User-steering Interpretable Visualization with Probabilistic PCA

Viet Minh Vu and Benoît Frénay

University of Namur, Belgium

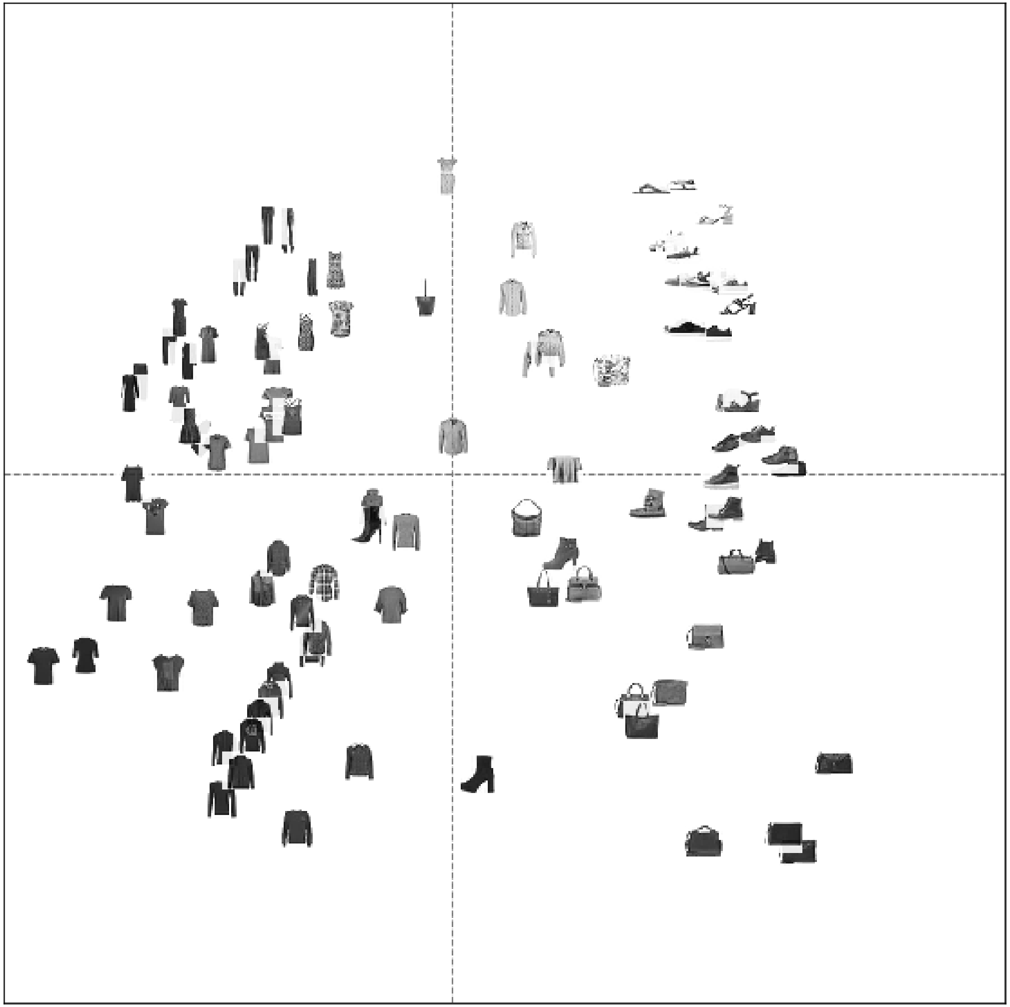
25/04/2019

Visualization of high dimensional data: Dimensionality Reduction (DR) problem

Sample of Fashion-MNIST dataset

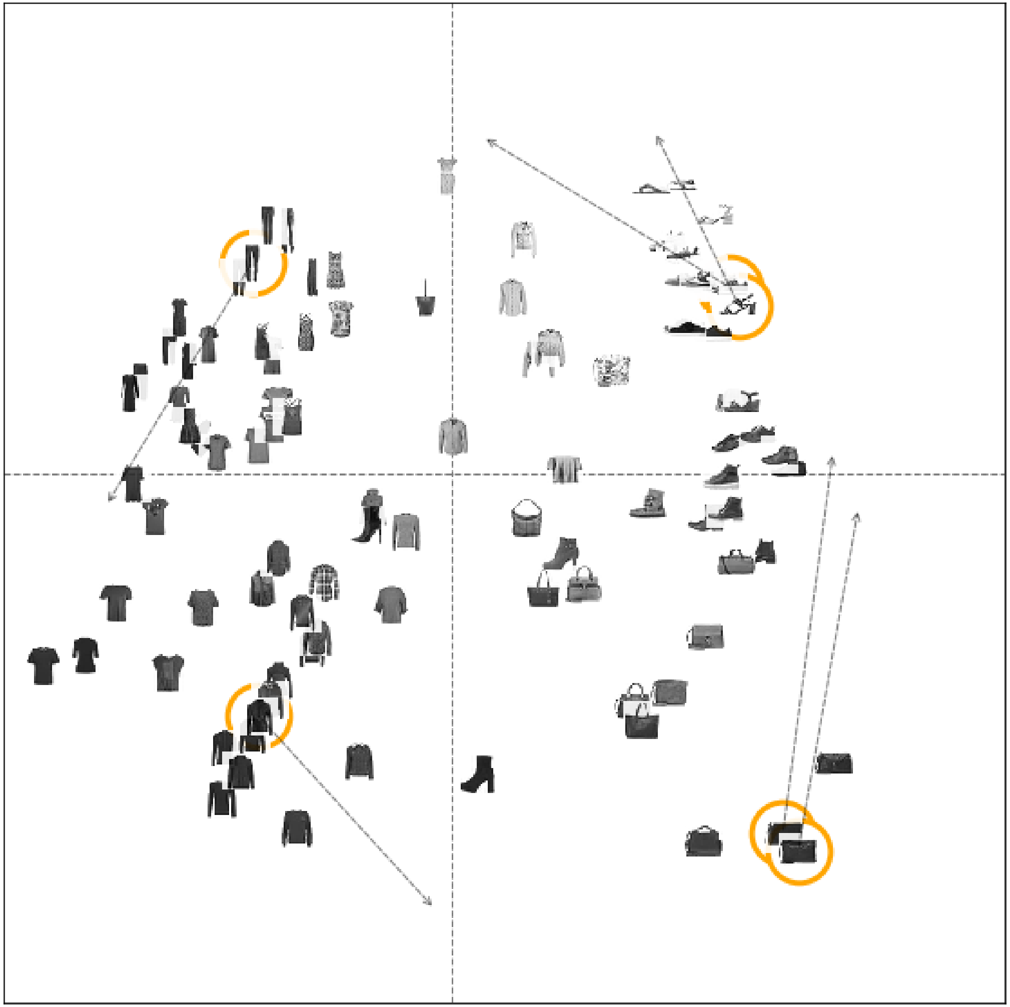
https://github.com/zalandoresearch/fashion-mnist

