ORIE 779: Functional Data Analysis

From last meeting

Review of Linear Algebra
- norms, inner products, orthonormal bases & projections
- singular value and eigen decompositions

Review of Multivariate Probability
- theoretical and empirical mean vectors
- theoretical and empirical covariance matrices

Mathematics behind PCA
- "Rotate data" using eigen-decomp. of covariance matrix
- Then optimization problem(s) are simple

PCA dual problem

Idea: Recall for HDLSS settings:

Sample size = n < d = dimension

So $rank(\hat{\Sigma}) \leq n$, and $\lambda_{n+1} = \lambda_d = 0$

Thus have "really only n dimensional eigen problem"

Can exploit this to boost computation speed

Again use notation: $\tilde{X} = \frac{1}{\sqrt{n-1}} \left(\underline{X}_1 - \underline{\overline{X}} \quad \cdots \quad \underline{X}_n - \underline{\overline{X}} \right)_{d \times n}$

PCA dual problem (cont.)

Recall: $\hat{\Sigma}_{d\times d} = \widetilde{X}\widetilde{X}^t$ has the eigenvalue decomp. $\hat{\Sigma} = BDB^t$

Study via Singular Value Decomposition of $\widetilde{X}$:

$$\widetilde{X} = USV^t, \quad \text{where} \quad U^tU = V^tV = I$$

giving:

$$\hat{\Sigma} = \widetilde{X}\widetilde{X}^t = \left(USV^t\right)\left(USV^t\right)^t = USV^tVS^tU^t = USS^tU^t$$

By uniqueness of eigen-analysis, have (except for order):

$$B = U \qquad\qquad D = SS^t$$

PCA dual problem (cont.)

For $n < d$:

$$S = \begin{pmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{d\times n} \quad \text{, so} \quad D = SS^t = \begin{pmatrix} s_1^2 & 0 & & \cdots & & 0 \\ 0 & \ddots & & & & \\ & & s_n^2 & \ddots & & \vdots \\ \vdots & & \ddots & 0 & & \\ & & & & \ddots & 0 \\ 0 & & \cdots & & 0 & 0 \end{pmatrix}_{d\times d}$$

PCA dual problem (cont.)

Thus:

- could do SVD of $\tilde{X}$, to compute Eigen-analysis
- i.e. replace $d \times d$ analysis by $d \times n$
- Singular Values are $\pm\sqrt{\ }$ of cov. matrix eigenvalues
- (usually taken as + square-root)
- Columns of U can be used for PCA projections
- since they *are* the eigenvectors (i.e. $B = U$)
- so PCA is *both*:
  - eigen-decomposition of covariance matrix
  - singular value decomposition of data matrix

PCA dual problem (cont.)

Improve to $n \times n$ analysis?

Can make U and V "change places" by considering $\widetilde{X}^t$

Singular Value Decomposition is:

$$\widetilde{X}^t = \left(USV^t\right)^t = VS^tU^t$$

Sizes are useful:

$$n \times d = n \times n \leftrightarrow n \times d \leftrightarrow d \times d$$

Return to an eigen representation as:

$$\widetilde{X}^t \widetilde{X} = VS^tU^t\left(USV^t\right) = VS^tSV^t$$

PCA dual problem (cont.)

Again for $n < d$:

$$S = \begin{pmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_n \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{d \times n}$$

, so define $D^* = S^t S = \begin{pmatrix} s_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_n^2 \end{pmatrix}_{n \times n}$

and define $B^* = V$, to give

$$\widetilde{X}^t \widetilde{X} = B^* D^* \left(B^*\right)^t$$

the "dual" eigen representation

PCA dual problem (cont.)

Have "dual eigen problem", where

- Treat $\widetilde{X}^t$ as "data matrix"
- i.e. rows of data matrix $\widetilde{X}$ are replaced by columns
- Based on "dual covariance matrix" $\hat{\Sigma}^*_{n \times n} = \widetilde{X}^t \widetilde{X}$
- "inner product of data matrix"
- compared to "outer product" for calculation of $\hat{\Sigma}_{d \times d} = \widetilde{X}\widetilde{X}^t$
- for $n < d$, have faster $n \times n$ eigen-calculation

PCA dual problem (cont.)

Now suppose know sol'n to dual problem, i.e. know B^* and D^*

How do we find B and D?

