Dimensionality Reduction

A Probabilistic Perspective

Aditya Ravuri

Overview of talk

- Motivation & Intro
- Background on DR – a statistical perspective
- Intro to ProbDR
  - t-SNE class algos
  - Wishart class algos
- Future directions

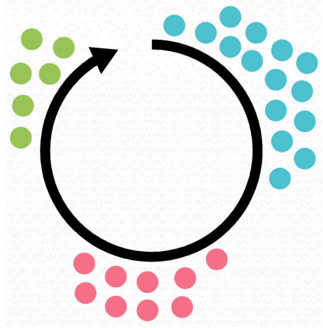
Algorithms ~> Inference

$$\hat{\mathbf{X}}_{MAP} = \mathbf{U}_{l\ maj}(\mathbf{\Lambda}_{l\ maj} - \hat{\sigma}^2\mathbf{I}_l)^{1/2}\mathbf{R}^T$$

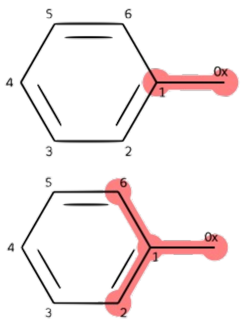
$$\underset{\mathbf{X}}{\operatorname{argmin}} -\sum_{ij} v_{ij} \log w_{ij}(\mathbf{X}_{i:}, \mathbf{X}_{j:})$$

Models enable new applications

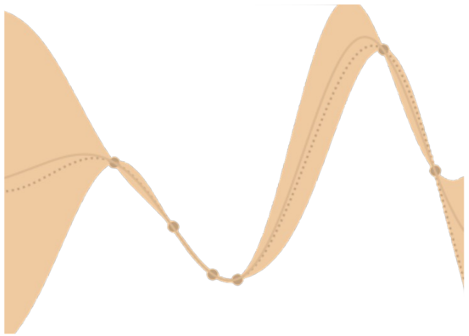
Probabilistic interpretations:



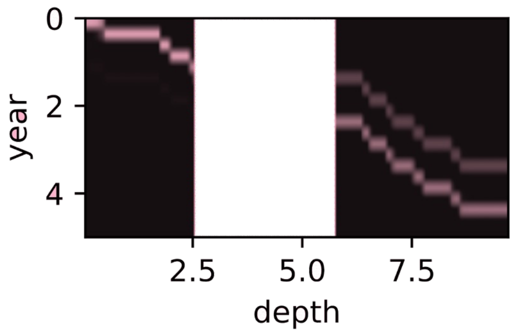
enable efficient reformulations
enable prior specification



enable downstream methods(e.g. bayesopt)



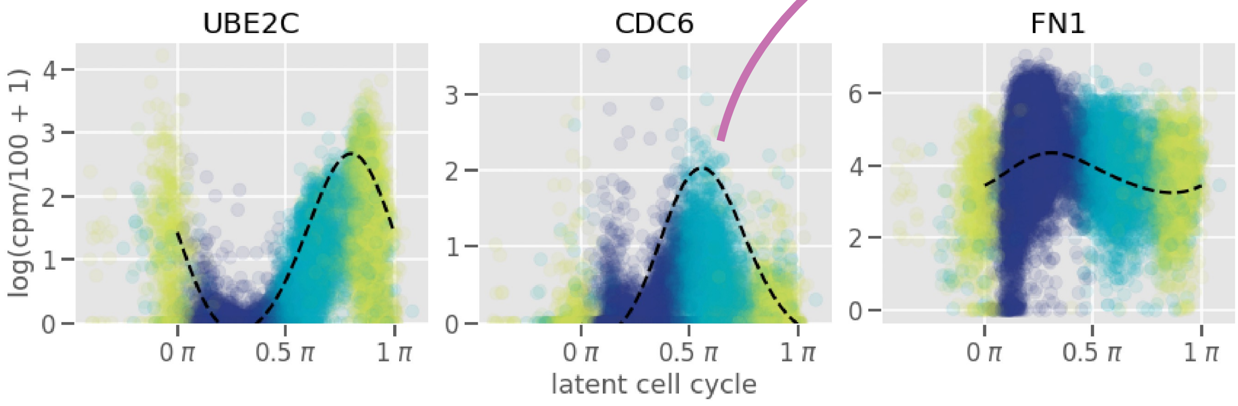
aid communication (model vs algo)
enable software cross-utilization

