Math 189: Matrix Algebra

I. Multivariate Data

Matrix Formulation

In multivariate analysis, we are interested in jointly analyzing multiple dependent variables.

- These variables can be represented using matrices and vectors.
- The benefits of matrix form:

1. Simplicity in notation
2. Expressing important formulas in readable format

Example 1: Multivariate Data of Infant Death

Data records several variables for a baby at birth:

1. Baby Weight (**bwt**): in ounces
2. Gestation (**gestation**): in days

Rows are babies, columns are variables

```
baby <- read.table("C:\\Users\\neide\\Documents\\GitHub\\ma189\\Data\\babies.dat",header=TRUE)
head(baby[,c(1,2)])
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
## 3 128       279
## 4 123       999
## 5 108       282
## 6 136       286
```

Notice anything weird?!

Source: The Child Health and Development Studies (CHDS) data are presented in *Stat Labs: Mathematical Statistics Through Applications* by Deborah Nolan and Terry Speed (Springer).

Math Notation

Let $\underline{x} = [x_1, x_2]'$ denote the column vector of **bwt** and **gestation**.

$$\underline{x} = \left[\begin{array}{c} x_1 \\ x_2 \end{array} \right] = \left[\begin{array}{c} \text{bwt} \\ \text{gestation} \end{array} \right].$$

So $\underline{x} \in \mathbb{R}^2$.

There are $n = 1236$ babies. Let $\underline{x}_i$ for $1 \leq i \leq 1236$ denote both variables for baby i, i.e.,

$$\underline{x}_1 = \left[\begin{array}{c} 120 \\ 284 \end{array} \right], \quad \underline{x}_2 = \left[\begin{array}{c} 113 \\ 282 \end{array} \right], \quad \underline{x}_3 = \left[\begin{array}{c} 128 \\ 279 \end{array} \right], \ldots$$

This is called the **record** for the ith baby (or unit). Let i index baby, and $j = 1, 2$ index variable. Then $x_{i,j}$ is **bwt** ($j = 1$) or **gestation** ($j = 2$) for baby i. Gestation for the fourth baby is $x_{4,2} = 999$.

A data matrix consists of all **records** as row vectors:

$$\begin{bmatrix} \underline{x}_1' \\ \underline{x}_2' \\ \underline{x}_3' \end{bmatrix} = \begin{bmatrix} 120 & 284 \\ 113 & 282 \\ 128 & 279 \end{bmatrix}.$$

```
x <- baby[1:3,1:2]
x
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
## 3 128       279
```

Each row is information for one baby; each column is all measurements (across babies) of either **bwt** or **gestation**

Definition of Matrix and Vector

A matrix is a two-dimensional array of numbers, symbols, or expressions, arranged in rows and columns.

- A vector is a matrix with either one row or one column.
- The dimension of a matrix is two numbers: number of rows (units) and number of columns (units). Dimension of baby is 1236, 7 and dimension of submatrix x is 3, 2.
- A square matrix is one for which the numbers of rows and number of columns are equal, e.g.

```
x <- baby[1:2,1:2]
dim(x)
```

```
## [1] 2 2
```

Notations of Matrix and Vector

- Usually in this course we only consider matrices and vectors whose entries are real numbers.
- A real-valued **matrix** $\mathbf{M}$ of dimension $n \times m$ is denoted $\mathbf{M} \in \mathbb{R}^{n \times m}$.
- A real **vector** $\underline{v}$ of length n is denoted $\underline{v} \in \mathbb{R}^n$.
- A row vector can be written $\underline{v} \in \mathbb{R}^{1 \times n}$, and a column vector can be written $\underline{v} \in \mathbb{R}^{n \times 1}$.
- A **scalar** is a real number, or a vector of length 1, or a 1×1 dimensional matrix.

```
x <- as.matrix(baby[1:3,1:2])
x
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
## 3 128       279
```

```
x[,1]
```

```
##   1   2   3
## 120 113 128
```

```
x[,1,drop=FALSE]
```

```
##   bwt
## 1 120
## 2 113
## 3 128
```

```
x[1,1]
```

```
## [1] 120
```

```
x[1,1,drop=FALSE]
```

```
##   bwt
## 1 120
```

