Reference (3)
Speech Materials

Jeff (CHI-HSUAN HO)

The subsequent slides only showcase a few parts of the work materials and results.

Multi-Omics Factor Analysis (MOFA)
A Framework for Unsupervised Integration of Multi-omics Data Sets

Jeff (CHI-HSUAN HO)

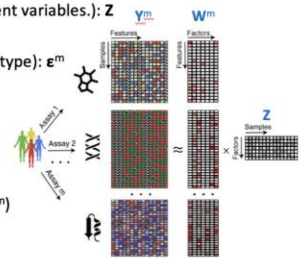
Materials

Methods

- MOFA models
  - Factor Analysis: Reduce the dimensionality of a (big) dataset into a small set of variables which are easier to interpret and visualize.
  - Sample size: N (n)
  - Feature numbers: D (d)
  - Data matrices, view: M (m)
  - Factor matrix (matrix that contains the low-dimensional latent variables.): **Z**
  - Loading matrix, weight matrix: $\mathbf{W}^m$
  - Residual noise (its form depends on the specific of the data type): $\boldsymbol{\varepsilon}^m$
  - The expectation of x under the distribution q. ($E_q[x]$): $\langle x \rangle$
  - Model:

$$\mathbf{Y}^m = \mathbf{Z}\mathbf{W}^{mT} + \boldsymbol{\epsilon}^m$$

  - A Bayesian framework is used. (Place prior dist. on **Z**, $\mathbf{W}^m$, $\boldsymbol{\varepsilon}^m$)



Methods

- *Likelihood* × *Prior* (Gaussian data)

$$p(\mathbf{Y}, \hat{\mathbf{W}}, \mathbf{S}, \mathbf{Z}, \boldsymbol{\Theta}, \boldsymbol{\alpha}, \boldsymbol{\tau}) = \prod_{m=1}^{M}\prod_{n=1}^{N}\prod_{d=1}^{D_m} \mathcal{N}\left(y_{nd}^m \,|\, \sum_{k=1}^{K} s_{dk}^m \hat{w}_{dk}^m z_{nk}, 1/\tau_d\right)$$

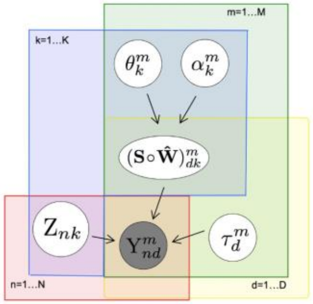
$$\prod_{m=1}^{M}\prod_{d=1}^{D_m}\prod_{k=1}^{K} \mathcal{N}\left(\hat{w}_{dk}^m \,|\, 0, 1/\alpha_k^m\right) \mathrm{Ber}(s_{d,k}^m | \theta_k^m)$$

$$\prod_{n=1}^{N}\prod_{k=1}^{K} \mathcal{N}\left(z_{nk} \,|\, 0, 1\right)$$

$$\prod_{m=1}^{M}\prod_{k=1}^{K} \mathrm{Beta}\left(\theta_k^m \,|\, a_0^\theta, b_0^\theta\right)$$

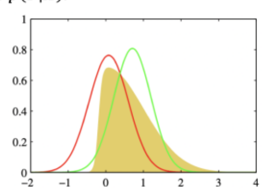
$$\prod_{m=1}^{M}\prod_{k=1}^{K} \mathcal{G}\left(\alpha_k^m \,|\, a_0^\alpha, b_0^\alpha\right)$$

$$\prod_{m=1}^{M}\prod_{d=1}^{D_m} \mathcal{G}\left(\tau_d^m \,|\, a_0^\tau, b_0^\tau\right).$$



Methods

- *Posterior* – model Inference for Gaussian data: Variational Bayes
  - Goal: In the set of the tractable distribution for $q(\boldsymbol{X})$ (or q), maximize $\mathcal{L}(q)$ (or minimize $KL(q||p)$) as possible so that $q(\boldsymbol{X})$ (or q) can provide a good approximation to $p(\boldsymbol{Y}|\boldsymbol{X})$.
  - Find the general form for the set of the tractable distribution for $q(\boldsymbol{X})$ (or q).
  - Mean-field theory assumption
    - Partition the elements of $\boldsymbol{X}$ into M disjoint groups, then $q(\boldsymbol{X}) = \prod_{i=1}^{M} q_i(\boldsymbol{x}_i)$