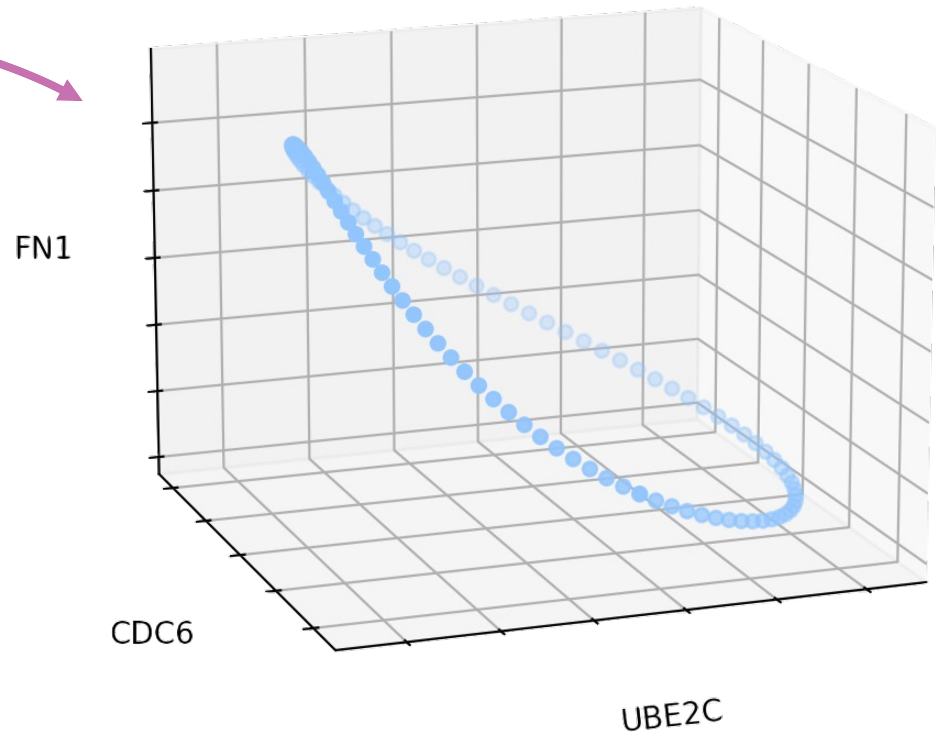


enable reasoning about missing data, uncertainties
enable model and inference extension

