Homework 2

36-462/662, Fall 2019

Due at 10 pm on Wednesday, 11 September 2019

AGENDA: Practice with pieces of linear algebra which are important for dimension reduction, especially for principal components analysis.

Reminders (which you can use without proving):

- The **trace** of a square matrix is the sum of its diagonal entries.
- The trace of a matrix is always equal to the sum of its eigenvalues.
- The eigenvalues of a symmetric matrix are all real (not complex) numbers, and all of the eigenvectors can be chosen to be orthogonal to each other.
- For any (scalar) random variables X and Y, $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + \mathrm{Cov}[X,Y]$. More generally, $\mathrm{Var}\left[\sum_{i=1}^{p} X_i\right] = \sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{Cov}[X_i, X_j]$.
- When $\vec{X}$ is a p-dimensional vector, $\mathrm{Var}\left[\vec{X}\right]$ is a $p \times p$ matrix, where $\mathrm{Var}\left[\vec{X}\right]_{ij} = \mathrm{Cov}[X_i, X_j]$.
- If $\mathbf{a}$ is a $q \times p$ matrix, then $\mathrm{Var}\left[\mathbf{a}\vec{X}\right] = \mathbf{a}\mathrm{Var}\left[\vec{X}\right]\mathbf{a}^T$.

Unless otherwise noted, you can assume:

- every vector $\vec{v}$ is a $p \times 1$ column matrix, so the inner or dot product $\vec{v} \cdot \vec{v}$ could also be written $\vec{v}^T \vec{v}$;
- $\mathbf{1}$ is the $p \times 1$ matrix whose entries are all 1.

1. *Online questions* (10) are online and due at 10 pm on Sunday, 8 September 2019

2. *Idempotency* (5) A matrix $\mathbf{w}$ is called **idempotent** when $\mathbf{w} = \mathbf{w}^2$. Show that every eigenvalue of an idempotent matrix is either 0 or 1. *Hint:* Pretend another eigenvalue was possible, and derive a contradiction.

3. *Projecting on to a line (experiment)*

   (a) (2) Generate 100 random two-dimensional vectors. The exact distribution you use doesn't matter, so long as it's not all concentrated on a single line. Give the command(s) you use to create the vectors, explain why that command generates vectors from that distribution, and a plot showing the locations of the vectors on the plane.

   (b) (3) Multiply all vectors by the matrix $\left[\begin{array}{cc} 1/2 & 1/2 \\ 1/2 & 1/2 \end{array} \right]$. Give the commands you used to do this, and plot the new vectors.

   (c) (2) Check numerically that the matrix from the previous part has two eigenvalues, 0 and 1, and that the eigenvector with eigenvalue 1 is (proportional to) $\left[\begin{array}{c} 1 \\ 1 \end{array} \right]$. *Hint:* `eigen()`.

   (d) (3) Explain, in words, where the new vectors are located, and the geometric relationship between the old vectors and the new. (It may help to draw lines between the old and new vectors.)

4. *Projecting on to a line (theory)*

   (a) (3) Let $\vec{u}$ be an arbitrary vector of length 1, and $\vec{v}$ be any vector, of any length. Explain (in your own words) the meaning of the following statement: "$\vec{u}\vec{u}^T\vec{v}$ projects $\vec{v}$ on to the line defined by $\vec{u}$".

   (b) (2) Show that the $p \times p$ matrix $\mathbf{w} \equiv \vec{u}\vec{u}^T$ is idempotent.

   (c) (2) Show that $\vec{u}$ is an eigenvector of $\mathbf{w}$ with eigenvalue 1.

   (d) (1) Are there any other eigenvectors of $\mathbf{w}$ with non-zero eigenvalues?

   (e) (2) Explain (in your own words) why it makes sense that projecting on to a line should be idempotent and have only one non-zero eigenvalue.

5. *Projecting on to a plane (experiment)* For plotting in three dimensions, you may find the `scatterplot3d` library helpful.

   (a) (3) Generate 100 random vectors in three dimensions. (The exact distribution doesn't matter very much.) Give the command(s) you used to create the vectors, explaining how they create the distribution you want. Plot the points in 3D space.

(b) (2) Multiply all your random vectors by the matrix $\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{bmatrix}$. Give the commands you used to do this, and plot the new vectors in 3D space.

(c) (2) Check numerically that of this matrix's three eigenvalues, two of them equal 1 and the other equals 0. Also check that two of the value-1 eigenvectors are (proportional to) $\begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}$, and $\begin{bmatrix} -1 \\ \sqrt{2} \\ -1 \end{bmatrix}$. (You might get very different-looking eigenvectors from `eigen()` — why?)

(d) (3) Explain, in words, where the new vectors are located, and the geometric relationship between the old vectors and the new. (It may help to draw lines between the old and new vectors.)

6. *Projecting on to a plane* For any two-dimensional plane, we can always find two vectors $\vec{u}_1$ and $\vec{u}_2$ which are in the plane, have length 1, and are orthogonal to each other, $\vec{u}_1 \cdot \vec{u}_2 = 0$. Let $\mathbf{u}$ be the $2 \times p$ matrix whose rows are $\vec{u}_1^T$ and $\vec{u}_2^T$. Finally, let $\mathbf{w} = \mathbf{u}^T \mathbf{u}$.

   (a) (4) Explain, in your own words, the meaning of the following statement: "$\mathbf{w}\vec{v}$ projects $\vec{v}$ on to the plane defined by $\vec{u}_1$ and $\vec{u}_2$".