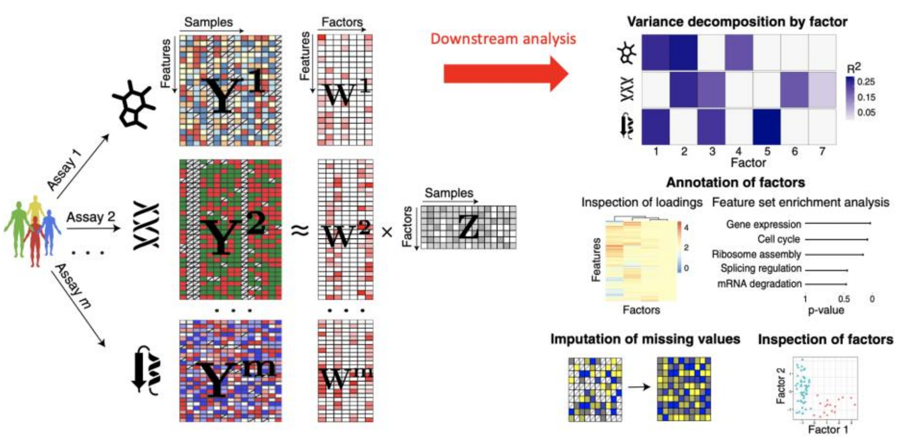


  - Maximize $\mathcal{L}(q)$ with respect to each of the latent variables and parameters in turn.
    - Dissect out the dependence on one of the latent variables and parameters $q_j(\boldsymbol{X}_j)$. (Keep the $\{q_{i\neq j}\}$ fixed, maximize $\mathcal{L}(q)$ respect to $q_j(\boldsymbol{X}_j)$).

$$\mathcal{L}(q) = \int \prod_{i=1}^{M} q_i(\boldsymbol{x}_i)\{\ln p(\boldsymbol{Y},\boldsymbol{X}) - \sum_{i=1}^{M} \ln q_i(\boldsymbol{x}_i)\} d\boldsymbol{X} = \int q_j \left\{ \int \ln p(\boldsymbol{Y},\boldsymbol{X}) \prod_{i\neq j} q_i \, d\boldsymbol{x}_i \right\} d\boldsymbol{x}_j - \int q_j \ln q_j d\boldsymbol{x}_j + const$$

$$= \int q_j \ln \tilde{p}(\boldsymbol{Y},\boldsymbol{x}_j) d\boldsymbol{x}_j - \int q_j \ln q_j d\boldsymbol{x}_j + const = -KL(q_j(\boldsymbol{x}_j) || \tilde{p}(\boldsymbol{Y},\boldsymbol{x}_j)), \text{ where } \ln\tilde{p}(\boldsymbol{Y},\boldsymbol{x}_j) = E_{q_{i\neq j}}[\ln p(\boldsymbol{Y},\boldsymbol{X})] + const.$$

Methods



https://github.com/jeff665547/SkillShare-MOFA/raw/master/MOFA.pdf

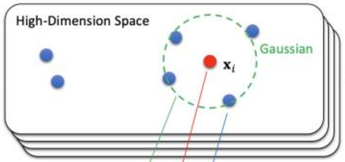
Dimension Reduction Techniques
From t-SNE to UMAP

Jeff (CHI-HSUAN HO)

Materials

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Measure how close between two different high-dimensional objects.



Gaussian: $\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x_j-x_i)^2}{2\sigma^2}\right\}$

Unscaled similarity

$$p_{ij} = \frac{\exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2\right)}{\sum_k \sum_{l \neq k} \exp(-\|\mathbf{x}_k - \mathbf{x}_l\|^2/2\sigma^2)},$$ Scaled similarity

σ *are all the same for each point*

- Similar (closer) points in the high-D space → Large p_{ij}.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

- Gradient of Kullback-Leibler divergences: $C = KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$
  - Let $d_{ij} = \|y_i - y_j\|$, $Z_i = \sum_{l \neq i}(1 + d_{il}^2)^{-1}$
  - $\frac{\partial C}{\partial y_i} = \sum_j \left(\frac{\partial C}{\partial d_{ij}} \cdot \frac{\partial d_{ij}}{\partial y_i} + \frac{\partial C}{\partial d_{ji}} \cdot \frac{\partial d_{ji}}{\partial y_i}\right) = \sum_j \left(\frac{\partial C}{\partial d_{ij}} + \frac{\partial C}{\partial d_{ji}}\right) \cdot \frac{y_i - y_j}{d_{ij}} = 2\sum_j \frac{\partial C}{\partial d_{ij}} \cdot \frac{y_i - y_j}{d_{ij}}$
  - $\frac{\partial C}{\partial d_{ij}} = -\sum_{k \neq i} p_{ik} \frac{\partial}{\partial d_{ij}} log(q_{ik}) = -\sum_{k \neq i} p_{ik} \frac{\partial}{\partial d_{ij}} (log(q_{ik} Z_i) - log Z_i)$

