Latent Variable Models

Bioinformatics Course Supplement

Introduction

- It is often helpful to introduce additional variables for understanding of the relationship between two observed variables.
- There are two possibilities
  - *Explanatory variables* which are actually observed.
  - *Hidden or latent variables* which are not observed.

Latent Variables

- Latent variable
  - A hypothetical construct invented for the purpose of understanding some research area of interest.
  - Can explain the structure in a set of correlated, observed variables.
  - Examples
    - *Social class*, *public opinion*, *extrovert personality*, *racial prejudice*, and so on in social and behavioral sciences

Latent Variable Model (1)

- Model
  - A simplified description of the structure of the observations.
  - Provide a simpler explanation of the phenomena under investigation that is consistent with the observations.
  - *data* = ***model*** + *residual*
- Latent variable model
  - Explain the statistical properties of the observed variables in terms of the *latent variables*.

Latent Variable Model (2)

- Can be classified according to the type of observed and latent variables.

	observed	
latent	continuous	discrete
continuous	***factor analysis***	***latent trait analysis***
discrete	***latent profile analysis***	***latent class analysis***

Latent Variable Model (3)

- General Formulation

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \int p(\mathbf{x} \mid \mathbf{z}) \, p(\mathbf{z}) d\mathbf{z}$$

$\mathbf{x} = (x_1, x_2, \ldots, x_n)$: vector of data variables

$\mathbf{z} = (z_1, z_2, \ldots, z_m)$: vector of observed variables

$p(\mathbf{x}, \mathbf{z})$: joint distribution of data and latent variables

$p(\mathbf{x} \mid \mathbf{z})$: conditional distribution of the data variables given the latent variables

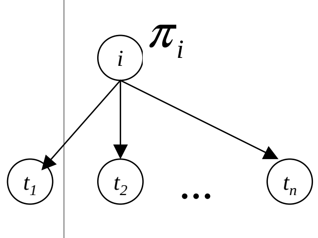
$p(\mathbf{x}), p(\mathbf{z})$: marginal distribution of the data variables and latent variables, respectively

Latent Variable Model (4)

- The task
  - Infer $p(\mathbf{x} \mid \mathbf{z})$ and $p(\mathbf{z})$
- General assumptions
  - *Conditional independence*
    - $p(\mathbf{x} \mid \mathbf{z}) = \prod_{i=1}^{n} p(x_i \mid \mathbf{z})$
    - It is the latent variables which produce the observed relationships amongst the data variables
  - *Density function of known form*
    - $p(\mathbf{x} \mid \mathbf{z}),\ p(\mathbf{z})$ are of known form, but dependent on a set of unknown parameter values.

Mixture Model (1)

- Mixture model
  - Probabilistic mixtures of M simpler parametric distributions
  - Can represent more complex distribution.
- Formulation



$$p(\mathbf{t}) = \sum_{i=1}^{M} \pi_i \, p(\mathbf{t} \mid i)$$

$$0 \le \pi_i \le 1 \quad (i = 1, \ldots, M)$$

$$\sum_i \pi_i = 1$$

Mixture Model (2)

- Model fitting
  - EM algorithm

$$L = \sum_{i=1}^{n} \ln \left\{ \sum_{j=1}^{m} \pi_j \, p(\mathbf{t} \mid i) \right\} \rightarrow L_c = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} \ln(\pi_j \, p(\mathbf{t} \mid i))$$

$\{z_{ij}\}$: missing data

z_{ij} : data *i* comes from component *j*

Factor Analysis (1)

- Origin
  - by psychologists with the aim to explain results from cognitive tests in terms the underlying organization of mental abilities.

- Formulation

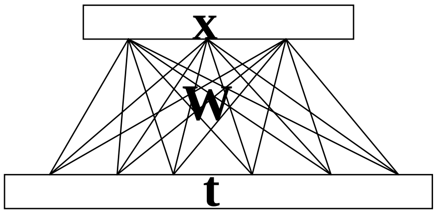
$$\mathbf{t} = \boldsymbol{\mu} + \mathbf{W}\mathbf{x} + \mathbf{e} \qquad \mathbf{t} = \mathbf{W}\mathbf{x} + \mathbf{e} \quad (\boldsymbol{\mu} = \mathbf{0})$$

$\mathbf{t}$ **:** ***D*-dimensional continuous variable**

$\mathbf{x}$ **:** ***L*-dimensional continuous latent variable**

$\mathbf{W}$ **:** ***D-by-L* matrix defining the linear function**

$\mathbf{e}$ **:** ***D*-dimensional vector representing the noise associated with each of the *D* observed variables**



Factor Analysis (2)

- Assumptions
  - $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_L)$

$$p(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{L/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right)$$

  - $\mathbf{e} \sim N(\mathbf{0}, \mathbf{\Psi})$, **x** and **e** are uncorrelated.

$$p(\mathbf{t} \mid \mathbf{x}, \mathbf{W}, \mathbf{\Psi}) = (2\pi)^{-D/2} \mid \mathbf{\Psi} \mid^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{t} - \mathbf{W}\mathbf{x})^T \mathbf{\Psi}^{-1}(\mathbf{t} - \mathbf{W}\mathbf{x})\right)$$

