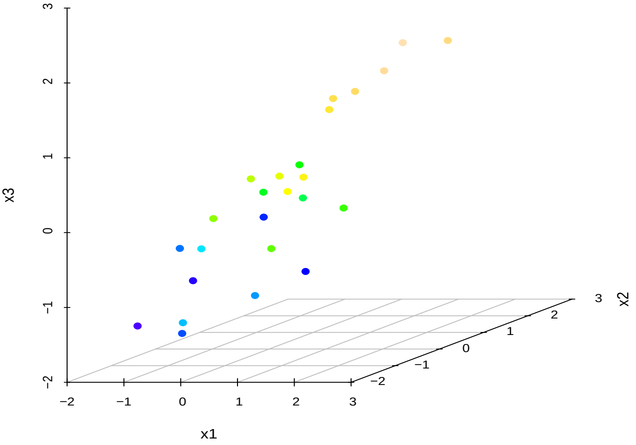
Regularization for Nonlinear Dimension Reduction by Subspace Constraint

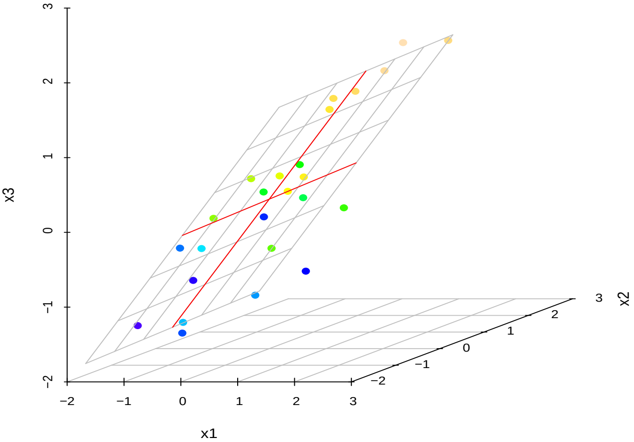
Billy Chang, Rafal Kustra
Dalla Lana School of Public Health
University of Toronto

Statistics Graduate Student Research Day
Fields Institute, University of Toronto
April 28, 2011

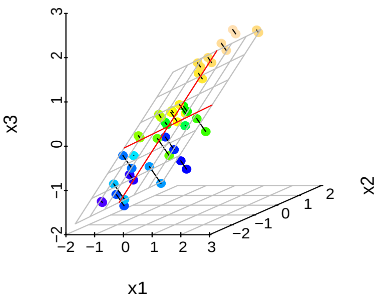
What is Dimension Reduction?

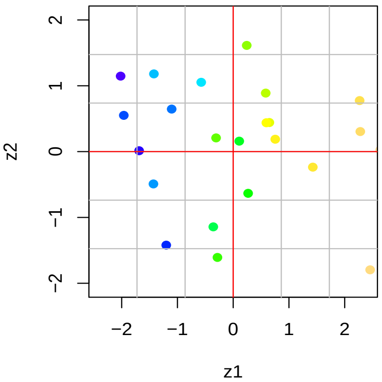


What is Dimension Reduction?



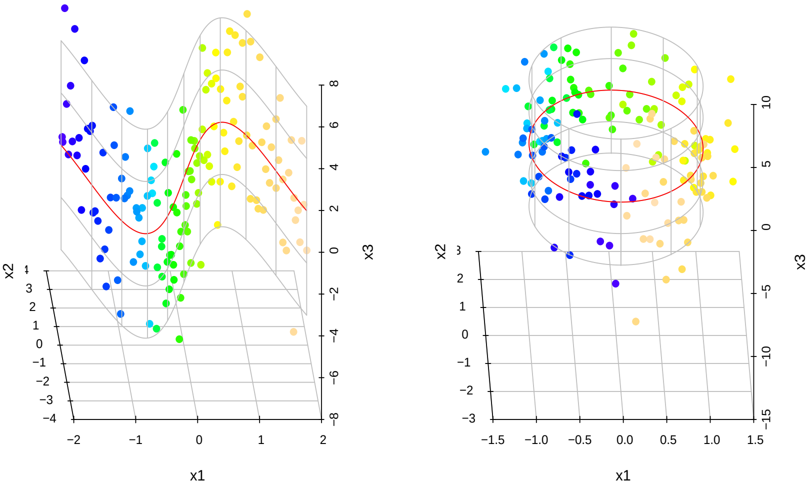
Why Dimension Reduction?