DR: Dimensionality Reduction

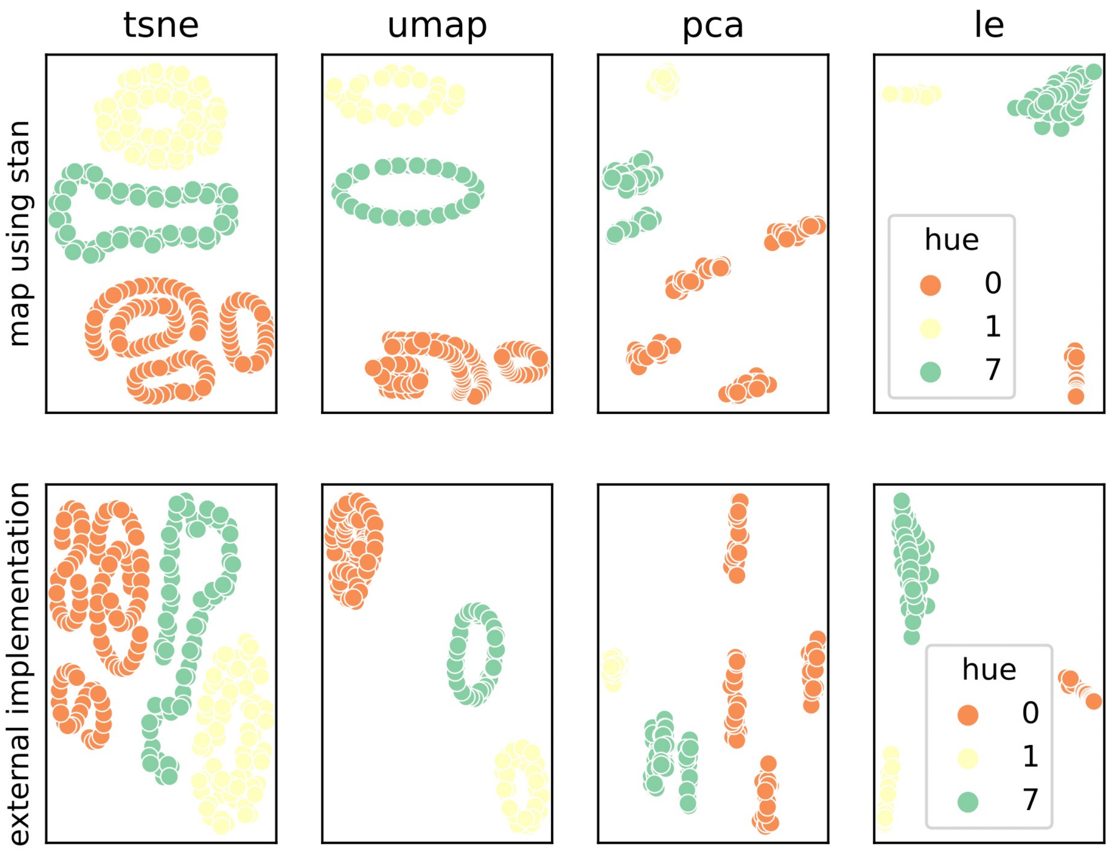
joint mean gene expression



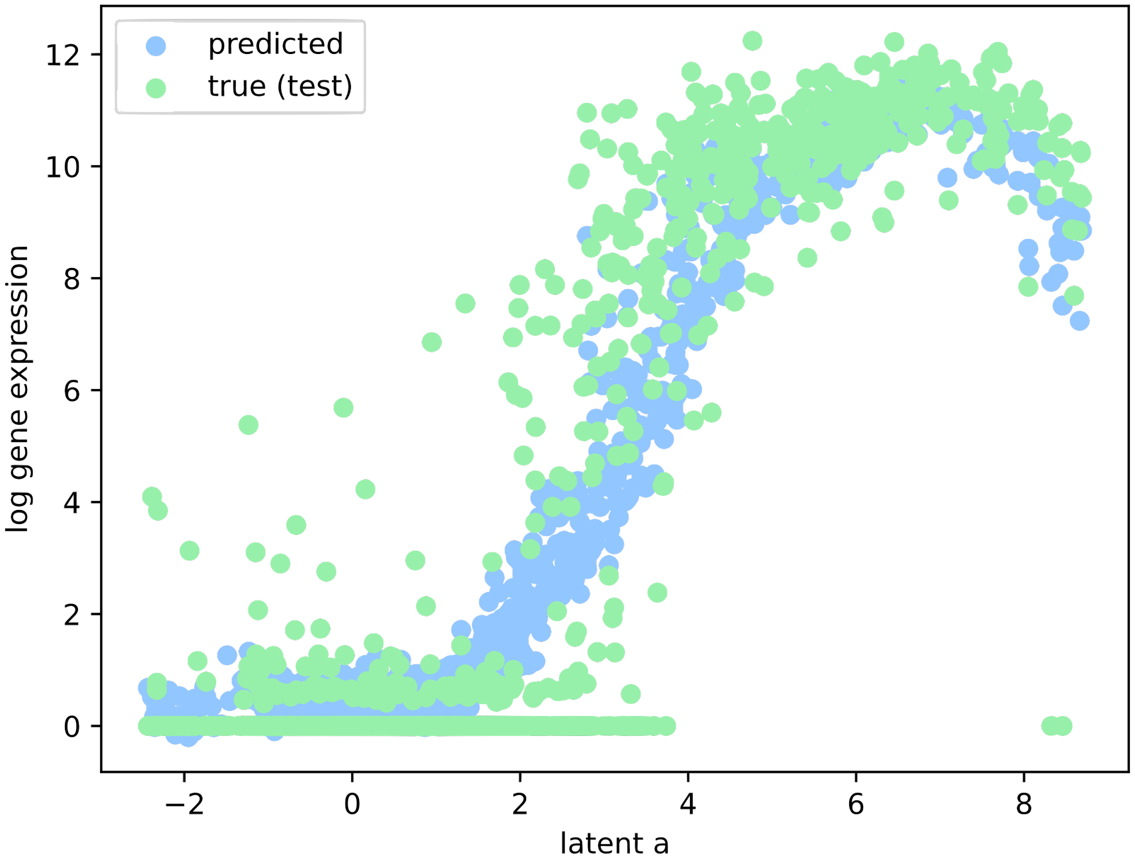


cell cycle driving gene expression

DR by automatic inference



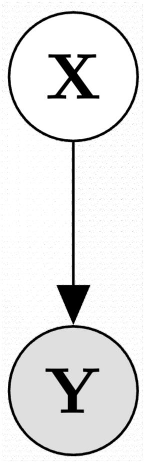
Generative models with DR



Primer on Probabilistic PCA

Existing Probabilistic DR Models

PCA



$$\mathbf{Y}|\mathbf{X} \sim \mathcal{MN}(\mathbf{0}, \mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}, \mathbf{I})$$

$$p(\mathbf{X}) \propto 1$$

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmax}} \log p(\mathbf{X}|\mathbf{Y})$$