Transpose of a Matrix

- Definition: flip the rows and columns by interchanging row and column index. Denoted by $\mathbf{M}'$.
- If $\mathbf{M}_{i,j}$ is row i, column j of matrix $\mathbf{M}$, then $\mathbf{M}'_{i,j} = \mathbf{M}_{j,i}$, so that now i is the column index, and j is the row index.

```
x
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
## 3 128       279
```

```
t(x)
```

```
##             1   2   3
## bwt       120 113 128
## gestation 284 282 279
```

Symmetric Matrices

- Definition: A square matrix $\mathbf{M}$ is symmetric if $\mathbf{M}' = \mathbf{M}$.
- Hence $\mathbf{M}_{i,j} = \mathbf{M}_{j,i}$.

```
m <- t(x) %*% x
m
```

```
##              bwt gestation
## bwt        43553    101658
## gestation 101658    238021
```

- Important examples of symmetric matrices in multivariate statistics include the variance-covariance matrix and the correlation matrix.

Linear Combination of Two Matrices

- Two matrices may be added if and only if they have the same dimensions. To add two matrices, add the corresponding elements.

```
y <- as.matrix(baby[5:7,1:2])
y
```

```
##   bwt gestation
## 5 108       282
## 6 136       286
## 7 138       244
```

```
x+y
```

```
##   bwt gestation
## 1 228       566
## 2 249       568
## 3 266       523
```

- If $\mathbf{A}$ and $\mathbf{B}$ are two $n \times m$ dimensional matrices, α and β are two scalars, then $\mathbf{C} = \alpha\,\mathbf{A} + \beta\,\mathbf{B} \in \mathbb{R}^{n\times m}$. A scalar multiples each entry of the matrix, i.e.,

$$\mathbf{C}_{i,j} = \alpha\,\mathbf{A}_{i,j} + \beta\,\mathbf{B}_{i,j}.$$

```
2*x + 3*y
```

```
##   bwt gestation
## 1 564      1414
## 2 634      1422
## 3 670      1290
```

Multiplication of Two Matrices

- The **dot product** of two vectors $\underline{x}$ and $\underline{y}$ is the sum of the product of their entries: $\underline{x} \cdot \underline{y} = \sum_{i=1}^{n} x_i\,y_i$.

```
sum(x[,1]*x[,2])
```

```
## [1] 101658
```

```
t(x[,1,drop=FALSE]) %*% x[,2,drop=FALSE]
```

```
##     gestation
## bwt    101658
```

- We multiply matrices by taking dot products of row and column vectors: $\mathbf{C} = \mathbf{A}\,\mathbf{B}$ is defined if number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$. Then

$$\mathbf{C}_{i,j} = \mathbf{A}_{i,-} \cdot \mathbf{B}_{-,j} = \sum_{k=1}^{p} \mathbf{A}_{i,k}\,\mathbf{B}_{k,j}.$$

  Here $\mathbf{A} \in \mathbb{R}^{n\times p}$, $\mathbf{B} \in \mathbb{R}^{p\times m}$, and $\mathbf{C} \in \mathbb{R}^{n\times m}$.

```
x %*% t(y)
```

```
##       5     6     7
## 1 93048 97544 85856
## 2 91728 96020 84402
## 3 92502 97202 85740
```

```
sum(x[1,]*t(y)[,3])
```

```
## [1] 85856
```

Identity Matrix

- Definition: An identity matrix $\mathbf{I}$ is a square matrix that has ones on its diagonal (from upper left to bottom right) and has zeros elsewhere.

```
eye <- diag(2)
eye
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

- When multiplying any square matrix, it leaves the matrix unchanged: $\mathbf{I}\,\mathbf{A} = \mathbf{A}$.