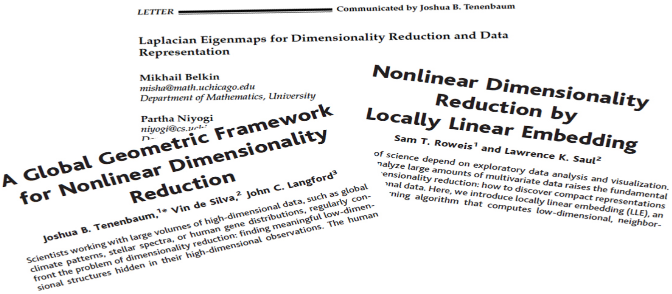


- Visualization, feature extraction for regression.
- Principal Component Analysis, Factor Analysis.
- Measures for model diagnosis: reconstruction error, likelihood.

What is Nonlinear Dimension Reduction?

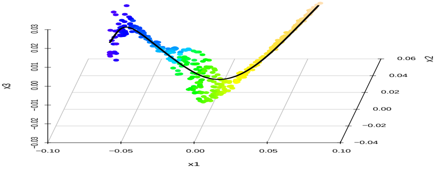


Challenges in Nonlinear Dimension Reduction

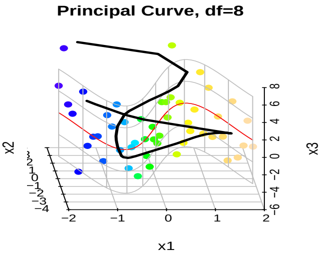


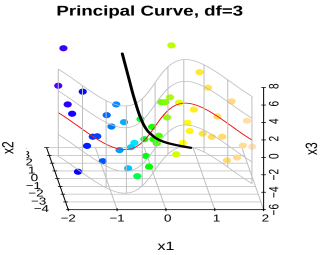
- No explicit mapping from original space to low-dimensional space.
- Model checking and parameter tuning difficult.

Principal Curve (Hastie et. al. 1989)



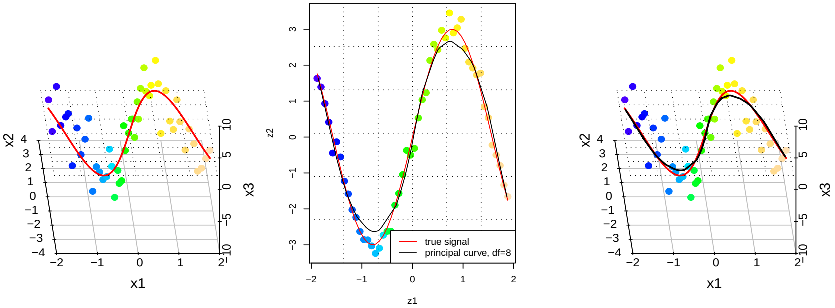
- A curve that passes through the centre of the data.
- Explicit mapping and reconstruction error available.
- Can over-fit, or under-fit.





Subspace Constraint

- Assume the nonlinearity lies on a lower-dimensional subspace.



- Reconstruction: embed the principal curve back into the original space.
- Question: How to find that constraint subspace?

