

Homework 2, Stat 541: Lin. Alg., Due Fri, Sept 23, 2010, 5pm

Your Name: ... (replace this)

September 24, 2011

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 "**hw02-Yourlastname-Yourfirstname.pdf**" to the usual class gmail address with "**Homework 2, 2011**" in the subject line:

stat541.at.wharton@gmail.com

You should not just answer the questions but give at least partial proofs where appropriate. You can discuss the homework with each other in general terms, but not with previous years' students of Stat 541. You must not consult solutions of similar homeworks of previous years. Finally, 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. Report here who you collaborated with and what sources you used:

My collaborators: ... (replace this)

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

The Problems

1. Consider two matrices \mathbf{X} and \mathbf{Y} , where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ is $n \times p$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q)$ is $n \times q$.
 - (a) What size is $\mathbf{X}^T \mathbf{Y}$?
 - (b) Write the (i, j) -entry $(\mathbf{X}^T \mathbf{Y})_{i,j}$ in terms of \mathbf{x}_i and \mathbf{y}_j .
 - (c) Interpretation?

- (d) Using this interpretation describe $\mathbf{X}^T \mathbf{Y}$ in English.
- (e) What would it mean if $\mathbf{X}^T \mathbf{Y} = \mathbf{0}$?
- (f) If this were the case and if both matrices were of full rank, what constraint would this impose on the dimensions n , p and q ?
- (g) Finally, write R code to calculate the entry $(\mathbf{X}^T \mathbf{Y})_{i,j}$ and to calculate the whole matrix $\mathbf{X}^T \mathbf{Y}$.

A:

- (a) The size of $\mathbf{X}^T \mathbf{Y}$ is $p \times q$.
- (b) the (i, j) -entry can be written as $(\mathbf{X}^T \mathbf{Y})_{i,j} = \mathbf{x}_i^T \mathbf{y}_j$.
- (c) Interpretation: $\mathbf{x}_i^T \mathbf{y}_j$ is the inner product of \mathbf{x}_i and \mathbf{y}_j .
- (d) The matrix $\mathbf{X}^T \mathbf{Y}$ contains all possible inner products between columns of \mathbf{X} and columns of \mathbf{Y} .
- (e) If $\mathbf{X}^T \mathbf{Y} = \mathbf{0}$, it would mean that the column spaces of \mathbf{X} and \mathbf{Y} are orthogonal to each other.
- (f) It would imply that $p + q \leq n$ because
- (g) R code for the (i, j) entry: `sum(X[,i]*Y[,j])`
for $\mathbf{X}^T \mathbf{Y}$: `t(X) %*% Y` or `crossprod(X,Y)`.

2. Definition: If $\mathbf{a} = (a_1, \dots, a_m)^T$ is a $m \times 1$ vector and $\mathbf{b} = (b_1, \dots, b_n)^T$ a $n \times 1$ vector (both $\neq \mathbf{0}$), then $\mathbf{M} = \mathbf{a}\mathbf{b}^T$ is called a rank-one matrix.

- (a) What size is \mathbf{M} ?
- (b) What is the (i, j) entry $\mathbf{M}_{i,j}$?
- (c) Based on this, describe this matrix in English.
- (d) What is the j 'th column of \mathbf{M} ?
- (e) What is the i 'th row?
- (f) Why is \mathbf{M} of rank one?
- (g) Is every rank-one matrix of this form?

- (h) If \mathbf{x} is a $n \times 1$ vector, what is $M\mathbf{x}$?
- (i) Describe $M\mathbf{x}$ in words.
- (j) If $N = \mathbf{xy}^T$ is another rank-one matrix, of size $n \times r$, what is MN (interpret the result)?

A:

- (a) The size of \mathbf{M} is $m \times n$.
 - (b) The (i, j) entry is $\mathbf{M}_{i,j} = a_i b_j$.
 - (c) The matrix \mathbf{M} contains all possible products of entries of \mathbf{a} and those of \mathbf{b} .
 - (d) The j 'th column is $\mathbf{a}b_j$.
 - (e) The i 'th row is $a_i \mathbf{b}^T$.
 - (f) The rank of \mathbf{M} is one because the column space is spanned by \mathbf{a} alone: all columns of \mathbf{M} are multiples of \mathbf{a} .
 - (g) Yes, all rank-one matrices are of this form: If the column space is one-dimensional, spanned by a vector \mathbf{a} , say, then each column is a multiple of \mathbf{a} , call it $\mathbf{a}b_j$. Hence the matrix is of the form $\mathbf{a}\mathbf{b}^T$.
 - (h) We have $M\mathbf{x} = (\mathbf{a}\mathbf{b}^T)\mathbf{x} = \mathbf{a}(\mathbf{b}^T\mathbf{x}) = (\mathbf{b}^T\mathbf{x})\mathbf{a}$, where the second equality holds due to associativity of matrix multiplication. The last equality holds because we are multiplying the vector \mathbf{a} with an inner product $\mathbf{b}^T\mathbf{x}$, which is just a number.
 - (i) $M\mathbf{x}$ equals \mathbf{a} multiplied with the inner product of \mathbf{b} and \mathbf{x} .
 - (j) We have $MN = (\mathbf{a}\mathbf{b}^T)(\mathbf{xy}^T) = \mathbf{a}(\mathbf{b}^T\mathbf{x})\mathbf{y}^T = (\mathbf{b}^T\mathbf{x})(\mathbf{a}\mathbf{y}^T)$, where the second equality uses associativity of matrix multiplication, and the third equality simply moves a scalar out of the remaining matrix product. The scalar $\mathbf{b}^T\mathbf{x}$ is the inner product of \mathbf{b} and \mathbf{x} , and $\mathbf{a}\mathbf{y}^T$ is a rank-one matrix of size $m \times r$.
3. Let the matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$ be of size $m \times p$ and the matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p)$ of size $n \times p$. Consider $\mathbf{M} = \mathbf{XY}^T$.
- (a) Of what size is \mathbf{M} ?
 - (b) Write \mathbf{M} as a sum of rank-one matrices.

- (c) If \mathbf{z} is a $n \times 1$ vector, expand \mathbf{Mz} using (b) and describe in words what happens.
- (d) Write R code that produces \mathbf{M} .

A:

- (a) The matrix \mathbf{M} is of size $m \times n$.
- (b) We have $\mathbf{M} = \mathbf{XY}^T = \sum_{j=1, \dots, p} \mathbf{x}_j \mathbf{y}_j^T$.
- (c) We have $\mathbf{Mz} = \mathbf{XY}^T \mathbf{z} = \sum_{j=1, \dots, p} (\mathbf{x}_j \mathbf{y}_j^T) \mathbf{z} = \sum_{j=1, \dots, p} \mathbf{x}_j (\mathbf{y}_j^T \mathbf{z}) = \sum_{j=1, \dots, p} (\mathbf{y}_j^T \mathbf{z}) \mathbf{x}_j$.
In words: We form a linear combination of the \mathbf{x}_j vectors using the inner products of \mathbf{z} and \mathbf{y}_j as coefficients.
- (d) The R code for \mathbf{M} : `X %*% t(Y)` or `tcrossprod(X, Y)`

4. If \mathbf{A} is a $n \times n$ matrix, when does it satisfy $\langle \mathbf{Ax}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{Ay} \rangle$ for all \mathbf{x} and all \mathbf{y} ?

A: $\langle \mathbf{Ax}, \mathbf{y} \rangle = (\mathbf{Ax})^T \mathbf{y} = \mathbf{x}^T \mathbf{A}^T \mathbf{y}$ and $\langle \mathbf{x}, \mathbf{Ay} \rangle = \mathbf{x}^T \mathbf{Ay}$, hence we have equality iff (“if and only if”) $\mathbf{A}^T = \mathbf{A}$, that is, if \mathbf{A} is symmetric. — Alternatively, you can plug in all basis vectors $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$ to see that $A_{ji} = A_{ij}$.

5. Write the Euclidean inner product as $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$ and the Euclidean squared norm as $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$. Define

$$\mathbf{y} \mapsto \mathbf{L}(\mathbf{y}) = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{x}\|^2} \mathbf{x} ?$$

- (a) Show the mapping \mathbf{L} is linear in \mathbf{y} .
- (b) Write down its matrix and call it \mathbf{P} .
- (c) Does this mapping remind you of anything you have seen in simple linear regression?
- (d) What happens to \mathbf{L} if we replace \mathbf{x} with any non-zero multiple $c\mathbf{x}$ ($c \neq 0$)?
- (e) What does (d) mean (1) geometrically and (2) in terms of a unit change, say, when \mathbf{x} is re-expressed in Euros instead of US-Dollars?

