Bayesian Manifold Learning : Locally Linear Latent Variable Model (LL-LVM)

Mijung Park, Zoltán Szabó, Ahmad Qamar, Lars Buesing, Maneesh Sahani
Gatsby Computational Neuroscience Unit, University College London

Manifold Learning
- Problems with high-dimensional data
  - Optimisation in high-dimensional space is computationally expensive and hard to find a global optimum
  - Good news: in many cases, the intrinsic dimensionality is actually low
  - datapoints are sampled from a low-dimensional manifold embedded in a high-dimensional space
  - example: swiss roll
- Manifold learning: attempts to uncover the manifold structure

Non-probabilistic prior work
- idea: preserve geometric properties of local neighbourhoods
  - limits:
    - sensitive to noise due to lack of explicit model
    - heuristic methods to evaluate manifold dimensionality
    - no measure of uncertainties in the estimates
    - out-of-sample extension requires extra approximations

GP-LVM
- idea: define a functional mapping from latent space to data space using GP
  - for data \( Y = [y_1, \ldots, y_n] \in \mathbb{R}^{n \times d} \) and latents \( X = [x_1, \ldots, x_n] \in \mathbb{R}^{n \times d} \),
    \[ p(Y|X) = \prod_{i=1}^{n} N(y_i|0, K_{xx} + \beta^{-1}I_d), \]
    where the \( i \)-th element of the covariance matrix is
    \[ K_{xx}(x_i, x_j) = \sigma_y^2 \exp \left( -\frac{1}{2} \sum_{k=1}^{d} r_k^2(x_{i,k} - x_{j,k})^2 \right), \]
  - prior on latents: assuming the neighbourhood latent variables are similar
    \[ -\frac{1}{2} \sum_{i=1}^{n} \alpha ||x_i||^2 + \frac{1}{2} \sum_{j=1}^{n} \eta_j ||x_j - x_i||^2 \]
    \[ \Rightarrow p(x|G, \alpha) = N(0, \Pi), \]
    where \( \alpha \) controls the expected scale, \( \Pi^{-1} = 2\Lambda \otimes I_d \) and \( \Pi = \alpha I_d + \Lambda^{-1} \).
  - prior on linear maps: similar
    \[ p(C|G, U) = MN(0, U, \Omega), \]
    where \( E[CC^\top] = \Omega \).
- likelihood: penalising the approximation error yields
  \[ -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_j ((y_i - y_j) - C_i(x_i - x_j))^2 \]
  \[ \Rightarrow p(y|C, x, V, G) = N(0, \Sigma_y), \]
  assuming \( V^{-1} = \gamma I \) and \( \gamma \) is a parameter.

Variational EM
- maximizing log marginal likelihood is intractable, instead maximise lower bound
  \[ \log p(y|G, \theta) \geq \int q(C, x) \log \frac{p(y, C, x|G, \theta)}{q(C, x)} \mathrm{d}x = \mathcal{F}(q(C, x), \theta), \]
  for computational tractability, assume \( q(C, x) = q(C)q(x|C) \).
  - variational expectation maximization algorithm
    - expectation step for computing \( q(C|\theta) \) by
      \[ q(C) \propto \exp \left( \int q(C) \log p(y, C, x|G, \theta) \mathrm{d}x \right) = N(c|\mu, \Sigma), \]
    - maximization step for estimating \( \theta \)
      \[ \theta = \arg\max_{\theta} \mathcal{F}(q(C, x), \theta). \]

Conclusion
A new probabilistic approach to manifold learning preserving local geometries in data and equipped with straightforward variational inference for learning the manifold.

References

Relation to GP-LVM
Integrating out \( C \) from likelihood yields
\[ p(y|x, G, \theta) = \int p(y|C, x, G, \theta)p(C|G, \theta)\mathrm{d}C, \]
\[ = \frac{1}{Z_{G\theta}} \exp \left( -\frac{1}{2} y^\top K_{G\theta}^{-1} y \right). \]
In contrast to GP-LVM, the precision matrix \( K_{G\theta}^{-1} \) depends on the Laplacian matrix.
- The functional form of precision is directly determined by the graph structure given the observations
  \[ K_{G\theta}^{-1} = (2\Lambda \otimes V^{-1}) - (W \otimes V^{-1})A(W \otimes V^{-1}), \]
  where \( W \) is a function in \( x \) and \( L \) and \( A \) is a function in \( x^\top \) and \( L \).

Illustration
- Mitigating short-circuiting problems
- Finding the optimal number of neighbours using variational lower bound

Figure: Two datapoints seem close to each other (A) but actually far in 2D space (B). Short-circuiting the two datapoints lower the bound (C)

Figure: A: 400 samples drawn from 3D Gaussian. B: LLE. C: GP-LVM. D (Left): The posterior mean of \( C \). D (Middle): posterior mean of \( x \). D (Right): Normalized variational lower bound.