Quantitative Foundations Project 4

Dimensionality Reduction & Classification

Instructor: Linwei Wang

Contents

1 Dimensionality Reduction

Why do we need dimensionality reduction?

Example 1: Digital images of handwritten numbers, each represented by a 28 by 28 image (784 by 1 data vector).

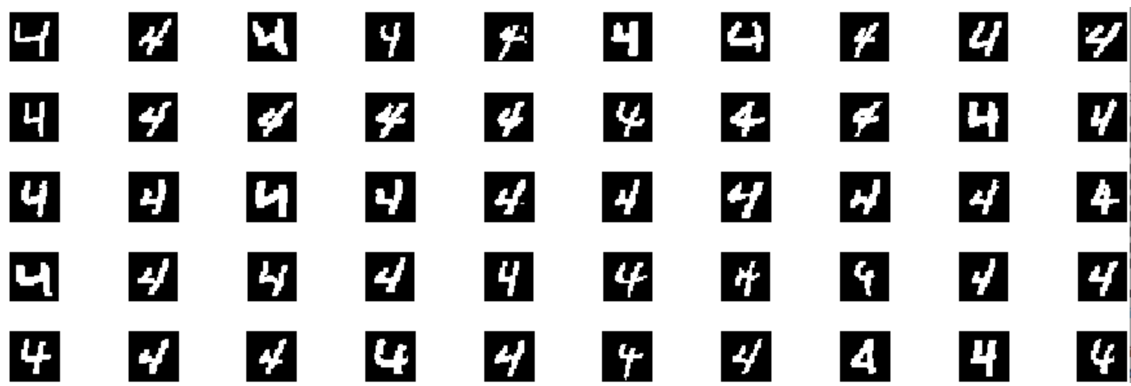

Figure 1.1: A set of examples of handwritten 4's, a subset from MNIST database (http://yann.lecun.com/exdb/mnist/)

Example 2: Face images, each represented by a 112 by 92 image or, a 10304 by 1 vector.

Figure 1.2: A set of face images, a subset of the face database provided by the Cambridge Universituy (http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html)

Given original data $\mathbf{d}$ in $\mathbb{R}^n$, principal components can be understood as a sequence of projections of the data, mutually uncorrelated and ordered in variance, that provide a a sequence of best linear approximations to that data. Often, a relatively small subset of q principal components ($q << n$) could serve as excellent lower-dimensional features for representing the high-dimensional data.

1.1 Basics

Definition (Linear Independence). The vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are said to be ***linearly independent*** if the euqation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{0}$$

implies that $c_1 = 0, c_2 = 0, \cdots, c_n = 0$. Otherwise the vectors are linearly dependent. In other words, the vectors are ***linearly dependent*** if there exists a set of numbers $\{c_1, c_2, \cdots, c_n\}$ not all zero such as $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n = \mathbf{0}$. This means one of the vectors can be written as a linear combination of the others.

Consider in 2-dimensional space the following two vectors $\mathbf{x}_1 = (x_1, y_1)^T$, $\mathbf{x}_2 = (x_2, y_2)^T$. What does it mean if the two vectors are linearly dependent?

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = 0$$

If $c_1, c_2 \neq 0$, we have $\mathbf{x}_1 = -\frac{c_2}{c_1}\mathbf{x}_2$. It means $\mathbf{x}_1$ and $\mathbf{x}_2$ are parallel. Two vectors in R^2 are linearly independent if and only if they are not parallel.

