FACTOR ANALYSIS

Machine Learning

IIIy CSE IDP and IDDMP

Refer : Stephen Marsland

Motivation

- High-dimensional data: Many observed variables (M) per sample
- Can we explain correlations via fewer latent variables (factors)?
- Common in psychology (IQ Tests), social sciences, and exploratory data analysis

What is Factor Analysis ?

- The idea of factor analysis is to ask whether the data that is observed can be explained by a **smaller number of uncorrelated factors or latent variables**

- The problem of factor analysis is to find independent factors of **underlying data sources**, and the **noise** that is inherent in the measurements of each factor

Model Assumptions

- Observations Matrix $X_{(N \times M)}$:

  Each row of **X** is M dimensional datapoint

  Let Covariance Matrix of **X** be **Σ**

- We center the data by subtracting off the mean of each variable (i.e., each column) $\mathbf{b}_j = \mathbf{x}_j - \boldsymbol{\mu}_j,\ j = 1 \ldots M$

- This leads to the Estimated mean of each column becoming zero $E[\mathbf{b}_i] = 0$

Model Equation:

- We can write the model that we are assuming as:

$$\mathbf{X} = \mathbf{WY} + \epsilon,$$

where **X** are the observations and **ε** is the noise

Since the factors $\mathbf{b}_i$ that we want to find should be independent, so $\mathrm{cov}(\mathbf{b}_i, \mathbf{b}_j) = 0$, if $i \neq j$

Assumptions continued ...

- Noise is assumed to be Gaussian with zero mean and some known variance: **Ψ**, with the variance of each element being

$$\Psi_i = \mathrm{var}(\epsilon_i)$$

- The covariance matrix of the original data, **Σ**, can now be broken down into cov (**Wb + ε**) = $\mathbf{W}\mathbf{W}^T + \Psi$

where, Ψ is the matrix of noise variances and we have used the fact that cov(b) = I, since the factors are uncorrelated

Aim of Factor Analysis :

- The aim of factor analysis is to try to find a set of
  - **factor loadings** $\mathbf{W}_{ij}$
  - values for the variance of the **noise parameters Ψ**

So that the data in X can be **reconstructed from the parameters**, or so that we can **perform dimensionality reduction**.

EM Framework and Likelihood

- Goal: Find **W (factor loadings)** and **Ψ (noise parameters)**
- EM algorithm treats latent factors x as missing data:

  E-step:

  Estimate expectations given current parameters

  M-step:

  Update W, Ψ to maximize expected complete-data log likelihood

Expectation – Maximisation Algorithm

The General Expectation-Maximisation (EM) Algorithm

- Initialisation
  - guess parameters $\hat{\boldsymbol{\theta}}^{(0)}$
- Repeat until convergence:
  - (E-step) compute the expectation $Q(\boldsymbol{\theta}', \hat{\boldsymbol{\theta}}^{(j)}) = E(f(\boldsymbol{\theta}'; D')|D, \hat{\boldsymbol{\theta}}^{(j)})$
  - (M-step) estimate the new parameters $\hat{\boldsymbol{\theta}}^{(j+1)}$ as $\max_{\boldsymbol{\theta}'} Q(\boldsymbol{\theta}', \hat{\boldsymbol{\theta}}^{(j)})$

Log Likelihood

- Step 1 : We define the log likelihood as

$$Q(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) = \int p(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}_{t-1}) \log(p(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}_t)p(\mathbf{x}))d\mathbf{x}.$$

- equation 1

(where θ is the data we are trying to fit)

- Step 2 : Replace terms with values and ignore terms that do not depend on ***θ***

$$Q(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) = \frac{1}{2}\int p(\mathbf{x}|\mathbf{y},\boldsymbol{\theta}_{t-1}) \log(\det(\Psi^{-1})) - (\mathbf{y} - W\mathbf{x})^T\Psi^{-1}(\mathbf{y} - W\mathbf{x}))\, d\mathbf{x}.$$

- equation 2

E – Step Details

Objective :
Compute the posterior moments of the **latent variables x** given the **observations y**, under the current parameters W and Ψ.

Posterior mean :

$$E(\mathbf{x}|\mathbf{y}) \quad = \quad \mathbf{W}^T(\mathbf{W}\mathbf{W}^T + \Psi)^{-1}\mathbf{b}$$

Posterior covariance :

$$E(\mathbf{x}\mathbf{x}^T|\mathbf{x}) - E(\mathbf{x}|\mathbf{y})E(\mathbf{x}|\mathbf{y})^T \quad = \quad \mathbf{I} - \mathbf{W}^T(\mathbf{W}\mathbf{W}^T + \Psi)^{-1}W.$$

Use in EM :
These two quantities "sufficient statistics” that get plugged into the M-step update formulas for W and Ψ.

M - Step Details

- Differentiate **equation 2** (slide - 10) with respect to W and the individual elements of Ψ, and apply some linear algebra, to get update rules

$$
\begin{aligned}
\mathbf{W}_{new} &= \left(\mathbf{y}E(\mathbf{x}|\mathbf{y})^T\right)\left(E(\mathbf{x}\mathbf{x}^T|\mathbf{y})\right)^{-1}, \\
\Psi_{new} &= \frac{1}{N}\mathrm{diagonal}\left(\mathbf{x}\mathbf{x}^T - WE(\mathbf{x}|\mathbf{y})\mathbf{y}^T\right),
\end{aligned}
$$

Summary

- Factor analysis decomposes covariance into common factors + unique noise
- EM algorithm with MLE finds optimal $\mathbf{W}$ and $\mathbf{\Psi}$ iteratively
- Useful for dimensionality reduction with noise modeling