HOMEWORK 1, STAT 961: Lin. Alg. 1
Due Wed 2019/09/18, 2:00pm

Your Name: ... (replace this)

September 17, 2019

Instructions: Edit this LaTex file by inserting your solutions after each problem statement. Generate a PDF file from it and e-mail the PDF in an attachment with filename

hw01-Yourlastname-Yourfirstname.pdf

to the following class Gmail address:

stat961.at.wharton@gmail.com

with **subject line exactly** as follows for easy gmail search:

Homework 1, 2019

Rules to be strictly followed under honor code:

1. You must write your own solutions and not copy from anyone. Verbatim copying from others or unlisted sources, no matter how minor, will result in zero points for the whole homework.
2. Subject to the previous item, you may explain the problems, but not the solutions, to each other in general terms.
3. Do not discuss the homework with previous years' students of Stat 961/541.
4. Do not consult solutions of similar homeworks of previous years.
5. Report here who you collaborated with and what sources you used. (You do not need to report help with LaTex and English language.)
   - **My collaborators:** ... (replace this)

- **The complete list of my sources is as follows:** ... (replace this)

Instructions for presentation and typesetting:

1. Give derivations where appropriate, but don't when instructed to give the answer without derivation.

2. LaTex math you will need: Symbols for matrices and vectors are locally defined, in particular $\mathbf{X}$, $\mathbf{Y}$ for general matrices, $\mathbf{H}$ for projection matrices, $\mathbf{I}_{...}$ for identity matrices, $\mathbf{0}_{...}$ for zero vectors and matrices, and $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{a}$, $\mathbf{b}$ for vectors.
   Geometry: $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T \mathbf{b}$ for inner products, $\|\mathbf{a}\|^2 = \mathbf{a}^T \mathbf{a}$ for squared Euclidean norms.
   Range space and null space of a linear map $\mathbf{H}$: $\mathbf{Range}(\mathbf{H})$, $\mathbf{Null}(\mathbf{H})$.

3. To mimic obvious R functions in LaTex math mode we use $\mathrm{cbind}(...)$, $\mathrm{sum}(\mathbf{a})$ and $\mathrm{mean}(\mathbf{a})$.

4. Write actual R code in inline verbatim mode: `abc %*% xyz`. This avoids conflicts between control characters of the R and LaTex languages, as in this example: R's matrix multiplication `%*%` would require backslashes \ to show the percentage signs in LaTex, but inside verbatim mode the percentage sign is shown without. Normally the percentage sign is a control character in LaTex to start a comment to the end of the line.
   Note that inline verbatim mode lets you choose the beginning and ending delimiters: `abc %*% xyz` does the same thing. (If you read the PDF, you can't see the difference; read the LaTex source.)

Problems for Review of Matrix Algebra, in Particular Projection Matrices

Homework 1, Stat 961, 2019C

1. Two interpretations of matrix products:

   (a) First, consider a $m \times p$ matrix $\mathbf{X} = \text{cbind}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p)$ and a $n \times p$ matrix $\mathbf{Y} = \text{cbind}(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_p)$, with columns $\mathbf{x}_j \in \mathbb{R}^m, \mathbf{y}_j \in \mathbb{R}^n$. Explain to yourself that the matrix product $\mathbf{X}\mathbf{Y}^T$ is well-defined and what its size is (nothing to write).

   Problems: Express $\mathbf{X}\mathbf{Y}^T$ in terms of the columns. The answer will be a sum of simpler matrices. What are these simple matrices called? What R function computes such simple matrices? (No derivations, please, just answers.)

      i. $\mathbf{X}\mathbf{Y}^T = ...$

      ii. The summand matrices are called ...

      iii. R code for summand matrix j using `X[,j]` and `Y[,j]`:

      $\ldots$

   (b) Next, consider a $n \times p$ matrix $\mathbf{X} = \text{cbind}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p)$ and an $n \times q$ matrix $\mathbf{Y} = \text{cbind}(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_q)$, with columns $\mathbf{x}_j, \mathbf{y}_k \in \mathbb{R}^n$. This time, explain to yourself that $\mathbf{X}^T\mathbf{Y}$ is well-defined and what its size is (nothing to write).