Example. Consider $\mathbf{x}_1 = (1, 1)^T$ and $\mathbf{x}_2 = (-1, 1)^T$. Are $\mathbf{x}_1$ and $\mathbf{x}_2$ independent?

$$c_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 0 \rightarrow \begin{cases} c_1 - c_2 = 0 \\ c_1 + c_2 = 0 \end{cases} \rightarrow \begin{cases} c_1 = 0 \\ c_2 = 0 \end{cases}$$

Example. Consider in R^3 if the following three vectors $\mathbf{x}_1 = (1, 2, 0)^T$, $\mathbf{x}_2 = (1, -1, 1)^T$ and $\mathbf{x}_3 = (2, 1, 1)^T$ are independent?

$$c_1\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + c_2\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + c_3\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 0 \rightarrow \begin{cases} c_1 + c_2 + 2c_3 = 0 \\ 2c_1 - c_2 + c_3 = 0 \\ 0c_1 + c_2 + c_3 = 0 \end{cases} \rightarrow \begin{cases} c_1 = 1 \\ c_2 = 1 \\ c_3 = -1 \end{cases}$$

Since $c_1, c_2, c_3 \neq 0$, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are not linearly independent. Three vectors in R^3 are linearly independent if and only if they do not lie in the same plane.

Definition (Basis). Suppose that $S = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ is a set of n vectors in R^n. The set S is called a basis for R^n if

- $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are linearly independent
- for every vector $\mathbf{x}$ in R^n there exists a unique set of scalars $\{c_1, c_2, \cdots, c_n\}$ so that $\mathbf{x}$ can be expressed as linear combination:

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n \tag{1.1}$$

Theorem. In R^n, any set of n linearly independent vectors form a basis of R^n. Each vector $\mathbf{x}$ in R^n is uniquely expressed as a linear combination of the basis vectors as shown in (1.1).

Example Consider the standard basis set in R^3:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Changing of Basis Vectors: Let $\mathbf{V} = [\mathbf{v}_1|\mathbf{v}_2|\cdots|\mathbf{v}_n]$ be a basis set in R^n. Any $\mathbf{x} \in R^n$ can be written as:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{Vc} \tag{1.2}$$

where c is called coordinates of $\mathbf{x}$ with respect to the basis $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. Now assume we change the basis vector to $\mathbf{U} = [\mathbf{u}_1|\mathbf{u}_2|\cdots|\mathbf{u}_n]$:

$$\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_n\mathbf{u}_n = \mathbf{Ud} \tag{1.3}$$

The new coordinates $\mathbf{d}$ of $\mathbf{x}$ with respect to the new basis $\mathbf{U}$ can be calculated by:

$$\mathbf{d} = \mathbf{U^{-1}Vc} \tag{1.4}$$

Examples (Coordinate Transfromation). Let $\mathbf{x} = 3\mathbf{e}_1 + 2\mathbf{e}_2$ with respect to the standard basis $(\mathbf{e}_1, \mathbf{e}_2)$ in R^2. What is the coordinate of $\mathbf{x}$ if the basis changes from $(\mathbf{e}_1, \mathbf{e}_2,)$to $(\mathbf{u}_1, \mathbf{u}_2)$?

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\mathbf{Ud} = \mathbf{Ec} \rightarrow \mathbf{d} = \mathbf{U}^{-1}\mathbf{Ec} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 2.5 \\ -0.5 \end{pmatrix} \tag{1.5}$$

Definition (Orthogonal). A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is said to be orthogonal provided that

$$\mathbf{v}_j'\mathbf{v}_k = 0 \quad \text{whenever} \quad j \neq k.$$

Definition (Orthonormal). If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a set of orthogonal vectors, then they are orthonormal if they are all of unit norm, that is:

$$\mathbf{v}_j'\mathbf{v}_k = 0 \quad \text{whenever} \quad j \neq k.$$
$$\mathbf{v}_j'\mathbf{v}_j = 1 \quad \text{for all} \quad j = 1, 2, \cdots, n.$$

Theorem. An orthonormal set of vectors is linearly independent.

Definition (Orthogonal Matrix). An $n \times n$ matrix $\mathbf{A}$ is said to be orthogonal provided that $\mathbf{A}'$ is the inverse of $\mathbf{A}$; that is,

$$\mathbf{A}'\mathbf{A} = \mathbf{I}.$$

In other words, $\mathbf{A}$ is orthogonal if and only if the columns (and rows) of $\mathbf{A}$ form a set of orthonormal vectors.

