Data Matrices and Linear Algebra

Eigenvectors and Eigenspaces

Ramesh Srinivasan

November 9, 2023

Geometry of Multivariate Data

n observations of p variables can be represented as a $n \times p$ matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \ddots & & \\ x_{n1} & \dots & & x_{np} \end{bmatrix}$$

n observations may represent n different participants in an experiment while p are different experimental variables observed in each participant.

In Neuroscience applications, n represents different samples in time, while p represents different locations in the brain.

$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \vdots \\ \mathbf{x_n} \end{bmatrix}$ The matrix a stack of n row vectors (of size p) of observations. **This is how data is collected.**

$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \dots \mathbf{x_p} \end{bmatrix}$ The matrix is a stack of p column vectors (of size n) of variables. **This is the useful way to think about data**

Centering Data

$\bar{\mathbf{X}}$ is a row vector of length p which is the coordinates given by averaging each column of $\mathbf{X}$.
The components of $\bar{\mathbf{X}}$ are

$$\bar{x}_k = \frac{1}{n}\sum_{j=1}^{n} x_{jk}, k = 1, 2, ...p$$

When performing data analysis, we should always center the data on the origin of the coordinate system by computing the *deviations*

$$\mathbf{d_k} = \mathbf{x_k} - \bar{x}_k, k = 1, 2, ...p$$

Example

$$\mathbf{X} = \begin{bmatrix} 4 & 1 \\ -1 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\bar{\mathbf{X}} = [2 \quad 3]$$

$$\mathbf{d}_1 = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$$

$$\mathbf{d}_2 = \begin{bmatrix} -2 \\ 0 \\ 2 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 2 & -2 \\ -3 & 0 \\ 1 & 2 \end{bmatrix}$$

D is a matrix of deviations, which is the original data matrix X now centered on the origin of the coordinate system.

Standard Deviation is a measure of length or norm

If we compute the squared length or norm of a deviation vector, we get a measure of variance,

$$\|\mathbf{d_k}\|^2 = \mathbf{d_k} \cdot \mathbf{d_k} = \sum_{j=1}^{n} d_{jk}^2 = \sum_{j=1}^{n} (x_{jk} - \bar{x}_k)^2 = n s_k^2$$

Therefore the length of the vector is proportional to standard deviation.

$$s_k = \sqrt{\frac{1}{n} \mathbf{d_k} \cdot \mathbf{d_k}} = \sqrt{\frac{1}{n} \|\mathbf{d_k}\|^2}$$

Covariance and Correlation Coefficient

Covariance is related to the dot product between two different deviation vectors

$$s_{kl} = \frac{1}{n}\mathbf{d_k} \cdot \mathbf{d_l}$$

Correlation coefficient is the dot product of unit vectors in the direction of the two data vectors.

$$r_{kl} = \frac{s_{kl}}{s_k s_l} = \frac{\mathbf{d_k} \cdot \mathbf{d_l}}{\|\mathbf{d_k}\| \, \|\mathbf{d_l}\|} = \mathbf{u_k} \cdot \mathbf{u_l}$$

Meaning of dot product

The dot product has a physical interpretation. The dot product is proportional to the cosine of the angle between two vectors.

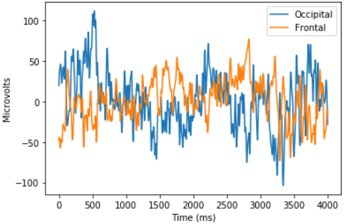
$$\mathbf{x} \cdot \mathbf{y} = \|x\| \, \|y\| \, cos(\theta)$$

If two vectors are parallel, the dot product is the product of their lengths. If the two vectors are perpendicular the dot product is zero. If we introduce this definition into the correlation coefficient,

$$r_{kl} = \frac{s_{kl}}{s_k s_l} = \frac{\mathbf{d_k} \cdot \mathbf{d_l}}{\|\mathbf{d_k}\| \, \|\mathbf{d_l}\|} = \frac{\|\mathbf{d_k}\| \, \|\mathbf{d_l}\| \, cos(\theta)}{\|\mathbf{d_k}\| \, \|\mathbf{d_l}\|} = cos(\theta)$$

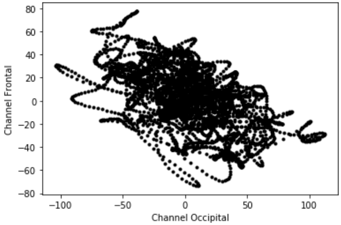
we find that correlation is related to the cosine of the angle between two vectors.