    $= -\sum_{k \neq i} p_{ik} \left(\frac{1}{q_{ik} Z_i} \frac{\partial}{\partial d_{ij}} (1 + d_{ik}^2)^{-1} - \frac{1}{Z_i} \frac{\partial Z_i}{\partial d_{ij}}\right)$, $\frac{\partial}{\partial d_{ij}}(1 + d_{ik}^2)^{-1}, \frac{\partial}{\partial d_{ij}} Z_i$ only have value when $k, l = j$

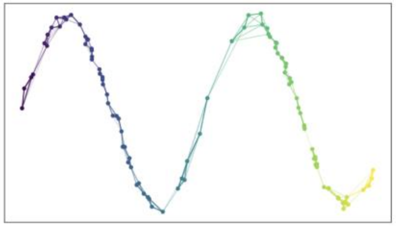
    $= p_{ij} \left(\frac{1}{q_{ij} Z_i} 2d_{ij} (1 + d_{ij}^2)^{-2}\right) + \sum_{k \neq i} \frac{-p_{ik}}{Z_i} \left((1 + d_{ij}^2)^{-2}\right) 2d_{ij}$

    $= 2d_{ij} \left(p_{ij} \frac{(1 + d_{ij}^2)^{-2}}{q_{ij} Z_i} - \frac{(1 + d_{ij}^2)^{-2}}{Z_i} \sum_{k \neq i} p_{ik}\right) = 2d_{ij} \left(p_{ij}(1 + d_{ij}^2)^{-1} - q_{ij}(1 + d_{ij}^2)^{-1} \sum_{k \neq i} p_{ik}\right)$

    $= 2d_{ij} \left(p_{ij}(1 + d_{ij}^2)^{-1} - q_{ij}(1 + d_{ij}^2)^{-1}\right) = 2d_{ij}(p_{ij} - q_{ij})(1 + d_{ij}^2)^{-1}$
  - $\frac{\partial C}{\partial y_i} = 4\sum_j (p_{ij} - q_{ij}) \cdot \left(1 + \|y_i - y_j\|^2\right)^{-1} \cdot (y_i - y_j)$

Uniform Manifold Approximation and Projection (UMAP)

A computation view of UMAP
- In the High-D space (manifold):



- Consider the symmetric matrix $B = A + A^T - A \circ A^T$, where $\circ$ is the Hadamard (or pointwise) product.
- In a meaningful way, the element in B represents $\mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$
- B is also the undirected weighted matrix.

Uniform Manifold Approximation and Projection (UMAP)

- Gradient of Cross entropy
  - Let $d_{ij} = \|y_i - y_j\|_2$

$$v_{ij} = \frac{1}{(1 + ad_{ij}^{2b})}, \qquad 1 - v_{ij} = \frac{ad_{ij}^{2b}}{(1 + ad_{ij}^{2b})}, \qquad \frac{\partial v_{ij}}{\partial d_{ij}} = -\frac{2abd_{ij}^{2b-1}}{(1 + ad_{ij}^{2b})^2}$$

$$\begin{aligned}
\frac{\partial C}{\partial y_i} &= \sum_j \left[-\frac{w_{ij}}{v_{ij}} \cdot \frac{\partial v_{ij}}{\partial d_{ij}} + \frac{1 - w_{ij}}{1 - v_{ij}} \cdot \frac{\partial v_{ij}}{\partial d_{ij}}\right] \frac{\partial d_{ij}}{\partial y_i} \\
&= \sum_j \left[\left(-w_{ij}(1 + ad_{ij}^{2b}) + \frac{(1 - w_{ij})(1 + ad_{ij}^{2b})}{ad_{ij}^{2b}}\right) \cdot \frac{\partial v_{ij}}{\partial d_{ij}}\right] \frac{\partial d_{ij}}{\partial y_i} \\
&= \sum_j \frac{2abd_{ij}^{2(b-1)}}{1 + ad_{ij}^{2b}} \cdot w_{ij} \cdot d_{ij} \cdot \frac{(y_i - y_j)}{d_{ij}} && \text{Attractive force} \\
&\quad - \frac{2b}{d_{ij}^2(1 + ad_{ij}^{2b})} \cdot (1 - w_{ij}) \cdot d_{ij} \cdot \frac{(y_i - y_j)}{d_{ij}} && \text{Repulsive force}
\end{aligned}$$

https://github.com/jeff665547/SkillShare-UMAP-and-t-SNE/raw/master/From%20t-SNE%20to%20UMAP.pdf

