Biostat 216 Homework 7

Due Dec 3 @ 11:59pm

Submit a PDF (scanned/photographed from handwritten solutions, or converted from RMarkdown or Jupyter Notebook) to Gracescope on BruinLearn.

Positive definite matrices

Q1

Suppose $\mathbf{C}$ is positive definite and $\mathbf{A}$ has independent columns. Apply the energy test to show that $\mathbf{A}'\mathbf{C}\mathbf{A}$ is positive definite.

Q2

Show that the diagonal entries of a positive definite matrix are positive.

Q3 (Rayleigh quotient)

Suppose $\mathbf{S}$ is positive definite with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ with corresponding orthonormal eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

1. What are the eigenvalues of the matrix $\lambda_1 \mathbf{I} - \mathbf{S}$? Is it positive semidefinite?

2. How does it follow that $\lambda_1 \mathbf{x}'\mathbf{x} \geq \mathbf{x}'\mathbf{S}\mathbf{x}$ for every $\mathbf{x}$?

3. Draw the conclusion: The maximum value of the Rayleigh quotient

$$R(\mathbf{x}) = \frac{\mathbf{x}'\mathbf{S}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$$

   is λ_1.

4. Show that the maximum value of the Rayleigh quotient subject to the condition $\mathbf{x} \perp \mathbf{u}_1$ is λ_2. Hint: expand $\mathbf{x}$ in eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

5. Show that the maximum value of the Rayleigh quotient subject to the conditions $\mathbf{x} \perp \mathbf{u}_1$ and $\mathbf{x} \perp \mathbf{u}_2$ is λ_3.

6. What is the maximum value of $\frac{1}{2}\mathbf{x}'\mathbf{S}\mathbf{x}$ subject to the constraint $\mathbf{x}'\mathbf{x} = 1$. Hint: write down the Lagrangian and set its derivative to zero.

SVD

Q4

Find the closest rank-1 approximation (in Frobenius norm or spectral norm) to these matrices

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Q5 (Moore-Penrose inverse)

1. With singular value decomposition $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$, verify that

$$\mathbf{X}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}' = \mathbf{V}_r\mathbf{\Sigma}_r^{-1}\mathbf{U}_r' = \sum_{i=1}^r \sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i',$$

   where $\mathbf{\Sigma}^+ = \text{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)$ and $r = \text{rank}(\mathbf{X})$, satifies the four properties of the Moore-Penrose inverse.

2. Show that $\text{rank}(\mathbf{X}^+) = \text{rank}(\mathbf{X})$.

3. Show that $\mathbf{X}^+\mathbf{X}$ is the orthogonal projector into $\mathcal{C}(\mathbf{X}')$ and $\mathbf{X}\mathbf{X}^+$ is the orthogonal projector into $\mathcal{C}(\mathbf{X})$.

4. Show that $\boldsymbol{\beta}^+ = \mathbf{X}^+\mathbf{y}$ is a minimizer of the least squares criterion $f(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$. Hint: check $\boldsymbol{\beta}^+$ satisfies the normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$.

5. Show that $\boldsymbol{\beta}^+ \in \mathcal{C}(\mathbf{X}')$.

6. Show that if another $\boldsymbol{\beta}^\star$ minimizes $f(\boldsymbol{\beta})$, then $\|\boldsymbol{\beta}^\star\| \geq \|\boldsymbol{\beta}^+\|$. This says that $\boldsymbol{\beta}^+ = \mathbf{X}^+\mathbf{y}$ is the least squares solution with smallest ℓ_2 norm. Hint: since both $\boldsymbol{\beta}^\star$ and $\boldsymbol{\beta}^+$ satisfy the normal equation, $\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^\star = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}^+$ and deduce that $\boldsymbol{\beta}^\star - \boldsymbol{\beta}^+ \in \mathcal{N}(\mathbf{X})$.

Q6

Let $\mathbf{B}$ be a submatrix of $\mathbf{A} \in \mathbb{R}^{m \times n}$. Show that the largest singular value of $\mathbf{B}$ is always less than or equal to the largest singular value of $\mathbf{A}$.

Q7

Show that all three matrix norms (ℓ_2, Frobenius, nuclear) are invariant under orthogonal transforms. That is

$$\|\mathbf{Q}_1\mathbf{A}\mathbf{Q}_2'\| = \|\mathbf{A}\| \text{ for orthogonal } \mathbf{Q}_1 \text{ and } \mathbf{Q}_2.$$

Optimization and multivariate calculus

Q8

1. Explain why the intersection $K_1 \cap K_2$ of two convex sets is a convex set.
2. Prove that the maximum F_3 of two convex functions F_1 and F_2 is a convex function. Hint: What is the set above the graph of F_3?

