Statistics 135

Chapter 2 - Linear Algebra

Chris Drake
Department of Statistics
University of California, Davis

Vectors

1 definition of vector, multiplication by a constant and addition of two vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad c\mathbf{x} = \begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

2 length or norm of a vector

$$L_x = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} = || \mathbf{x} || \quad L_{cx} = c \times L_x = | c | \cdot || \mathbf{x} ||$$

3 unit vector: $L_x = || \mathbf{x} || = 1$

4 inner product of two vectors $\mathbf{x}'\mathbf{y} = \sum_{i=1}^{n} x_i y_i$

5 angle θ between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$cos(\theta) = \left[\mathbf{x}'\mathbf{y}\right] / \sqrt{\mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}}$$

6 def: a set of m-tuples (vectors) and all other vectors that can obtained as as linear combinations of the form $\mathbf{x} = \sum_{i=1}^{k} a_i \mathbf{x_k}$ is called a vector space

7 def: the linear span of a set of vectors $\mathbf{x_i}, .., \mathbf{x_k}$ is the set of all linear combinations $\mathbf{x} = \sum_{i=1}^{k} a_i \mathbf{x_k}$

8 def: a set of vectors is linearly independent if $\mathbf{x} = \sum_{i=1}^{k} a_i \mathbf{x_k} = 0$ implies $a_1, .., a_k = 0$

9 def: a set of m linearly independent m-tuples (vectors) is called a basis for the space of m-tuples.

10 fact: any vector can be expressed uniquely in terms of a given basis

11 fact: $\mathbf{x}$ and $\mathbf{y}$ are orthogonal (perpendicular) if $\mathbf{x}'\mathbf{y} = 0$

12 fact: the projection of $\mathbf{x}$ on $\mathbf{y}$ is

$$P_{\mathbf{y}} = \frac{\mathbf{x}'\mathbf{y}}{|| \mathbf{y} ||^2} \mathbf{y}$$

Matrices

A matrix $\mathbf{A}$ is an array of real numbers with n rows and p columns. Its transpose $\mathbf{A}'$ is the array with rows and columns exchanged.

$$\mathbf{A} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad \textit{and} \quad \mathbf{A}' = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{np} \end{pmatrix}$$

1 the multiplication of a matrix by a constant is the element by element multiplication by this constant, similar to a vector multiplication by a constant.

2 two matrices of the same dimension (n rows, p columns) are added by summing element by element; $(a+b)_{ij} = a_{ij} + b_{ij}$.

3 the product of two matrices $\mathbf{A}\cdot\mathbf{B}$ is defined if the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$; then $(ab)_{ij} = \sum_{l=1}^{k} a_{il} b_{lj}$ and $\mathbf{A} \cdot \mathbf{B}$ has r_A rows and c_B columns.

4 def: a square matrix has the same number of rows and columns; a symmetric matrix is a square matrix that equals its transpose.

5 fact: if $\mathbf{A}$ and $\mathbf{B}$ are square matrices then $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{B} \cdot \mathbf{A}$ are both defined but need not be equal; matrix multiplication is not commutative in general.

6 def: an inverse $\mathbf{B}$ for a square matrix $\mathbf{A}$ is defined by that fact that $\mathbf{B} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{B} = \mathbf{I}$.

7 fact: a square matrix $\mathbf{A}$ has an inverse if the columns (rows) are linearly independent.

8 def: a square matrix $\mathbf{Q}$ is called orthogonal if $\mathbf{Q} \cdot \mathbf{Q}' = \mathbf{Q}' \cdot \mathbf{Q} = \mathbf{I}$.

9 def: an eigenvalue of a square matrix $\mathbf{A}$ is denoted by λ the solution to the following equation: $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. The vector $\mathbf{x}$ is called an eigenvector. The number of nonzero eigenvalues equals the number of linearly independent columns (rows); eigenvectors belonging to two different eigenvalues are orthogonal; if a matrix has an eigenvalue of multiplicity > 1 the associated eigenvectors are orthogonal but not unique.

10 def: a square $n \times n$ matrix $\mathbf{A}$ is called of full rank if it has n linearly independent columns (rows); the number of independent columns (rows) is called the rank of $\mathbf{A}$; row and column rank of a matrix are always equal.

11 def: the spectral decomposition of a square matrix $\mathbf{A}$ is given by the following $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ where $\mathbf{P}$ is orthogonal and $\mathbf{\Lambda}$ is a diagonal matrix with

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

The number of nonzero eigenvalues equals the rank of $\mathbf{A}$.

12 def: an equation $\mathbf{x}'\mathbf{A}\mathbf{x}$ with $\mathbf{A}$ an $n \times n$ square matrix and $\mathbf{x}$ a vector (n-tuple) is called a quadratic form (note that this form is an equation with only quadratic terms in the x_i and second order terms $x_i x_j$.

13 def: a quadratic form and matrix $\mathbf{A}$ are called positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

14 def: if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ then $\mathbf{A}$ and the quadratic form are called nonnegative definite.

