

# STAT151A Homework 1 (prerequisites review).

Your name here

This homework is due on Gradescope on **Friday September 13th at 9pm**.

## 1 Linear systems

Write the following system of equations in matrix form. Say whether each system has no solutions, a single solution, or an infinite number of solutions, and how you know.

**Solution note.** The purpose of this exercise, in addition to reminding yourself how to write out equations in matrix form, is to remember that if there are  $P$  unknowns and  $Q$  equations:

- If  $P > Q$  there are an infinite number of solutions
- If  $P = Q$  there are either an infinite number or one solution
- If  $P < Q$  there are either an infinite number, one, or no solutions.

The key to note is that  $Q$  may be made “artificially large” by repeating equations or including redundant equations.

Formally, for a system of equations of the form  $\mathbf{A}\mathbf{b} =$ , the singular value decomposition of  $\mathbf{A}$  is the easiest way to characterize the number of solutions, but that’s beyond the scope of the class.

(1)

$$a_1 + 2a_2 = 1$$

$$a_1 + 3a_2 = 0$$

**Solution** One solution. (You can difference the equations to solve for  $a_2$  then plug in.)

The matrix solution is

$$\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

I won't write out all the matrix solutions since they're obvious once you get the pattern. The coefficients go into the first matrix, then a vector of the unknowns, then a vector of the right hand side terms.

(2)

$$a_1 + 2a_2 + 3a_3 = 1$$

$$a_1 + 3a_2 + 3a_3 = 0$$

$$2a_1 + 4a_2 + 3a_3 = 5$$

**Solution** This code helps show that there is one solution with  $a_3 \neq 0$ .

```
a <- matrix(c(1, 1, 2, 2, 3, 4, 3, 3, 3), ncol=3)
det(a)
solve(a, c(c(1, 0, 5)))
```

(3)

$$a_1 + 2a_2 + 3a_3 = 1$$

$$a_1 + 3a_2 + 3a_3 = 0$$

**Solution** There are two equations and three unknowns and so an infinite number of solutions. (E.g., starting from the solution to the previous one, add  $x$  to  $a_3$  and  $-x/3$  to  $a_1$ .)

(4)

$$a_1 + 2a_2 = 1$$

$$a_1 + 3a_2 = 0$$

$$2a_1 + 4a_2 = 5$$

**Solution** There are three equations and two unknowns. There are no solutions. This is the same as (2) but with  $a_3$  removed, and the solution to (2) had  $a_3 \neq 0$ .

(5)

$$a_1 + 2a_2 = 1$$

**Solution** There is one equation and two unknowns, so an infinite number of solutions.

(6)

$$a_1 + 2a_2 = 1$$

$$a_1 + 2a_2 = 1$$

$$a_1 + 2a_2 = 1$$

$$a_1 + 2a_2 = 1$$

$$a_1 + 2a_2 = 1$$

**Solution** Although there are five equations and two unknowns, they are all the same equation repeated, so there is still an infinite number of solutions.

(7)

$$a_1 + 2a_2 = 1$$

$$a_1 + 2a_2 = 2$$

$$a_1 + 2a_2 = 3$$

$$a_1 + 2a_2 = 4$$

$$a_1 + 2a_2 = 5$$

**Solution** There are no solutions. The difference with the previous example is that the right hand side is different.

(8)

$$a_1 + 2a_2 = 1$$

$$a_1 + 3a_2 = 1$$

$$a_1 + 4a_2 = 1$$

$$a_1 + 5a_2 = 1$$

$$a_1 + 6a_2 = 5$$

**Solution** There are no solutions. For the first four, we'd have to have  $a_2 = 0$ , but then we'd have to have  $a_1 = 5$  and  $a_1 = 1$ , which is impossible. It's interesting to see that if the last right-hand-side term were 1 then there would in fact be a solution.

(9)

$$5a_1 + 2a_2 = 1$$

$$10a_1 + 4a_2 = 1$$

**Solution** Although there are two equations and two unknowns there are no solutions. Note that two times the first row minus the second gives  $0 = 1$ .

(10)

$$5a_1 + 2a_2 = 1$$

$$10a_1 + 4a_2 = 2$$

**Solution** Now there are an infinite number of solutions, since the bottom equation is just two times the first, so these are actually a single equation.