{Next time: this can be more cleanly done using SVD:
$X^t = VS^tU^t \Rightarrow X^tU = VS^t$, then use structure of S to "invert"}

A heuristic approach:

i. want B so that

$$D = B^t \hat{\Sigma} B = B^t \widetilde{X} \widetilde{X}^t B$$

PCA dual problem (cont.)

ii. choose B to introduce form $\tilde{X}^t \tilde{X} = \Sigma^*$,
i.e. $B = \tilde{X} C$ (for some C), then

$$D = C^t \tilde{X}^t \left(\tilde{X} \tilde{X}^t \right) \tilde{X} C = C^t \left(\tilde{X}^t \tilde{X} \right) \left(\tilde{X}^t \tilde{X} \right) C = C^t \Sigma^* \Sigma^* C$$

iii. choose C to relate to $\Sigma^* = B^* D^* B^{*t}$, i.e. $B^{*t} \Sigma^* B^* = D^*$
i.e. $C = B^* R$ (for some R), then

$$D = C^t \Sigma^* \left(B^* B^{*t} \right) \Sigma^* C = \left(R^t B^{*t} \right) \Sigma^* B^* B^{*t} \Sigma^* \left(B^* R \right)$$

$$D = R^t \left(B^{*t} \Sigma^* B^* \right) \left(B^{*t} \Sigma^* B^* \right) R = R^t D^* D^* R$$

PCA dual problem (cont.)

iv. Choose R to "preserve energy",

i.e. "make B orthonormal",

i.e. "make B a rotation matrix",

i.e. choose $R = \left(D^*\right)^{-1/2}$, then

$D = D^*$, i.e. same (nonzero) eigenvalues!

PCA dual problem (cont.)

Heuristic summary: Want $B = \widetilde{X}C = \widetilde{X}\left(B^{*}R\right) = \widetilde{X}B^{*}\left(D^{*}\right)^{-1/2}$

Technical problems:

- dimensions wrong: $B_{d \times d}$, $\widetilde{X}_{d \times n}$, $B^{*}_{n \times n}$, $D^{*}_{n \times n}$
- $D^{*}_{n \times n}$ not full rank?, thus:
    - root inverse doesn't exist?
    - B is not a basis matrix

PCA dual problem (cont.)

Solution: Assume $D^* = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ is of full rank,

i.e. $\lambda_1 \geq \cdots \geq \lambda_n > 0$

Then let $\breve{B}_{d \times n} = \widetilde{X} B^* \left(D^* \right)^{-1/2}$,

Where

$$\left(D^* \right)^{-1/2} = \begin{pmatrix} \lambda_1^{-1/2} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{-1/2} \end{pmatrix}$$

PCA dual problem (cont.)

And “fill out the rest of B” with “columns in null space”,

I.e. let $G_{d\times(d-n)}$ be $d-n$ orthonormal column vectors,

that are orthogonal to $\widetilde{X}$ (compute by Gram-Schmidt process)

Thus “pad $\breve{B}$ out to a basis matrix”, by defining:

$$B = \begin{pmatrix} \breve{B} & G \end{pmatrix}$$

PCA dual problem (cont.)

Check orthonormality:

$$B^t B = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \begin{pmatrix} \breve{B} & G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \breve{B} & 0 \\ 0 & I \end{pmatrix}$$

but

$$\breve{B}^t \breve{B} = \left(\widetilde{X} B^* \left(D^*\right)^{-1/2}\right)^t \left(\widetilde{X} B^* \left(D^*\right)^{-1/2}\right) = \left(\left(D^*\right)^{-1/2} B^{*t} \widetilde{X}^t\right)\left(\widetilde{X} B^* \left(D^*\right)^{-1/2}\right)$$
$$\breve{B}^t \breve{B} = \left(D^*\right)^{-1/2} B^{*t} \left(\widetilde{X}^t \widetilde{X}\right) B^* \left(D^*\right)^{-1/2} = \left(\left(D^*\right)^{-1/2} B^{*t}\right) \Sigma^* \left(B^* \left(D^*\right)^{-1/2}\right)$$
$$\breve{B}^t \breve{B} = \left(D^*\right)^{-1/2} \left(B^{*t} \Sigma^* B^*\right) \left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} D^* \left(D^*\right)^{-1/2} = I$$

so B is orthonormal.

PCA dual problem (cont.)

Check diagonalization:

$$B^t \hat{\Sigma} B = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \hat{\Sigma} \begin{pmatrix} \breve{B} & G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \\ G^t \end{pmatrix} \begin{pmatrix} \hat{\Sigma}\breve{B} & \hat{\Sigma} G \end{pmatrix} = \begin{pmatrix} \breve{B}^t \hat{\Sigma} \breve{B} & \breve{B}^t \hat{\Sigma} G \\ G^t \hat{\Sigma} \breve{B} & G^t \hat{\Sigma} G \end{pmatrix}$$

but

$$\breve{B}^t \hat{\Sigma} \breve{B} = \left(D^*\right)^{-1/2} B^{*t} \widetilde{X}^t \hat{\Sigma} \widetilde{X} B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} B^{*t} \widetilde{X}^t \widetilde{X} \widetilde{X}^t \widetilde{X} B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} B^{*t} \Sigma^* \Sigma^* B^* \left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} B^{*t} \Sigma^* \left(B^* B^{*t}\right) \Sigma^* B^* \left(D^*\right)^{-1/2} =$$
$$= \left(D^*\right)^{-1/2} \left(B^{*t} \Sigma^* B^*\right)\left(B^{*t} \Sigma^* B^*\right)\left(D^*\right)^{-1/2} = \left(D^*\right)^{-1/2} D^* D^* \left(D^*\right)^{-1/2} = D^*$$

PCA dual problem (cont.)