1.2 Eigenvalues and Eigenvectors

1.2.1 Definition

Let $\mathbf{A}$ be a square matrix of dimension $n \times n$ and $\mathbf{x}$ a vector of dimension n. The product $\mathbf{y} = \mathbf{A}\mathbf{x}$ can be viewed as a linear transformation from n-dimensional space into itself. The nonzero vector $\mathbf{x}$ is an ***eigenvecotor*** of $\mathbf{A}$ if there is a scale factor λ such that:

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{1.6}$$

and the scalar λ is the ***eigenvalue*** corresponding to the eigenvector $\mathbf{x}$. The linear opeartor $\mathbf{A}$ maps $\mathbf{x}$ onto the multiple $\lambda\mathbf{x}$, that is, the eigenvecotors are not changed by the operator $\mathbf{A}$ except for the scale factor λ. *This means that eigenvectors are invariant with respect to the operator* $\mathbf{A}$.

The above equation can be rewritten in the form of a standard linear system:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{1.7}$$

The system has a nontrivial solution if and only if matrix $\mathbf{A} - \lambda\mathbf{I}$ is singular.

Matrix Determinant: *Determinant* of $\mathbf{A}$, denoted as $det(\mathbf{A})$, is an indication of the *sigularity* or invertability of $\mathbf{A}$. A square matrix $\mathbf{A}$ is said to be *nonsingular* if it has an inverse; that is if there exists a matrix denoted by $\mathbf{A}^{-1}$ such that:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{1.8}$$

The existence of $\mathbf{A}^{-1}$ is directly related to $det(\mathbf{A})$.

- Definition of Determinant: $det(\mathbf{A}) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}M_{ij}$ for any $i = 1, 2, \ldots, n$ where M_{ij} is the determinant of the matrix obtained by deleting the row and column containing a_{ij}.
- Matrix $\mathbf{A}$ is nonsingular if $det(\mathbf{A}) \neq 0$.

Special cases:

- $\mathbf{A} = (a_{11}), det(\mathbf{A}) = a_{11}$.
- $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.
  - $det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$
  - Considering what situation does the following system has nontrivial solution?

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. $\rightarrow det(\mathbf{A}) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}))$

Example. Find the determinant of matrix:

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 1 & 2 \\ -1 & 2 & 4 \end{pmatrix}$$

$$\begin{aligned} det(\mathbf{A}) &= (-1)^{1+1}(1) \times \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} + (-1)^{2+1}(2) \times \begin{vmatrix} 3 & 2 \\ -1 & 4 \end{vmatrix} + (-1)^{3+1}(1) \times \begin{vmatrix} 3 & 1 \\ -1 & 2 \end{vmatrix} \\ &= 1 \times (4-4) - 2 \times (12+2) + 1 \times (6+1) = -21 \end{aligned}$$

Back to the eigenvalue problem, for matrix $\mathbf{A} - \lambda\mathbf{I}$ to be singular we need:

$$det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{1.9}$$

This determinant can be written in the form:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_1n \\ a_{21} & a_{22}-\lambda & \cdots & a_2n \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0$$

which is a polynomial of degree n denoted by $P(\lambda)$ and called the ***characteristic polynomial*** of matrix $\mathbf{A}$:

$$P(\lambda) = det(\mathbf{A} - \lambda\mathbf{I}) = (-1)^n(\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_{n-1}\lambda + c_n) \tag{1.10}$$

- This characteristic polynomial has n roots, not necessarily distinct, that represent the eigenvalues of $\mathbf{A}$.
- Each root λ can be substituted into (1.7) to obtain a corresponding nontrivial eigenvector $\mathbf{x}$. They can be real or complex.

Definition (Eigenvalue). If $\mathbf{A}$ is an $n \times n$ real matrix, then its n eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the real and complex roots of the characteristic polynomial (1.10).

Definition (Eigenvector). If λ is an eigenvalue of $\mathbf{A}$ and the nonzero vector $\mathbf{v}$ has the property that:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \tag{1.11}$$

then $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ corresponding to the eigenvalue λ.