## 2 Dimensions of linear algebra expressions

For this problem, I will use the following definitions.

- $\mathbf{X}$  denotes an  $N \times P$  matrix
- $\mathbf{y}$  denotes an  $N$ -vector (i.e. an  $N \times 1$  matrix)
- $\mathbf{1}$  denotes an  $N$ -vector containing all ones

- $\beta$  denotes a  $P$ -vector

I will take  $N > P > 1$ . A transpose is denoted with a superscript  $\top$ , and an inverse by a superscript  $-1$ . A matrix trace is denoted  $\text{trace}()$ . You may assume that each matrix is full column rank.

For each expression, write the dimension of the result, or write “badly formed” if the expression is not a valid matrix expression.

**Tip:** For this assignment, and throughout the class, it can be very helpful to write the dimensions of a matrix or vector underneath to help make sure your matrix expressions are valid. For example, we can write  $\underset{PN}{\mathbf{X}^\top} \underset{NP}{\mathbf{X}}$ , which we can see is valid because the  $N$ 's are next to one another. Similarly, we can see immediately that the expression  $\underset{NP}{\mathbf{X}} \underset{NP}{\mathbf{X}}$  is invalid because  $P$  is next to  $N$  in the matrix multiplication, which is not allowed.

- (1)  $\mathbf{X}^\top \mathbf{y}$
- (2)  $\mathbf{X}^\top \mathbf{X}$
- (3)  $\mathbf{X} + \mathbf{y}$
- (4)  $\beta \beta^\top$
- (5)  $\beta^\top \beta$
- (6)  $\mathbf{y}^\top \mathbf{y}$
- (7)  $\mathbf{X}^\top \mathbf{y}$  (Duplicate — that's okay, just repeat the other answer)
- (8)  $(\mathbf{X}^\top \mathbf{X})^\top$
- (9)  $(\mathbf{X}^\top \mathbf{y})^\top$
- (10)  $(\mathbf{X}^\top \mathbf{X})^{-1}$
- (11)  $(\mathbf{X}^\top \mathbf{y})^{-1}$
- (12)  $\mathbf{X}^{-1}$
- (13)  $\mathbf{y}^{-1} \mathbf{y}$
- (14)  $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$
- (15)  $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$
- (16)  $((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y})^\top$
- (17)  $\mathbf{y}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}$
- (18)  $\mathbf{X} \beta$
- (19)  $\mathbf{y} - \mathbf{X} \beta$
- (20)  $\mathbf{y} - \mathbf{X}^\top \beta$
- (21)  $\mathbf{y}^\top - \beta^\top \mathbf{X}^\top$
- (22)  $\beta - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$
- (23)  $\mathbf{X} (\beta - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y})$
- (24)  $\mathbf{1}^\top \mathbf{y}$
- (25)  $\mathbf{y} - (\mathbf{1}^\top \mathbf{y}) \mathbf{y}$
- (26)  $\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$
- (27)  $\mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}$
- (28)  $\mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}$  (Duplicate — that's okay, just repeat the other answer)
- (29)  $\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta$

- (30)  $(\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta)^{-1}$
- (31)  $\text{trace}(\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta)$
- (32)  $\text{trace}((\mathbf{X}^\top \mathbf{X})^{-1} \beta \beta^\top)$