Factor Analysis (3)

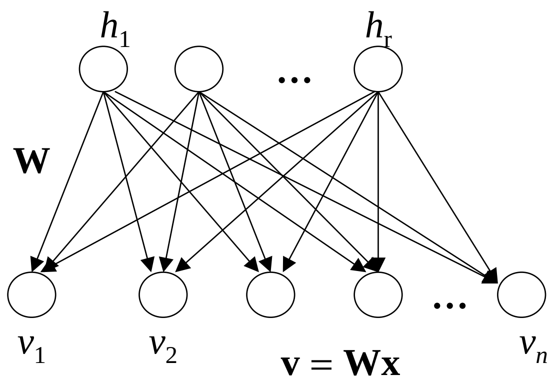
- Covariance matrix

$$\begin{aligned} E(\mathbf{t}\mathbf{t}^T) = \mathbf{\Sigma} &= E[(\mathbf{W}\mathbf{x} + \mathbf{e})(\mathbf{W}\mathbf{x} + \mathbf{e})^T] \\ &= E[\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T + \mathbf{W}\mathbf{x}\mathbf{e}^T + \mathbf{e}\mathbf{x}^T\mathbf{W}^T + \mathbf{e}\mathbf{e}^T] \\ &= \mathbf{W}E(\mathbf{x}\mathbf{x}^T)\mathbf{W}^T + \mathbf{W}E(\mathbf{x}\mathbf{e}^T) + E(\mathbf{e}\mathbf{x}^T)\mathbf{W}^T + E(\mathbf{e}\mathbf{e}^T) \\ &= \mathbf{W}\mathbf{W}^T + E(\mathbf{e}\mathbf{e}^T) \\ &= \mathbf{W}\mathbf{W}^T + \mathbf{\Psi} \end{aligned}$$

- Estimation procedure
  - Lawley, 1940, Joreskog, 1967
  - Rubin and Thayer, *EM algorithm*, 1982
  - Neal and Dayan, *wake-sleep algorithm*, 1997

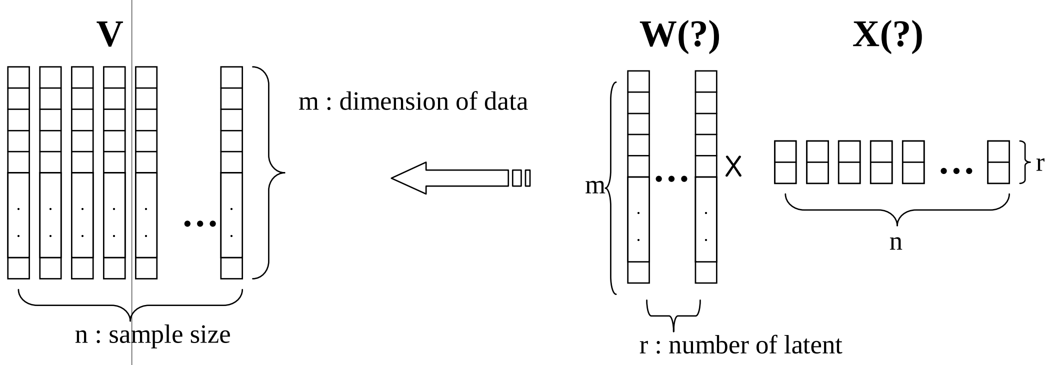
Non-negative Matrix Factorization (1)

- Formulation
  - $\mathbf{T} \approx \mathbf{W}\mathbf{X} \quad \mathbf{T}_{ij},\ \mathbf{W}_{ij},\ \mathbf{X}_{ij} \geq 0$
  - NMF as a latent variable model



Non-negative Matrix Factorization (2)

- NMF as a latent variable model



Non-negative Matrix Factorization (3)

- Modeling as Poisson distribution

$$p(\mathbf{T}_{ij}, \mathbf{X}_j \mid \mathbf{W}) = \frac{\lambda^{\mathbf{T}_{ij}} e^{-\lambda}}{\mathbf{T}_{ij}!}$$

$$\lambda = \sum_{a=1}^{r} (\mathbf{W}_{ia} \mathbf{X}_{aj})$$

- Objective Function

$$F = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\mathbf{T}_{ij} \log(\mathbf{WX})_{ij} - (\mathbf{WX})_{ij} \right]$$

Non-negative Matrix Factorization (4)

- Iterative algorithm for fitting

$$\mathbf{W}_{ia} = \mathbf{W}_{ia} \sum_j \frac{\mathbf{V}_{ij}}{(\mathbf{WX})_{ij}} \mathbf{X}_{aj}, \qquad \mathbf{W}_{ia} = \frac{\mathbf{W}_{ia}}{\sum_j \mathbf{W}_{ja}}$$

$$\mathbf{H}_{aj} = \mathbf{H}_{aj} \sum_i \frac{\mathbf{V}_{ij}}{(\mathbf{WX})_{ij}} \mathbf{W}_{ia}$$