A:

- (a) $\mathbf{L}(\mathbf{y}_1 + \mathbf{y}_2) = \mathbf{L}(\mathbf{y}_1) + \mathbf{L}(\mathbf{y}_2)$ and $\mathbf{L}(c\mathbf{y}) = c\mathbf{L}(\mathbf{y})$ are obvious because \mathbf{y} sits inside an inner product.
- (b) $\mathbf{P} = \mathbf{xx}^T / \|\mathbf{x}\|^2$

- (c) The scalar $\langle \mathbf{y}, \mathbf{x} \rangle / \|\mathbf{x}\|^2$ looks like the formula for the slope in a simple linear regression.
 - (d) \mathbf{L} does not change if \mathbf{x} is replaced with $c\mathbf{x}$ because c cancels out in the definition of \mathbf{L} .
 - (e) Geometrically this means \mathbf{L} only depends on the direction, not the size of \mathbf{x} . It also means \mathbf{L} is unchanged under unit changes.
6. Continuation of the previous problem: Simplify by normalizing \mathbf{x} to unit length and assume from now on $\|\mathbf{x}\|^2 = 1$.
- (a) Write \mathbf{P} again after the simplification and describe it in terms introduced earlier.
 - (b) Show that \mathbf{P} is “**idempotent**”: $\mathbf{P}^2 = \mathbf{P}$.
 - (c) Show that \mathbf{P} is symmetric.
 - (d) Intuitively what does idempotence mean? Reason in terms of the column space of \mathbf{P} .
 - (e) Show that $\mathbf{P}\mathbf{y}$ and $\mathbf{y} - \mathbf{P}\mathbf{y}$ are orthogonal to each other.
 - (f) \mathbf{P} is called an “**orthogonal projection**”. What does \mathbf{P} “project” onto?
 - (g) Why “orthogonal”?

A:

- (a) $\mathbf{P} = \mathbf{x}\mathbf{x}^T$, which is a rank-one matrix.
 - (b) $\mathbf{P}^2 = (\mathbf{x}\mathbf{x}^T)(\mathbf{x}\mathbf{x}^T) = \mathbf{x}(\mathbf{x}^T\mathbf{x})\mathbf{x}^T = \mathbf{x}\mathbf{x}^T = \mathbf{P}$ because $\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2 = 1$ by assumption.
 - (c) $P_{i,j} = x_i x_j = x_j x_i = P_{j,i}$
 - (d) Idempotence means that elements $\mathbf{P}\mathbf{y}$ in the column space of \mathbf{P} stay fixed under \mathbf{P} : $\mathbf{P}(\mathbf{P}\mathbf{y}) = \mathbf{P}\mathbf{y}$. That is, the column space is a fixed-point space.
 - (e) Orthogonality: $\langle \mathbf{P}\mathbf{y}, \mathbf{y} - \mathbf{P}\mathbf{y} \rangle = (\mathbf{P}\mathbf{y})^T(\mathbf{y} - \mathbf{P}\mathbf{y}) = \mathbf{y}^T\mathbf{P}(\mathbf{y} - \mathbf{P}\mathbf{y}) = \mathbf{y}^T\mathbf{P}\mathbf{y} - \mathbf{y}^T\mathbf{P}^2\mathbf{y} = 0$.
 - (f) \mathbf{P} projects onto the one-dimensional space spanned by \mathbf{x} (that is, the set of multiples of \mathbf{x}).
 - (g) “Orthogonal” because what is “taken out” from \mathbf{y} , namely $\mathbf{y} - \mathbf{P}\mathbf{y}$, is orthogonal to $\mathbf{P}\mathbf{y}$.
7. Assume \mathbf{x} and \mathbf{y} are both unit length and orthogonal to each other. Define $\mathbf{P} = \mathbf{x}\mathbf{x}^T + \mathbf{y}\mathbf{y}^T$.
- (a) Is \mathbf{P} symmetric? Do you need unit length and orthogonality?

(b) Is \mathbf{P} idempotent? Do you need unit length and orthogonality?

A:

(a) Symmetry of \mathbf{P} holds always, without unit length and orthogonality are not needed (too trivial to show).

(b) Idempotence of \mathbf{P} :

$$\begin{aligned}\mathbf{P}\mathbf{P} &= \mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T + \mathbf{x}\mathbf{x}^T\mathbf{y}\mathbf{y}^T + \mathbf{y}\mathbf{y}^T\mathbf{x}\mathbf{x}^T + \mathbf{y}\mathbf{y}^T\mathbf{y}\mathbf{y}^T \\ &= \mathbf{x}\mathbf{x}^T + \mathbf{x}(\mathbf{x}^T\mathbf{y})\mathbf{y}^T + \mathbf{y}(\mathbf{y}^T\mathbf{x})\mathbf{x}^T + \mathbf{y}\mathbf{y}^T \\ &= \mathbf{P} + \mathbf{x}(\mathbf{x}^T\mathbf{y})\mathbf{y}^T + \mathbf{y}(\mathbf{y}^T\mathbf{x})\mathbf{x}^T \\ &= \mathbf{P} + \langle \mathbf{x}, \mathbf{y} \rangle (\mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T)\end{aligned}$$

We used unit length, $\mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{y} = 1$, to get here, and with orthogonality we see that $\langle \mathbf{x}, \mathbf{y} \rangle (\mathbf{x}\mathbf{y}^T + \mathbf{y}\mathbf{x}^T)$ vanishes. Hence idempotence holds if we assume unit lengths and orthogonality.