Example. Find eigenvalues and eigenvectors for matrix.

$$\mathbf{A} = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}$$

- Characteristic polynomial

  $$P(\lambda) = det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} 3-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 3-\lambda \end{vmatrix} = -\lambda^3 + 8\lambda^2 - 19\lambda + 12 = 0.$$

  which can be written as $-(\lambda-1)(\lambda-3)(\lambda-4) = 0$ Therefore we have three eigenvalues: $\lambda_1 = 1, \lambda_2 = 3, \lambda_3 = 4$.

- Eigenvectors:
  - $\lambda_1 = 1$

    $$\begin{array}{rrrl} 2x_1 & -x_2 & & = 0 \\ -x_1 & +x_2 & -x_3 & = 0 \\ & -x_2 & +2x_3 & = 0 \end{array} \rightarrow \mathbf{v}_1 = a \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

  - $\lambda_2 = 3$

    $$\begin{array}{rrrl} & -x_2 & & = 0 \\ -x_1 & -x_2 & -x_3 & = 0 \\ & -x_2 & & = 0 \end{array} \rightarrow \mathbf{v}_2 = b \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

  - $\lambda_3 = 4$

    $$\begin{array}{rrrl} -x_1 & -x_2 & & = 0 \\ -x_1 & -2x_2 & -x_3 & = 0 \\ & -x_2 & -x_3 & = 0 \end{array} \rightarrow \mathbf{v}_3 = c \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

  - An eigenvector is unique only up to a constant multiple.

1.2.2 Properties

Existence of Eigenvectors

The characteristic polynomial (1.10) can be factored into:

$$P(\lambda) = (-1)^n(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k} \tag{1.12}$$

where m_j is the *multiplicity* of the eigenvalue λ_j. The sum of all multiplicities is n: $n = m_1 + m_2 + \cdots + m_k$.

Theorem.

- For each *distinct* eigenvalue λ there exists *at least one* eigenvector $\mathbf{v}$ corresponding to λ.
- If λ has multiplicity r, there exist *at most* r linearly independent eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r$ that correspond to λ.

Independency of Eigenvectors

Theorem. Supposed that $\mathbf{A}$ is a square matrix and $\lambda_1, \lambda_2, \cdots, \lambda_k$ are distinct eigenvalues of $\mathbf{A}$, with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$, respectively; then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is a set of linearly independent vectors.

Theorem. If the eigenvalues of the $n \times n$ matrix $\mathbf{A}$ are all distinct, then there exists n linearly independent eigenvectors $\mathbf{v}_j$, for $j = 1, 2, \cdots, n$.

Example.

- Are eigenvectors of the previous example linearly independent?
  (Consider if $det([\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]) = 0$ or not.)
- Are they orthogonal?

1.2.3 Matrix Diagonalization

Let $\mathbf{A}$ be an $n \times n$ matrix with n eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. We have:

$$\mathbf{AV} = \mathbf{V\Lambda}$$

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n]$, and

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{1.13}$$

The above equation is valid regardless of whether the eigenvectors are linearly independent or not.

If the n eigenvalues are distinct, the eigenvectors are linearly independent and thus $\mathbf{V}$ is of full rank. In this case, we have:

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} \tag{1.14}$$

or

$$\mathbf{\Lambda} = \mathbf{V}^{\mathbf{-1}}\mathbf{A}\mathbf{V} \tag{1.15}$$

Theorem (Diagonalization). The matrix $\mathbf{A}$ is called ***diagonalizable*** if it is similar to a diagonal matrix. $\mathbf{A}$ is similar to a diagonal matrix Λ if and only if it has n linearly independent eigenvectors:

$$\begin{aligned} \mathbf{V}^{\mathbf{-1}}\mathbf{A}\mathbf{V} = \mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_n) \\ \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n] \end{aligned} \tag{1.16}$$

where $\lambda_j, \mathbf{v}_j$ are eigenpairs for $j = 1, 2, \cdots, n$.

• Relating back to the property of eigenvectors:

- If the $n \times n$ $\mathbf{A}$ has n distinct eigenvalues $\rightarrow$ it is always diagonalizable.
- If $\mathbf{A}$ has n independent eigenvectors $\rightarrow$ it is always diagonalizable regardless whether it has distinct or repeat eigenvalues.
- If $\mathbf{A}$ has less than n independent eigenvectors, it is not diagonalizable.

