CS4070 – PART 2
EXERCISES RELATED TO LECTURE 1

(1) Review some concepts from linear algebra such as: eigenvalue decomposition of a matrix, best approximation theorem, null/column space, spanned subspace of a set of vectors, invertibility of a matrix, trace of a matrix.

(2) Suppose the square matrix A has eigenvalue decomposition $A = V\Lambda V^T$, where Λ is a diagonal matrix containing the eigenvalues and V is an orthogonal matrix containing eigenvectors as columns. Show that $\mathrm{tr}A = \mathrm{tr}\Lambda$ and hence that the trace of A is the sum of its eigenvalues.

(3) Show that $\mathcal{N}(A) = \mathcal{N}(A^T A)$, where $\mathcal{N}(A) = \{x : Ax = 0\}$.

(4) If

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

then the mean-vector and covariance matrix are defined by

$$\mathbb{E}Y = \begin{bmatrix} \mathbb{E}Y_1 \\ \vdots \\ \mathbb{E}Y_n \end{bmatrix}$$

and

$$\mathrm{Cov}Y = \mathbb{E}[(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T],$$

respectively.

(a) Check that element $[i, j]$ of the matrix $\mathrm{Cov}(Y)$ is given by $\mathrm{Cov}(Y_i, Y_j)$.

(b) Suppose that A is a $k \times n$ matrix. Verify that $\mathbb{E}[AY] = A\mathbb{E}Y$. *Hint: note that for $j \in \{1, \dots, k\}$, the j-th element of the vector AY satisfies $(AY)[j]$ $= \sum_{i=1}^n A_{ji} Y_i$.*

(c) Verify that

$$\mathrm{Cov}(AY) = A\,(\mathrm{Cov}Y)\,A^T.$$

(5) Suppose U and V are independent random variables, each with the $N(0, 1)$-distribution. for $\rho \in [-1, 1]$, define

$$\begin{aligned} X_1 &= \sqrt{1-\rho^2}U + \rho V \\ X_2 &= \qquad\qquad\quad V \end{aligned}$$

Verify that the correlation between X_1 and X_2 equals ρ.

(6) Consider the linear model

$$\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(0, \Sigma).$$

Assume Σ is nonsingular and known.

Date: November 12, 2023.

(a) Derive an expression for the maximum-likelihood estimator for $\boldsymbol{\theta}$, if it exists. Also explain when the maximum-likelihood estimator is unique.
(b) Derive an expression for the Hessian matrix of the loglikelihood (this is the matrix containing the second-order derivatives of the loglikelihood).

1. Solutions

(1) –

(2) Use that $\operatorname{tr}(ABC) = \operatorname{tr}(CAB)$ (for compatible matrices). So you can cyclically permute the matrices within the trace. This gives

$$\operatorname{tr} A = \operatorname{tr}(V\Lambda V^T) = \operatorname{tr}(\Lambda V^T V) = \operatorname{tr}(\Lambda) = \sum_i \lambda_i.$$

The third equality holds since V is an orthogonal matrix.

(3) Suppose $x \in \mathcal{N}(A)$, that is $Ax = 0$. Then $A^T Ax = A^T(Ax) = A^T 0 = 0$ and hence $x \in \mathcal{N}(A^T A)$. This shows that $\mathcal{N}(A) \subseteq \mathcal{N}(A^T A)$.
Now for the other way around, suppose that $x \in \mathcal{N}(A^T A)$, that is $A^T Ax = 0$. This implies $x^T A^T Ax = 0$ which is equivalent to $\|Ax\|^2 = 0$. Now the norm of a vector can only be zero if the vector itself is zero. Hence $Ax = 0$. But this says $x \in \mathcal{N}(A)$. Therefore, $\mathcal{N}(A^T A) \subseteq \mathcal{N}(A)$.

Check yourself carefully the meaning of 0 in this derivation (at some places it is the number 0, at some places it is the zero-vector).

(4) (a) For a vector a, the matrix aa^T has as its $[i,j]$-th element $a_i a_j$. Hence, the $[i,j]$-th element of Cov$Y$ is given by $\mathbb{E}(Y_i - \mathbb{E}Y_i)(Y_j - \mathbb{E}Y_j)$ which is $\operatorname{Cov}(Y_i, Y_j)$.

(b)

$$\mathbb{E}[AY] = \mathbb{E}\begin{bmatrix} \sum_{i=1}^n A_{1i} Y_i \\ \vdots \\ \sum_{i=1}^n A_{ki} Y_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n A_{1i} \mathbb{E} Y_i \\ \vdots \\ \sum_{i=1}^n A_{ki} \mathbb{E} Y_i \end{bmatrix} = \begin{bmatrix} \langle \operatorname{Row}_1 A, \mathbb{E}Y \rangle \\ \vdots \\ \langle \operatorname{Row}_k A, \mathbb{E}Y \rangle \end{bmatrix} = A\mathbb{E}Y$$

(here $\langle x, y \rangle = x^T y$ for vectors x and y of equal length).

(c) Using the just shown linearity of expectation we get

$$\begin{aligned} \operatorname{Cov}(AY) &= \mathbb{E}(AY - \mathbb{E}(AY))(AY - \mathbb{E}(AY))^T \\ &= \mathbb{E}\left[A(Y - \mathbb{E}Y)(A(Y - \mathbb{E}Y))^T\right] \\ &= A\mathbb{E}\left[(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T A^T\right] \\ &= A\mathbb{E}\left[(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T\right] A^T. \end{aligned}$$

(5) Firt note that both X_1 and X_2 have variance equal to 1 so that $\operatorname{corr}(X_1, X_2) = \operatorname{Cov}(X_1, X_2)$. Next, using half-linearity of the covariance operator (it is linear in one component, if the other component is fixed) we get

$$\begin{aligned} \operatorname{corr}(X_1, X_2) &= \operatorname{Cov}\Big(\sqrt{1-\rho^2}U + \rho V, V\Big) \\ &= \sqrt{1-\rho^2}\operatorname{Cov}(U, V) + \rho \operatorname{Cov}(V, V). \end{aligned}$$

Now the first term is zero, as U and V are assumed to be independent. The second term equals $\rho \operatorname{Var}(V) = \rho$.

(6) (a) The loglikelihood $\ell(\theta$ is given by

$$\ell(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\boldsymbol{y} - X\boldsymbol{\theta})^T \Sigma^{-1}(\boldsymbol{y} - X\boldsymbol{\theta}).$$

So a maximiser is found by setting the gradient with respect to $\boldsymbol{\theta}$ equal to zero. This gives

$$\nabla \ell(\theta) = X^T \Sigma^{-1}(\boldsymbol{y} - X\boldsymbol{\theta}) = 0.$$

That is

$$X^T \Sigma^{-1} X \boldsymbol{\theta} = X^T \Sigma^{-1} \boldsymbol{y}.$$

If $\theta \mapsto \ell(\theta)$ is concave, then the stationary point is a maximiser (this is the case, see part (b)). A unique estimator exists when $X^T \Sigma^{-1} X$ is invertible.

(b) We need to take the gradient of each element of $\nabla \ell(\theta)$ and put the resulting vectors together in a matrix. This yields $H = -X^T \Sigma^{-1} X$. As this matrix is negative semi-definite, this shows $\theta \mapsto \ell(\theta)$ is concave.