   Problems: Express the element (j,k) of $\mathbf{X}^T\mathbf{Y}$ in terms of the column vectors, explain its meaning, and write R code to compute it. (Again, no derivations, just answers.)

      i. Express the element (j,k) of $\mathbf{X}^T\mathbf{Y}$ in terms of the columns:

      $$(\mathbf{X}^T\mathbf{Y})_{j,k} = ...$$

      ii. What is the geometric meaning of $(\mathbf{X}^T\mathbf{Y})_{j,k}$?

      ...

      iii. R code for computing $(\mathbf{X}^T\mathbf{Y})_{j,k}$ using `X[,j]` and `Y[,k]`:

      $\ldots$

2. With a $n \times p$ matrix $\mathbf{X} = \text{cbind}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p)$ and a $n \times q$ matrix $\mathbf{Y} = \text{cbind}(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_q)$ as in the previous problem (b), what do the following conditions mean?

   (a) $\mathbf{X}^T\mathbf{Y} = \mathbf{0}_{p\times q}$: ...

(b) $\mathbf{X}^T\mathbf{X} = \mathbf{I}_{p\times p}$: ...

3. Let $\mathbf{x}$ and $\mathbf{y}$ be n-vectors, and consider simple regression through the origin of $\mathbf{y}$ on $\mathbf{x}$, no intercept, that is, $\mathbf{X} = \mathbf{x}$ is of size $n \times 1$. For this special case do the following (no derivations, just answers):

   (a) Write down the triple-$\mathbf{X}$ matrix. Indicate the size of the triple-$\mathbf{X}$ matrix in the form $a \times b$. Use squared norm notation.

   $$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T ... \qquad \text{Size: } ...$$

   (b) Write the single OLS coefficient/slope estimate $\hat{\beta}$ by matrix-multiplying the triple-$\mathbf{X}$ matrix with the response vector $\mathbf{y}$. Show the result using inner product and squared norm notation.

   $$\hat{\beta} = ...$$

   (c) Write down the quadruple-$\mathbf{X}$ matrix, again using squared norm notation. Of what type is this matrix?

   $$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = ... \qquad \text{Type: } ...$$

   (d) Think like a physicist: What are the units of the elements of this quadruple-$\mathbf{X}$ matrix? Explain. (To make it concrete, you may assume the units of $\mathbf{x}$ are US dollars.) What happens to it if the units are changed, from US dollars to Euros, say?

      i. The elements ...

      ii. Explanation: ...

      iii. Under a change of units of $\mathbf{x}$, the quadruple-$\mathbf{X}$ matrix ...

   (e) Construct another vector $\tilde{\mathbf{x}}$ from $\mathbf{x}$ that has the same units as the quadruple-$\mathbf{X}$ matrix and also generates the same quadruple-$\mathbf{X}$ matrix (no derivations).

   $$\tilde{\mathbf{x}} = ...$$

   Quadruple-$\mathbf{X}$ matrix in terms of $\tilde{\mathbf{x}}$: ...

   (f) Consider next the case that $\mathbf{x} = \mathbf{e} = (1, 1, ..., 1)^T$. What is the OLS coefficient estimate $\hat{\beta}$ for the response $\mathbf{y}$ when the single regressor is $\mathbf{x} = \mathbf{e}$? Give a short derivation in which you translate geometric concepts to R functions (but in LaTex math mode, not verbatim mode).

$$\hat{\beta} = ... \quad \text{because} \quad \|\mathbf{e}\|^2 = ...$$

(g) What is $\tilde{\mathbf{e}}$, the special case of $\tilde{\mathbf{x}}$ where $\mathbf{x} = \mathbf{e}$, from the question before?

$$\tilde{\mathbf{e}} = ...$$

4. Back to multiple regression: We now assume $\mathbf{X} = \text{cbind}(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_p)$ is of size $n \times (p+1)$ and the first column is $\mathbf{x}_0 = \tilde{\mathbf{e}}$ for a rescaled intercept coefficient. Consider the case that $\mathbf{X}^T\mathbf{X} = \mathbf{I}_{(p+1)\times(p+1)}$.