Example Stan model

```
model {
    Y ~ multi_gp(gp_dot_prod_cov(X, s), ones);
}
```

Primer on Probabilistic DR

Existing Probabilistic DR Models

PCA

FA

GMM

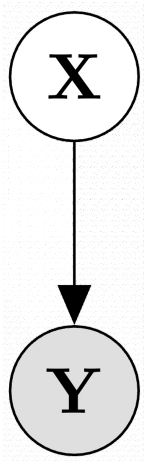
NMF

LDA

ICA

GPLVM

VAE



$$\mathbf{y}|\mathbf{x} \sim \mathrm{ExpFam}(\mathbf{x})$$

$$\mathbf{x} \sim \mathrm{prior}(.)$$

Do highly used DR algos do inference?

Existing Probabilistic DR Models

PCA

FA

GMM

NMF

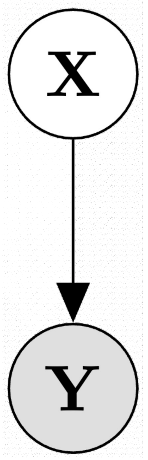
LDA

ICA

GPLVM

VAE

DRTree



V. Influential DR Algo.s

SNE

t-SNE

UMAP

CMDS

LLE

LE

MVU

DM

kPCA

Isomap

Yes.