Probabilistic Principle Component Analysis (PPCA)



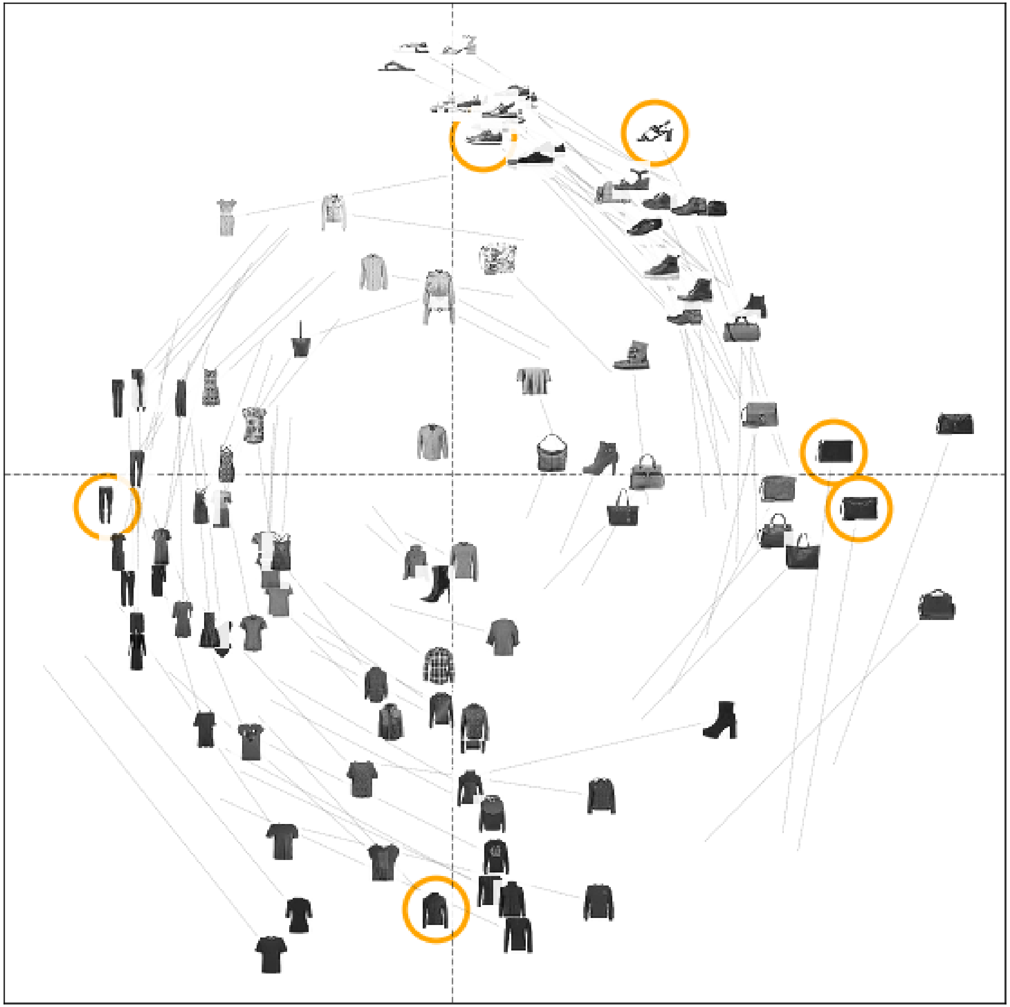
Having an initial visualization with PPCA model, ...

Interactive PPCA (iPPCA)



