STAT151A Homework 1 (prerequisites review).

$$\text{Your name here}$$

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.

(1)
$$a_1 + 2a_2 = 1$$
$$a_1 + 3a_2 = 0$$

(2)
$$a_1 + 2a_2 + 3a_3 = 1$$
$$a_1 + 3a_2 + 3a_3 = 0$$
$$2a_1 + 4a_2 + 3a_3 = 5$$

```
a <- matrix(c(1, 1, 2, 2, 3, 4, 3, 3, 3), ncol=3)
det(a)
solve(a, c(c(1, 0, 5)))   # Note that a_3 \ne 0
```

(3)
$$a_1 + 2a_2 + 3a_3 = 1$$
$$a_1 + 3a_2 + 3a_3 = 0$$

(4)
$$a_1 + 2a_2 = 1$$
$$a_1 + 3a_2 = 0$$
$$2a_1 + 4a_2 = 5$$

(5)
$$a_1 + 2a_2 = 1$$

(6)
$$a_1 + 2a_2 = 1$$
$$a_1 + 2a_2 = 1$$
$$a_1 + 2a_2 = 1$$
$$a_1 + 2a_2 = 1$$
$$a_1 + 2a_2 = 1$$

(7)
$$a_1 + 2a_2 = 1$$
$$a_1 + 2a_2 = 2$$
$$a_1 + 2a_2 = 3$$
$$a_1 + 2a_2 = 4$$
$$a_1 + 2a_2 = 5$$

(8)
$$a_1 + 2a_2 = 1$$
$$a_1 + 3a_2 = 1$$
$$a_1 + 4a_2 = 1$$
$$a_1 + 5a_2 = 1$$
$$a_1 + 6a_2 = 5$$

(9)
$$5a_1 + 2a_2 = 1$$
$$10a_1 + 4a_2 = 1$$

(10)
$$5a_1 + 2a_2 = 1$$
$$10a_1 + 4a_2 = 2$$

2 Dimensions of linear algebra expressions

For this problem, I will use the following definitions.

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

- $\boldsymbol{\beta}$ denotes a P–vector

I will take $N > P > 1$. A transpose is denoted with a superscript T, and an inverse by a superscipt -1. A matrix trace is denoted 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 $\underset{P N}{\boldsymbol{X}^{\mathsf{T}}} \underset{N P}{\boldsymbol{X}}$, which we can see is valid because the N's are next to one another. Similarly, we can see immediately that the expression $\underset{N P}{\boldsymbol{X}} \underset{N P}{\boldsymbol{X}}$ is invalid because P is next to N in the matrix multiplication, which is not allowed.

(1) $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}$

(2) $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}$

(3) $\boldsymbol{X} + \boldsymbol{y}$

(4) $\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}}$

(5) $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta}$

(6) $\boldsymbol{y}^{\mathsf{T}}\boldsymbol{y}$

(7) $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}$

(8) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{\mathsf{T}}$

(9) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y})^{\mathsf{T}}$

(10) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}$

(11) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y})^{-1}$

(12) $\boldsymbol{X}^{-1}$

(13) $\boldsymbol{y}^{-1}\boldsymbol{y}$

(14) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}$

(15) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}$

(16) $((\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y})^{\mathsf{T}}$

(17) $\boldsymbol{y}^{\mathsf{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}$

(18) $\boldsymbol{X}\boldsymbol{\beta}$

(19) $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$

(20) $\boldsymbol{y} - \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\beta}$

(21) $\boldsymbol{y}^{\mathsf{T}} - \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{X}^{\mathsf{T}}$

(22) $\boldsymbol{\beta} - (\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}$

(23) $\boldsymbol{X}\left(\boldsymbol{\beta} - (\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}\right)$

(24) $\mathbf{1}^{\mathsf{T}}\boldsymbol{y}$

(25) $\boldsymbol{y} - (\mathbf{1}^{\mathsf{T}}\boldsymbol{y})\boldsymbol{y}$

(26) $\boldsymbol{X}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}$

(27) $\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}$

(28) $\boldsymbol{X}^{\mathsf{T}}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}$

(29) $\boldsymbol{\beta}^{\mathsf{T}}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{\beta}$

(30) $\left(\boldsymbol{\beta}^{\mathsf{T}}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{\beta}\right)^{-1}$

(31) $\operatorname{trace}\left(\boldsymbol{\beta}^{\mathsf{T}}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{\beta}\right)$

(32) $\operatorname{trace}\left((\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}}\right)$

3 Orthonormal vectors and bases

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

(1) Write an expression for the length of $\boldsymbol{a}$ in terms of its entries a_1 and a_2.