### Solutions

- (1)  $\mathbf{X}^\top \mathbf{y} : P \times 1$
- (2)  $\mathbf{X}^\top \mathbf{X} : P \times P$
- (3)  $\mathbf{X} + \mathbf{y} : \text{Invalid}$
- (4)  $\beta \beta^\top : P \times P$
- (5)  $\beta^\top \beta : 1 \times 1$
- (6)  $\mathbf{y}^\top \mathbf{y} : 1 \times 1$
- (7)  $\mathbf{X}^\top \mathbf{y}$  (Duplicate)
- (8)  $(\mathbf{X}^\top \mathbf{X})^\top : P \times P$
- (9)  $(\mathbf{X}^\top \mathbf{y})^\top : 1 \times P$
- (10)  $(\mathbf{X}^\top \mathbf{X})^{-1} : P \times P$
- (11)  $(\mathbf{X}^\top \mathbf{y})^{-1} : \text{Invalid}$  (only square matrices are invertible)
- (12)  $\mathbf{X}^{-1} : \text{Invalid}$  (only square matrices are invertible)
- (13)  $\mathbf{y}^{-1} \mathbf{y} : \text{Invalid}$  (only square matrices are invertible)
- (14)  $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top : P \times N$
- (15)  $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} : P \times 1$
- (16)  $((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y})^\top : 1 \times P$
- (17)  $\mathbf{y}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} : 1 \times P$
- (18)  $\mathbf{X} \beta : N \times 1$
- (19)  $\mathbf{y} - \mathbf{X} \beta : N \times 1$
- (20)  $\mathbf{y} - \mathbf{X}^\top \beta : \text{Invalid}$
- (21)  $\mathbf{y}^\top - \beta^\top \mathbf{X}^\top : 1 \times N$
- (22)  $\beta - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} : P \times 1$
- (23)  $\mathbf{X} (\beta - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}) : N \times 1$
- (24)  $\mathbf{1}^\top \mathbf{y} : 1 \times 1$
- (25)  $\mathbf{y} - (\mathbf{1}^\top \mathbf{y}) \mathbf{y} : N \times 1$
- (26)  $\mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top : N \times N$
- (27)  $\mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} : \text{Invalid}$
- (28)  $\mathbf{X}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}$  (Duplicate)
- (29)  $\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta : 1 \times 1$
- (30)  $(\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta)^{-1} : 1 \times 1$
- (31)  $\text{trace}(\beta^\top (\mathbf{X}^\top \mathbf{X})^{-1} \beta) : 1 \times 1$
- (32)  $\text{trace}((\mathbf{X}^\top \mathbf{X})^{-1} \beta \beta^\top) : 1 \times 1$

### 3 Orthonormal vectors and bases

For this problem, assume that  $\mathbf{u} = (u_1, u_2)^\top$  and  $\mathbf{v} = (v_1, v_2)^\top$  are orthonormal. That is,  $\mathbf{u}^\top \mathbf{u} = \mathbf{v}^\top \mathbf{v} = 1$ , and  $\mathbf{u}^\top \mathbf{v} = 0$ . Let  $\mathbf{a} = (a_1, a_2)^\top$  denote a generic 2-dimensional vector.

- (1) Write an expression for the length of  $\mathbf{a}$  in terms of its entries  $a_1$  and  $a_2$ .
- (2) Write an expression for the length of  $\mathbf{a}$ , using only matrix operations (i.e., without explicit reference to the entries of  $\mathbf{a}$ ).
- (3) Write an explicit expression for a vector pointing in the same direction as  $\mathbf{a}$  but with unit length. (Hint: show that, for a scalar  $\alpha$ , the length of  $\alpha \mathbf{a}$  is  $\alpha$  times the length of  $\mathbf{a}$ , and then make a clever choice for  $\alpha$ .)
- (4) Write an explicit expression for a vector pointing in the same direction as  $\mathbf{v}$  but with the same length as  $\mathbf{a}$ .
- (5) Suppose that I tell you that  $\mathbf{a} = \alpha \mathbf{u} + \gamma \mathbf{v}$ . Find an explicit expression for  $\alpha$  in terms of  $\mathbf{a}$  and  $\mathbf{u}$  alone.
- (6) Find explicit expressions for scalars  $\alpha$  and  $\gamma$  such that  $\mathbf{a} = \alpha \mathbf{u} + \gamma \mathbf{v}$ .
- (7) Let  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} := \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$  denote the  $2 \times 2$  matrix with  $\mathbf{u}$  in the first column and  $\mathbf{v}$  in the second column. Show that  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^\top = \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^{-1}$ .
- (8) Observe that  $\alpha \mathbf{u} + \gamma \mathbf{v} = \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$ . Using this, write an explicit expression for the coefficient vector  $(\alpha, \gamma)^\top$  in terms of  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}$  and  $\mathbf{a}$ .
- (9) Show that  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^\top$ .
- (10) Suppose I tell you that that  $\mathbf{u} = (1, 0)^\top$ . In terms of  $a_1$  and  $a_2$ , what is  $\alpha$  in the decomposition  $\mathbf{a} = \alpha \mathbf{u} + \gamma \mathbf{v}$ ?