The user wants to maninpulate the visualization by moving some points

iPPCA result



The result of interactive model is explainable to human

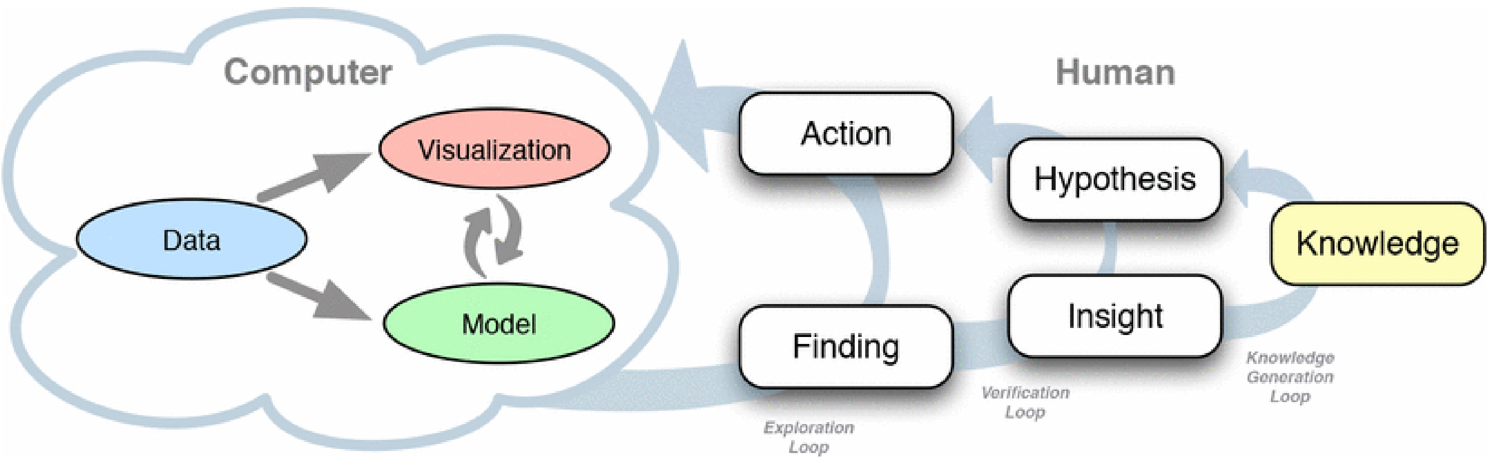
Motivation

User interaction in model design and analysis

Motivation

User interaction in model design and analysis

- Visual analytic with Human-in-the-loop

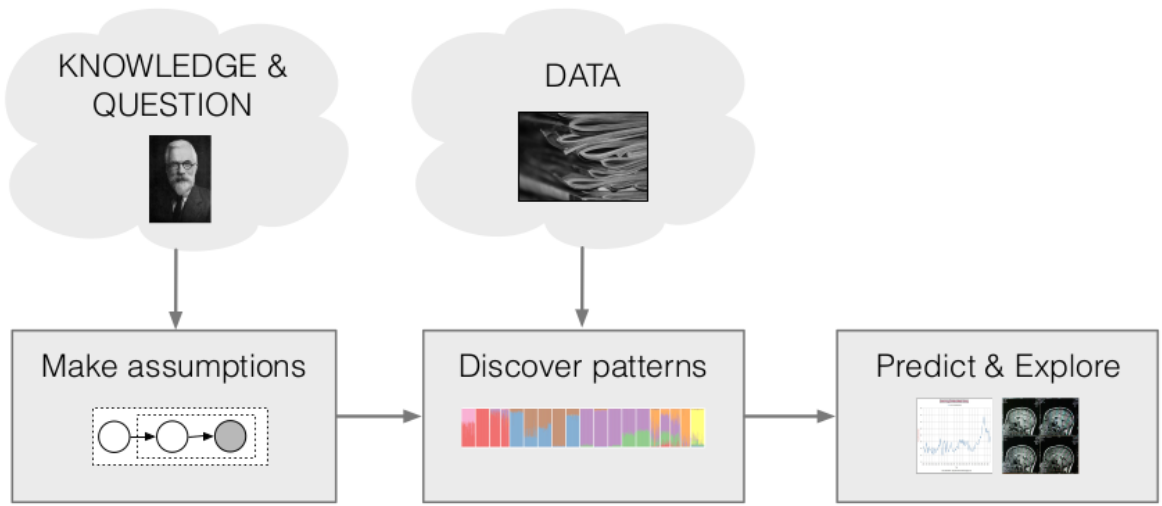


Sacha, Dominik, et al. "Knowledge generation model for visual analytics." IEEE TVCG 2014

Motivation

User interaction in model design and analysis

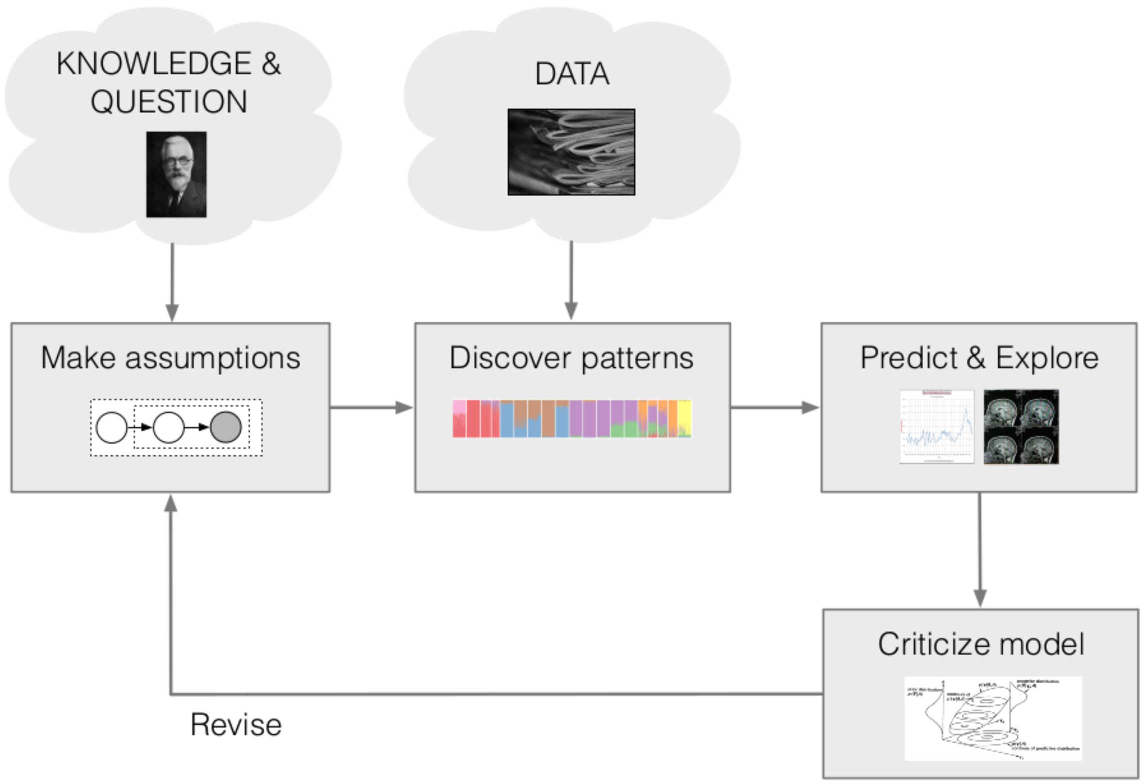
- Probabilistic model pipeline



Motivation

User interaction in model design and analysis

- Probabilistic model pipeline with revising



Motivation

User interaction in model design and analysis

- Visual analytic with Human-in-the-loop
- Probabilistic model pipeline with revising

Allow the user to manipulate the visualization

- to express their needs
- to create a more understandable / explainable visualization
- without loss of quality

Existing approaches

Intergrating user's feedbacks into existing DR methods

Some (complicated) examples.

