Machine Learning in Astrophysics
Making Sense of Simulation Data Through Probabilistic
Multi-Manifold Learning

Peter Tino
School of Computer Science
University of Birmingham
UK
Special thanks to...

- Marco Canducci
- Michele Mastropietro
- Sven De Rijcke
- Reynier Peletier
- Rory Smith
- Petra Awad
- Abolfazl Taghribi
- Kerstin Bunte
- Mohammad Mohammadi
- ...

...
Emerging low-dim structures

- **Galaxies disrupted in a hostile environment**
  - e.g. dwarf galaxy entering a galaxy cluster (e.g. Fornax)
  - ram pressure
  - jellyfish-like structures
  - a ”head” and tidal streams emitting from the head

- **”Bubbles” - e.g. supernova signatures**
  - rapidly expanding ”spherical voids”

- **Cosmic web**
  - on the cosmological scale, distribution of matter is not uniform
  - filaments, walls, nodes
Jellyfish galaxies

Credit: COURTESY OF HUBBLE SPACE TELESCOPE, NASA, ESA AND HUBBLE HERITAGE TEAM (STScI/AURA); ACKNOWLEDGMENT: M. SUN University of Alabama in Huntsville.

Bubbles

Detected bubbles in a snapshot of particle simulation of a jellyfish galaxy evolution.
Cosmic web

Dark matter particles in a cosmological simulation.
Multi-manifold learning

- **detection**
  - find multiple manifolds scattered in a noisy background
  - unknown number
  - unknown dimensionalities (can vary across the manifold set)
  - background noise can be substantial

- **modelling**
  - for each manifold construct its abstract discretized model
  - ”smooth” embedding into the ambient space
  - generative probabilistic model as a constrained mixture of manifold aligned multivariate Gaussians
Variety of manifolds

(a)  (b)  (c)  
(d)  (e)  (f)
Synthetic data
Dimensionality index

- for each point - local PCA (characteristic scale)

- normalized eigenspectra (live on a standard simplex)

\[ \tilde{\lambda}_{i,j} = \frac{\lambda_{i,j}}{\sum_{k=1}^{d} \lambda_{i,k}}, \]

- "ideal" 1D, 2D, 3D etc. normalized eigenspectra:
  
  \((1, 0, 0, 0, ...), (1/2, 1/2, 0, 0, ...), (1/3, 1/3, 1/3, 0, ...), ... \)

- cluster normalized spectra w.r.t. to such cluster representatives

  Riemannian geodesic Fisher distance:

  \[ d_J(\tilde{\Lambda}_k, \tilde{\Lambda}_l) = 2 \arccos \left( \sum_{j=1}^{d} \sqrt{(\tilde{\lambda}_{kj} \cdot \tilde{\lambda}_{lj})} \right) \]
Spatial smoothing of the estimated dimensionality indices helps! Gaussian kernel - characteristic scale
Capturing individual manifolds

- crawl on diffused data - Structure Aware Filtering [Wu et al., IEEE TPAMI 2018]

- probabilistic model built on the original ”noisy” data
The trick - abstract latent space
Crawling the Moebius strip
Probabilistic model

- "smooth" graph embedding

\[ \overline{v} = f(v; W) = W \Phi(v) \]

- flat mixture

\[ p(t|W, \zeta) = \frac{1}{K} \sum_{i=1}^{K} p(t|v_i, \zeta_i \Sigma_i, W), \]

\[ p(t|v_i, \zeta_i \Sigma_i, W) = \frac{1}{[(2\pi \zeta_i)^d|\Sigma_i|]^{\frac{1}{2}}} \exp \left\{ - \frac{\Delta t^\top \Sigma_i^{-1} \Delta t}{2 \zeta_i} \right\} \]

- free parameters \( W \) and \( \{\zeta_i\} \) are trained using E-M
Captured noisy manifolds
Figure: Gas particles of the simulated dwarf galaxy falling into the halo of the Fornax Galaxy Cluster.
Star formation

- **temperature**

- **density** (of gas particles)

- **metallicity** - abundance of elements heavier than hydrogen and helium

- **neutral fraction** - fraction of neutral gas particles (mostly hydrogen) obtained by excluding the ionized ones (positively charged due to electron loss)
Studying the streams I
Studying the streams II
Cavity signatures
Probabilistic models of cavity boundaries
What is the data telling us?
Cosmic web
Studying the filament structures

Picture credit: The Eagle Project - http://icc.dur.ac.uk/Eagle/index.php