8. Assume \mathbf{P} ($n \times n$) is idempotent and symmetric.

(a) Is $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ also idempotent and symmetric?

(b) Show that $\mathbf{P}\mathbf{Q} = \mathbf{0}$.

(c) Show that $\mathbf{P}\mathbf{y}$ and $\mathbf{Q}\mathbf{y} = \mathbf{y} - \mathbf{P}\mathbf{y}$ are orthogonal to each other.

(d) If the column space of \mathbf{P} is of dimension p , then \mathbf{P} is called a “ p -dimensional orthogonal projection”. Again, what is \mathbf{P} projecting onto, and why “orthogonal”?

(e) Show that the null space and the column space of \mathbf{P} are orthogonal to each other.

(f) Show that the null space of \mathbf{P} is the column space of \mathbf{Q} .

(g) Show that the null space of \mathbf{Q} is the column space of \mathbf{P} .

(h) Show that any vector that is orthogonal to the column space of \mathbf{P} is in the null space of \mathbf{P} .

A:

(a) Symmetry of \mathbf{Q} is trivial: $(\mathbf{I} - \mathbf{P})^T = \mathbf{I}^T - \mathbf{P}^T = \mathbf{I} - \mathbf{P}$.

Idempotence holds also: $(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}\mathbf{P} = \mathbf{I} - \mathbf{P}$.

- (b) $\mathbf{PQ} = \mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{P} - \mathbf{P}^2 = \mathbf{0}$.
- (c) $\langle \mathbf{P}\mathbf{y}, \mathbf{Q}\mathbf{y} \rangle = \langle \mathbf{Q}\mathbf{P}\mathbf{y}, \mathbf{y} \rangle = \langle \mathbf{0}, \mathbf{y} \rangle = 0$
- (d) \mathbf{P} projects onto its column space, whatever that is. “Orthogonal” because what \mathbf{P} “takes out” from \mathbf{y} is $\mathbf{Q}\mathbf{y} = \mathbf{y} - \mathbf{P}\mathbf{y}$, which is orthogonal to \mathbf{y} .
- (e) If \mathbf{y}_0 is in the null space, that is, $\mathbf{P}\mathbf{y}_0 = \mathbf{0}$, and if $\mathbf{z} = \mathbf{P}\mathbf{y}_1$ is in the column space, then $\langle \mathbf{y}_0, \mathbf{z} \rangle = \langle \mathbf{y}_0, \mathbf{P}\mathbf{y}_1 \rangle = \langle \mathbf{P}\mathbf{y}_0, \mathbf{y}_1 \rangle = \langle \mathbf{0}, \mathbf{y}_1 \rangle = 0$.
- (f) \mathbf{y} is in the null space of \mathbf{P} iff $\mathbf{P}\mathbf{y} = \mathbf{0}$ iff $\mathbf{y} - \mathbf{P}\mathbf{y} = \mathbf{y}$ iff $\mathbf{Q}\mathbf{y} = \mathbf{y}$ iff \mathbf{y} is in the column space of \mathbf{Q} .
- (g) \mathbf{y} is in the column space of \mathbf{P} iff $\mathbf{P}\mathbf{y} = \mathbf{y}$ iff $\mathbf{y} - \mathbf{P}\mathbf{y} = \mathbf{0}$ iff $\mathbf{Q}\mathbf{y} = \mathbf{0}$ iff \mathbf{y} is in the null space of \mathbf{Q} .
- (h) If \mathbf{y} is orthogonal to the column space of \mathbf{P} , then it is in particular orthogonal to $\mathbf{P}\mathbf{y}$: $0 = \langle \mathbf{y}, \mathbf{P}\mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{P}\mathbf{P}\mathbf{y} \rangle = \langle \mathbf{P}\mathbf{y}, \mathbf{P}\mathbf{y} \rangle = \|\mathbf{P}\mathbf{y}\|^2$, hence $\mathbf{P}\mathbf{y} = \mathbf{0}$, hence \mathbf{y} is in the null space of \mathbf{P} .