Existing approaches

Intergrating user's feedbacks into existing DR methods

$$\mathbf{Y} = \underset{[\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(n)}]}{\arg\min} \sum_{i<j\leq n} \boxed{\rho\Big|d_{\omega}(i,j) - d_{\mathrm{Y}}(i,j)\Big|} + \boxed{(1-\rho)\Big|d_{\omega_F}(i,j) - d_{\mathrm{Y}}(i,j)\Big|}$$

Existing approaches

Intergrating user's feedbacks into existing DR methods

$$\mathbf{Y} = \underset{[\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(n)}]}{\arg\min} \sum_{i<j\leq n} \boxed{\rho \Big| d_{\omega}(i,j) - d_{\mathrm{Y}}(i,j) \Big|} + \boxed{(1-\rho) \Big| d_{\omega_F}(i,j) - d_{\mathrm{Y}}(i,j) \Big|}$$

$$\mathbf{A} = \underset{\mathbf{A}}{\arg\min} \frac{1}{2} \Big(\boxed{\sum_{i,j} (\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 \widetilde{W}_{ij}} \boxed{+ \sum_{ML'} (\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 - \sum_{CL'} (\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 \Big)}$$

Existing approaches

Intergrating user's feedbacks into existing DR methods

$$\mathbf{Y} = \underset{[\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(n)}]}{\arg\min} \sum_{i<j\leq n} \boxed{\rho\left|d_\omega(i,j) - d_{\mathrm{Y}}(i,j)\right|} + \boxed{(1-\rho)\left|d_{\omega_F}(i,j) - d_{\mathrm{Y}}(i,j)\right|}$$

$$\mathbf{A} = \underset{\mathbf{A}}{\arg\min} \frac{1}{2}\left(\boxed{\sum_{i,j}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 \widetilde{W}_{ij}} \boxed{+ \sum_{ML'}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 - \sum_{CL'}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2\right)}$$

$$J(\mathbf{A}) = \boxed{\frac{1}{2n^2}\sum_{i,j}\|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2} \boxed{\begin{aligned}&+\frac{\alpha}{2n_{CL}} \sum_{(\mathbf{x}^{(i)},\mathbf{x}^{(j)})\in CL} \|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2 \\ &- \frac{\beta}{2n_{ML}} \sum_{(\mathbf{x}^{(i)},\mathbf{x}^{(j)})\in ML} \|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2\end{aligned}}$$

Existing approaches

Intergrating user's feedbacks into existing DR methods

$$\mathbf{Y} = \underset{[\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(n)}]}{\arg\min} \sum_{i<j\leq n} \boxed{\rho\left|d_\omega(i,j) - d_{\mathrm{Y}}(i,j)\right|} + \boxed{(1-\rho)\left|d_{\omega_F}(i,j) - d_{\mathrm{Y}}(i,j)\right|}$$

$$\mathbf{A} = \underset{\mathbf{A}}{\arg\min}\, \frac{1}{2}\Big(\boxed{\sum_{i,j}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 \widetilde{W}_{ij}} \boxed{+ \sum_{ML'}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2 - \sum_{CL'}(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})^2\Big)}$$

$$J(\mathbf{A}) = \boxed{\frac{1}{2n^2}\sum_{i,j}\|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2} \boxed{+\frac{\alpha}{2n_{CL}}\sum_{(\mathbf{x}^{(i)},\mathbf{x}^{(j)})\in CL}\|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2 - \frac{\beta}{2n_{ML}}\sum_{(\mathbf{x}^{(i)},\mathbf{x}^{(j)})\in ML}\|\mathbf{A}^T\mathbf{x}^{(i)} - \mathbf{A}^T\mathbf{x}^{(j)}\|^2}$$

- User's feedbacks $\Longrightarrow$ Regularization term
- Jointly optimized with the Objective function of the basic DR method.

Existing approaches

Intergrating user's feedbacks into existing DR methods as a regularization term

Existing approaches and ours

Intergrating user's feedbacks into $\underbrace{\text{existing DR methods}}_{\Downarrow}$ ~~as a regularization term~~

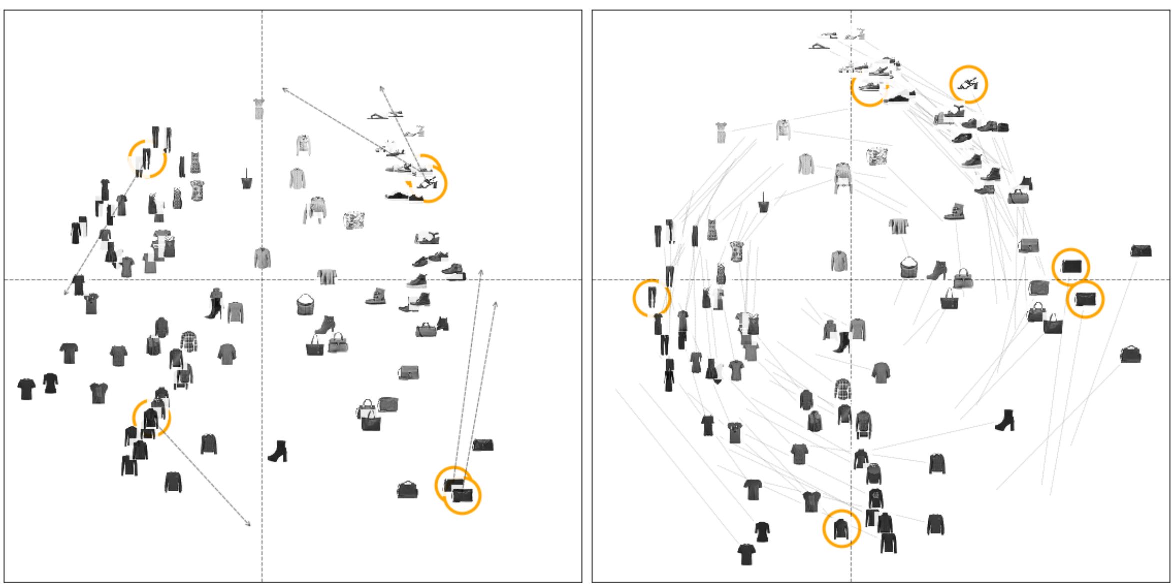
$$\text{a probabilistic dimensionality reduction model}$$