Q9

Show that these functions are convex:

1. Entropy $x \log x$.
2. Log-sum-exp: $\log(e^x + e^y)$.
3. ℓ_p norm: $\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p)^{1/p}, p \geq 1$.
4. Largest eigenvalue: $\lambda_{\max}(\mathbf{S})$ as a function of the symmetric matrix $\mathbf{S}$. Hint: Q3.6 and Q8.2.

Q10

Minimize $f(x_1, x_2) = \frac{1}{2}\mathbf{x}'\mathbf{S}\mathbf{x} = \frac{1}{2}x_1^2 + 2x_2^2$ subject to the constraint $\mathbf{A}'\mathbf{x} = x_1 + 3x_2 = b$.

1. What is the Lagrangian $L(\mathbf{x}, \lambda)$ for this problem.
2. What are the three equations “derivative of L=zero”?
3. Solve these equations to find $\mathbf{x}^\star = (x_1^\star, x_2^\star)$ and the multiplier $\lambda^\star$.
4. Verify that the derivative of the minimum cost is $\partial f^\star / \partial b = -\lambda^\star$.

Q11 (MLE of multivariate normal)

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^p$ be iid samples from a p-dimensional multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, where the mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and covariance $\boldsymbol{\Omega} \in \mathbb{R}^{p \times p}$ are unkonwn parameters. The log-likelihood is

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Omega}) = -\frac{n}{2}\log\det\boldsymbol{\Omega} - \frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) - \frac{n}{2}\log 2\pi.$$

Show that the maximum likelihood estimate (MLE) is

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} \mathbf{y}_i}{n}$$
$$\widehat{\boldsymbol{\Omega}} = \frac{\sum_{i=1}^{n} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})'}{n}.$$

That is to show that $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Omega}}$ maximize ℓ.

Hint: Use the first order optimality condition to find $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Omega}}$. To check the optimality of $\widehat{\boldsymbol{\Omega}}$, use its Cholesky factor.

Q12 (Smallest matrix subject to linear constraints)

Find the matrix $\mathbf{X}$ with the smallest Frobenius norm subject to the constraint $\mathbf{X}\mathbf{U} = \mathbf{V}$, assuming $\mathbf{U}$ has full column rank.

Hint: Write down the optimization problem and use the method of Lagrange multipliers.

Q13 (Minimizing a convex quadratic form over manifold)

$\mathbf{A} \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix. Find a matrix $\mathbf{U} \in \mathbb{R}^{n \times r}$ with orthonomal columns that maximizes $\operatorname{tr}(\mathbf{U}'\mathbf{A}\mathbf{U})$. That is to

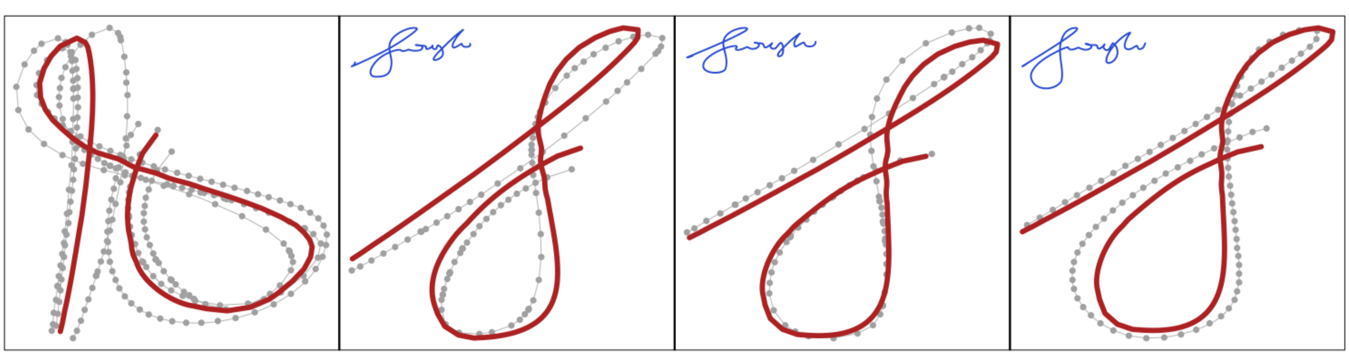
$$\begin{array}{ll} \text{maximize} & \operatorname{tr}(\mathbf{U}'\mathbf{A}\mathbf{U}) \\ \text{subject to} & \mathbf{U}'\mathbf{U} = \mathbf{I}_r. \end{array}$$