#### Solutions

- (1)  $\sqrt{a_1^2 + a_2^2}$
- (2)  $\sqrt{\mathbf{a}^\top \mathbf{a}}$
- (3)  $\mathbf{a} / \sqrt{\mathbf{a}^\top \mathbf{a}}$ . From now on,  $\|\mathbf{a}\| := \sqrt{\mathbf{a}^\top \mathbf{a}}$
- (4)  $\|\mathbf{a}\| \mathbf{v}$  (since  $\mathbf{v}$  has unit length already)
- (5)  $\alpha = \mathbf{a}^\top \mathbf{u}$
- (6)  $\gamma = \mathbf{a}^\top \mathbf{v}$  and  $\alpha = \mathbf{a}^\top \mathbf{u}$
- (7)  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^\top \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{u}^\top \mathbf{u} & \mathbf{u}^\top \mathbf{v} \\ \mathbf{v}^\top \mathbf{u} & \mathbf{v}^\top \mathbf{v} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ . Since inverses are unique, the transpose must be the inverse.
- (8)  $\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix}^\top \mathbf{a} = \begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$
- (9) Repeated question (sorry)
- (10)  $\alpha = a_1$

## 4 Eigenvalues and eigenvectors of square symmetric matrices

Let  $\mathbf{A}$  denote a  $P \times P$  symmetric, square matrix. Recall that an “eigenvalue”  $\lambda_k$  of and its associated “eigenvector”  $\mathbf{u}_k$  of  $\mathbf{A}$  satisfy  $\mathbf{A}\mathbf{u}_k = \lambda_k\mathbf{u}_k$ . (In this definition,  $\mathbf{u}_k$  must be non-degenerate, i.e., it must have some nonzero entries.)

Let  $\mathbf{U} = (\mathbf{u}_1 \dots \mathbf{u}_P)$  denote the  $P \times P$  matrix with eigenvector  $\mathbf{u}_k$  in the  $k$ -th column, and let  $\Lambda$  denote the  $P \times P$  diagonal matrix with  $\lambda_k$  in the  $k$ -th diagonal entry. Let  $\mathbf{a}$  denote a generic  $P$ -vector.

- If  $\mathbf{A}$  is the identity matrix (i.e., the matrix with ones on the diagonal and zeros elsewhere), what are its eigenvalues?

**Solution** They are all 1.

- If  $\mathbf{A}$  is the zero matrix (i.e., the matrix containing only zeros), what are its eigenvalues?

**Solution** They are all 0.

- If  $\mathbf{A}$  is a diagonal matrix with the entries  $a_1, \dots, a_P$  on the diagonal, what are its eigenvalues?

**Solution** They are the values  $a_p$ . (The eigenvectors are vectors with exactly one non-zero element.)

- Let us prove that the eigenvectors can be taken to be unit vectors without loss of generality.
  - Show that, if  $\mathbf{u}_k$  is an eigenvector, then  $\alpha\mathbf{u}_k$  is also an eigenvector with the same eigenvalue as  $\mathbf{u}_k$  for any scalar  $\alpha \neq 0$ .
  - In particular, show that  $\mathbf{u}'_k = \mathbf{u}_k / \sqrt{\mathbf{u}_k^\top \mathbf{u}_k}$  is also an eigenvector with eigenvalue  $\lambda_k$ , and that  $(\mathbf{u}'_k)^\top \mathbf{u}'_k = 1$ .

**Solution**

$\mathbf{A}(\alpha\mathbf{u}_k) = \alpha\mathbf{A}\mathbf{u}_k = \alpha\lambda_k\mathbf{u}_k = \lambda_k(\alpha\mathbf{u}_k)$ . Then take  $\alpha = 1/\|\mathbf{u}_k\|$ .