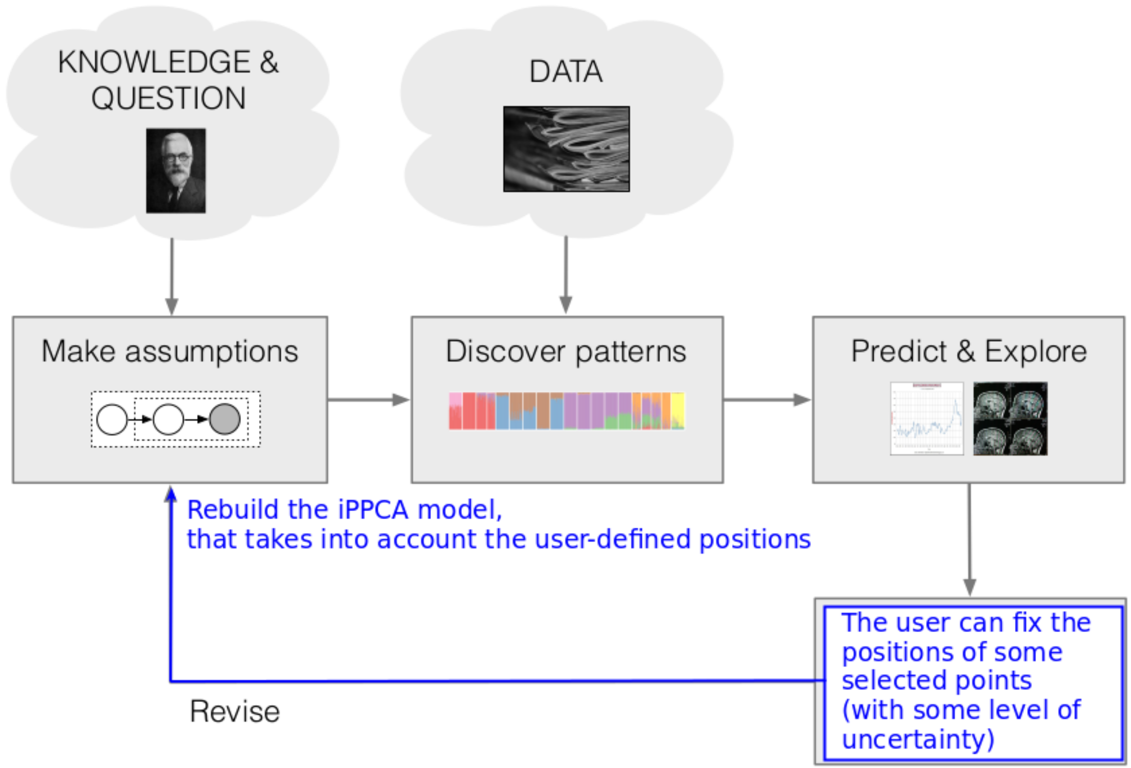
Existing approaches and ours

Intergrating user's feedbacks into $\underbrace{\text{existing DR methods}}_{\Downarrow}$ ~~as a regularization term~~

$$\text{a probabilistic dimensionality reduction model}$$

- Probabilistic PCA (PPCA) as a simple basic model to work with
- User's feedbacks $\approx$ prior knowledge to (re)construct the model.





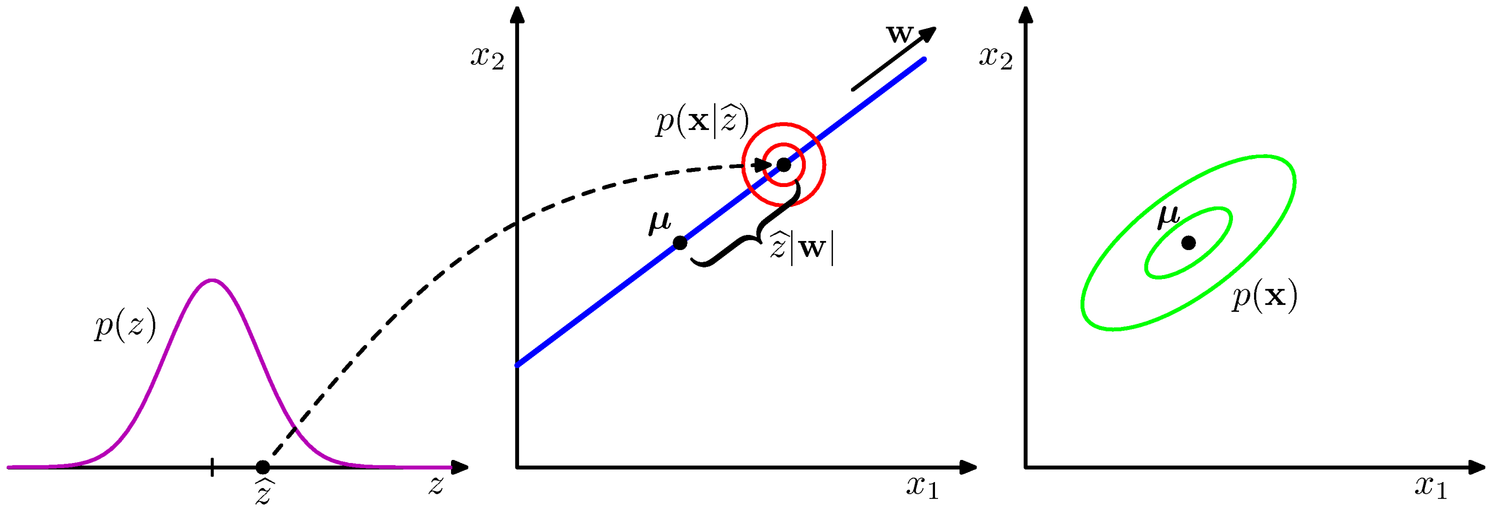
The user-indicated postion of selected points is modelled directly

in the prior distribution of the PPCA model.

A closer look at the PPCA model

A generative view of the probabilistic PCA model.

