ST758, Homework 5

Due Oct 21, 2014

1. (Maggie's question) Assume $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ has full column rank. Gram-Schmidt or modified Gram-Schmidt algorithm yields $\boldsymbol{X} = \boldsymbol{Q}_1 \boldsymbol{R}_1$ and Household algorithm (without pivoting) yields $\boldsymbol{X} = \boldsymbol{Q}_2 \boldsymbol{R}_2$, where $\boldsymbol{Q}_1, \boldsymbol{Q}_2 \in \mathbb{R}^{n \times p}$, $\boldsymbol{Q}_1^T \boldsymbol{Q}_1 = \boldsymbol{Q}_2^T \boldsymbol{Q}_2 = \boldsymbol{I}_p$, and $\boldsymbol{R}_1, \boldsymbol{R}_2 \in \mathbb{R}^{p \times p}$ are upper triangular with positive diagonal entries. Show that $\boldsymbol{Q}_1 = \boldsymbol{Q}_2$ and $\boldsymbol{R}_1 = \boldsymbol{R}_2$.

2. (Ridge regression revisited) In ridge regression, we minimize

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2,$$

where $\lambda \geq 0$ is a tuning parameter.

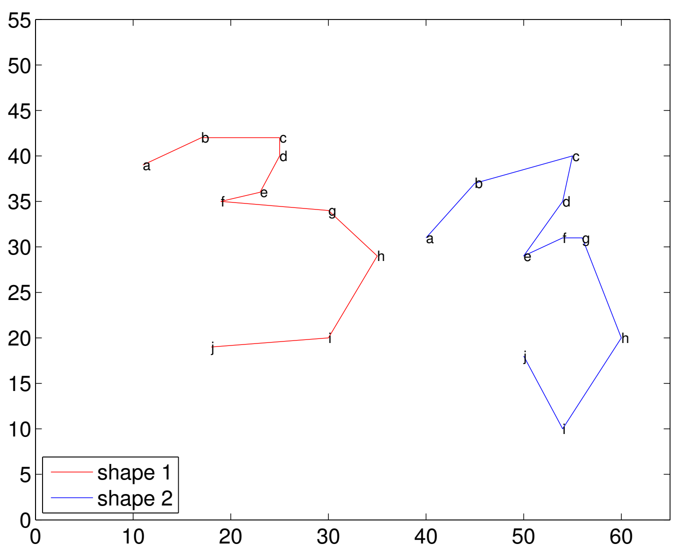
   (a) Express ridge solution $\widehat{\boldsymbol{\beta}}(\lambda)$ in terms of the singular value decomposition (SVD) of $\boldsymbol{X}$.

   (b) Show that (i) the ℓ_2 norms of ridge solution $\|\widehat{\boldsymbol{\beta}}(\lambda)\|_2$ and corresponding fitted values $\|\hat{\boldsymbol{y}}(\lambda)\|_2 = \|\boldsymbol{X}\hat{\boldsymbol{\beta}}(\lambda)\|_2$ are non-increasing in λ and (ii) the ℓ_2 norm of the residual vector $\|\boldsymbol{y} - \hat{\boldsymbol{y}}(\lambda)\|_2$ is non-decreasing in λ.

   (c) Re-compute and plot the ridge solution for the Longley data in HW4 at $\lambda = 5, 10, 15, 20, \ldots, 100$ using the SVD approach.

   (d) Comment on the computation efficiency of SVD approach compared to the approach you used in HW4.

3. (Matching images) Below figure displays two 3s my son wrote on a piece of paper and I want to properly align them.



Let matrices $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{n \times p}$ record n points on the two shapes. In this case $n = 10$ and $p = 2$. Mathematically we consider the problem

$$\text{minimize}_{\beta, \boldsymbol{O}, \boldsymbol{\mu}} \quad \|\boldsymbol{X} - (\beta \boldsymbol{Y}\boldsymbol{O} + \mathbf{1}_n \boldsymbol{\mu}^T)\|_{\text{F}}^2,$$

where $\beta > 0$ is a scaling factor, $\boldsymbol{O} \in \mathbb{R}^{p \times p}$ is an orthogonal matrix, and $\boldsymbol{\mu} \in \mathbb{R}^p$ is a vector of shifts. Here $\|\boldsymbol{M}\|_{\mathrm{F}}^2 = \sum_{i,j} m_{ij}^2$ is the squared Frobenius norm. Intuitively we want to rotate, stretch, and shift the shape $\boldsymbol{Y}$ to match the shape $\boldsymbol{X}$ as much as possible.

(a) Let $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ be the column mean vectors of the matrices and $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Y}}$ be the versions of these matrices with means removed. Show that the solution $(\hat{\beta}, \widehat{\boldsymbol{O}}, \hat{\boldsymbol{\mu}})$ satisfies

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} - \hat{\beta}\widehat{\boldsymbol{O}}^T \bar{\boldsymbol{y}}.$$

Therefore we can center each matrix at its column centroid and then ignore the location completely.

(b) Derive the solution to

$$\text{minimize}_{\beta, \boldsymbol{O}} \quad \|\tilde{\boldsymbol{X}} - \beta \tilde{\boldsymbol{Y}} \boldsymbol{O}\|_{\mathrm{F}}^2$$

using the SVD of $\tilde{\boldsymbol{Y}}^T \tilde{\boldsymbol{X}}$.

(c) Implement your method and solve the alignment problem in the figure. Display your solution together with the original two 3s.