And using the orthogonality of the columns of $\widetilde{X}$ and G

$$\breve{B}^t \hat{\Sigma} G = \breve{B}^t \left(\widetilde{X} \widetilde{X}^t \right) G = \left(\breve{B}^t \widetilde{X} \right) \left(\widetilde{X}^t G \right) = \left(\breve{B}^t \widetilde{X} \right) 0 = 0_{n \times (d-n)}$$

$$G^t \hat{\Sigma} \breve{B} = G^t \left(\widetilde{X} \widetilde{X}^t \right) \breve{B} = \left(G^t \widetilde{X} \right) \left(\widetilde{X}^t \breve{B} \right) = 0 \left(\widetilde{X}^t \breve{B} \right) = 0_{(d-n) \times n}$$

$$G^t \hat{\Sigma} G = G^t \left(\widetilde{X} \widetilde{X}^t \right) G = \left(G^t \widetilde{X} \right) \left(\widetilde{X}^t G \right) = 0 \cdot 0 = 0_{(d-n) \times (d-n)}$$

PCA dual problem (cont.)

Thus:

$$B^t \hat{\Sigma} B = \begin{pmatrix} D^* & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & & \cdots & & 0 \\ 0 & \ddots & & & & \\ & & \lambda_n & \ddots & & \vdots \\ \vdots & & \ddots & 0 & & \\ & & & & \ddots & 0 \\ 0 & & \cdots & & 0 & 0 \end{pmatrix} = D$$

PCA dual problem (cont.)

Aside about orthogonal component G:

Usually don't need to compute,

Since only care about "eigenvectors for non-zero eigenvalues"

Statistics of PCA

Above "optimization of directions" approach to PCA:

- gives useful insights
- shows can compute for *any* point cloud

But there are other views.

Statistics of PCA (cont.)

Alternate View 1: Gaussian likelihood

When data are multivariate Gaussian

PCA finds "major axes of elliptical contours"

of Probability density (maximum likelihood estimate)

Mistaken idea: PCA only useful for Gaussian data

Statistics of PCA (cont.)

Simple check for Gaussian distribution:

Standardized parallel coordinate plot

1. Subtract coordinate wise median (robust version of mean)

   (not good as "point cloud center", but now only looking at coordinates)

2. Divide by MAD / MAD($N(0,1)$)

   (put on same scale as "standard deviation")

3. See if data stays in range –3 to +3

Statistics of PCA (cont.)

Check for Gaussian dist'n: Standardized parallel coordinate plot

E.g. Cornea data (recall image view of data)

- several data points > 20 "s.d.s" from the center
- distribution clearly *not* Gaussian
- strong kurtosis
- but PCA still gave strong insights

Statistics of PCA (cont.)

Alternate View 2: Dimension reduction

An approach to HDLSS data: try to reduce dimensionality

PCA approach:

- keep only largest eigenvalue projections
- optimal reduction (in sense of Sums of Squares)

Statistics of PCA (cont.)

Alternate View 3: Data compression (e.g. PKzip)

Loss-less: delete components with 0 eigenvalues

With loss: PCA gives optimal compression

(in sense of Sums of Squares)

PCA for shapes

New Data Set: Corpus Callosum data

- "window" between right and left halves of the brain
- from a vertical slice MR image of head
- "segmented" (ie. found boundary)
- shape is resulting closed curve
- have sample from $n = 71$ people
- Feature vector of $d = 80$ coefficients from

  Fourier boundary representation (closed curve)

PCA for shapes (cont.)

Modes of shape variation?

To do later (???):

1. PCA for Corpus Callosum Data Fourier & M-Rep
2. PCA variations (correlation matrix)
3. PCA and Clusters – Mass Flux Data – SiZer
4. Revisit Paul's Toy M-rep examples (from cluster viewpoint)
5. PCA time series – chemometrics data
6. Independent Component Analysis
7. In vector space, orthogonal basis introduction
8. Fourier basis
9. Legendre basis
10. Tensor product Fourier Legendre basis
11. Zernike basis
12. Revisit cornea data?   (compare "raw image" with "fit images", fiddle with Cornean power map? (do this at home?), use Figure from LMTZ paper, see directories D:\DellInspiron7000\SW30\Docs\Steve and D:\DellInspiron7000\SW30\Pictures)
13. Elliptical Fourier bases

14. Complex plane representation (no simple real valued basis)
15. Corpora Collosa Approximation
16. Discrimination – Corpus Collosum Data
17. Fisher Linear Discrimination
18. High dimensional geometry?
19. Support Vector Machines
20. Polynomial Embedding
21. Micro-Array Data analysis
22. Normal KerCli discrimination (in Cornean/demo)