- 2-dimensional data $p(\mathbf{x})$
- generated from 1-dimensional latent variable $p(\mathbf{z})$



Bishop's PRML Figure. 12.9

Proposed interactive PPCA model

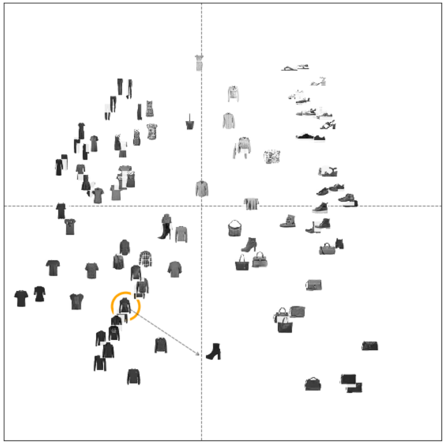
- $\mathbf{X} = \{\mathbf{x}_n\}$: observed dataset of N data points of D-dimensions.
- The embedded points in the 2D visualization imply the corresponding latent variables $\mathbf{Z} = \{\mathbf{z}_n\}$.
- The moved points in the visualization are modelled in the prior distribution of $\mathbf{Z}$

$$\boxed{\boldsymbol{z}_n \sim \begin{cases} \mathcal{N}(\boldsymbol{z}_n \mid \boldsymbol{\mu}_n,\ \sigma^2_{\text{fix}}) & \text{if } \boldsymbol{z}_n \text{ is fixed by user,} \\ \mathcal{N}(\boldsymbol{z}_n \mid \mathbf{0}, \mathbf{1}) & \text{otherwise.} \end{cases}}$$

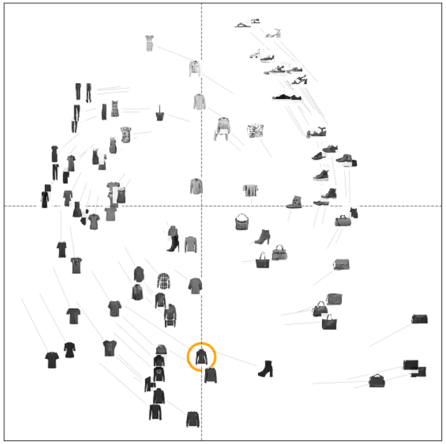
- The iPPCA model: $\mathbf{x}_n \mid \mathbf{z}_n \sim \mathcal{N}(\mathbf{x}_n \mid \mathbf{W}\mathbf{z}_n,\ \sigma^2\mathbf{I}_D)$.
- The inference problem: $\theta_{MAP} = \text{argmax}_\theta \log p(\theta \mid \mathbf{X})$ where θ represents all model's parameters.
- The MAP estimate of the latent variables $\mathbf{Z}$ is found by following the partial gradient $\nabla_{\mathbf{Z}} \log p(\theta, \mathbf{X})$ to its local optima.

How the user prior is handled?

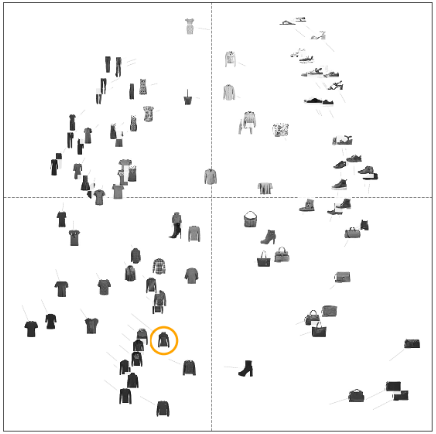
- The user can fix the position of several interested points, with some level of uncertainty (σ_{fix})
- A very small of uncertainty $\Longrightarrow$ the user is very certain.
- A large enough of uncertainty $\Longrightarrow$ the user is not sure.



