Guided Tour of Machine Learning in Finance

Week 4: Reinforcement Learning

4-1-1-Latent Variable models

Igor Halperin

NYU Tandon School of Engineering, 2017

Mixtures of Gaussians

A mixture of Gaussians:

$$p(y \mid \Theta) = \sum_{k-1}^{K} \pi_k \, p(y \mid \theta_k)$$

Here each component is a Gaussian with $\theta_k = (\mu_k, \Sigma_k)$ for its mean and variance. The component weights $0 \leq \pi_k \leq 1$ can be described by a **latent (hidden)** variable s such that $s = k$ if the data point was generated by component k. Then we can write it as

$$p(y \mid \Theta) = \sum_{k-1}^{K} P(s = k \mid \pi) \, p(y \mid s = k, \theta)$$

Estimation of model now reduces to estimation of parameters $\theta_k = (\mu_k, \Sigma_k)$, as well as inference of the hidden variable s.

This can be done using the **EM algorithm**!

There are other **Latent Variable Models** that can be estimated using the EM algorithms.

Factor analysis

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}, \; i = 1,\ldots,N$
Factor analysis seeks a decomposition of the observable signal as a weighted sum of hidden (latent) uncorrelated Gaussian variables x with zero means and unit variances, plus a N-dimensional white noise ε with a diagonal correlation matrix Ψ :

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon$$

Here x is a K-dimensional vector, and Λ is a **factor loading matrix** of size $N \times K$.
The vector of model parameters is $\theta = (\Lambda, \Psi)$

Factor analysis

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}, \; i = 1,\ldots,N$
Factor analysis seeks a decomposition of the observable signal as a weighted sum of hidden (latent) uncorrelated Gaussian variables x with zero means and unit variances, plus a N-dimensional white noise ε with a diagonal correlation matrix Ψ :

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon$$

Here x is a K-dimensional vector, and Λ is a **factor loading matrix** of size $N \times K$.
The vector of model parameters is $\theta = (\Lambda, \Psi)$
The probability of data can be obtained by integrating out the factors:

$$p(y \mid \Theta) = \int p(x \mid \theta) p(y \mid x, \theta) dx = N\left(0, \Lambda\Lambda^T + \Psi\right)$$

Here $N\left(\mu, \Sigma\right)$ is a multivariate Gaussian density.

Factor analysis

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}$, $i = 1,\ldots,N$
Factor analysis seeks a decomposition of the observable signal as a weighted sum of hidden (latent) uncorrelated Gaussian variables x with zero means and unit variances, plus a N-dimensional white noise ε with a diagonal correlation matrix Ψ :

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon$$

Here x is a K-dimensional vector, and Λ is a **factor loading matrix** of size $N \times K$.
The vector of model parameters is $\theta = (\Lambda, \Psi)$
The probability of data can be obtained by integrating out the factors:

$$p(y \mid \Theta) = \int p(x \mid \theta) p(y \mid x, \theta) dx = N\left(0, \Lambda\Lambda^T + \Psi\right)$$

Here $N\left(\mu, \Sigma\right)$ is a multivariate Gaussian density.
Properties of Factor Analysis:

- Reduces the number of parameters to estimate from $O\left(N^2\right)$ to $O\left(N(K+1)\right)$
- Can compute the posterior $p(x \mid y, \theta)$ to provide a low-dimensional representation of the data

Factor analysis

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}$, $i = 1,\dots,N$
Factor analysis seeks a decomposition of the observable signal as a weighted sum of hidden (latent) uncorrelated Gaussian variables x with zero means and unit variances, plus a N-dimensional white noise ε with a diagonal correlation matrix Ψ :

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon$$

Here x is a K-dimensional vector, and Λ is a **factor loading matrix** of size $N \times K$.
The vector of model parameters is $\theta = (\Lambda, \Psi)$

Non-uniqueness of Λ and x: take an orthogonal matrix U with $UU^T = I$. Then

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon = \Lambda U U^T x^{(t)} + \varepsilon = \left(\Lambda U\right)\left(U^T x^{(t)}\right) + \varepsilon = \hat{\Lambda}\hat{x}^{(t)} + \varepsilon$$

Factor analysis

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}, \; i = 1,\ldots,N$
Factor analysis seeks a decomposition of the observable signal as a weighted sum of hidden (latent) uncorrelated Gaussian variables x with zero means and unit variances, plus a N-dimensional white noise ε with a diagonal correlation matrix Ψ :

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon$$

Here x is a K-dimensional vector, and Λ is a **factor loading matrix** of size $N \times K$.
The vector of model parameters is $\theta = (\Lambda, \Psi)$

Non-uniqueness of Λ and x: take an orthogonal matrix U with $UU^T = I$. Then

$$y^{(t)} = \Lambda x^{(t)} + \varepsilon = \Lambda U U^T x^{(t)} + \varepsilon = \left(\Lambda U \right)\left(U^T x^{(t)} \right) + \varepsilon = \hat{\Lambda} \hat{x}^{(t)} + \varepsilon$$

Enforce uniqueness by fixing Λ to be an orthogonal matrix
Connection to the PCA: when $\Psi = \sigma^2 \mathrm{I}_\mathrm{N}$, Λ is made of the first K eigenvectors of covariance matrix of $y^{(t)}$, stored column-wise.

EM algorithm: E-step: estimate x with parameters from the last step. M-step: adjust parameters by maximize the low bound on the log-likelihood.

Probabilistic PCA

The data: T values of a N-dimensional vector $y = \left\{ y_i^{(t)} \right\}_{t=1}^{T}$, $i = 1,\ldots,N$
For Probabilistic PCA, the setting is the same as for Factor Analysis, except that now the correlation matrix Ψ of N-dimensional white noise ε is **isotropic**:

$$\Psi = \sigma^2 \mathrm{I}_{\mathrm{N}}$$

The MLE: find $\theta = (\Lambda, \Psi)$ for which the Gaussian model with covariance $\Lambda\Lambda^T + \Psi$ has the maximum likelihood.

The deterministic PCA is obtained in the limit $\sigma^2 \to 0$
Probabilistic PCA enables simulation of data, or impute missing values in data.

Control question

Select all correct answers

1. Probabilistic PCA is obtained by randomly re-ordering the eigenvectors in the regular PCA. This helps to mitigate the impact of outliers and noise in the data.
2. In Factor Analysis, if in addition to x and ε being Gaussian random variables, also all elements of the factor loading matrix Λ are Gaussian random variables, then the observed variable y is still Gaussian, but its variance is multiplied by an additional factor $e^{-\Lambda^2/2}$
3. Gaussian Mixtures, Factor Analysis, and Probabilistic PCA are all examples of Latent Variable Models, which can all be estimated using the EM algorithm.
4. All Latent Variables Models presented so far assume that data y is i.i.d. (independent and identically distributed).
5. Data in Finance is always i.i.d. as long as it is collected from at least four independent brokers.

Correct answers: 3, 4