PPCA Extraction + Prediction in validation set

Alejo Perez

February 2022

Here below my implementation will be outlined.

1 Factor analysis

Let us assume a gaussian vector x_i of dimension $K \times 1$, assume $x_i \times \mathcal{N}(\mu, \Sigma)$

$$x_i = \mu + Wz_i + \varepsilon_i$$

μ is dimension $K \times 1$, ε is dimension $K \times 1$ and $\varepsilon \sim \mathcal{N}(0, \Psi)$

$$x_i \mid z_i \sim \mathcal{N}(\mu + Wz_i, \Psi)$$

$$z_i \sim \mathcal{N}(0, I)$$

Then, integrating the latent variable z out we yield

$$x_i \sim \mathcal{N}(\mu, WW^T + \Psi)$$

2 PPCA

We add this constraint to the Factor Analysis Model to obtain the PPCA model

$$\Psi = \sigma^2 I$$

PPCA assumes that the data points are independent of each other given latent variables.

$$x_i \mid z_i \sim \mathcal{N}(\mu + Wz_i, \quad \sigma^2 I)$$

$$x_i \sim \mathcal{N}(\mu, WW^T + \sigma^2 I)$$

z_i is a latent variable of dimension $Q \times 1$, $Q \leq K$, and W is loading matrix of dimension $K \times Q$.

We assume prior $z_i \sim N(\mu_o, \Sigma_o)$, to simplify we can promote $z_i \sim N(0, I)$. Once we have estimated μ *and* W from the **ppca method** we are interested in characterising the predictive distribution.

The distribution of the latent variable given the observed data can be derived using Bayes' Theorem to give:

$$z_i \mid x_i \sim \mathcal{N}((WW^T + \sigma^2 I)^{-1} W^T (x_i - \mu), \quad \sigma^2 (WW^T + \sigma^2 I)^{-1})$$

3 BPCA

Following the approach of Shigeyuki Oba et al. (2003) [2], our BPCA promotes priors for $p(W, \mu, \sigma^2)$ over the parameters of the model. Then, effective dimensionality over the matrix W is intented to be obtained by means of the introduction of a hierarchical prior $p(W \mid \alpha)$ where $\alpha = \{\alpha_1, ...\alpha_q\}$. The maximum number of α hyperparameters is $q = d - 1$, where d is the number of features of the observed vector x_i, ie. Multipath Components. α_i controls the inverse variance of the corresponding W_i. The larger the α_i, the smaller the W_i, eventually switching it off and hence reducing latent space dimensionality.

Finally we can derive the posterior estimate of W, W_{MP}

$$ln\, p(W \mid) = L - \frac{1}{2} \sum_{i=1}^{d-1} \alpha_i \|w_i\|^2 + const$$

4 Dealing with missing data through inference on latent variables

We follow the approach in Jenelius & Koutsopoulos (2017) [1]. For each sample x_i ($[K \times 1]$) we divide the Multipath Components into $x_{observed}$ (x_{obs}) and $x_{missing}$ (x_{mis}). We distinguish between 2 vectors of indices that discern between the **observed** features and the **missing** ones (namely Idx_{obs}, Idx_{mis} with dimensions $[K_{obs} \times 1]$ and $[K_{mis} \times 1]$) these correspond to the frequency band with received pilot and the ones with no observed pilot respectively. Therefore, we define $x_{obs} = x[Idx_{obs}]$ and $x_{mis} = x[Idx_{mis}]$.

$$\mathbf{X_{obs}} = \begin{bmatrix} x_{Idx_{obs}}[1] \\ . \\ . \\ . \\ x_{Idx_{obs}}[K_{obs}] \end{bmatrix} \quad \mathbf{X_{mis}} = \begin{bmatrix} x_{Idx_{mis}}[1] \\ . \\ . \\ . \\ x_{Idx_{mis}}[K_{mis}] \end{bmatrix}$$

The loading matrix W can be similarly split into the $K_{obs} \times Q$ matrix W_{obs} and the $K_{mis} \times Q$ matrix W_{mis}, and μ into μ_{obs} and μ_{mis} (where Q is the number of PPCA extracted).

We use the properties of Multivariate Normal Distribution to obtain our estimates $x_{mis} \mid x_{obs} \sim \mathcal{N}(\hat{x}_{mis|obs}, \Sigma_{mis|obs})$, such that

$$\hat{x}_{mis|obs} = \mu_{mis} + W_{mis}(W_{obs}^T W_{obs} + \sigma^2 I)^{-1} W_{obs}^T (x_{obs} - \mu_{obs})$$

$$\Sigma_{mis|obs} = \sigma^2 W_{mis}(W_{obs}^T W_{obs} + \sigma^2 I)^{-1} W_{mis}^T + \sigma^2 I$$

The mean vector $\hat{\mu}_{mis}$ serves as point estimate for the missing Multipath Components, whereas $\Sigma_{mis|obs}$ describes the variability around those predictions. μ_{mis}, in turn, establish a baseline for the prediction to be added to the product of the updated loadings $W_{mis}(W_{obs}^T W_{obs} + \sigma^2 I)^{-1} W_{obs}^T$ times the difference vector between observed MP components and MP means $(x_{obs} - \mu_{obs})$

References

[1] Erik Jenelius and Haris N. Koutsopoulos. "Urban Network Travel Time Prediction Based on a Probabilistic Principal Component Analysis Model of Probe Data". In: *IEEE Transactions on Intelligent Transportation Systems* 19.2 (2018), pp. 436–445. DOI: 10.1109/TITS.2017.2703652.

[2] Shigeyuki Oba et al. "A Bayesian missing value estimation method for gene expression profile data". In: *Bioinformatics* 19.16 (Nov. 2003), pp. 2088–2096. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btg287. eprint: https://academic.oup.com/bioinformatics/article-pdf/19/16/2088/667802/btg287.pdf. URL: https://doi.org/10.1093/bioinformatics/btg287.