Principal component analysis

Questions and answers

Miguel Rodo

Questions and answers

- What does multivariate analysis help uncover?
  - Patterns and relationships within datasets containing multiple variables
- What is an example of multivariate analysis in meteorology?
  - Analyzing temperature, humidity, wind speed, and more to understand weather systems
- What are linear combinations used as a basis for multivariate analysis?
  - $\mathbf{x}_1\mathbf{u}_1 + \mathbf{x}_2\mathbf{u}_2 + ... + \mathbf{x}_p\mathbf{u}_p$
- What are some advantages of using linear combinations?
  - Simplicity, strong theoretical results, robustness, speed and interpretability
- What are two main types of multivariate analysis based on variables involved?
  - $\mathbf{X} \sim \mathbf{X}$: PCA, factor analysis, correspondence analysis
  - $\mathbf{Y} \sim \mathbf{X}$: Multiple/multivariate regression, canonical correlation analysis, discrimination and classification
- What does principal component analysis (PCA) allow you to do?
  - Represent data more compactly and interpret relationships between variables
- What are some example applications of PCA?
  - Market analysis, image compression, bioinformatics
- What is the formula for projecting vector $\mathbf{b}$ onto vector $\mathbf{a}$?
  - $\text{proj}_{\mathbf{a}}(\mathbf{b}) = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$
- If $\mathbf{a}$ is a unit vector, what is the projection of $\mathbf{b}$ onto $\mathbf{a}$?

  - $\text{proj}_{\mathbf{a}}(\mathbf{b}) = (\mathbf{a}^T\mathbf{b})\mathbf{a}$

- In a two-variable height and weight example, what does projecting onto the vector $[1/\sqrt{2}, 1/\sqrt{2}]$ represent?
  - A general "size" variable combining height and weight equally
- What do the original variable axes represent in terms of projections?
  - Height is the projection onto [1,0], weight is the projection onto [0,1]
- How can height and weight be represented using two orthogonal projection directions?
  - Projecting onto $[1/\sqrt{2}, 1/\sqrt{2}]$ and $[1/\sqrt{2}, -1/\sqrt{2}]$ gives an alternative coordinate system
- For the k-th principal component, what is the objective function being maximized?
  - $\text{Var}(\mathbf{a}_k'\mathbf{X}) = \mathbf{a}_k^T\mathbf{\Sigma}\mathbf{a}_k$ subject to $||\mathbf{a}_k|| = 1$ and $\mathbf{a}_k^T\mathbf{a}_j = 0$ for $j < k$
- What does Theorem 1 state about maximizing quadratic forms $\mathbf{x}'\mathbf{B}\mathbf{x}/\mathbf{x}'\mathbf{x}$ for points $\mathbf{x}$ on the unit sphere?
  - The maximum is attained at the largest eigenvalue λ_1 of $\mathbf{B}$ when $\mathbf{x}$ equals the corresponding eigenvector $\mathbf{e}_1$
- For a random vector $\mathbf{X}$ with covariance matrix $\mathbf{\Sigma}$, what is the i-th principal component Y_i according to Theorem 2?
  - $Y_i = \mathbf{e}_i'\mathbf{X}$ where $\mathbf{e}_i$ is the eigenvector of $\mathbf{\Sigma}$ corresponding to its i-th largest eigenvalue λ_i
- What is the variance of the i-th principal component Y_i?
  - $\text{Var}[Y_i] = \lambda_i$ where λ_i is the i-th largest eigenvalue of $\mathbf{\Sigma}$
- What is the sum of the variances of the original variables X_i equal to, according to Theorem 3?
  - The sum of the eigenvalues of the covariance matrix $\mathbf{\Sigma}$, which equals the sum of variances of the principal components Y_i
- What does a scree plot show?
  - The proportion of total variance accounted for by each principal component
- What is the formula for the correlation between the i-th principal component Y_i and the k-th original variable X_k?
  - $\rho_{Y_i,X_k} = \frac{\sqrt{\lambda_i}e_{ik}}{\sqrt{\sigma_{kk}}}$ where e_{ik} is the k-th element of the i-th eigenvector of $\mathbf{\Sigma}$
- How does standardizing variables to have unit variance before PCA change the results?

    - The principal components will be different linear combinations that don't depend on the original variable scales

- When estimating PCA from sample data, what is used to estimate the covariance matrix $\mathbf{\Sigma}$?
    - The sample covariance matrix $\mathbf{S} = n^{-1}\mathbf{X}'\mathbf{X}$ where $\mathbf{X}$ is mean-centered
- What are the two main uses of approximating a data matrix $\mathbf{X}$ with a lower rank matrix $\hat{\mathbf{X}}$?
    - Data compression and understanding key patterns while minimizing information loss
- What does the singular value decomposition (SVD) provide the best rank s approximation to a matrix $\mathbf{X}$ according to Theorem 5a?
    - $\hat{\mathbf{X}} = \mathbf{U}\mathbf{D}\mathbf{J}_s\mathbf{V}' = \sum_{i=1}^{s} d_i \mathbf{u}_i \mathbf{v}_i'$ where $d_i, \mathbf{u}_i, \mathbf{v}_i$ are the i-th singular value, left singular vector, right singular vector
- How is the PCA rank s approximation $\tilde{\mathbf{X}}$ related to the SVD approximation $\hat{\mathbf{X}}$?
    - They are equivalent, $\tilde{\mathbf{X}} = \mathbf{Y}\mathbf{J}_k\mathbf{P}' = \mathbf{U}\mathbf{D}\mathbf{J}_k\mathbf{V}' = \hat{\mathbf{X}}$
- What is a biplot?
    - A graphical display of both the observations (as points) and variables (as vectors) based on a low-rank PCA approximation to the data matrix

Johnson, Richard A., and Dean W. Wichern. 2007. Applied Multivariate Statistical Analysis. 6th ed. Upper Saddle River, NJ: Pearson Education Inc.