(2) Write an expression for the length of $\boldsymbol{a}$, using only matrix operations (i.e., without explicit reference to the entries of $\boldsymbol{a}$).

(3) Write an explicit expression for a vector pointing in the same direction as $\boldsymbol{a}$ but with unit length. (Hint: show that, for a scalar α, the length of $\alpha\boldsymbol{a}$ is α times the length of $\boldsymbol{a}$, and then make a clever choice for α.)

(4) Write an explicit expression for a vector pointing in the same direction as $\boldsymbol{v}$ but with the same length as $\boldsymbol{a}$.

(5) Suppose that I tell you that $\boldsymbol{a} = \alpha\boldsymbol{u} + \gamma\boldsymbol{v}$. Find an explicit expression for α in terms of $\boldsymbol{a}$ and $\boldsymbol{u}$ alone.

(6) Find explicit expressions for scalars α and γ such that $\boldsymbol{a} = \alpha\boldsymbol{u} + \gamma\boldsymbol{v}$.

(7) Let $\begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix} := \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \end{pmatrix}$ denote the 2×2 matrix with $\boldsymbol{u}$ in the first column and $\boldsymbol{v}$ in the second column. Show that $\begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}^{-1}$.

(8) Observe that $\alpha\boldsymbol{u} + \gamma\boldsymbol{v} = \begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}\begin{pmatrix} \alpha \\ \gamma \end{pmatrix}$. Using this, write an explicit expression for the coefficient vector $(\alpha, \gamma)^{\mathsf{T}}$ in terms of $\begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}$ and $\boldsymbol{a}$.

(9) Show that $\begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{u} & \boldsymbol{v} \end{pmatrix}^{\mathsf{T}}$.

(10) Suppose I tell you that that $\boldsymbol{u} = (1, 0)^{\mathsf{T}}$. In terms of a_1 and a_2, what is α in the decomposition $\boldsymbol{a} = \alpha\boldsymbol{u} + \gamma\boldsymbol{v}$?

4 Eigenvalues and eigenvectors of square symmetric matrices

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

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

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

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

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

- Let us prove that the eigenvectors can be taken to be unit vectors without loss of generality.

    - Show that, if $\boldsymbol{u}_k$ is an eigenvector, then $\alpha \boldsymbol{u}_k$ is also an eigenvector with the same eigenvalue as $\boldsymbol{u}_k$ for any scalar $\alpha \neq 0$.
    - In particular, show that $\boldsymbol{u}'_k = \boldsymbol{u}_k / \sqrt{\boldsymbol{u}_k^\mathsf{T} \boldsymbol{u}_k}$ is also an eigenvector with eigenvalue λ_k, and that $(\boldsymbol{u}'_k)^\mathsf{T} \boldsymbol{u}'_k = 1$.

- 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 $\boldsymbol{u}_k^\mathsf{T} \boldsymbol{u}_j = 0$. Carefully justify each step in the following proof.

    - We have $\boldsymbol{u}_j^\mathsf{T} \boldsymbol{A} \boldsymbol{u}_k = (\boldsymbol{u}_j^\mathsf{T} \boldsymbol{A} \boldsymbol{u}_k)^\mathsf{T}$. (Why?)
    - We also have $(\boldsymbol{u}_j^\mathsf{T} \boldsymbol{A} \boldsymbol{u}_k)^\mathsf{T} = \boldsymbol{u}_k^\mathsf{T} \boldsymbol{A}^\mathsf{T} \boldsymbol{u}_j$. (Why?)
    - We then have $\boldsymbol{u}_j^\mathsf{T} \boldsymbol{A} \boldsymbol{u}_k = \boldsymbol{u}_k^\mathsf{T} \boldsymbol{A} \boldsymbol{u}_j$. (Why?)
    - We then have $\lambda_k \boldsymbol{u}_j^\mathsf{T} \boldsymbol{u}_k = \lambda_j \boldsymbol{u}_k^\mathsf{T} \boldsymbol{u}_j$. (Why?)
    - We then have $\boldsymbol{u}_k^\mathsf{T} \boldsymbol{u}_j = 0$. (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 $\alpha \boldsymbol{u}_k + \gamma \boldsymbol{u}_j$ is also an eigenvector with eigenvalue λ_k for any α and γ.

- Suppose that each $\lambda_k \neq 0$. Show that the inverse of Λ is given by the diagonal matrix with $1/\lambda_k$ in the k–th diagonal entry and zero elsewhere.