   (b) (2) Show that $\mathbf{w}$ is idempotent. *Hint:* Use the properties of the vectors $\vec{u}_1$ and $\vec{u}_2$.

   (c) (3) Show that $\mathbf{w}$ has two non-zero eigenvalues, and that $\mathbf{u}_1$ and $\mathbf{u}_2$ are both eigenvectors of eigenvalue 1.

   (We built the projection matrix for *one* choice of two orthogonal vectors lying in plane. Can you show that we get the *same* projection matrix no matter what vectors we start with, so long as they've got length 1, are orthogonal to each other, and lie in the plane?)

7. *Projecting on to a linear subspace* (5) Suppose we want to project on to a q-dimensional linear subspace, $q < p$. Explain how to construct the matrix to do this projection, and why the matrix will be idempotent and have q non-zero eigenvalues.

8. *Linear regression is projection* Suppose we linearly regress an $n \times 1$ vector $\vec{y}$ on an $n \times p$ matrix of regressors $\mathbf{x}$ by ordinary least squares. As you recall from your regression class, the vector of fitted values is $\vec{m} = \mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\vec{y}$. The matrix $\mathbf{x}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T$ is called the **hat matrix**, or $\mathbf{h}$.

   (a) (2) Show that $\mathbf{h}$ is idempotent.

   (b) (3) Show that $\mathbf{h}$ has p non-zero eigenvalues. (Remember that $\mathbf{x}$ is $n \times p$.)

(c) (4) Explain, in your own words, the meaning of the following statement: "The actual response vector is n dimensional, but ordinary least squares projects it into a p-dimensional subspace".

9. *Variance of linear combinations and positive-semi-definite matrices* $X_1, X_2, \ldots X_p$ are random variables, with $\mathrm{Cov}\left[X_i, X_j\right] = c_{ij}$ (and so $\mathrm{Var}\left[X_i\right] = c_{ii}$). We say that the collection of variables has variance (or covariance) matrix $\mathbf{c}$, of dimension $p \times p$.

   (a) (1) Show that $\mathrm{Var}\left[\sum_{i=1}^{p} X_i\right] = \mathbf{1}^T \mathbf{c} \mathbf{1}$.

   (b) (2) Show that $\mathrm{Var}\left[\sum_{i=1}^{p} a_i X_i\right] = \vec{a}^T \mathbf{c} \vec{a}$, for any vector of coefficients $\vec{a}$.

   (c) (3) A matrix $\mathbf{b}$ is called **positive-semi-definite** when for any vector $\vec{v}$, $\vec{v}^T \mathbf{b} \vec{v} \geq 0$. (Some people say **non-negative-definite** instead.) Explain why every *covariance* matrix must be positive-semi-definite.

   (d) (2) If a matrix $\mathbf{b}$ is non-negative definite, and in addition $\vec{v}^T \mathbf{b} \vec{v} > 0$ for all vectors $\vec{v}$, the matrix is called **positive-definite**. Give an example of random variables where the covariance matrix is non-negative definite, but not positive definite. *Hint:* Consider negative correlations between some of the variables, and find a linear combination of the variables with zero variance.

   (e) (2) Show that if a matrix is non-negative definite, then all its eigenvalues are ≥ 0.

   (f) (2) Show that if a matrix is positive-definite, then all its eigenvalues are > 0.

   (g) (1) Suppose every linear combination of the variables has a variance > 0. Show that all the eigenvalues are > 0.

10. *Correlating and decorrelating by matrix multiplication* Suppose a matrix $\mathbf{b}$ is a symmetric, positive-definite matrix with eigendecomposition $\mathbf{b} = \mathbf{v}\mathbf{d}\mathbf{v}^T$, where $\mathbf{d}$ is the diagonal matrix of eigenvalues, and $\mathbf{v}$ is the matrix whose columns are the eigenvectors.

   (a) (3) Define $\mathbf{b}^{1/2} = \mathbf{v}\mathbf{d}^{1/2}$, where $\mathbf{d}^{1/2}$ is the diagonal matrix with the (positive) square roots of eigenvalues of $\mathbf{b}$ on the diagonal. Show that $\mathbf{b} = (\mathbf{b}^{1/2})(\mathbf{b}^{1/2})^T$.

   (b) (3) Suppose that $\vec{Z}$ is a p-dimensional random vector where each dimension has variance 1 and the dimensions are uncorrelated, so $\mathrm{Var}\left[\vec{Z}\right] = \mathbf{I}$. Show that $\mathrm{Var}\left[\mathbf{b}^{1/2}\vec{Z}\right] = \mathbf{b}$.

   (c) (3) Suppose that $\mathrm{Var}\left[\vec{X}\right] = \mathbf{c}$, where $\mathbf{c}$ is positive-definite. Show that $\mathrm{Var}\left[\mathbf{b}^{-1/2}\vec{X}\right] = \mathbf{I}$.

RUBRIC (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. Questions which ask for a plot or table are answered with both the figure itself and the command (or commands) use to make the plot. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision. Code is properly integrated with a tool like R Markdown or knitr, and both the knitted file and the source file are submitted. The code is indented, commented, and uses meaningful names. All code is relevant; there are no dangling or useless commands. All parts of all problems are answered with actual coherent sentences, and raw computer code or output are only shown when explicitly asked for.