   (a) Triple-$\mathbf{X}$ matrix = ...

   (b) $\hat{\beta}_j = ...$

   (c) Quadruple-$\mathbf{X}$ matrix = ...

5. Back to a general $\mathbf{X}$ matrix, only assuming it has full rank. Show that its quadruple-$\mathbf{X}$ matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is an orthogonal projection, that is, it is idempotent and symmetric.

$$\begin{aligned}
\mathbf{HH} &= ... \\
&= ... \\
&= ... \\
&= \mathbf{H}
\end{aligned}$$

$$\begin{aligned}
\mathbf{H}^T &= ... \\
&= ... \\
&= ... \\
&= ... \\
&= \mathbf{H}
\end{aligned}$$

6. Show that if $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$ satisfy $\mathbf{X}_1^T\mathbf{X}_2 = \mathbf{0}_{p_1\times p_2}$, then the sum of the corresponding hat matrices, $\mathbf{H}_1 + \mathbf{H}_2$, is also an orthogonal projection.

Do not show any calculations; explain the reasons in English. (You may use inline notation for simple expressions such as $\mathbf{H}_1\mathbf{H}_2$ in your explanations.)

Idempotence: ...

Symmetry: ...

Make a Guess: Is $\mathbf{H}_1 + \mathbf{H}_2$ also a quadruple-$\mathbf{X}$ matrix? If so, what is this $\mathbf{X}$? (No derivations, just a guess.)

Answer: : ...

7. If $\mathbf{H}$ $(n \times n)$ is idempotent and symmetric, does the same hold for $\mathbf{I} - \mathbf{H}$? Show derivations.

   Idempotence: ...

   Symmetry: ...

8. If $\mathbf{H}$ and $\mathbf{I} - \mathbf{H}$ are matrix-multiplied in either order, what's the result?

$$\mathbf{H}(\mathbf{I} - \mathbf{H}) = ...$$
$$(\mathbf{I} - \mathbf{H})\mathbf{H} = ...$$

9. Geometrically, idempotence means projection: dropping/projecting a point $\mathbf{x}$ into a subspace, $\mathbf{x} \mapsto \mathbf{H}\mathbf{x} \in \mathbf{Range}(\mathbf{H})$, and if the dropping/projecting is repeated, then the point $\mathbf{H}\mathbf{x}$ does not move because it's already in that subspace. Call $\mathbf{x} - \mathbf{H}\mathbf{x} = (\mathbf{I} - \mathbf{H})\mathbf{x}$ the (reverse) projection direction. Show that this direction is in the null space $\mathbf{Null}(\mathbf{H}) = \{\mathbf{z} | \mathbf{H}\mathbf{z} = \mathbf{0}\}$. No derivation, just an argument based on the previous question.

   Answer: ...

10. The mystery that symmetry and orthogonality of projections are equivalent: Intuitively, orthogonality of a projection means that points get dropped into a subspace such that the direction of dropping/projecting is orthogonal to the subspace. Now, the subspace that $\mathbf{H}$ projects onto is $\mathbf{Range}(\mathbf{H})$, and the (reverse) direction of dropping/projecting for a point $\mathbf{x}$ is $\mathbf{x} - \mathbf{H}\mathbf{x}$. Thus to understand orthogonality of projection, we have to understand what it means that $\mathbf{x} - \mathbf{H}\mathbf{x} = (\mathbf{I} - \mathbf{H})\mathbf{x}$ is orthogonal to *all* points in $\mathbf{Range}(\mathbf{H})$, that is,

$(\mathbf{I}-\mathbf{H})\mathbf{x} \perp \mathbf{H}\mathbf{z}$ for *all* $\mathbf{x}$ and *all* $\mathbf{z}$ in $\mathbb{R}^n$:

$$\begin{aligned} 0 &= ... \\ &= ... \\ &= ... \end{aligned}$$

It follows that $\mathbf{H} =$ .... But any matrix of the form ... is symmetric! Hence $\mathbf{H}$ is symmetric.