Review Packet 2

1. What is the inverse of a diagonal matrix, $\mathbf{D} = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$?

$$D^{-1} = diag\left\{\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_r}\right\}$$

2. What is the effect of multiplying a matrix, $\mathbf{X}$ by a diagonal matrix on the right (as in $\mathbf{XD}$)? on the left?

XD - scales the columns of X by corresponding diagonal element of D

DT - scales the rows of X "

3. Combining the previous two problems, what happens when we multiply a data matrix, $\mathbf{X}$, by $\mathbf{D}^{-1}$ on the right if $\mathbf{D} = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ (as in $\mathbf{XD}^{-1}$)? (Can you guess why we might be using σ's as the diagonal elements?) It would divide each element in column j by σ_j. (This is what standardization/normalization looks like)

4. For a general matrix $\mathbf{A}_{m \times n}$ describe what the following products will provide. Also give the size of the result (i.e. "$n \times 1$ vector" or "scalar"). *Hint: If you cannot see these effects in the general sense, try using a simple 3×3 matrix $\mathbf{A}$ as an example first.*

   a. $\mathbf{Ae}_j$ $(m\times 1)$ - j^{th} column of A
   b. $\mathbf{e}_i^T\mathbf{A}$ $(1\times n)$ - i^{th} row of A
   c. $\mathbf{e}_i^T\mathbf{Ae}_j$ (scalar) - A_{ij}
   d. $\mathbf{Ae}$ $(m\times 1)$ - row sums of A **sums across the rows, not sum OF the rows**
   e. $\mathbf{e}^T\mathbf{A}$ $(1\times n)$ - column sums of A **sums down the columns, not sum OF the columns**
   f. $\frac{1}{n}\mathbf{e}^T\mathbf{A}$ $(1\times n)$ - column averages of A

   Should be m!!

5. Write the vector $\mathbf{v}$ as a linear combination of each given $\mathbf{x}$ and $\mathbf{y}$, if possible.

$$\mathbf{v} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

   a. $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\mathbf{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ $v = 2x + 3y$

   b. $\mathbf{x} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ $\mathbf{y} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ $v = -2x - 3y$

   c. $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\mathbf{y} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ not possible

$$\left(\begin{array}{cc|c} 1 & 2 & 2 \\ 1 & 2 & 3 \end{array}\right)$$ is inconsistent.

6. (*True/False*) For a set of vectors, $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, the linear combination

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = V\alpha$$

can be written as a matrix vector product. If true, define the matrix and vector which should be multiplied together to achieve this sum. TRUE

$$V = (V_1 \mid V_2 \mid V_3 \mid \dots \mid V_n) \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

7. Prove that the products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ will be symmetric for any matrix $\mathbf{A}$.

$$(A^TA)^T = A^T(A^T)^T = A^TA$$

transpose equal to original $\Rightarrow$ symmetric

8. Suppose that I take a matrix of data, $\mathbf{X}_{n\times p}$, and decompose it into the product of two factors, $\mathbf{F}_{n\times r}$ and $\mathbf{C}_{r\times p}$:

$$\mathbf{X} = \mathbf{F}\mathbf{C}$$

Using the (i,j)-notation you've learned, show how the 1st column (i.e. variable) of the data matrix can be represented as a linear combination of columns from the matrix $\mathbf{F}$.

columns

$$\begin{pmatrix} | & | & & | \\ X_1 & X_2 & & X_p \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ F_1 & F_2 & \dots & F_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & & \vdots \\ \vdots & \vdots & & \vdots \\ C_{r1} & C_{r2} & \cdots & C_{rp} \end{pmatrix} = C_{11}F_1 + C_{21}F_2 + \cdots + C_{r1}F_r$$

scalars

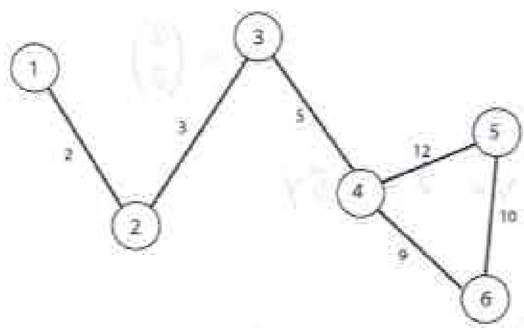
Can you also write the first row (i.e. observation) as a linear combination of rows from the matrix $\mathbf{C}$?

