ETC3250: Dimension reduction

Semester 1, 2019

Professor Di Cook

Econometrics and Business Statistics
Monash University

Week 4 (b)

Outline

- LDA
  - Compare with PCA

Discriminant space: is the low-dimensional space where the class means are the furthest apart relative to the common variance-covariance.

The discriminant space is provided by the eigenvectors after making an eigen-decomposition of $\Sigma^{-1}\Sigma_B$, where

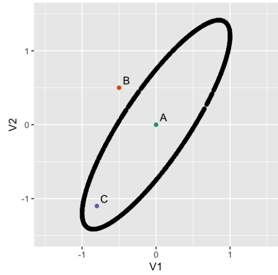
$$\Sigma_B = \frac{1}{K}\sum_{i=1}^{K}(\mu_i - \mu)(\mu_i - \mu)'$$

and

$$\Sigma = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i=1}^{n_k}(x_i - \mu_k)(x_i - \mu_k)'$$

Outline

- LDA
  - Compare with PCA
  - Mahalanobis distance



Which points are closest according to Euclidean distance?

Which points are closest relative to the variance-covariance?

Outline

- LDA
  - Compare with PCA
  - Mahalanobis distance
  - Discriminant space

Both means the same. Two different variance-covariance matrices. Discriminant space depends on the variance-covariance matrix.

Outline

- LDA
- PP vs PCA

Projection pursuit (PP) generalises PCA

PCA:

$$\underset{\phi_{11},\ldots,\phi_{p1}}{\text{maximize}} \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{p}\phi_{j1}x_{ij}\right)^2 \text{ subject to } \sum_{j=1}^{p}\phi_{j1}^2 = 1$$

PP:

$$\underset{\phi_{11},\ldots,\phi_{p1}}{\text{maximize}} f\left(\sum_{j=1}^{p}\phi_{j1}x_{ij}\right) \text{ subject to } \sum_{j=1}^{p}\phi_{j1}^2 = 1$$

Outline

- LDA
- PP vs PCA
- MDS

Multidimensional scaling (MDS) finds a low-dimensional layout of points that minimises the difference between distances computed in the p-dimensional space, and those computed in the low-dimensional space.

$$\text{Stress}_D(x_1,\ldots,x_N) = \left(\sum_{i,j=1;i\neq j}^{N}(d_{ij}-d_k(i,j))^2\right)^{1/2}$$

where D is an $N\times N$ matrix of distances (d_{ij}) between all pairs of points, and $d_k(i,j)$ is the distance between the points in the low-dimensional space.

Outline

- LDA
- PP vs PCA
- MDS

- Classical MDS is the same as PCA
- Metric MDS incorporates power transformations on the distances, d_{ij}^r.
- Non-metric MDS incorporates a monotonic transformation of the distances, e.g. rank

Outline

- LDA
- PP vs PCA
- MDS
- nonlinear

- T-distributed Stochastic Neighbor Embedding (t-SNE): similar to MDS, except emphasis is placed on grouping observations into clusters. Observations within a cluster are placed close in the low-dimensional representation, but clusters themselves are placed far apart.

- Local linear embedding (LLE): Finds nearest neighbours of points, defines interpoint distances relative to neighbours, and preserves these proximities in the low-dimensional mapping. Optimisation is used to solve an eigen-decomposition of the knn distance construction.

- Self-organising maps (SOM): First clusters the observations into $k \times k$ groups. Uses the mean of each group laid out in a constrained 2D grid to create a 2D projection.

Outline

- LDA
- PP vs PCA
- MDS
- nonlinear

- **T-distributed Stochastic Neighbor Embedding (t-SNE)**: similar to MDS, except emphasis is placed on grouping observations into clusters. Observations within a cluster are placed close in the low-dimensional representation, but clusters themselves are placed far apart.

- **Local linear embedding (LLE)**: Finds nearest neighbours of points, defines interpoint distances relative to neighbours, and preserves these proximities in the low-dimensional mapping. Optimisation is used to solve an eigen-decomposition of the knn distance construction.

- **Self-organising maps (SOM)**: First clusters the observations into $k \times k$ groups. Uses the mean of each group laid out in a constrained 2D grid to create a 2D projection.

Outline

- LDA
- PP vs PCA
- MDS
- nonlinear

- **T-distributed Stochastic Neighbor Embedding (t-SNE)**: similar to MDS, except emphasis is placed on grouping observations into clusters. Observations within a cluster are placed close in the low-dimensional representation, but clusters themselves are placed far apart.

- **Local linear embedding (LLE)**: Finds nearest neighbours of points, defines interpoint distances relative to neighbours, and preserves these proximities in the low-dimensional mapping. Optimisation is used to solve an eigen-decomposition of the knn distance construction.

- **Self-organising maps (SOM)**: First clusters the observations into $k \times k$ groups. Uses the mean of each group laid out in a constrained 2D grid to create a 2D projection.

👩‍💻 Made by a human with a computer

Slides at https://monba.dicook.org.

Code and data at https://github.com/dicook/Business_Analytics.

Created using R Markdown with flair by **xaringan**, and **kunoichi** (female ninja) style.