Homework 4

Due: Thursday, October 26, 2023, 1:30pm on Gradescope

Please upload your answers timely to Gradescope. Start a new page for every problem. For the programming/simulation questions you can use any reasonable programming language. Comment your source code and include the code and a brief overall explanation with your answers.

1. **10 pts** As we discussed in class, all symmetric square real matrices have real eigenvalues and n orthogonal eigenvectors.

   a) (**5 pts**) Give an example of a square real matrix which has no real eigenvalues.

   b) (**5 pts**) Give an example of a square real matrix whose eigenvectors cannot be chosen to be orthogonal.

2. (**10 pts**) Exercise 3.8 in text.

Exercise 3.8 **(a)** Let $[K] = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$. Show that 1 and $1/2$ are eigenvalues of $[K]$ and find the normalized eigenvectors. Express $[K]$ as $[Q\Lambda Q^{-1}]$, where $[\Lambda]$ is diagonal and $[Q]$ is orthonormal.

(b) Let $[K'] = \alpha[K]$ for real $\alpha \neq 0$. Find the eigenvalues and eigenvectors of $[K']$. Do not use brute force – think!

(c) Find the eigenvalues and eigenvectors of $[K^m]$, where $[K^m]$ is the mth power of $[K]$.

3. (**15 pts**) Missing Proofs from the KL Expansion Discussion

We will first prove the following preliminary results:

(i) (**1 pts**) Prove that for two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times n}$, $\text{Trace}(AB) = \text{Trace}(BA)$.

(ii) (**2 pts**) Suppose $V \in \mathbb{R}^{n\times k}$, $k \leq n$, and $V^T V = I_k$, that is the columns of V are orthonormal in $\mathbb{R}^n$. (Note that V is not $n \times n$ so VV^T is not equal to the identity matrix.) Show that

$$||V^T \mathbf{a}||^2 \leq ||\mathbf{a}||^2 \qquad \text{for all} \qquad \mathbf{a} \in \mathbb{R}^n.$$

By choosing the vector $\mathbf{a} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith column of the identity matrix, show that $0 \leq (VV^T)_{ii} \leq 1$ for all $1 \leq i \leq n$.

(iii) (**1 pts**) Given a non-negative integer k, and a positive integer $n \geq k$, let $P = \{\mathbf{x} = (x_1, \ldots, x_n) : 0 \leq x_i \leq 1, \sum_i x_i = k\}$ (P is the subset of the n-dimensional unit cube on which the coordinates sum to k.) Suppose $c_1 \geq \cdots \geq c_n$ are real numbers. Show that

$$\max_{\mathbf{x}\in P} \sum_{i=1}^{n} c_i x_i = \sum_{i=1}^{k} c_i.$$

We will next turn to proving one of the claims we stated in the lecture. Suppose $K \in \mathbb{R}^{n\times n}$ is a real and symmetric matrix. We know that such a matrix can be written as $K = U\Lambda U^T$ where U is orthonormal and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_i \in \mathbb{R}$. Without loss of generality, assume that we have permuted the rows and columns of K so that $K_{11} \geq K_{22} \geq \cdots \geq K_{nn}$, and we have indexed the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

(a) (**2 pts**) Show that $\max_{\mathbf{v}\in\mathbb{R}^n:\mathbf{v}^T\mathbf{v}=1} \mathbf{v}^T K \mathbf{v} = \lambda_1$.

(b) (**2 pts**) Show that $K_{11} \leq \lambda_1$.

(c) (**3 pts**) Show that for any $k = 1, \ldots, n$,

$$\max_{V\in\mathbb{R}^{n\times k}:V^T V=I_k} \text{Trace}(V^T K V) = \max_{V\in\mathbb{R}^{n\times k}:V^T V=I_k} \text{Trace}(V^T \Lambda V) = \sum_{i=1}^{k} \lambda_i.$$

[Hint: You may find the preliminary results (i), (ii) and (iii) useful.]

(d) (**2 pts**) Show that for any $k = 1, \ldots, n$,

$$\sum_{i=1}^{k} K_{ii} \leq \sum_{i=1}^{k} \lambda_i,$$

and that equality holds when $k = n$.

(e) (**2 pts**) Finally, we will connect the result we proved above to the notation we used in class. Let $V \in \mathbb{R}^{n \times k}$ be such that $V^T V = I_k$ as above. Let $\mathbf{Z} = [Z_1 \dots Z_k]^T = V^T \mathbf{X}$ where $\mathbf{X} \in \mathbb{R}^n$ is a random vector with covariance matrix $K_{\mathbf{X}}$. Show that

$$\sum_{i=1}^{k} \mathrm{Var}(Z_i) = \mathrm{Trace}(V^T K_{\mathbf{X}} V).$$

4. (**15 pts**) You are interested in estimating the m by m covariance matrix K of a m-dimensional random vector $\mathbf{X}$. Your data is $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, drawn independently from the same distribution as $\mathbf{X}$. For simplicity, assume that all entries of $\mathbf{X}$ take values in the interval $[a, b]$.

   For simplicity we will first assume that $\mathbf{X}$ has zero mean.

   a) (**5 pts**) Suppose $m = 1$, i.e. the data are scalars $X_1, \ldots, X_n$.

      i. (**2 pts**) What does K become in this case?

      ii. (**3 pts**) Propose an estimator $\hat{K}_n$ of K, computed from the data, and show that it satisfies two properties:

         - It is unbiased, i.e. $E[\hat{K}_n] = K$.
         - $\hat{K}_n$ converges in probability to K as $n \to \infty$, i.e. for any $\epsilon > 0$

$$\mathbb{P}(|\hat{K}_n - K| \geq \epsilon) \to 0 \qquad \text{as} \quad n \to \infty$$

.

   b) (**5 pts**) Now let us consider the general case for $m > 1$ while still assuming that $\mathbf{X}$ has zero mean. Propose an estimator $\hat{K}_n$ of K and show that it satisfies two properties:

      i. It is unbiased, i.e. $E[\hat{K}_n] = K$. (Recall that the expectation of a random matrix is just taking the expectation of each entry of the matrix.)

      ii. Each entry of $\hat{K}_n$ converges in probability to the corresponding entry of K as $n \to \infty$.

      iii. $\hat{K}_n$ converges in probability to K as $n \to \infty$. This means that all the entries of $\hat{K}_n$ uniformly converge to the corresponding entries of K, i.e.

$$\mathbb{P}(\exists (i,j) \quad s.t. \quad |(\hat{K}_n)_{i,j} - K_{i,j}| \geq \epsilon) \to 0 \qquad \text{as} \quad n \to \infty$$

   c) (**5 pts**) Now suppose $\mathbf{X}$ has non-zero mean and you do not know the mean.

      i. (**1 pts**) Propose an unbiased estimator for the mean from the data. Show that your estimator is unbiased.

      ii. (**4 pts**) Propose an unbiased estimator $\hat{K}_n$ for K and show that it is unbiased. Hint: Assume that you first use the data to estimate the mean of $\mathbf{X}$ and then use your estimate for the mean to modify the estimator you suggested in part (b). Check whether it is unbiased for $m = 1$ before looking at the $m > 1$ case. If it turns out to be biased, can you scale it by a small factor and make it unbiased?