PCA

FA

GMM

NMF

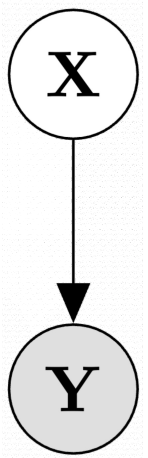
LDA

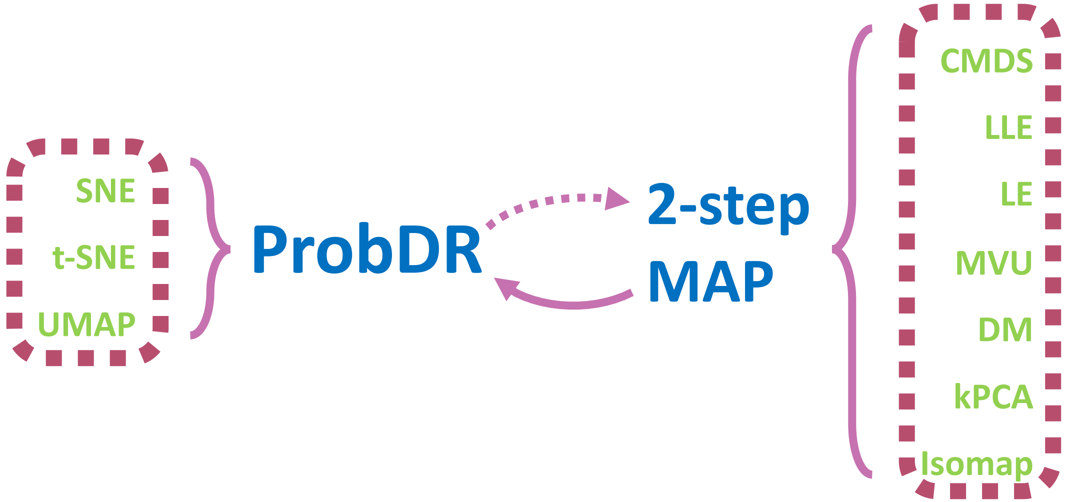
ICA

GPLVM

VAE

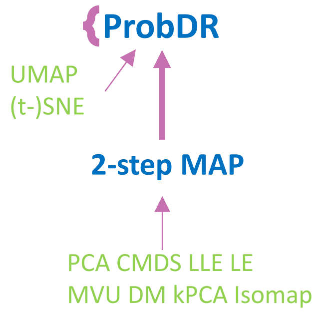
DRTree



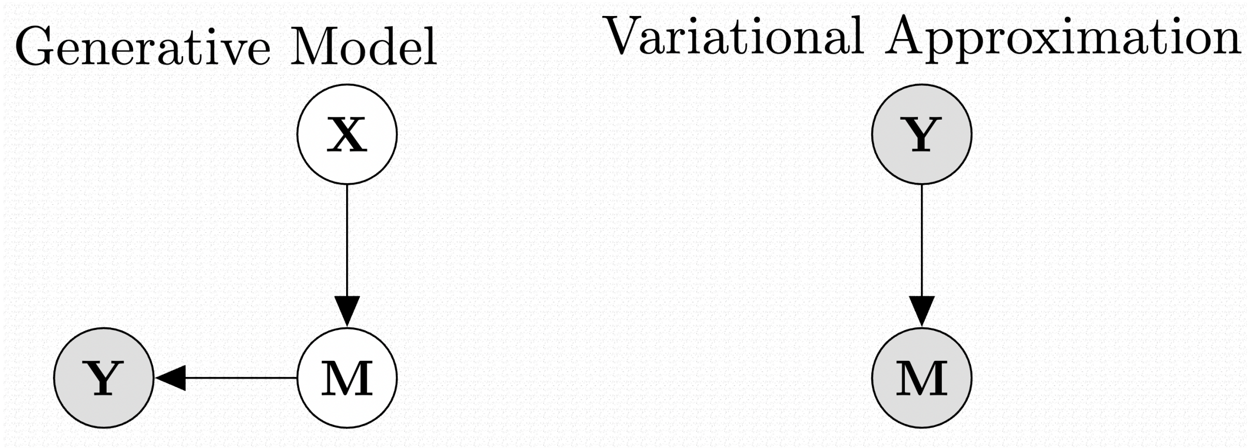


Dimensionality Reduction as Probabilistic Inference

ProbDR



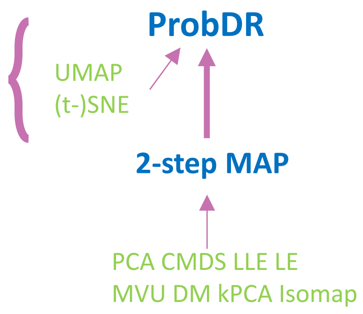
model

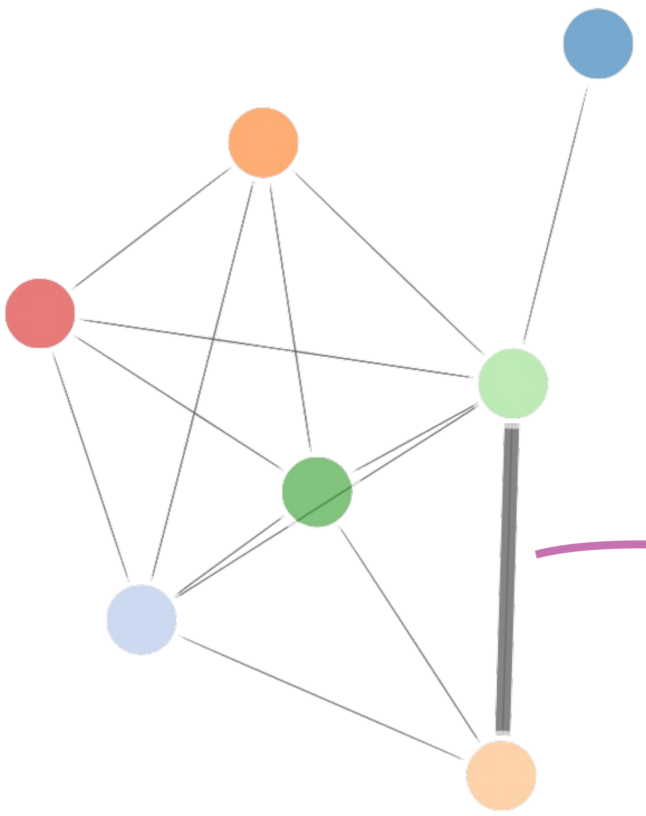


inference

$$\arg\max_{\mathbf{X},\theta} \mathbb{E}_{q(\mathbf{M}|\mathbf{Y})}[\log p_\theta(\mathbf{Y}|\mathbf{M})] - \mathrm{KL}(q(\mathbf{M}|\mathbf{Y})||p(\mathbf{M}|\mathbf{X}))$$