user's uncertainty σ_{fix}

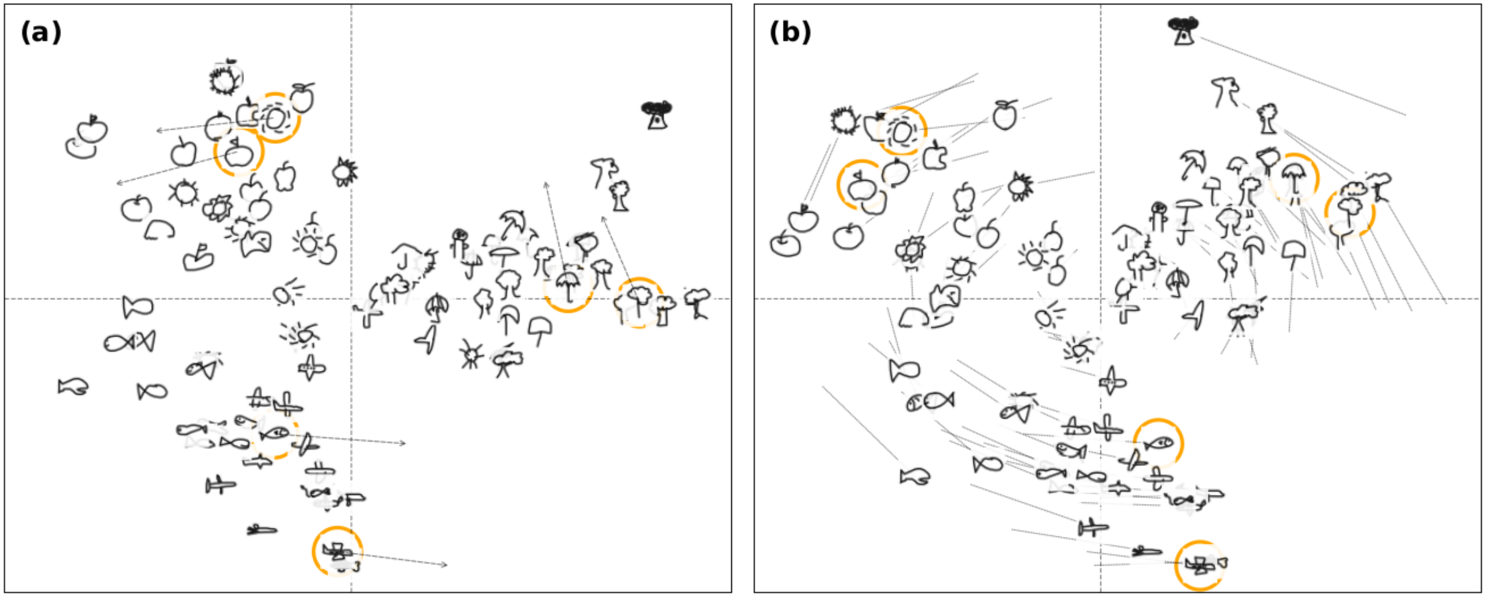


Very small $\sigma_{fix} = 1e-4$: very sure



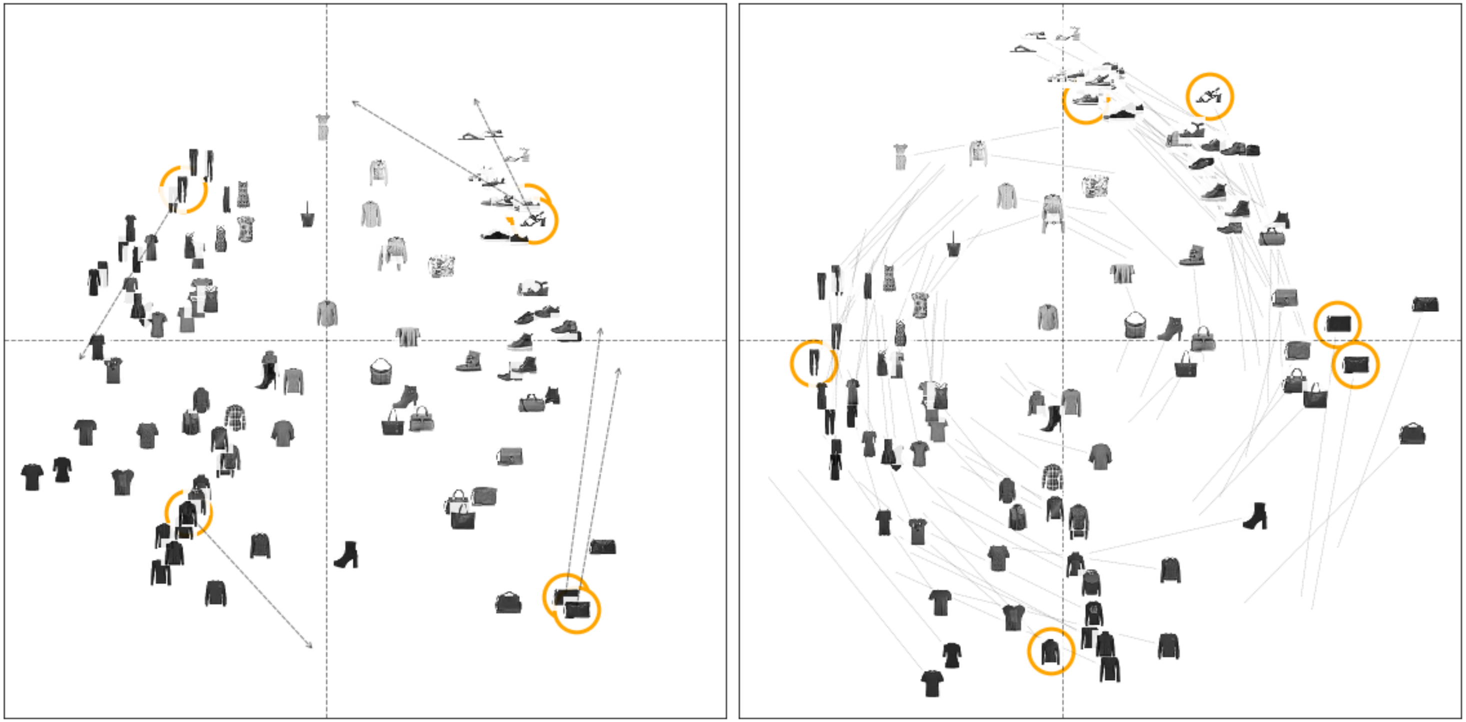
Large $\sigma_{fix} = 0.2$: very uncertain

Some working examples



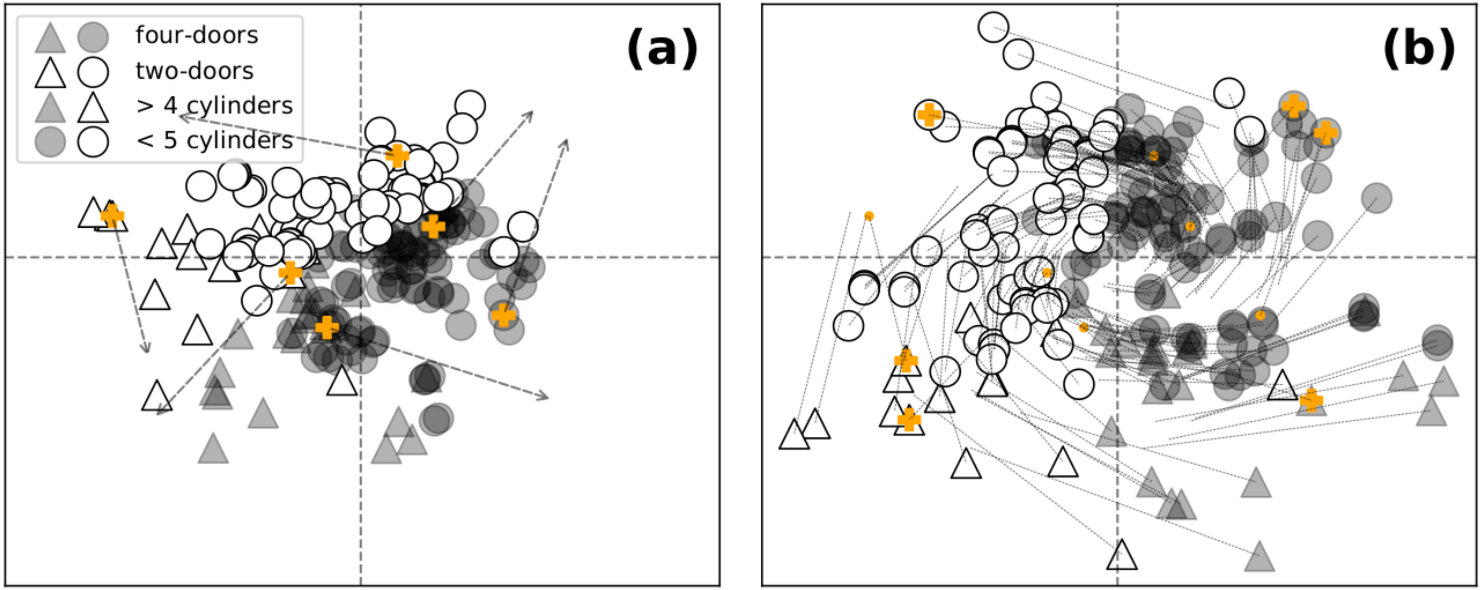
90 images of Quickdraw dataset

Some working examples



100 images of Fashion dataset

Some working examples



203 datapoints in Automobile dataset

More discussion on the motivation of this approach

- Problem of PPCA's ambiguous-rotation:

