Lecture 9

Independent Component Analysis (ICA)

$\underset{\sim}{X}$ mean vector $\underset{\sim}{\mu}$, covariance matrix C

Model: $\underset{\sim}{X} = \underset{\sim}{\mu} + A\underset{\sim}{Y}$ where $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix}$ with $y_1, \cdots, y_p$ independent $E(y_i)=0$, $Var(y_i)=1$

$C = AA^T$

Assumption: At most one of $y_1, \cdots, y_p$ are normal

Given data $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_n$, estimate mixing matrix A and independent components $\underset{\sim}{Y}_1, \cdots, \underset{\sim}{Y}_n$.

Fast ICA

① Rewhitening: Transform center data $\underset{\sim}{X}_1 - \overline{\underset{\sim}{X}}, \cdots, \underset{\sim}{X}_n - \overline{\underset{\sim}{X}}$ to make variables uncorrelated.

$$\underset{\sim}{Z}_i = \underset{p\times p}{L}(\underset{\sim}{X}_i - \overline{\underset{\sim}{X}})$$

$$\frac{1}{n-1}\sum_{i=1}^n \underset{\sim}{Z}_i \underset{\sim}{Z}_i^T = I$$

- use principle components for L.

$$\hat{C} = \frac{1}{n-1}\sum_{i=1}^n (\underset{\sim}{X}_i - \overline{\underset{\sim}{X}})(\underset{\sim}{X}_i - \overline{\underset{\sim}{X}})^T$$
$$= V\Lambda V^T$$

estimated covariance matrix

$$L = \Lambda^{-\frac{1}{2}} V^T$$

② Try to find projections of $\{\underset{\sim}{Z}_i\}$ that are as normal as possible, vector $\underset{\sim}{w}$ with $\underset{\sim}{w}^T\underset{\sim}{w}=1$

$\Rightarrow$ maximize non-normality of $\{\underset{\sim}{w}^T\underset{\sim}{Z}_i\}$

Find $\underset{\sim}{w}_1, \cdots, \underset{\sim}{w}_p$ with $\underset{\sim}{w}_i^T\underset{\sim}{w}_j = \begin{cases} 1 & \text{if } i \neq j \\ 0 & \text{ow} \end{cases}$ with $\{\underset{\sim}{w}_i^T\underset{\sim}{Z}_i\}$ as non-normal as possible

Two methods:

Find $\underset{\sim}{w}_1, \cdots, \underset{\sim}{w}_p$ in one shot

$\Rightarrow$ simutaneous

Sequentials. Find $\underset{\sim}{w}_1$ then $\underset{\sim}{w}_2 \Rightarrow$ deflations.

Fast ICA-deflation

Prewhitened data $\{\underset{\sim}{Z}_j\}$. Find $\underset{\sim}{w}$ to minimize $J(\underset{\sim}{w}) = \frac{1}{n}\sum_{j=1}^n H(\underset{\sim}{w}^T\underset{\sim}{Z}_i)$ over $\underset{\sim}{w}$ with $\underset{\sim}{w}^T\underset{\sim}{w}=1$

Add Lagrange multiplier:

Minimize $J(\underset{\sim}{w}) + \lambda[\underset{\sim}{w}^T\underset{\sim}{w} - 1] = J^*(\underset{\sim}{w})$

$$\nabla_{\underset{\sim}{w}} J^*(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^n H(\underset{\sim}{w}^T\underset{\sim}{Z}_i)\underset{\sim}{Z}_i + 2\lambda\underset{\sim}{w} \quad (= \underset{\sim}{0})$$

Newton-Raphson algorithm

One dimension: $g(x) = 0$

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

Need matrix of derivatives:

$$\nabla^2_{\underset{\sim}{w}} J^*(\underset{\sim}{w}) = \frac{1}{n}\sum_{i=1}^n H''(\underset{\sim}{w}^T \underset{\sim}{z}_i)\, \underset{\sim}{z}_i \underset{\sim}{z}_i^T + 2\lambda I$$

$$\nabla^2_{\underset{\sim}{w}} J^*(\underset{\sim}{w}) = [\underbrace{\frac{1}{n}\sum_{i=1}^n H''(\underset{\sim}{w}^T \underset{\sim}{z}_i) + 2\lambda}_{\text{scalar}}] I$$

Modified N-R update

$$\underset{\sim}{w}_{k+1} = \underset{\sim}{w}_k - \frac{\frac{1}{n}\sum_{i=1}^n H'(\underset{\sim}{w}_k^T \underset{\sim}{z}_i)\underset{\sim}{z}_i + 2\lambda \underset{\sim}{w}_k}{\frac{1}{n}\sum_{i=1}^n H''(\underset{\sim}{w}_k^T \underset{\sim}{z}_i) + 2\lambda}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^n H''(\underset{\sim}{w}_k^T \underset{\sim}{z}_i)\underset{\sim}{w}_k - \frac{1}{n}\sum_{i=1}^n H'(\underset{\sim}{w}_k^T \underset{\sim}{z}_i)\underset{\sim}{z}_i}{\frac{1}{n}\sum_{i=1}^n H''(\underset{\sim}{w}_k^T \underset{\sim}{z}_i)\underset{\sim}{w}_k + 2\lambda} \Longrightarrow = \underset{\sim}{v}_k$$

$\|w_{k+1}\| = 1$

$$\underset{\sim}{w}_{k+1} = \frac{\underset{\sim}{v}_k}{\|\underset{\sim}{v}_k\|} = \frac{\underset{\sim}{v}_k}{\sqrt{\underset{\sim}{v}_k^T \underset{\sim}{v}_k}}$$

Now iterate until convergence.

How to compute other $\underset{\sim}{w}$'s? Add orthogonality conditions $\longrightarrow$ extra Lagrange multipliers

<u>Examples</u> Track records. (see Blackboard)

8 variables (national records in 8 events)

55 countries

- R package: fast ICA.

Use deflation method using standardized variables

<u>PCA</u>: First 2 PCs very intepretable.

<u>ICA</u>: Mixing matrix quite incomprehensible!

- outliers are much more apparent in ICA.
- for example, ICA identifies South Korea (KOR) as a clear outlier.

<u>Summary</u>: PCA vs ICA

- similar goals, different results

<u>PCA</u>: Maximize variance of one-dimensional projection subject to orthogonality covariantes.

$$\left.\begin{matrix}\hat{C}\\ \hat{R}\end{matrix}\right\} = V\Lambda V^T$$

- scaling of variables is important - when in doubt, use $\hat{R}$!

<u>ICA</u>: $\underset{\sim}{x} = \underset{\sim}{\mu} + A\underset{\sim}{y}$ $\leftarrow$ indp comp

$\uparrow$ mixing matrix

$$\mathrm{Cov}(\underset{\sim}{x}) = C = AA^T$$

- in theory, scaling of variables doesn't matter.
  - defining criteria for independence.
  - components are interchangeable
  - scaling can matter!

Implementation of ICA is very complicated.