Examples. Test in matlab the previous example of matrix $\mathbf{A}$ is diagnoalizable.

Now think about what happens if we select the eigenvectors of an $n \times n$ matrix $\mathbf{A}$ as the basis of $\mathcal{R}^n$?

Recall that the eigenpair $\lambda_i, \mathbf{v}_i$ has the property that the linear mapping $\mathbf{y} = \mathbf{A}\mathbf{x}$ maps $\mathbf{v}_i$ onto the multiple of $\lambda_i\mathbf{v}_i$. If a basis of $\mathcal{R}^n$ is selected that consists of the eigenvectors of $\mathbf{A}$, it is easier to visualize how space is transformed by the mapping $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Theorem. Suppose that $\mathbf{y} = \mathbf{A}\mathbf{x}$ is an $n \times n$ matrix that possesses n linearly independent eigenpairs $\lambda_j, \mathbf{v}_j$ for $j = 1, 2, \cdots, n$; then any vector $\mathbf{x}$ in R^n has a unique representation as a linear combination of the eigenvectors:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \tag{1.17}$$

The linear transformation $\mathbf{y} = \mathbf{A}\mathbf{x}$ maps $\mathbf{x}$ onto the vector:

$$\mathbf{y} = \mathbf{A}\mathbf{x} = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_n\lambda_n\mathbf{v}_n \tag{1.18}$$

Example. Suppose that 3×3 matrix $\mathbf{A}$ has eigenvalues $\lambda_1 = 2, \lambda_2 = -1, \lambda_3 = 4$, which correspond to eigenvectors $\mathbf{v}_1 = [1 \quad 2 \quad -2]'$, $\mathbf{v}_2 = [-2 \quad 1 \quad 1]'$, and $\mathbf{v}_3 = [1 \quad 3 \quad -4]'$, respectively. If $\mathbf{x} = [-1 \quad 2 \quad 1]'$, find the image of $\mathbf{x}$ under the mapping $\mathbf{y} = \mathbf{A}\mathbf{x}$.

- First, express x as a linear combination of the eigenvectors, which equals to solving the linear system:

$$c_1 \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} + c_2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$$

  we have $c_1 = 2, c_2 = 1, c_3 = -1$.

- Second, $\mathbf{Ax}$ is found by:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{A}(2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3) \\ &= 2\mathbf{A}\mathbf{v}_1 + \mathbf{A}\mathbf{v}_2 - \mathbf{A}\mathbf{v}_3 \\ &= 2(2\mathbf{v}_1) - \mathbf{v}_2 - 4\mathbf{v}_3 = [2 \quad -5 \quad 7]^T \end{aligned}$$

1.2.4 Virtues of Symmetry

A real symmetric matrix has n real eigenvectors and for each eigenvalue of multiplicity m_i there corresponds m_i linearly independent eigenvectors. Hence every real symmetric matrix is diagonalizable.

Theorem. If $\mathbf{A}$ is a real symmetric matrix, there exists an orthogonal matrix $\mathbf{K}$ such that

$$\mathbf{K}'\mathbf{A}\mathbf{K} = \mathbf{K}^{-1}\mathbf{A}\mathbf{K} = \mathbf{\Lambda} \tag{1.19}$$

where Λ is a diagonal matrix consisting of the eigenvalues of $\mathbf{A}$.

Corollary. If $\mathbf{A}$ is an $n \times n$ real symmetric matrix, there exist n linearly independent eigenvectors for $\mathbf{A}$ and they form an orthogonal set.

Corollary. The eigenvalues of a real symmetric matrix are all real numbers.

Theorem. A symmetric matrix $\mathbf{A}$ is ***positive definite*** if and only if all the eigenvalues of $\mathbf{A}$ are positive.

