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 $rank\left(\hat{\Sigma}\right) \leq n$, 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: $\tilde{X} = \frac{1}{\sqrt{n-1}} \left(\underline{X}_1 - \overline{\underline{X}} \quad \cdots \quad \underline{X}_n - \overline{\underline{X}} \right)_{d \times n}$

PCA dual problem (cont.)

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

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

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

giving:

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

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

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

PCA dual problem (cont.)

For $n < d$:

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

[go to next page]