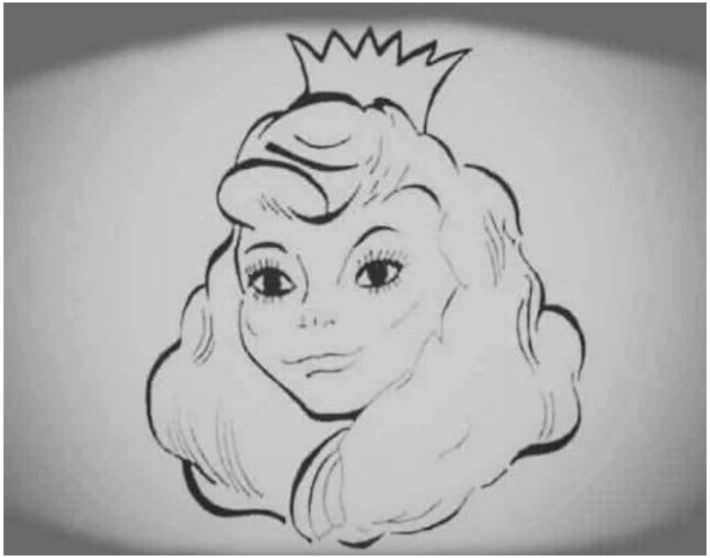
More discussion on the motivation of this approach

- Problem of PPCA's ambiguous-rotation:

The *wife in my eyes*

More discussion on the motivation of this approach

- Problem of PPCA's ambiguous-rotation:



The *wife in others' eyes*

More discussion on the motivation of this approach

Problem of PPCA's ambiguous-rotation:

- the principle components in PPCA are not necessarily orthogonal
- the visualization can thus be in any rotation around the origin

Role of user interaction:

- can be considered as a complement to solve the above ambiguity
- the user'decision is subjective, but can help to communicate the analytical result
- creating a more understandable, easily explainable visualization

Advantage of probabilistic approach

Combination of solid theoretical models and morden powerful inference toolboxes

- take any old-class model or modern generative model
- plug into a modern probability framework (Stan, PyMC3, Pyro, Tensorflow Probability)

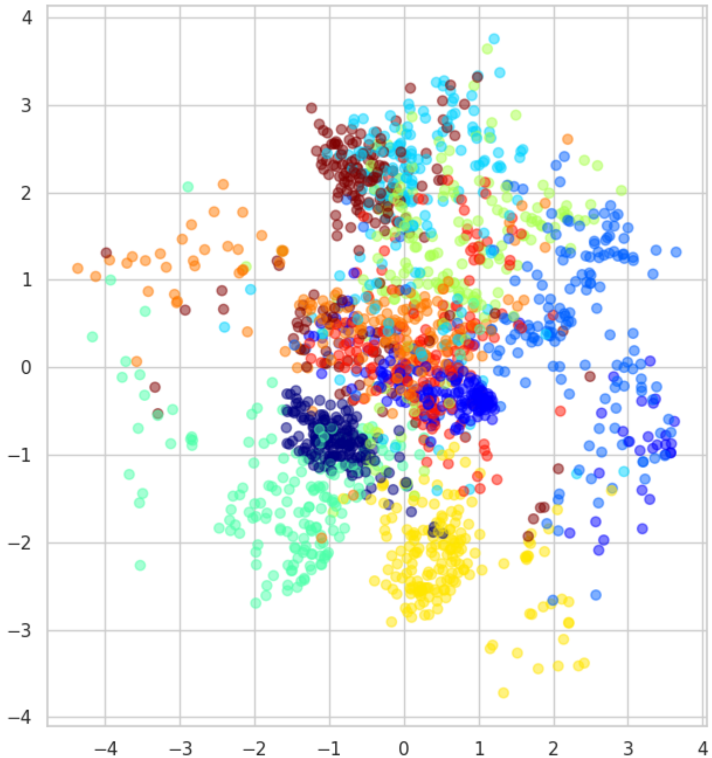
Can easily extend the classic models

- extend the general generative process:

$$\mathbf{x}_n \mid \mathbf{z}_n \sim \mathcal{N}(f(\mathbf{z}_n), \sigma^2 \mathbf{I})$$

  - in PPCA model, $f(\mathbf{z}_n) = \mathbf{W}\mathbf{z}_n$
  - but $f(\mathbf{z}_n)$ can be any high-capacity representation function (a neural net)
- take advantages of the modern inference methods, e.g., Stochatic Variational Inference

Advantage of probabilistic approach



Embedding of 1797 digits with modified PPCA in which a decoder $f(\mathbf{z})$ is a simple neural network with one hidden layer of 50 unit and a sigmoid activation function. The inference is done with pyro's built-in SVI optimizer.

Recap

- [Why we do that] The proposed interactive PPCA model allows the user to control the visualization
  - in order to create an explainable one (for communicating the analytical result)
  - explore the visualization (kind of "what-if" analysis)
- [Technique] The user's feedbacks can be efficiently integrated into a probabilistic model via the prior distribution of the latent variable.
- [Potential] The probabilistic model is flexible to extend and can be easily optimized by the blackbox inference methods.
- [Future work] We can thus focus on the problem of modeling the user's feedback without worrying about the complex optimization procedure.

User-steering Interpretable Visualization with Probabilistic PCA

Viet Minh Vu and Benoît Frénay

University of Namur, Belgium

-