```
z <- as.matrix(baby[1:2,1:2])
z
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
```

```
eye %*% z
```

```
##      bwt gestation
## [1,] 120       284
## [2,] 113       282
```

Matrix Inverse

- Definition: The inverse of a square matrix $\mathbf{A}$, if it exists (the inverse of zero does not exist!?), is denoted by $\mathbf{A}^{-1}$, and satisfies
$$\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{A}\,\mathbf{A}^{-1} = \mathbf{I}.$$

```
a <- t(x) %*% x
a_inv <- solve(a)
a_inv
```

```
##                    bwt    gestation
## bwt        0.007396631 -0.003159077
## gestation -0.003159077  0.001353433
```

```
a %*% a_inv
```

```
##           bwt    gestation
## bwt         1 2.842171e-14
## gestation   0 1.000000e+00
```

```
a_inv %*% a
```

```
##                    bwt    gestation
## bwt       1.000000e+00 1.136868e-13
## gestation 2.842171e-14 1.000000e+00
```

What is a numerical zero in R?

- If the inverse exists, we say the matrix is **invertible**.

Matrix Trace

- Definition: The trace of an $n \times n$ matrix $\mathbf{A}$ is the sum of its diagonal elements:

$$\mathrm{tr}\mathbf{A} = \sum_{i=1}^{n} \mathbf{A}_{i,i}.$$

```r
sum(diag(a))
```

```
## [1] 281574
```

- For two matrices $\mathbf{A} \in \mathbb{R}^{n\times p}$, $\mathbf{B} \in \mathbb{R}^{p\times n}$,

$$\mathrm{tr}\mathbf{A}\,\mathbf{B} = \mathrm{tr}\mathbf{B}\,\mathbf{A}.$$

```r
x
```

```
##   bwt gestation
## 1 120       284
## 2 113       282
## 3 128       279
```

```r
z <- t(y)
z
```

```
##             5   6   7
## bwt       108 136 138
## gestation 282 286 244
```

```r
sum(diag(x %*% z))
```

```
## [1] 274808
```

```r
sum(diag(z %*% x))
```

```
## [1] 274808
```

Eigenvalues and Eigenvectors

- Given a symmetric matrix $\mathbf{A}$, then any scalar λ and vector $\underline{v}$ such that

$$\mathbf{A}\,\underline{v} = \lambda\,\underline{v}$$

are an **eigenvalue** and **eigenvector** (together called an **eigenpair**).

```r
eig <- eigen(a)
a %*% eig$vector[,1]
```

```
##               [,1]
## bwt       110594.8
## gestation 258821.1
```

```r
eig$values[1] * eig$vector[,1]
```

```
## [1] 110594.8 258821.1
```

. - A $n \times n$ symmetric matrix has n real eigenvalues, denoted $\lambda_1, \ldots, \lambda_n$.

- Fact: $\mathrm{tr}\mathbf{A} = \sum_{i=1}^{n} \lambda_i$.

```r
eig$values
```

```
## [1] 281459.6687    114.3313
```

```r
sum(eig$values)
```

```
## [1] 281574
```

```r
sum(diag(a))
```

```
## [1] 281574
```

- Fact: the determinant is the product of the eigenvalues: $\det \mathbf{A} = \prod_{i=1}^{n} \lambda_i$.

Positive Definite Matrix

- A symmetric $n \times n$ dimensional matrix $\mathbf{A}$ is **positive definite** if $\underline{v}'\mathbf{A}\underline{v} > 0$ for all non-zero length n vectors $\underline{v}$. Compactly written $\mathbf{A} > 0$.

```r
v <- rnorm(2)
t(v) %*% a %*% v
```

```
##          [,1]
## [1,] 12128.92
```

- A symmetric $n \times n$ dimensional matrix $\mathbf{A}$ is **positive semi-definite** if $\underline{v}'\mathbf{A}\underline{v} \geq 0$ for all length n vectors $\underline{v}$. Compactly written $\mathbf{A} \geq 0$.
- Fact: a symmetric $n \times n$ dimensional matrix is **positive definite** if and only if all of its eigenvalues are positive. (Remember, they must be real by symmetry!)
- Fact: a **positive definite** matrix is invertible.