(t-)SNE & UMAP



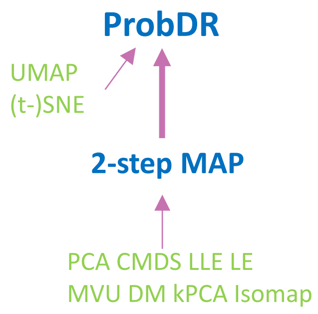


KL

algo	$q(\mathbf{A}'\mid\mathbf{Y})$	$p(\mathbf{A}'\mid\mathbf{X})$	$\mathrm{KL}(q\|\|p)$
UMAP	$\prod_{i\neq j}^{n} \text{Bernoulli}(\mathbf{A}'_{ij}\mid v^{U}_{ij}(\mathbf{Y}))$	$\prod_{i\neq j}^{n} \text{Bernoulli}(\mathbf{A}'_{ij}\mid w^{U}_{ij}(\mathbf{X}_{i:,j:}))$	$\mathcal{C}_{\text{UMAP}}$

$$w_{ij} = (1 + a\|\mathbf{X}_{i:} - \mathbf{X}_{j:}\|^{2b})^{-1}$$

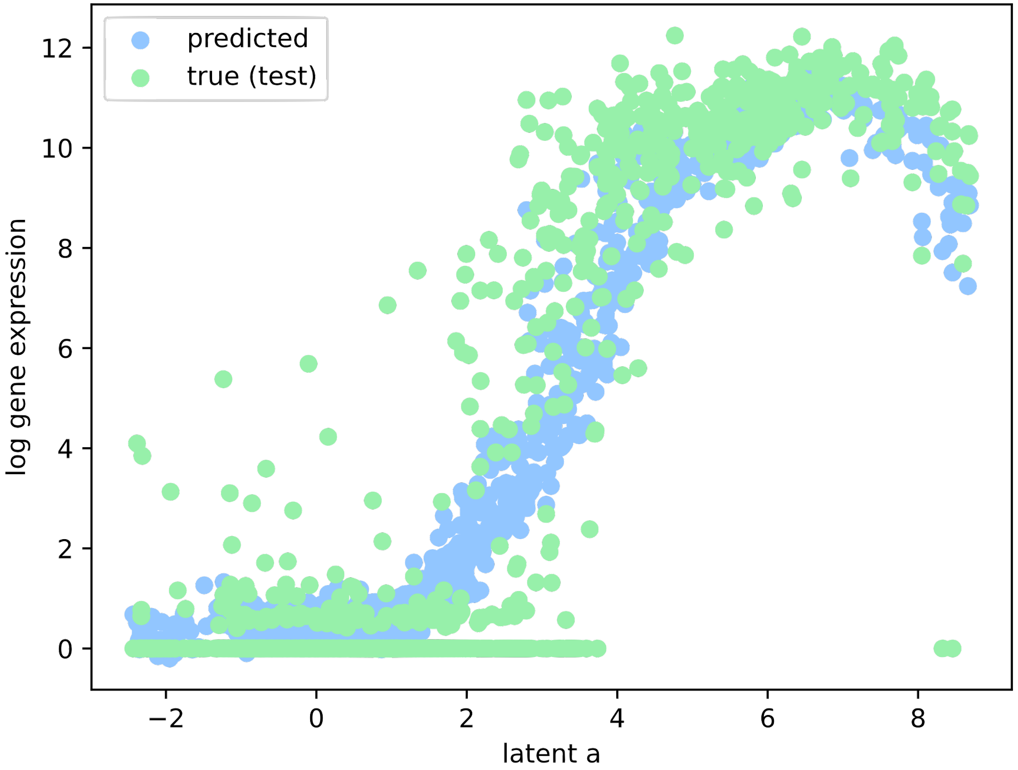
UMAP model in Stan

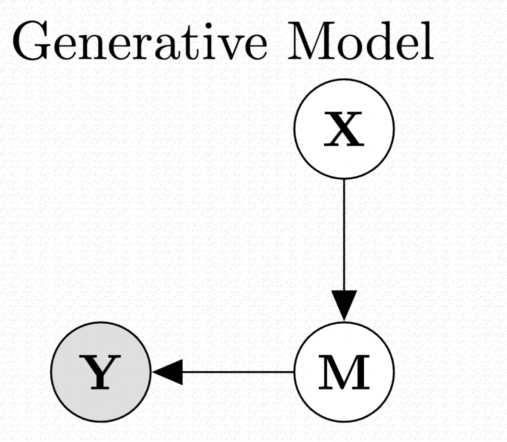



```
transformed parameters {
    matrix[n, n] W = 1 ./ (1 + 2 * squared_distances(X));
}
model {
    // umap
    for(i in 2:n)
        for(j in 1:(i - 1))  // lower triangle
            M_int[i, j] ~ binomial(rho, W[i, j]);
}
```