1.3 Singular Value Decomposition (SVD)

Theorem (SVD). Let $\mathbf{A}$ be an $n \times m$ matrix with rank r. Then there exists unitary matrices $\mathbf{U}(n \times n)$ and $\mathbf{V}(m \times m)$ such that:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{1.20}$$

where Σ is an $n \times m$ matrix with entries that are singular values of $\mathbf{A}$.

Interpreting SVD. And how does it related to eigenvalue and eigenvectors?

- Look at $\mathbf{B} = \mathbf{A}\mathbf{A}^\mathbf{T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathbf{T} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\mathbf{T} \rightarrow \mathbf{B}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$. $\rightarrow$ The left singular vectors $\mathbf{U}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^T$.
- Consider $n \times n$ matrix $\mathbf{B} = \mathbf{A}\mathbf{A}^T$ a rank r, symmetric matrix.
  - Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be the set of orthonormal eigenvectors with the associated eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_r\}$ for $\mathbf{B}$: $\mathbf{B}\mathbf{u}_i = \lambda_i \mathbf{u}_i$
  - $\sigma_i = \sqrt{\lambda_i}$
  - Let $\mathbf{\Sigma}$ be the $n \times m$ diagonal matrix with entries

    $$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

    where$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the ordered set of σ_i. Because $\mathbf{V}$ is orthogonal, (1.30) can be written as:

    $$\mathbf{U}^T\mathbf{A} = \mathbf{\Sigma}\mathbf{V}^T \tag{1.21}$$

    In other words, we have:

    $$\mathbf{u}_i^T\mathbf{A} = \sigma_i \mathbf{v}_i^T \tag{1.22}$$

    where $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a set of orthonormal vectors. In other words, $\mathbf{U}^T$ is a change of basis from $\mathbf{A}$ to $\mathbf{\Sigma}\mathbf{V}^T$, or, the orthonormal basis $\mathbf{U}^T$ transform column vectors of $\mathbf{X}$ to orthonormal vectors (spans the column space of $\mathbf{X}$).
- Similarly consider $\mathbf{C} = \mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^T$. $\rightarrow$ The right singular vectors $\mathbf{V}$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$.
- Consider $\mathbf{C} = \mathbf{A}^\mathbf{T}\mathbf{A}$ a rank r, symmetric $m \times m$ matrix.
  - Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be the set of orthonormal eigenvectors with the associated eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_r\}$ for $\mathbf{C}$: $\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i$
  - $\sigma_i = \sqrt{\lambda_i}$

- Similarly, because $\mathbf{U}$ is unitary from (1.30) we have:

$$\mathbf{AV} = \mathbf{U\Sigma} \tag{1.23}$$

  Namely we have:

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i \tag{1.24}$$

  In other words, $\mathbf{V}^T$ is a change of basis from $\mathbf{A}^T$ to $\mathbf{\Sigma U}^T$, or, the orthonormal basis $\mathbf{V}^T$ transform row vectors of $\mathbf{A}$ to orthonormal vectors (spans the row space of $\mathbf{A}$).

- Recall the virtue of symmetry $\rightarrow \mathbf{U}, \mathbf{V}$ are orthornormal. σ are real.

Definition: Matrix Conditional Number. Condition number $c(\mathbf{A})$ of matrix $\mathbf{A}$:

$$c(\mathbf{A}) = ||\mathbf{A}||||\mathbf{A}^{-1}|| = \sqrt{\frac{\lambda_{max}(\mathbf{A}^T\mathbf{A})}{\lambda_{min}(\mathbf{A}^T\mathbf{A})}} = \sigma_{max}/\sigma_{min} \tag{1.25}$$

where σ_{max} and σ_{min} are the maximum and minimum singular values of $\mathbf{A}$.

Property. If $c(\mathbf{A})$ is small, the relative error in the solutions to $\mathbf{Ax} = \mathbf{b}$ due to a small perturbation of $\mathbf{A}$ is small. If $c(\mathbf{A})$ is large, $\mathbf{A}$ is said to be ill-conditioned and the relative error in the solutions to $\mathbf{Ax} = \mathbf{b}$ using matrix inversion maybe large, even for a small error in $\mathbf{A}$.