Example



The above example shows EEG traces at two channels, one over the occipital lobe and one over the frontal lobe. The correlation coefficient between the two signals is r = -0.5. This would correspond to an angle of 120 degrees in a 4000 dimensional space, with each dimension corresponding to one of the time points.

Variables as Dimensions



Another useful way to think about our two EEG time series is to plot them in a plane, where the two dimensions correspond to each one of the channels.
The negative correlation between the channels is visible in the geometry of the cloud of points. Each point is a joint observation of the EEG at the two channels. Of course, in reality we observe EEG at many channels, so the dimensionality of our space corresponds to the number of variables we simultaneously observe. The remainder of today's lecture is focused on the geometry of this space

Matrix Properties

1. Matrix A is of dimension m x n if it has m rows and n column. We refer to the elements of A as $a_{ij},\ i = 1, .., m,\ j = 1, .., n$
2. Matrix Addition
$$C = A + B \iff c_{ij} = a_{ij} + bij$$
3. Matrix Multiplication
$$C = A \times B \iff c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj}$$
In words, matrix multiplication is carried out by the dot product between the rows of matrix A and the columns of matrix B to obtain each element of matrix C. Matrix multiplication can only be applied if the inner dimension of the two matrices are identical.
$$A = (m, n),\ B = (n, p) \implies C = (m, p)$$
4. Matrix/Vector multiplication Matrix and vectors can be multiplied as long as the inner dimensions are matched. For example matrix A of size (m,n) can be multipled to a vector of size (n,1). The result is a vector of size (m,1)
5. The transpose of a matrix is obtained by flipping the matrix so that the rows of the matrix become the columns of the matrix. If A is of size (n,m) then A^T is of size (m,n)

Rotations in the Plane

Lets consider a point (a,b) in the plane.
We would like to rotate the point around the origin by an angle θ, to obtain a new point (c,d).
Since the operation is purely a rotation, its useful to express the location of (a,b) in terms of polar coordinates

$$(a,b) = (r\cos(\alpha), r\sin(\alpha))$$

Since we want to rotate the point around the origin by an angle θ, the new coordinates can be written as

$$(c,d) = (r\cos(\alpha+\theta), r\sin(\alpha+\theta))$$

Rotation Matrix

We can use the trig rules for sum of angles to note that

$c = rcos(\alpha + \theta) = rcos(\alpha)cos(\theta) - rsin(\alpha)sin(\theta)$
$d = rsin(\alpha + \theta) = rcos(\alpha)sin(\theta) + rsin(\alpha)cos(\theta)$

which simplifies to

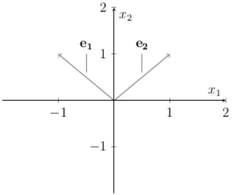
$c = acos(\theta) - bsin(\theta)$
$d = asin(\theta) + bcos(\theta)$

In matrix form

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Basis of a Vector Space

In 2-D, unit vectors along the coordinate axes $\mathbf{u_1} = (1, 0)$ and $\mathbf{u_2} = (0, 1)$ are orthonormal vectors.
Any vector $\mathbf{x}$ in the plane can be written as a linear combination, $\mathbf{x} = x_1\mathbf{u_1} + x_2\mathbf{u_2}$
Thus, together $\{\mathbf{u_1}, \mathbf{u_2}\}$ *span* the vector space of all vectors in a plane, and form a **basis** of the vector space.



In 2-dimensional space, any 2 linearly independent vectors can form an basis.
If n-dimensional space, any n linearly independent vectors can form a basis.
Linearly independent vectors have dot product of zero.

Properties of the Rotation Matrix

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

The rotation matrix is an example of an orthogonal matrix.

An orthogonal matrix has columns (or rows) that form an orthonormal basis.