Predict at unseen X using ProbDR

ProbDR
UMAP
(t-)SNE
2-step MAP
PCA CMDS LLE LE
MVU DM kPCA Isomap

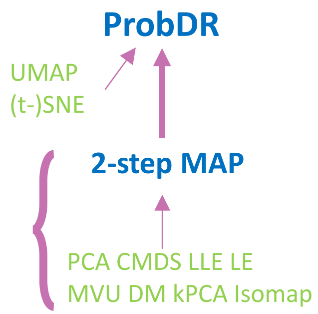




observation model

$$\forall i : \mathbf{Y}_{:i}|\mathbf{L} \sim \mathcal{N}\left(\mathbf{0}, \begin{cases} [\mathbf{L}+\beta\mathbf{I}]^{-1} & \text{Matérn-1 case} \\ \exp[-t\mathbf{L}] & \text{Matérn-}\infty\text{ case} \end{cases}\right)$$

2-step MAP: Introduction



Many DR algos are specified as:

1. estimate a PSD matrix

PCA	$\hat{\mathbf{S}}(\mathbf{Y}) = \mathbf{Y}\mathbf{Y}^T/d$
CMDS MVU kPCA Isomap	$\hat{\mathbf{S}}(\mathbf{Y}) = \mathbf{HKH}$
SE	$\hat{\mathbf{\Gamma}}(\mathbf{Y}) = \mathbf{L}$
LLE LE	$\hat{\mathbf{S}}(\mathbf{Y}) = \mathbf{H}(\tilde{\mathbf{L}} + \gamma\mathbf{I})^{-1}\mathbf{H}$
DM	$\hat{\mathbf{S}}(\mathbf{Y}) = \mathbf{P}$

2. set embedding to eigencomponents of the matrix

2-step MAP = ProbDR Inference

ProbDR

UMAP
(t-)SNE

2-step MAP

PCA CMDS LLE LE
MVU DM kPCA Isomap

- Step 2 is MAP inference.

$$\hat{\mathbf{S}} * d | \mathbf{X} \sim \mathcal{W}\left(\mathbf{X}\mathbf{X}^T + \sigma^2 \mathbf{I}_n, d\right) \quad \Rightarrow \quad \hat{\mathbf{X}}_{MAP} = \mathbf{U}_{l\ maj}(\mathbf{\Lambda}_{l\ maj} - \hat{\sigma}^2 \mathbf{I}_l)^{1/2} \mathbf{R}^T$$

$$\hat{\mathbf{\Gamma}} * d | \mathbf{X} \sim \mathcal{W}\left((\mathbf{X}\mathbf{X}^T + \beta \mathbf{I}_n)^{-1}, d\right) \quad \Rightarrow \quad \hat{\mathbf{X}}_{MAP} = \mathbf{U}_{l\ min}(\mathbf{\Lambda}_{l\ min}^{-1} - \hat{\beta} \mathbf{I}_l)^{1/2} \mathbf{R}^T$$

- This 2-step MAP inference is equivalent to ProbDR.