9. Remedial material about basis changes and associated coordinate transformations, and the meaning of similar matrices:

Consider two bases of n -space: $\mathbf{e}_1^{old}, \dots, \mathbf{e}_n^{old}$ and $\mathbf{e}_1^{new}, \dots, \mathbf{e}_n^{new}$. For any vector \mathbf{y} we have coordinates in both bases:

$$\mathbf{y} = \sum_i y_i^{new} \mathbf{e}_i^{new} = \sum_j y_j^{old} \mathbf{e}_j^{old} \quad (1)$$

Collect the respective coordinates in two coordinate vectors: $\mathbf{y}^{new} = (y_1^{new}, \dots, y_n^{new})^T$ and $\mathbf{y}^{old} = (y_1^{old}, \dots, y_n^{old})^T$. To link the two types of coordinates to each other, we express the old basis in terms of the new basis as follows:

$$\mathbf{e}_j^{old} = \sum_i T_{i,j} \mathbf{e}_i^{new} \quad (2)$$

Collect the coefficients in a $n \times n$ matrix $\mathbf{T} = (T_{i,j})$.

- (a) What does \mathbf{T} contain in its j 'th column?
- (b) Express \mathbf{y}^{new} in terms of \mathbf{T} and \mathbf{y}^{old} by substituting (2) in (1).
- (c) A linear map $\mathbf{z} = A(\mathbf{y})$ is expressed in terms of different matrices \mathbf{A}^{old} and \mathbf{A}^{new} in the two coordinate systems: $\mathbf{z}^{old} = \mathbf{A}^{old}\mathbf{y}^{old}$ and $\mathbf{z}^{new} = \mathbf{A}^{new}\mathbf{y}^{new}$. How are \mathbf{A}^{old}

and \mathbf{A}^{new} related to each other? (If two matrices are in this relation to each other, they are called “similar”.)

A:

- (a) \mathbf{T} contain in its j 'th column the coordinates of \mathbf{e}_j^{old} in the new basis $\mathbf{e}_1^{new}, \dots, \mathbf{e}_n^{new}$.
- (b) $\mathbf{y} = \sum_j y_j^{old} \mathbf{e}_j^{old} = \sum_j y_j^{old} (\sum_i T_{i,j} \mathbf{e}_i^{new}) = \sum_i (\sum_j T_{i,j} y_j^{old}) \mathbf{e}_i^{new} = \sum_i y_i^{new} \mathbf{e}_i^{new}$. Comparison yields $y_i^{new} = \sum_j T_{i,j} y_j^{old}$.
- (c) Comparison of $\mathbf{z}^{new} = \mathbf{A}^{new} \mathbf{y}^{new} = \mathbf{A}^{new} \mathbf{T} \mathbf{y}^{old}$ and $\mathbf{z}^{new} = \mathbf{T} \mathbf{z}^{old} = \mathbf{T} \mathbf{A}^{old} \mathbf{y}^{old}$ yields $\mathbf{A}^{new} = \mathbf{T} \mathbf{A}^{old} \mathbf{T}^{-1}$.

10. We know from two problems ago that if \mathbf{P} is an orthogonal projection then the null space is the orthogonal complement of the column space. Assume \mathbf{P} is of rank p , that is, the column space is of dimension p . Now introduce an orthonormal basis $\mathbf{e}_1, \dots, \mathbf{e}_p, \mathbf{e}_{p+1}, \dots, \mathbf{e}_n$ as follows: $\mathbf{e}_1, \dots, \mathbf{e}_p$ forms a basis of the column space of \mathbf{P} , and $\mathbf{e}_{p+1}, \dots, \mathbf{e}_n$ forms a basis of its null space. In this basis, what is the matrix of \mathbf{P} ? You don't need to do complex typesetting; write down what the effect of \mathbf{P} on the basis vectors is, and describe in English what the matrix looks like. [Note that you really don't need the previous material of basis changes. We are not changing from an old to a new basis but simply introduce a specific basis from scratch and reason what the matrix of \mathbf{P} must look like in this basis.]

A: $\mathbf{P} \mathbf{e}_1 = \mathbf{e}_1, \dots, \mathbf{P} \mathbf{e}_p = \mathbf{e}_p, \mathbf{P} \mathbf{e}_{p+1} = \dots = \mathbf{P} \mathbf{e}_n = \mathbf{0}$. Hence the matrix is diagonal with values +1 down the first p entries of the diagonal, and the rest all zeros.