- Let us prove that, for a general symmetric matrix, eigenvectors with distinct eigenvalues are orthogonal. That is, if  $\lambda_k \neq \lambda_j$  for some  $k \neq j$ , we will show that  $\mathbf{u}_k^\top \mathbf{u}_j = 0$ . Carefully justify each step in the following proof.
  - We have  $\mathbf{u}_j^\top \mathbf{A}\mathbf{u}_k = (\mathbf{u}_j^\top \mathbf{A}\mathbf{u}_k)^\top$ . (Why?)
  - We also have  $(\mathbf{u}_j^\top \mathbf{A}\mathbf{u}_k)^\top = \mathbf{u}_k^\top \mathbf{A}^\top \mathbf{u}_j$ . (Why?)
  - We then have  $\mathbf{u}_j^\top \mathbf{A}\mathbf{u}_k = \mathbf{u}_k^\top \mathbf{A}\mathbf{u}_j$ . (Why?)
  - We then have  $\lambda_k \mathbf{u}_j^\top \mathbf{u}_k = \lambda_j \mathbf{u}_k^\top \mathbf{u}_j$ . (Why?)
  - We then have  $\mathbf{u}_k^\top \mathbf{u}_j = 0$ . (Why?)

- **Note (not graded):** If the eigenvalues are not distinct, then the corresponding eigenvectors may not be orthogonal. However, one can always find an orthogonal set of eigenvectors for each repeated eigenvalue, though we won't prove this here. See a linear algebra text for the proof, or try yourself! Hint: if  $\lambda_k = \lambda_j$  for some  $k \neq j$ , then  $\alpha \mathbf{u}_k + \gamma \mathbf{u}_j$  is also an eigenvector with eigenvalue  $\lambda_k$  for any  $\alpha$  and  $\gamma$ .

## Solutions

- The transpose of a scalar is the same scalar
- In general,  $(\mathbf{RS})^\top = \mathbf{S}^\top \mathbf{R}^\top$  (transposes reverse the order of matrix multiplication). Apply this rule twice.
- Combine the first two results
- $\mathbf{u}_k$  and  $\mathbf{u}_j$  are eigenvectors
- Because  $\lambda_k \neq \lambda_j$ , and so the only way the preceding equation can be true is if  $\mathbf{u}_k^\top \mathbf{u}_j = 0$ .
- Suppose that each  $\lambda_k \neq 0$ . Show that the inverse of  $\Lambda$  is given by the diagonal matrix with  $1/\lambda_k$  in the  $k$ -th diagonal entry and zero elsewhere.

## Solution

$$\begin{pmatrix} 1/\lambda_1 & 0 & \dots & 0 \\ 0 & 1/\lambda_2 & 0 & \dots & 0 \\ & & \ddots & & \\ & & & \dots & 0 & 1/\lambda_P \end{pmatrix} \Lambda = \begin{pmatrix} \lambda_1/\lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2/\lambda_2 & 0 & \dots & 0 \\ & & \ddots & & \\ & & & \dots & 0 & \lambda_P/\lambda_P \end{pmatrix} = \mathbf{I}$$

- Suppose that  $\lambda_k = 0$  for some  $k$ . Show that  $\Lambda$  is not invertible. Hint: find a vector  $\mathbf{b}$  such that  $\Lambda \mathbf{a} = \mathbf{b}$  has no solution. From this it would follow that  $\Lambda$  is not invertible by contradiction for, if  $\Lambda$  were invertible,  $\mathbf{a} = \Lambda^{-1} \mathbf{b}$  would be a solution.

**Solution** Set  $\mathbf{b}$  to be zeros except in the  $k$ -th entry. Since  $\Lambda \mathbf{a}$  is zero in the  $k$ -th entry no matter what  $\mathbf{a}$  is, and so not equal to  $\mathbf{b}$ ,  $\Lambda$  cannot be invertible.