$$\arg\max_{\mathbf{X}} \log p(\hat{\mathbf{M}}(\mathbf{Y}) * d | \mathbf{X})$$

$$p(\mathbf{M} | g(\mathbf{X})) = \mathcal{W}(\mathbf{M} | g(\mathbf{X}), d)$$

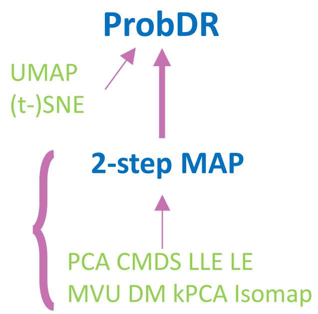
$$\longleftrightarrow$$

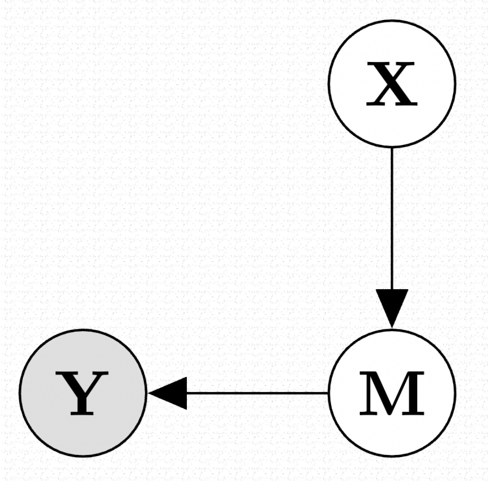
$$\arg\min_{\mathbf{X}} KL(q(\mathbf{M} | \hat{\mathbf{M}}(\mathbf{Y})) \| p(\mathbf{M} | g(\mathbf{X}))))$$

$$\textit{model (law of p)} : \mathbf{M} | g(\mathbf{X}) \sim \mathcal{W}(g(\mathbf{X}), d),$$

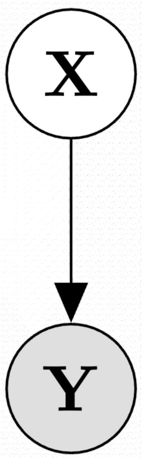
$$\textit{variational approx (law of q)} : \mathbf{M} | \hat{\mathbf{M}}(\mathbf{Y}) \sim \mathcal{W}(\hat{\mathbf{M}}(\mathbf{Y}), d)$$

The Wishart generative models ≈ GPs





$\rho \to \infty$



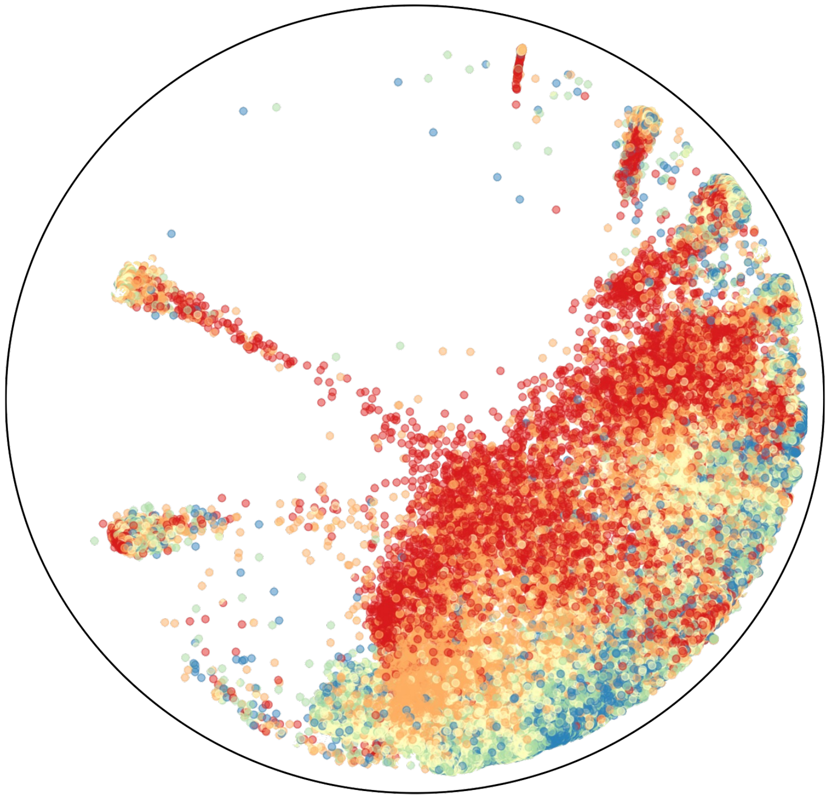
$$\mathbf{y}|\mathbf{S} \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{\rho}\mathbf{S}\right),$$
$$\mathbf{S}|\mathbf{X} \sim \mathcal{W}(\mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}, \rho),$$

/

$$\mathbf{y}|\mathbf{S} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{S} * (\rho - n + 1)\right),$$
$$\mathbf{S}|\mathbf{X} \sim \mathcal{W}^{-1}\left(\mathbf{X}\mathbf{X}^T + \beta\mathbf{I}, \rho\right)$$

$$\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{X}\mathbf{X}^T + \sigma^2\mathbf{I}).$$

Future directions



Do variational assumptions correspond to semantic assumptions?

If yes, then ProbDR ≈ GPLVMs?

E.g.
Model assumption: data lies on hyperbolic space
Semantic assumption: data has tree structure