Introduction to Linear Algebra

1. Matrix, vector and array

   Scalar – a non-dimensional quantity
   Vector – one-dimensional array of numbers
   Matrix – two-dimensional array of numbers

   Here is an example of a 2x3 matrix, $A = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 7 & 3 \end{pmatrix}$

   Matrix A has 2 rows and 3 columns, can be indexed as A(i, j), where i is the row number and j is the column number. For example, A(1,2) = 4, A(2,3) = 3.

   A vector can be a row vector or a column vector, needs only one index.
   Row vector: $B = \begin{pmatrix} 2 & 4 & 8 \end{pmatrix}$
   Column vector: $C = \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix}$

   Square matrices are those with the same number of row and column. m = n. For example,
   $A = \begin{pmatrix} 2 & 4 \\ 1 & 5 \end{pmatrix}$, $b = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 7 & 3 \\ 2 & 5 & 6 \end{pmatrix}$

   Diagonal matrices are square matrices with only the values along the main diagonal are non-zero. For example,

$$C = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

   Identity matrices are diagonal matrices where all the non-zero values are 1. For example,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. Matrix transposition
   Transposition flips rows and columns; each row of the original matrix becomes the corresponding column of the new matrix.

$$A = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 7 & 3 \end{pmatrix}, \quad A^T = \begin{pmatrix} 2 & 1 \\ 4 & 7 \\ 8 & 3 \end{pmatrix}$$

3. Matrix addition
   Addition is an operation that is defined for two matrices or two vectors of the same dimensionality. Adding matrices algebraically is adding corresponding components to form a new matrix. Thus, each of the two matrices or two vectors being added must contain only elements that correspond with those in the other. For example,
   $$A = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 7 & 3 \end{pmatrix}, B = \begin{pmatrix} 5 & 4 & -1 \\ 2 & 3 & 7 \end{pmatrix}, A + B = \begin{pmatrix} 7 & 8 & 7 \\ 3 & 10 & 10 \end{pmatrix}$$

   Because addition is defined only for cases where the two values being added have the same dimensionality, cases where the dimensionality differ would be termed undefined or meaningless.

4. Scalar multiplication
   When a matrix is multiplied by a scalar value, each element of the matrix is simply multiplied by that number. For example,
   $$A = \begin{pmatrix} 2 & 4 & 8 \\ 1 & 7 & 3 \end{pmatrix}, \qquad 2A = \begin{pmatrix} 4 & 8 & 16 \\ 2 & 14 & 6 \end{pmatrix}$$

5. Matrix multiplication
   When you multiply two matrices together, AB, each element of the resulting matrix, C, is the sum of the corresponding row elements of A times the corresponding column elements of B. For example,
   $$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1*2+2*1 & 1*1+2*(-1) \\ 2*2+3*1 & 2*1+3*(-1) \end{pmatrix} = \begin{pmatrix} 4 & -1 \\ 7 & -1 \end{pmatrix}$$

   Another example,
   $$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1*2+2*1 \\ 2*2+3*1 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$$

   Note: For two matrices A and B, the number of columns in A must match the number of rows in B for the product AB to be defined. If A has m rows and n columns, and B has n rows and p columns, then AB has m rows and p columns.

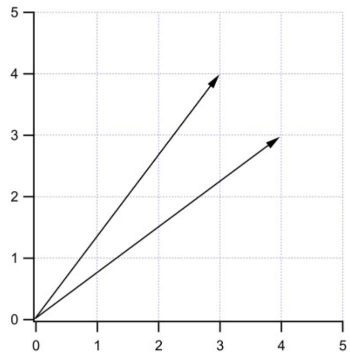
   Note: Matrix multiplication is not communicative. In another word, AB is not the same as BA.

   The inverse of a matrix D, D^{-1}, is the matrix that when multiplied with the original matrix results in the identity matrix:
   $$D\, D^{-1} = I$$
   Note that D must be a square matrix.
   If the determinant of a matrix is zero, det(D) = 0, the matrix is singular and there is no inverse matrix. For a 2x2 matrix,
   $$D = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \det(D) = ad - bc$$

6. Geometrical interpretation of matrix multiplication

   The vector $B = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ can be plotted on the Cartesian plan. Multiplication of a matrix with the vector is simply rotating the vector.

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, AB = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$



7. Eigenvalues and eigenvectors

   For many square matrices A, there exists corresponding vectors B such that

$$AB = \lambda B$$

where λ is a scalar. Geometrically, this means that for a given matrix A, there is a vector B that does not rotate when multiplied by A. The scalar λ is called an eigenvalue of the matrix A. The invariant vector B is called an eigenvector of the matrix A, and each eigenvector B is associated with a particular eigenvalue λ.

Obviously, B = $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is a solution to AB = λB, but we are not interested in this trivial one. We are interested in the non-zero eigenvectors to a square matrix.

$$AB = \lambda B$$
$$AB = \lambda IB$$
$$AB - \lambda IB = 0$$
$$(A - \lambda I)B = 0$$

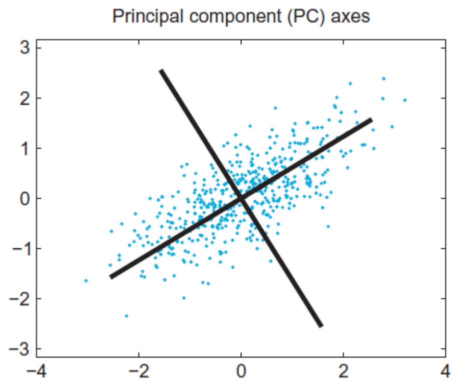
If (A - λI) has an inverse,

$$(A - \lambda I)^{-1} (A - \lambda I)B = 0$$
$$IB = 0$$

Here the zero vector B is exactly the trivial answer that we don't want. Therefore, if the non-trivial solutions for B exist, (A - λI) must not have an inverse. That means det(A - λI) = 0. For matrix A = $\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$, you can find λ to be -1 or 3. For λ = -1, B = $\begin{pmatrix} X \\ -2X \end{pmatrix}$. For λ = 3, B = $\begin{pmatrix} X \\ 2X \end{pmatrix}$.

8. Eigendecomposition theorem and principal component analysis (PCA)
   Eigendecomposition theorem states:
   For any nxn matrix A with distinct eigenvalues, $A = VDV^{-1}$, where V is the square matrix whose columns are the eigenvectors of A, and D is the square diagonal matrix containing the eigenvalues of A along the diagonal.

   PCA provides a means of identifying the independent axes responsible for major sources of variability in a multivariate sample. For example, if you want to compress the data as seen in the figure below, i.e., reducing the data from 2-dimension to 1-dimension. By rotating the axes, you can see that most of the variance is on one of the new axes.



   Based on the original dataset, a covariance matrix Σ is computed. The eigendecomposition theorem states:

$$\Sigma = VDV^{-1}$$
$$\Sigma V = VDV^{-1}V$$
$$\Sigma V = VD$$
$$V^{-1}\Sigma V = V^{-1}VD$$
$$V^{-1}\Sigma V = D$$

   Thus, by rotating the original covariance matrix Σ, it can be transformed into a diagonal matrix containing just the eigenvalues. Each eigenvector is a new axe called principal component (PC), and its associated eigenvalue indicates the relative variance along the PC.