- Assume that  $\mathbf{A}$  has  $P$  orthonormal eigenvectors (it always does; we have proved some parts of this assertion above). We will prove that we then have  $\mathbf{A} = \mathbf{U} \Lambda \mathbf{U}^\top$ . Carefully justify each step in the following proof.
  - For any  $\mathbf{a}$ , we can write  $\mathbf{a} = \sum_{p=1}^P \alpha_p \mathbf{u}_p$  for some scalars  $\alpha_p$ . (Why?)
  - Using the previous expansion, we thus have  $\mathbf{A} \mathbf{a} = \sum_{p=1}^P \lambda_p \alpha_p \mathbf{u}_p$ . (Why?)
  - Let  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_P)^\top$ . Then  $\mathbf{a} = \mathbf{U} \boldsymbol{\alpha}$ . (Why?)
  - Similarly, we have  $\mathbf{A} \mathbf{a} = \mathbf{U} \Lambda \boldsymbol{\alpha}$ . (Why?)
  - We also have  $\mathbf{u}_p^\top \mathbf{a} = \alpha_p$ . (Why?)
  - We then have  $\boldsymbol{\alpha} = \mathbf{U}^\top \mathbf{a}$ . (Why?)
  - Combining, we have  $\mathbf{A} \mathbf{a} = \mathbf{U} \Lambda \mathbf{U}^\top \mathbf{a}$ . (Why?)



- Since the preceding expression is true for any  $\mathbf{a}$ , we must have  $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ . (Why? Hint: if you take  $\mathbf{a}$  to be 1 in the first entry and zeros elsewhere, then we have that the first columns match. Continue in this fashion to show that each entry of the matrices are the same.)

## Solutions

- The  $P$  eigenvectors span  $P$  dimensional space, so every  $P$ -vector can be written as a linear combination of them.
- $\mathbf{A}\mathbf{a} = \mathbf{A}\sum_{p=1}^P \alpha_p \mathbf{u}_p = \sum_{p=1}^P \alpha_p \mathbf{A}\mathbf{u}_p = \sum_{p=1}^P \alpha_p \lambda_p \mathbf{u}_p$  because the  $\mathbf{u}_p$  are eigenvectors.
- This is just a matrix way of writing the the first result  $\sum_{p=1}^P \alpha_p \mathbf{u}_p$ .
- This is just a matrix way of writing the the second result (since  $\mathbf{\Lambda}$  is diagonal, its diagonal entries just multiply each corresponding row of  $\mathbf{a}$ ).
- We have  $\mathbf{u}_p^\top \mathbf{a} = \sum_{q=1}^P \alpha_q \mathbf{u}_p^\top \mathbf{u}_q = \alpha_p$  by orthogonality
- This is the matrix version of the preceding result stacked row-wise
- Combining,  $\mathbf{A}\mathbf{a} = \mathbf{U}\mathbf{\Lambda}\mathbf{\alpha} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{a}$ .
- If  $\mathbf{a}$  is 1 in row  $k$  and zeros elsewhere, it says that the  $k$ -th column of  $\mathbf{A}$  must be the same as the  $k$ -th column of  $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ . Since this is true for each  $k$ , the entire matrices must be the same.

## 5 Statistical asymptotics

For this problem, suppose that  $x_n$  are independent and identically distributed random variables, with  $\mathbb{E}[x_n] = 3$ , and  $\text{Var}(x_n) = 4$ . You should not assume that the  $x_n$  are normally distributed.

Recall that  $\text{Var}(x_n) = \mathbb{E}[x_n^2] - \mathbb{E}[x_n]^2$ . Also recall that, for any scalar  $\alpha$ ,  $\mathbb{E}[\alpha x_n] = \alpha \mathbb{E}[x_n]$ , and  $\text{Var}(\alpha x_n) = \mathbb{E}[(\alpha x_n)^2] - \mathbb{E}[\alpha x_n]^2 = \alpha^2 \text{Var}(x_n)$ .

For each limiting statement, state clearly whether your answer is a constant, a random variable, or that the limit does not exist. If the limit is a random variable, give its distribution. If it is a constant, give its value. If it does not exist, argue why. You may refer to known results from probability theory, particularly the law of large numbers and the central limit theorem.

- (1) For any particular (finite) value of  $N$ , is  $\frac{1}{N} \sum_{n=1}^N x_n$  a constant or a random variable?
- (2) For any particular (finite) value of  $N$ , is  $\frac{1}{N} \sum_{n=1}^N \mathbb{E}[x_n]$  a constant or a random variable?
- (3) Compute the expectation  $\mathbb{E}[x_1 + x_2]$ .
- (4) Compute the expectation  $\mathbb{E}\left[\frac{1}{2}(x_1 + x_2)\right]$ .
- (5) Compute the expectation  $\mathbb{E}\left[\frac{1}{3}(x_1 + x_2 + x_3)\right]$ .
- (6) Compute the expectation  $\mathbb{E}\left[\frac{1}{N} \sum_{n=1}^N x_n\right]$ .
- (7) Compute the variance  $\text{Var}(x_1 + x_2)$ .
- (8) Compute the variance  $\text{Var}\left(\frac{1}{2}(x_1 + x_2)\right)$ .