15 def: the square root of a matrix $\mathbf{A}$ is given by $\mathbf{A}^{\mathbf{1/2}} = \mathbf{P}\boldsymbol{\Lambda}^{\mathbf{1/2}}\mathbf{P}'$ where

$$\boldsymbol{\Lambda}^{\mathbf{1/2}} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}$$

16 fact:

$\mathbf{A}^{\mathbf{1/2}}$ is symmetric

$\mathbf{A}^{\mathbf{1/2}}\mathbf{A}^{\mathbf{1/2}} = \mathbf{A}$

$\mathbf{A}^{-(\mathbf{1/2})} = \mathbf{P}\boldsymbol{\Lambda}^{-(\mathbf{1/2})}\mathbf{P}'$

17 a full rank square matrix $\mathbf{A}$ has inverse $\mathbf{A}^{-\mathbf{1}} = \mathbf{P}\boldsymbol{\Lambda}^{-\mathbf{1}}\mathbf{P}'$ where $\boldsymbol{\Lambda}^{-\mathbf{1}}$ has i^{th} diagonal element λ_i^{-1}.

Random Vectors and Matrices

$\mathbf{X} = (X_1, .., X_n)'$ is a vector of random variables. Similarly, a random matrix is a $(n \times p)$ array of random variables. The mean of a random vector and matrix are obtained by taking expectation for each element.

$$E(\mathbf{X}) = \begin{pmatrix} E(X_{11}) & E(X_{12}) & \dots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \dots & E(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \dots & E(X_{np}) \end{pmatrix}$$

1 fact: $E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$

2 fact: $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$ for **A, B** constant matrices and $\mathbf{X}$ a random matrix.

3 fact: the behavior of random vectors and matrices is generally described by their probability distribution; generally, the variables are not independent.

4 fact: if $\mathbf{X}$ is a p-variate random vector with density $f(x_1, x_2, .., x_p)$ then $(X_1, .., X_p)$ independent if and only if $f(x_1, x_2, .., x_p) = f(x_1) \times ... \times f(x_p)$. Independence implies zero correlation but the converse is not true.

5 variance of a random vector

$$Var(\mathbf{X}) = \mathbf{\Sigma} = E(\mathbf{X} \times \mathbf{X}') - E(\mathbf{X}) \times E(\mathbf{X})'$$

note that the $(i, j)^{th}$ element of the first matrix is $E(X_i \times X_j)$ and of the second matrix it is $E(X_i) \times E(X_j) = \mu_i \times \mu_j$

6 the off-diagonal elements of the variance-covariance matrix are denoted by σ_{ij} and the diagonal elements by σ_{ii}.

7 the $(i, j)^{th}$ correlation is given by $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii} \times \sigma_{jj}}$

8 if $\mathbf{V^{1/2}}$ denotes the diagonal matrix with $\sqrt{\sigma_{ii}}$ on the diagonal and $\mathbf{R}$ the correlation matrix, then $\mathbf{\Sigma} = \mathbf{V^{1/2} R V^{1/2}}$

9 $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$ is a partition of $\mathbf{X}$; the mean vector is partitioned accordingly into $\mu = (\mu_{\mathbf{1}}, \mu_{\mathbf{2}})$ and the variance-covariance matrix is partitioned into

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Note, $\boldsymbol{\Sigma_{12}} = \boldsymbol{\Sigma'_{21}}$

10 linear combinations of random variables (vector notation); $\mathbf{X}$ a random vector (p-tuple); $\mathbf{c}$ a vector of length p of constants then $\mathbf{c}'\mathbf{X}$ is a linear combination of the $X's$ with mean $\mathbf{c}'\mu$ and variance $\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$.

11 extended Cauchy-Schwarz inequality: **b, d** are $p \times 1$ vectors and $\mathbf{B}$ a positive definite matrix; then

$$(\mathbf{b}'\mathbf{d})^{\mathbf{2}} \leq (\mathbf{b}'\mathbf{B}\mathbf{b})(\mathbf{d}'\mathbf{B}^{-\mathbf{1}}\mathbf{d}')$$

For $\mathbf{B} = \mathbf{I}$ we get the usual Cauchy-Schwarz inequality.

12 maximization lemma: $\mathbf{B_{p\times p}}$ positive definite and $\mathbf{d_{p\times 1}}$ a vector; then for an arbitrary nonzero vector $\mathbf{x}$ we have

$$\begin{matrix} max \\ \mathbf{x} \neq 0 \end{matrix} \quad \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\ \ \mathbf{x}} = \mathbf{d}'\mathbf{B}\ \ \mathbf{d}$$

for $\mathbf{B}_{p\times p}$ positive definite with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ we get

$$\begin{matrix} max \\ \mathbf{x} \neq 0 \end{matrix} \quad \frac{\mathbf{x}'\mathbf{B}\ \ \mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1$$

$$\begin{matrix} min \\ \mathbf{x} \neq 0 \end{matrix} \quad \frac{\mathbf{x}'\mathbf{B}\ \ \mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p$$

The maximum is achieved when $\mathbf{x}$ equals $\mathbf{e}_1$ the eigenvector associated with λ_1; similarly for the minimum.