Subspace Search: Kernel PCA and Maximum Eigenvalue Maximization

- Kernel PCA (Schölkopf et. al. 1998) solves:

$$\max_{g \in \mathcal{H}_K : \|g\|_{\mathcal{H}_K}=1} \hat{var}(g(\mathbf{x})), \tag{1}$$

  where $\mathcal{H}_K$ is the Hilbert Space of functions induced by the kernel K (Hastie et. al. 2009).
- Assuming the objective (1) is a measure of nonlinearity, we search for a projection matrix $\mathbf{H}$ such that:

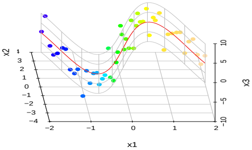
$$\max_{\mathbf{H}:\mathbf{H}^T\mathbf{H}=\mathbf{I}} \left\{ \max_{g \in \mathcal{H}_K : \|g\|_{\mathcal{H}_K}=1} \hat{var}(g(\mathbf{H}^{\mathbf{T}}\mathbf{x})) \right\}$$

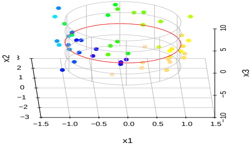
- The term inside {} is proportional to the largest eigenvalue of the Gaussian kernel matrix $\mathbf{K}^{\mathbf{H}}$ for the projected data:

$$\mathbf{K}^{\mathbf{H}}_{ij} = e^{-\frac{\|\mathbf{H}^T(\mathbf{x}_i - \mathbf{x}_j)\|^2}{\sigma}}$$

- Further dimension reduction can be performed on $(\mathbf{I} - \mathbf{H}\mathbf{H}^T)\mathbf{x}$.

Simulation

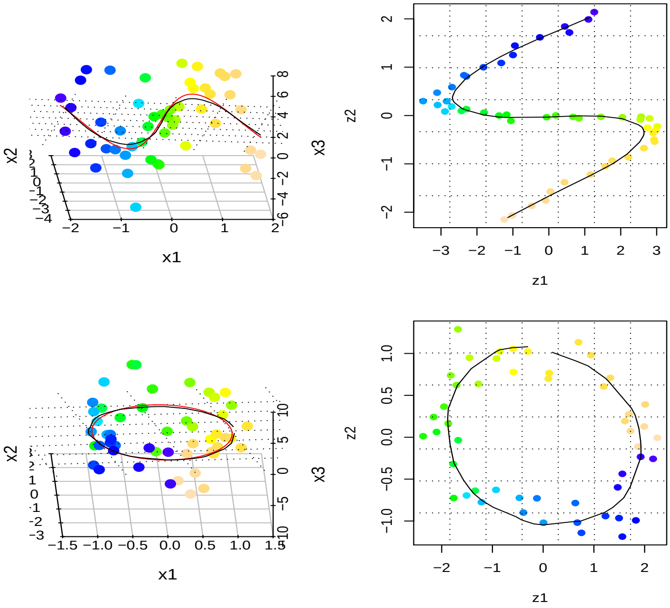




Sinusoidal	Circular
$x_1 = (-0.6\pi ... 0.6\pi)$	$\theta = (-\pi ... \pi)$
$x_2 = 3sin(2x_1) + \epsilon$	$x_1 = sin(\theta) + \epsilon_1$
$x_3 \sim N(0, \sigma = 3)$	$x_2 = 2cos(\theta) + \epsilon_2$
	$x_3 \sim N(0, \sigma = 3)$

- 50 training samples, 500 validation samples.
- d.f. for principal curve and σ for Gaussian Kernel chosen by reconstruction error on the validation set.

Results (—- True Curve, —- Fitted Principal Curve)



Conclusion

- Regularization for nonlinear dimension reduction: roughness-penalty is not enough.
- Subspace constraint regularizes by controlling the direction principal curve can move.
- Future work: a better measure of nonlinearity.

Reference

- Schölkopf, B, Smola, A. & Muller, K.R. (1998) Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation 10, 1299-1319.
- Hastie, T. & Stuetzle, W. (1989) Principal Curves. JASA. 84(406):502-516.
- Hastie, T, Tibshirani, R. & Friedman, J. (2009) The Elements of Statistical Learning, 2nd Ed. Springer.