- (9) Compute the variance  $\text{Var}\left(\frac{1}{3}(x_1 + x_2 + x_3)\right)$ .
- (10) Compute the variance  $\text{Var}\left(\frac{1}{N} \sum_{n=1}^N x_n\right)$ .
- (11) As  $N \rightarrow \infty$ , what (if anything) does  $\frac{1}{N} \sum_{n=1}^N x_n$  converge to?
- (12) As  $N \rightarrow \infty$ , what (if anything) does  $\frac{1}{N} \sum_{n=1}^N (x_n - 3)$  converge to?
- (13) As  $N \rightarrow \infty$ , what (if anything) does  $\frac{1}{\sqrt{N}} \sum_{n=1}^N (x_n - 3)$  converge to?
- (14) As  $N \rightarrow \infty$ , what (if anything) does  $\frac{1}{\sqrt{N}} \sum_{n=1}^N x_n$  converge to?
- (15) As  $N \rightarrow \infty$ , what (if anything) does  $\sum_{n=1}^N x_n$  converge to?
- (16) As  $N \rightarrow \infty$ , what (if anything) does  $\sum_{n=1}^N (x_n - 3)$  converge to?
- (17) **(Bonus question — will not be graded)** What is the limit as  $N \rightarrow \infty$  of  $\frac{1}{N} \sum_{n=1}^N (4 \cdot (-1)^n + 3)$ ? In contrast, suppose that  $x_n$  is generated by randomly flipping a coin and setting  $x_n = 3 + 4$  if the coin comes up heads and  $x_n = 3 - 4$  if it comes up tails. What is different about the convergence of  $\frac{1}{N} \sum_{n=1}^N (4 \cdot (-1)^n + 3)$  and the convergence of  $\frac{1}{N} \sum_{n=1}^N x_n$ ?

### Solutions

- (1) Random variable
- (2) Constant ( $\mathbb{E}[x_n]$  is a constant, not a random variable)
- (3)  $3 + 3 = 6$
- (4)  $\frac{3+3}{2} = 3$
- (5)  $\frac{3+3+3}{3} = 3$
- (6)  $\frac{N3}{N} = 3$
- (7)  $4 + 4 = 8$
- (8)  $\frac{4+4}{2^2} = 8/4 = 2$
- (9)  $\frac{4+4+4}{3^2} = 12/9 = 4/3$
- (10)  $\frac{N4}{N^2} = 4/N$
- (11) 3
- (12) 0
- (13)  $\mathcal{N}(0, 4)$
- (14)  $\frac{1}{\sqrt{N}} \sum_{n=1}^N x_n = \frac{1}{\sqrt{N}} \sum_{n=1}^N (x_n - 3) + \frac{N}{\sqrt{N}} 3 \rightarrow \infty$ .
- (15)  $\sum_{n=1}^N x_n = N(\frac{1}{N} \sum_{n=1}^N x_n) \rightarrow \infty$
- (16)  $\sum_{n=1}^N (x_n - 3) = \sqrt{N} \frac{1}{\sqrt{N}} \sum_{n=1}^N (x_n - 3)$  is  $\sqrt{N}$  times a quantity converging to a normal random variable. This does not converge to anything, and indeed may take arbitrarily large positive and negative values with non-vanishing probability.