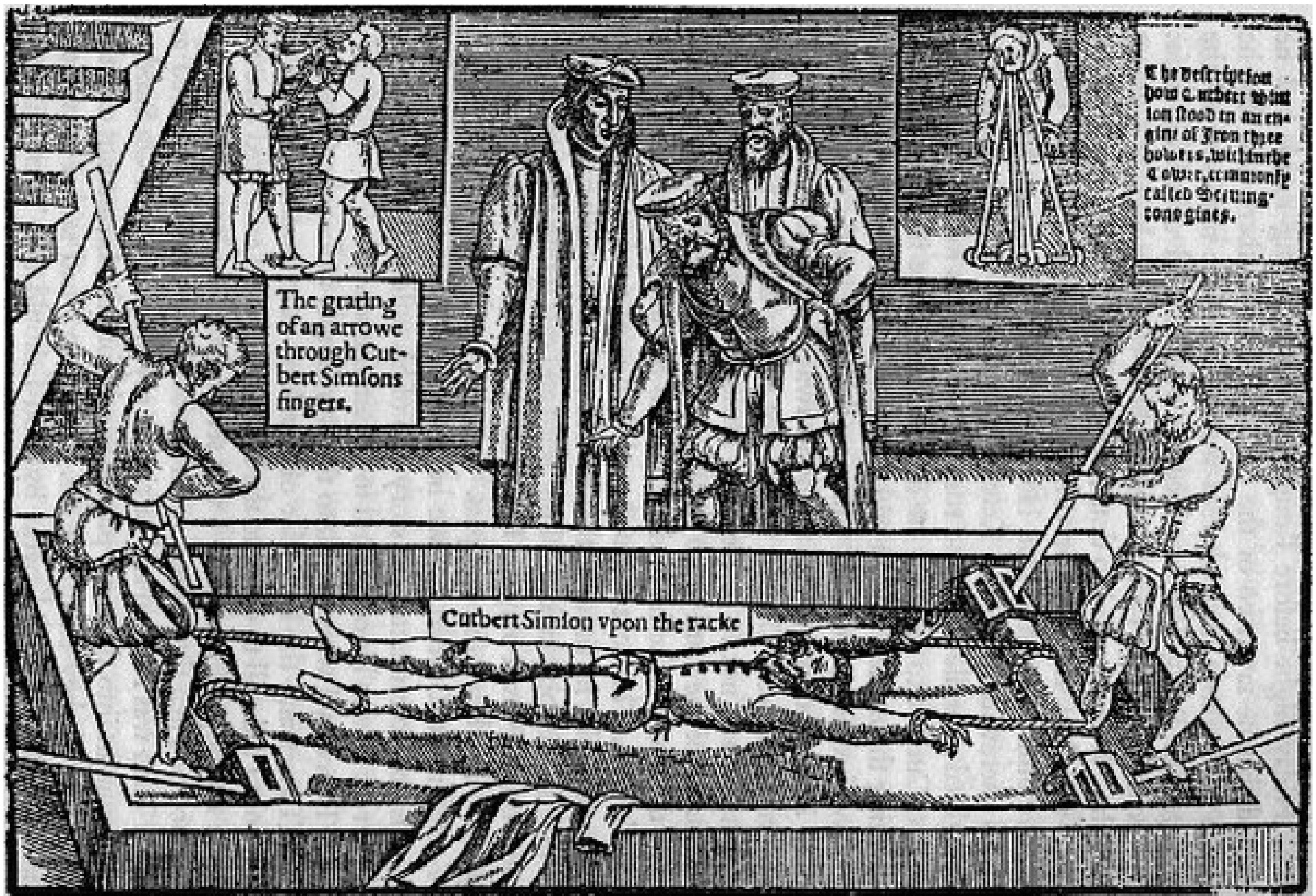
This result generalizes the fact that the top eigenvector of $\mathbf{A}$ maximizes $\mathbf{u}'\mathbf{A}\mathbf{u}$ subject to constraint $\mathbf{u}'\mathbf{u} = 1$.

Hint: Use the method of Lagrange multipliers.

Q14 (Procrustes rotation)

The Procrustes problem studies how to properly align images.





'A true description of the racking and cruell handeling of Cutbert Simson in the Tower.'

