Review Packet 2

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

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?

3. Combining the previous two problems, what happens when we multiply a data matrix, $\mathbf{X}$, by $\mathbf{D}^{-1}$ on the right if $\mathbf{D} = diag\{\sigma_1, \sigma_2, \dots, \sigma_r\}$ (as in $\mathbf{XD}^{-1}$)? (Can you guess why we might be using σ's as the diagonal elements?)

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. $\mathbf{A}\mathbf{e}_j$
   b. $\mathbf{e}_i^T\mathbf{A}$
   c. $\mathbf{e}_i^T\mathbf{A}\mathbf{e}_j$
   d. $\mathbf{A}\mathbf{e}$
   e. $\mathbf{e}^T\mathbf{A}$
   f. $\frac{1}{n}\mathbf{e}^T\mathbf{A}$

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. $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

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

   c. $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

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

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

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

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

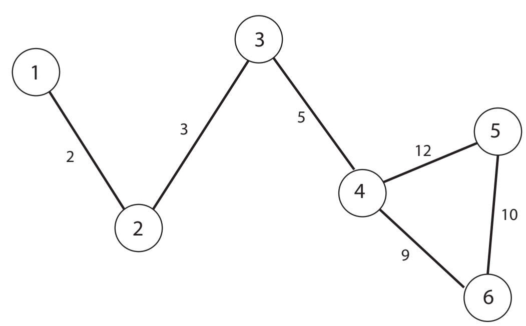
8. Suppose that I take a matrix of data, $\mathbf{X}_{n\times p}$, and decompose it into the product of two factors, $\mathbf{F}_{n\times r}$ and $\mathbf{C}_{r\times p}$:

$$\mathbf{X} = \mathbf{FC}$$

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

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

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. 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.

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.)

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

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

    a. $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

    b. $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

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

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

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:

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

b. Sample covariance:

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

c. Correlation coefficient:

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

14. Write a *matrix formula* for the covariance matrix, $\mathbf{\Sigma}$, using a matrix of *centered* data,

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

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

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 $s_i = \|\mathbf{x}_i\|$.

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

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