## 6 Bad coding

Suppose you work at Spotify, and a colleague told you that high energy songs are the most popular. You asked for their code, and this is what they sent you. The dataset they load refers to [the Spotify dataset found here](#).

```
data_location <- "../datasets"
df <- read.csv(file.path(data_location, "spotify_songs.csv"))
# run analysis
xx <- df[,c(4,12,13,14,15,16,17,18,19,20,21,22,23)]; xxx =df[,4]
z=sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,2]-sum(xx[,2])/nrow(xx)))/t(xx[,2]-sum(xx[,2])/nrow(xx))
z2=sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,3]-sum(xx[,3])/nrow(xx)))/t(xx[,3]-sum(xx[,3])/nrow(xx))
z3=sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,4]-sum(xx[,4])/nrow(xx)))/t(xx[,4]-sum(xx[,4])/nrow(xx))
z4<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,5]-sum(xx[,5])/nrow(xx)))/t(xx[,5]-sum(xx[,5])/nrow(xx))
z5<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,6]-sum(xx[,6])/nrow(xx)))/t(xx[,6]-sum(xx[,6])/nrow(xx))
z6<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,7]-sum(xx[,7])/nrow(xx)))/t(xx[,7]-sum(xx[,7])/nrow(xx))
z7<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,8]-sum(xx[,8])/nrow(xx)))/t(xx[,8]-sum(xx[,8])/nrow(xx))
z8<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,9]-sum(xx[,9])/nrow(xx)))/t(xx[,9]-sum(xx[,9])/nrow(xx))
z9<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,10]-sum(xx[,10])/nrow(xx)))/t(xx[,10]-sum(xx[,10])/nrow(xx))
z10 <- sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,11]-sum(xx[,11])/nrow(xx)))/t(xx[,11]-sum(xx[,11])/nrow(xx))
z11 <- sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,12]-sum(xx[,12])/nrow(xx)))/t(xx[,12]-sum(xx[,12])/nrow(xx))
yy=(names(df)[c(12,13,14,15,16,17,17,19,20,21,22,23)][order(c(z, z2,z3,z4, z5, z6, z7, z8, z9, z10, z11))])
print(sprintf("The best is %s",yy))
```

- What statistical technique did they use to attempt to answer this question?
- Did they make any mistakes?
- Is it easy to read and think critically about their analysis?
- Would it be easy to re-run their analysis with a slightly different method or dataset?
- Write, in R, a better version of the same analysis that is error-free, more readable, and more reusable.

### Solutions

- They ran simple linear regression with an intercept and tried to find the largest coefficient.
- Yes, several. The sorting is incorrect. The slopes are not invariant to the scale of the regressor. There is an index error in **z6**. (There may be others.)
- It is very difficult to read and understand what they did, much less discover the mistakes.
- No, a slight change in the column order or meaning would invalidate the whole analysis.

### Better code:

Here is an example of better code (but not necessarily better analysis). We will grade flexibly, there are many different ways you can do this.

```

# What matters most for popularity? Explore this question with linear regression.

# Regress track_popularity ~ regressor + 1 for each regressor.
regressors <- c("danceability", "energy", "loudness", "speechiness",
               "acousticness", "instrumentalness", "liveness", "tempo")

# Check that all regressors are present
stopifnot(all(regressors %in% names(df)))

# Run a univariate regression  $y \sim x + 1$  and return the coefficient for  $x$ .
#
# Arguments:
# - regressor: The name of a column in the data frame `df` to use as  $x$ 
# - response: The name of a column in the data frame `df` to use as  $y$ 
# - df: A dataframe containing the data to analyze.
#
# Returns:
# - The estimate of the coefficient of  $x$  in the regression  $y \sim x + 1$ .
RunRegression <- function(regressor, response, dataframe) {
  stopifnot(regressor %in% names(dataframe))
  stopifnot(response %in% names(dataframe))

  x <- df[[regressor]] # Regressor
  y <- df[[response]] # Response

  # Subtract the means to account for regression on a constant
  x_centered <- x - mean(x)
  y_centered <- y - mean(y)

  # Regression estimate using the standard OLS formula
  betahat <- mean(x_centered * y_centered) / mean(x_centered^2)

  return(betahat)
}

# Run the regression for each regressor.
results <- data.frame()
for (regressor in regressors) {
  betahat <- RunRegression(
    regressor=regressor, response="track_popularity", dataframe=df)
  result <- data.frame(regressor=regressor, betahat=betahat)
  results <- bind_rows(results, result)
}

```

```
}
```

```
# Sort by the regression coefficient and summarize the results.  
results <- arrange(results, desc(betahat))  
print(results)  
print(sprintf("%s is the regressor with the largest track popularity coefficient.",  
              results[1, "regressor"]))
```