Let matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n \times p}$ record n points on the two shapes. Mathematically we consider the problem

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

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

1. Let $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ be the column mean vectors of the matrices and $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ be the versions of these matrices centered by column means. Show that the solution $(\hat{\beta}, \hat{\mathbf{O}}, \hat{\mu})$ satisfies

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

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

2. Derive the solution to

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

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

Q15 (Ridge regression)

One popular regularization method in machine learning is the ridge regression, which estimates regression coefficients by minimizing a penalized least squares criterion

$$\frac{1}{2}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2.$$

Here $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{R}^{n \times p}$ are fixed data. $\boldsymbol{\beta} \in \mathbb{R}^p$ are the regression coefficients to be estimated.

1. Show that, regardless the shape of $\mathbf{X}$, there is always a unique global minimum for any $\lambda > 0$ and the ridge solution is given by

$$\widehat{\boldsymbol{\beta}}(\lambda) = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}.$$

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

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

4. (**Optional**) Let's address how to choose the optimal tuning parameter λ. Let $\widehat{\boldsymbol{\beta}}_k(\lambda)$ be the solution to the ridge problem

$$\text{minimize } \frac{1}{2}\|\mathbf{y}_{-k} - \mathbf{X}_{-k}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2,$$

where $\mathbf{y}_{-k}$ and $\mathbf{X}_{-k}$ are the data with the k-th observation taken out. The optimal λ can to chosen to minimize the cross-validation square error

$$C(\lambda) = \frac{1}{n}\sum_{k=1}^{n}[y_k - \mathbf{x}_k^T\widehat{\boldsymbol{\beta}}_k(\lambda)]^2.$$

However computing n ridge solutions $\widehat{\boldsymbol{\beta}}_k(\lambda)$ is expensive. Show that, using SVD $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$,

$$C(\lambda) = \frac{1}{n}\sum_{k=1}^{n}\left[\frac{y_k - \sum_{j=1}^{r} u_{kj}\tilde{y}_j\left(\frac{\sigma_j^2}{\sigma_j^2+\lambda}\right)}{1 - \sum_{j=1}^{r} u_{kj}^2\left(\frac{\sigma_j^2}{\sigma_j^2+\lambda}\right)}\right]^2,$$

where $\tilde{\mathbf{y}} = \mathbf{U}^T\mathbf{y}$. Therefore, in practice, we only need to do SVD of $\mathbf{X}$ and then find the optimal value λ that minimizes the leave-one-out (LOO) cross-validation error.

Q16 (Optional) (Factor analysis)

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^p$ be iid samples from a multivariate normal distribution $N(\mathbf{0}_p, \mathbf{F}\mathbf{F}' + \mathbf{D})$, where $\mathbf{F} \in \mathbb{R}^{p \times r}$ and $\mathbf{D} \in \mathbb{R}^{p \times p}$ is a diagonal matrix with positive entries. We estimate the factor matrix $\mathbf{F}$ and diagonal matrix $\mathbf{D}$ by maximizing the log-likelihood function

$$\ell(\mathbf{F}, \mathbf{D}) = -\frac{n}{2} \ln \det(\mathbf{F}\mathbf{F}' + \mathbf{D}) - \frac{n}{2} \operatorname{tr}\left[(\mathbf{F}\mathbf{F}' + \mathbf{D})^{-1}\mathbf{S}\right] - \frac{np}{2} \ln 2\pi,$$

where $\mathbf{S} = n^{-1} \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i'$.

1. We first show that, for fixed $\mathbf{D}$, we can find the maximizer $\mathbf{F}$ explicitly using SVD by the following steps.
    - Step 1: Take derivative with respect to $\mathbf{F}$ and set to 0 to obtain the first-order optimality condition.
    - Step 2: Reparameterize $\mathbf{H} = \mathbf{D}^{-1/2}\mathbf{F}$ and $\tilde{\mathbf{S}} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$, and express the first-order optimality condition in terms of $\mathbf{H}$ and $\tilde{\mathbf{S}}$.
    - Step 3: Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$ be its SVD. Show that columns of $\mathbf{U}$ must be r eigenvectors of $\tilde{\mathbf{S}}$.
    - Step 4: Identify which r eigenvectors of $\tilde{\mathbf{S}}$ to use in $\mathbf{U}$ and then derive the solution to $\mathbf{F}$.
2. Show that, for fixed $\mathbf{F}$, we can find the maximizer $\mathbf{D}$ explicitly. (Hint: first-order optimality condition.)

Combining 1 and 2, a natural algorithm for finding the MLE of factor analysis model is to alternately update $\mathbf{F}$ and $\mathbf{D}$ until convergence.