Its easy to see that the dot product of the columns of the rotation matrix is zero.

$$cos(\theta)(-sin(\theta)) + sin(\theta)cos(\theta) = 0$$

Its also easy to see that the length of each column of the rotation matrix is 1.

$$cos^2(\theta) + sin^2(\theta) = 1$$

Let $\theta = \frac{\pi}{4}$

$$\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

The Covariance Matrix

n observations of p variables can be represented as a $n \times p$ matrix

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \ddots & & \\ x_{n1} & \dots & & x_{np} \end{bmatrix}$$

$\mathbf{X} = \left[\mathbf{x_1} \dots \mathbf{x_p}\right]$ The matrix is a stack of p column vectors (of size n) of variables. **This is the useful way to think about data** When performing data analysis, we should always center the data on the origin of the coordinate system by computing the *deviations*

$$\mathbf{d_k} = \mathbf{x_k} - \bar{x}_k, k = 1, 2, ...p$$

We can define the **covariance** matrix Σ of the multivariate data as the matrix with elements

$$\Sigma_{lk} = \frac{1}{n} \mathbf{d_l} \cdot \mathbf{d_k}$$

Note that the diagonal entries of the matrix are variances, σ_j^2

Covariance Matrix

We can write the covariance matrix as a matrix multiplication

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

where $\mathbf{X}^T$ is the transpose of X. The covariance matrix has a number of important properties.

1. If X is of size (n,p) then Σ is of size (p,p), where p is the number of variables.
2. Σ is a square symmetric matrix, i.e., $\Sigma_{lk} = \Sigma_{kl}$

Similarly, we can define a correlation matrix, **R** with elements that are correlation coefficients, obtained by normalizing the covariance as

$$R_{lk} = \frac{\Sigma_{lk}}{\|\mathbf{d_k}\| \, \|\mathbf{d_l}\|}$$

Eigenvectors

Given a square symmetric matrix A, there exists a set of vectors **x**, known as **eigenvectors** that have the property that

$$\mathbf{Ax} = \lambda \mathbf{x}$$

The values λ are the eigenvalues that capture the scale of the eigenvectors.
A is of dimension (p,p) then there can be up to p eigenvectors with p distinct eigenvalues.
Eigenvectors corresponding to non-zero eigenvalues are linearly independent and have length or norm of 1.
If a matrix of dimension (p,p) has p distinct non-zero eigenvalues, the eigenvectors are all linearly independent, and **form a vector basis**.

Eigenvectors form a Rotation Matrix

Let's the define the matrix **M** (known as the modal matrix) whose columns are eigenvectors of the matrix square symmetric (p,p) matrix **A**.

Since eigenvectors are linearly independent vectors of unit length, the matrix M is an orthogonal matrix like a rotation matrix.(In fact it is a rotation matrix!)

Let Λ be a matrix with the eigenvalues λ corresponding to each eigenvector along the diagonal, and zero at all off-diagonal values.

$$\mathbf{M}^T A\mathbf{M} = \Lambda$$

Thus the matrix **A** can be **diagonalized** by the model matrix **M**.

Eigenvectors of the Covariance Matrix

Consider the covariance matrix

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

of dimensions (p,p) corresponding to the p variables in an experiment.

The covariance matrix has eigenvectors which can be used to form a modal matrix **V**, whose columns are the eigenvectors. V diagonalizes the covariance matrix Σ such that

$$\mathbf{V}^T\Sigma\mathbf{V} = \Lambda$$

The notion of a diagonal covariance matrix has significant meaning. Covariance (or correlation) expresses the linear relationship between two variables. The covariance matrix contains these relationships between all pairs of variables.
If the covariance matrix is diagonalized, it means **there is no correlation between the variables**

Latent Variables

We compute the covariance from the data matrix as

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

$$\frac{1}{n}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \Lambda$$

$$\frac{1}{n}(\mathbf{XV})^T\mathbf{XV} = \Lambda$$

This rotation matrix can be used to compute a new data matrix $\mathbf{Y} = \mathbf{XV}$ which has the property of having a diagonal correlation matrix, i.e., **these new variables are uncorrelated, with covariance equal to zero**.
These *latent* variables are weighted mixtures of the original variables, with the weights given by the eigenvectors.