rows

$$F_{11}C_1 + F_{12}C_2 + \dots + F_{1r}C_r$$

scalars

9. Refer to the network/graph shown below. This particular network has 6 numbered vertices (the circles) and edges which connect the vertices. Each edge has a certain *weight* (perhaps reflecting some level of association between the vertices) which is given as a number.



$$A = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 & 12 & 9 \\ 0 & 0 & 0 & 12 & 0 & 10 \\ 0 & 0 & 0 & 9 & 10 & 0 \end{pmatrix}$$

a. Write down the adjacency matrix, $\mathbf{A}$, for this graph where $\mathbf{A}_{ij}$ reflects the weight of the edge connecting vertex i and vertex j.

b. The **degree** of a vertex is defined as the sum of the weights of the edges connected to that vertex Create a vector $\mathbf{d}$ such that d_i is the degree of node i.

$$d = Ae = \begin{pmatrix} 2 \\ 5 \\ 8 \\ 26 \\ 22 \\ 19 \end{pmatrix}$$

10. Suppose I want to compute the matrix product $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where $\mathbf{U}$ is $n \times r$, $\mathbf{D}$ is an $r \times r$ diagonal matrix, $\mathbf{D} = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$, and $\mathbf{V}^T$ is $r \times p$. *(Side note: we will quite often want to compute such a matrix product – this is the form of the singular value decomposition (SVD)! The following exercise is not just for fun - what you end up with in part b is exactly how we will want to write the SVD to best understand how it works.)*

a. Using what you know about multiplication by diagonal matrices, if we view the matrix $\mathbf{U}$ as a collection of columns,

$$\mathbf{U} = (\mathbf{u}_1|\mathbf{u}_2|\mathbf{u}_3|\ldots|\mathbf{u}_r)$$

then how would I write the same partition of the matrix $\mathbf{UD}$?

$$\mathbf{UD} = (?|?|?|\ldots|?)$$

Keep in mind that when multiplying matrices/vectors by scalars, it is always preferable to write the scalar first ($\sigma\mathbf{x}$ rather than $\mathbf{x}\sigma$)

b. Now, using the above representation for $\mathbf{UD}$, what happens when I multiply by the matrix $\mathbf{V}^T$, viewed as a collection of rows,

$$\mathbf{V}^T = \begin{pmatrix} \mathbf{v}_1^T \\ \hline \mathbf{v}_2^T \\ \hline \mathbf{v}_3^T \\ \hline \vdots \\ \hline \mathbf{v}_r^T \end{pmatrix} ?$$

(Hint: your answer should be a sum. Each term in the sum should be an outer product.)

$$\mathbf{UDV}^T = ?$$

see solution from WKST 14
(primer)

11. Determine the unit vector that points in the same direction of the following vectors:

a. $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\quad \|v_1\| = \sqrt{2} \quad u_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a unit vector in same direction

b. $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ $\quad u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

12. Suppose you have survey data where individuals scored 10 statements about travel preferences on a Likert scale from 1-10 where 1='strongly disagree' and 10= 'strongly agree'. Let the vector **a** contain the numerical responses from person A and vector **b** contain the numerical responses from person B (So $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{10}$). Explain in words how to interpret the quantity

$$\|\mathbf{a} - \mathbf{b}\|_{\infty}.$$

The maximum "disagreement" that person A and person B ever had on the questionaire responses

13. **Statistical Formulas Using Linear Algebra Notation.** Almost every statistical formula can be written in a more compact fashion using linear algebra. Most of the elementary formulas involve vector inner products or the Euclidean norm. To begin, we'll introduce the concept of *centering* the data. **Centering** the data means that the mean of a variable is subtracted from each observation. For example, if we have some variable, **x**, and 3 observations on that variable:

$$\mathbf{x} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

then obviously, $\bar{x} = 3$. The **centered** version of **x** would then be

$$\mathbf{x} - \bar{x}\mathbf{e} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} - \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

We simply subtract the mean from every observation so that the new mean of the variable is 0.