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

- Assume that $\boldsymbol{A}$ has P orthonormal eigenvectors (it always does; we have proved some parts of this assertion above). We will prove that we then have $\boldsymbol{A} = \boldsymbol{U} \Lambda \boldsymbol{U}^\mathsf{T}$. Carefully justify each step in the following proof.

    - For any $\boldsymbol{a}$, we can write $\boldsymbol{a} = \sum_{p=1}^{P} \alpha_p \boldsymbol{u}_p$ for some scalars α_p. (Why?)
    - Using the previous expansion, we thus have $\boldsymbol{A} \boldsymbol{a} = \sum_{p=1}^{P} \lambda_p \alpha_p \boldsymbol{u}_p$. (Why?)
    - Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_P)^\mathsf{T}$. Then $\boldsymbol{a} = \boldsymbol{U} \boldsymbol{\alpha}$. (Why?)
    - Similarly, we have $\boldsymbol{A} \boldsymbol{a} = \boldsymbol{U} \Lambda \boldsymbol{\alpha}$. (Why?)

- We also have $\boldsymbol{u}_p^\mathsf{T}\boldsymbol{a} = \alpha_p$. (Why?)
- We then have $\boldsymbol{\alpha} = \boldsymbol{U}^\mathsf{T}\boldsymbol{a}$. (Why?)
- Combining, we have $\boldsymbol{A}\boldsymbol{a} = \boldsymbol{U}\Lambda\boldsymbol{U}^\mathsf{T}\boldsymbol{a}$. (Why?)
- Since the preceding expression is true for any $\boldsymbol{a}$, we must have $\boldsymbol{A} = \boldsymbol{U}\Lambda\boldsymbol{U}^\mathsf{T}$. (Why? Hint: if you take $\boldsymbol{a}$ 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.)

5 Statistical asymptotics

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

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

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 $\mathrm{Var}(x_1 + x_2)$.

(8) Compute the variance $\mathrm{Var}\left(\frac{1}{2}(x_1 + x_2)\right)$.

(9) Compute the variance $\mathrm{Var}\left(\frac{1}{3}(x_1 + x_2 + x_3)\right)$.

(10) Compute the variance $\mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)$.

(11) As $N \to \infty$, what (if anything) does $\frac{1}{N}\sum_{n=1}^{N} x_n$ converge to?

(12) As $N \to \infty$, what (if anything) does $\frac{1}{N}\sum_{n=1}^{N}(x_n - 3)$ converge to?

(13) As $N \to \infty$, what (if anything) does $\frac{1}{\sqrt{N}}\sum_{n=1}^{N}(x_n - 3)$ converge to?

(14) As $N \to \infty$, what (if anything) does $\frac{1}{\sqrt{N}}\sum_{n=1}^{N} x_n$ converge to?

(15) As $N \to \infty$, what (if anything) does $\sum_{n=1}^{N} x_n$ converge to?

(16) As $N \to \infty$, 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 $N \to \infty$ 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 $\frac{1}{N}\sum_{n=1}^{N} x_n$?

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
z2=sum((xx[,1]-sum(xx[,1])/nrow(xx))* (xx[,3]-sum(xx[,3])/nrow(xx)))/t(xx[,3]-sum(xx[,3])/nro
z3=sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,4]-sum(xx[,4])/nrow(xx)))/t(xx[,4]-sum(xx[,4])/nrow
z4<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,5]-sum(xx[,5])/nrow(xx)))/t(xx[,5]-sum(xx[,5])/nro
z5<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,6]-sum(xx[,5])/nrow(xx)))/t(xx[,6]-sum(xx[,6])/nro
z6<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,7]-sum(xx[,7])/nrow(xx)))/t(xx[,7]-sum(xx[,2])/nro
z7<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,8]-sum(xx[,8])/nrow(xx)))/t(xx[,8]-sum(xx[,8])/nro
z8<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,9]-sum(xx[,9])/nrow(xx)))/t(xx[,9]-sum(xx[,9])/nro
z9<-sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,10]-sum(xx[,10])/nrow(xx)))/t(xx[,10]-sum(xx[,10])
z10 <- sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,11]-sum(xx[,11])/nrow(xx)))/t(xx[,11]-sum(xx[,1
z11 <- sum((xx[,1]-sum(xx[,1])/nrow(xx))*(xx[,12]-sum(xx[,12])/nrow(xx)))/t(xx[,12]-sum(xx[,1
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
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.