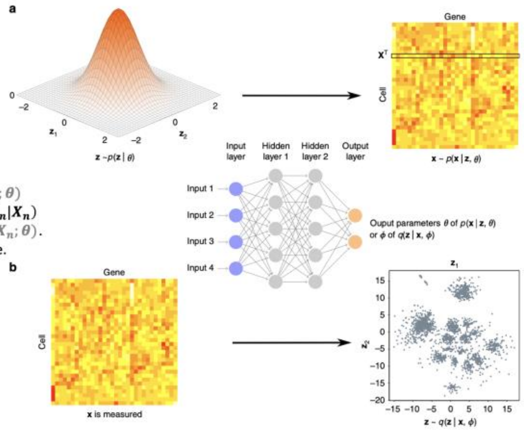
CENTRILLION

Interpretable Dimensionality Reduction of Single Cell Transcriptome Data with Deep Generative Models (SCVIS)

Jeff (CHI-HSUAN HO)

Materials

SCVIS



- Interest: Find $P(Z_n|X_n;\theta)$
  → Find an optimal $q(Z_n|X_n)$ to approximate $P(Z_n|X_n;\theta)$.
  → Variational Inference.

SCVIS

- Regularizer:

$$\sum_i \mathbb{KL}\left(p_{\cdot|i}||q_{\cdot|i}\right) = \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} p_{j|i}\log\frac{p_{j|i}}{q_{j|i}}$$

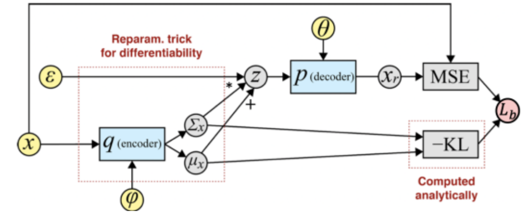
- Final objective function:

$$\arg\min_{\theta,\phi}\left(-\sum_{n=1}^{N}\mathrm{ELBO}_n + \alpha\sum_{n=1}^{N}\mathbb{KL}\left(p_{\cdot|n}||q_{\cdot|n}\right)\right)$$

  - α is set to the dimensionality of the input high-dimensional data.
    (ELBO scales with the dimensionality of the input data)

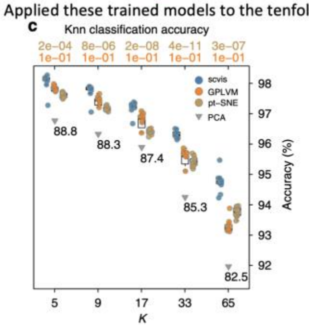
Variational Autoencoder (VAE)

- Relationship with the Autoencoder
  - Cost function: $L_b(q(Z_n|X_n;\phi),\theta) = E_{z\sim q_\phi}[\log P(X_n|Z_n;\theta)] - KL[q(Z_n|X_n;\phi)||P(Z_n)]$
    $= E_{\varepsilon\sim N(0,1)}[\log P(X_n|g(\varepsilon,X_n,\phi);\theta)] - KL[q(Z_n|X_n;\phi)||P(Z_n)]$
    - $q(Z_n|X_n;\phi)$: Encoder, $P(X_n|g(\varepsilon,X_n,\phi);\theta)$: Decoder, $\mu_\phi(X_n)=\mu_x$, $\sigma_\phi(X_n)=\Sigma_x$
    - $E_{\varepsilon\sim N(0,1)}[\log P(X_n|g(\varepsilon,X_n,\phi);\theta)]$: MSE, (Reconstruction Error), it based on the encoder and the decoder.
    - $KL[q(Z_n|X_n;\phi)||P(Z_n)]$ only based on the encoder.
      (A penalty for Σ_x (if Σ_x close to 0 → autoencoder), we want it to close 1 (same as prior).)



Simulation

- Model performance on the new data embedding (pt-SNE, GPLVM, PCA, and SCVIS)
  - Train SCVIS, GPLVM, pt-SNE and KNN classifiers on the original 2200 simulated data.
  - Applied these trained models to the tenfold dataset (22000 simulated data.)



- SCVIS performs significantly better for different Ks.
- For a larger K, SCVIS assigns the outliers to the six genuine clusters (performance decrease).
- PCA is worse than other methods.

https://github.com/jeff665547/SkillShare-SCVIS/raw/master/SCVIS%20A%20VAE-based%20approach.pdf