Most multivariate textbooks start by saying *"all variable vectors in this textbook are assumed to be centered to have mean zero unless otherwise specified"*. Looking at the most common statistical formulas helps us see why. Try to re-write the following formulas using linear algebra notation, using the vectors **x** and **y** to represent centered data:

$$\mathbf{x} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}$$

For this exercise, keep in mind the following linear algebra constructs, which you should be very familiar with by now:

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

$$\mathbf{a}^T\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 + \cdots + a_nb_n$$

a. Sample standard deviation:

$$s = \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{\sqrt{n-1}} = \boxed{\frac{\|x\|}{\sqrt{n-1}}}$$

b. Sample covariance:

$$covariance(\mathbf{x},\mathbf{y}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \boxed{\frac{1}{n-1}\ x^T y}$$

c. Correlation coefficient:

* the covariance for the standardized data.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \boxed{\frac{x^T y}{\|x\|\ \|y\|}}$$

14. Write a *matrix formula* for the covariance matrix, Σ, using a matrix of *centered* data,

$$\mathbf{X} = (\mathbf{x}_1|\mathbf{x}_2|\dots|\mathbf{x}_p),$$

where $\Sigma_{ij} = cov(\mathbf{x}_i, \mathbf{x}_j)$.

$$\boxed{\Sigma = \frac{1}{n-1} X^T X}$$

* diagonal elements give variances of each variable

* Capital Sigma is generally used to represent a covariance matrix.

15. Write a *matrix formula* for the correlation matrix, $\mathbf{C}$, using a matrix of *centered* data,

$$\mathbf{X} = (\mathbf{x}_1|\mathbf{x}_2|\dots|\mathbf{x}_p),$$

where $\mathbf{C}_{ij} = r_{ij}$ is Pearson's correlation measure between variables $\mathbf{x}_i$ and $\mathbf{x}_j$. To do this, we need more than an inner product, we need to first divide each column by the corresponding standard deviation $s_i = \|\mathbf{x}_i\|$.

let $D = diag\{s_1, s_2, \dots, s_p\}$

then XD^{-1} is standardized data.

Since correlation is merely covariance of standardized data,

$$C = (XD^{-1})^T(XD^{-1}) = \boxed{1/(n-1)\ D^{-1} X^T X D^{-1}}$$

16. **List of Key Words.** You should be completely comfortable with the following terminology.

linear	outer product	$(\mathbf{A}+\mathbf{B})(\mathbf{C}+\mathbf{D}) = ?$
matrix	matrix inverse	Associative Property
vector	systems of equations	Transpose of Product
scalar	row operations	$(\mathbf{ABC})^T = ?$
A_{ij}	row-echelon form	$(\alpha\mathbf{A})^T = ?$
$\mathbf{A}_{\star j}$	pivot element	Inverse of Transpose, $\mathbf{A}^{-T} = ?$
$\mathbf{A}_{i\star}$	Gaussian elimination	Partitioned Matrix
dimensions	Gauss-Jordan elimination	Multiply Partitioned Matrices
diagonal element	reduced row-echelon form	Vector Norm
square matrix	rank	Magnitude/Length
rectangular matrix	unique solution	2-norm
network	infinitely many solutions	$\\|\mathbf{x}\\|_2$
graph	inconsistent	$\sqrt{\mathbf{x}^T\mathbf{x}}$
adjacency matrix	back-substitution	Euclidean Norm
correlation matrix	residual error	Euclidean Distance
transpose	least squares	Unit vector
symmetric matrix	normal equations	Create unit vector
trace	least squares solution	1-norm
diagonal matrix	parameter estimate	$\\|\mathbf{x}\\|_1$
identity matrix	linearly independent	Manhattan distance
upper triangular matrix	linearly dependent	Taxicab distance
lower triangular matrix	full rank	Cityblock distance
matrix addition	perfect multicollinearity	$\\|\mathbf{x}\\|_\infty$
matrix subtraction	severe multicollinearity	Max Distance
scalar multiplication	invertible	Mahalanobis distance
inner product	nonsingular	
matrix product	Distributive Property	
linear combination	$\mathbf{A}(\mathbf{B}+\mathbf{C}) = ?$	