Chapter 2 Rough review of linear algebra and linear regression

AMA4680
Fu Guanxing
guanxing.fu@polyu.edu.hk

Linear algebra

Vectors

An array $\mathbf{x}$ of n real numbers $x_1, \ldots, x_n$ is called a *vector*, and it is written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{or} \quad \mathbf{x} = [x_1, x_2, \cdots, x_n]'.$$

- Here the prime denotes the operation of transposing a column to a row.
- The number n is referred to as the *dimension* of the vector $\mathbf{x}$.
- The *coordinates* $x_1, \ldots, x_n$ can also be complex numbers, in which case $\mathbf{x}$ is a complex vector. However in this subject we consider ONLY real numbers and real vectors.
- The set of all the real vectors of dimension n, is denoted by $\mathbb{R}^n$.
- Convention. Without otherwise stated, all vectors are column ones.

Operations of vectors

- Multiplication: One can scale a vector $\mathbf{x}$ by multiplying it by a constant c.

$$\mathbf{x} = [x_1, x_2, \cdots, x_n]' \quad \Rightarrow \quad c\mathbf{x} = [cx_1, cx_2, \cdots, cx_n]'.$$

- Addition: Vectors can be added.

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}.$$

- Inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^{n} x_i y_i = \langle \mathbf{y}, \mathbf{x} \rangle = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}.$$

- Length, or Norm:

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

- For any constant c,

$$\langle c\mathbf{x}, \mathbf{y} \rangle = c \langle \mathbf{x}, \mathbf{y} \rangle ;$$
$$\|c\mathbf{x}\| = \sqrt{\langle c\mathbf{x}, c\mathbf{x} \rangle} = \sqrt{c^2 \langle \mathbf{x}, \mathbf{x} \rangle} = |c|\sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = |c|\|\mathbf{x}\|.$$

- Triangle inequality: (proof left as a warming up exercise)

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

Matrix

- Matrix of dimension $m \times n$: a table of (real) numbers.

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{bmatrix}.$$

- $A_{i,j}$ is called the (i, j)-entry of the matrix A.
- Transpose: A' is a $n \times m$ matrix, defined by

$$A' = \begin{bmatrix} A_{1,1} & A_{2,1} & \cdots & A_{m,1} \\ A_{1,2} & A_{2,2} & \cdots & A_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,n} & A_{2,n} & \cdots & A_{m,n} \end{bmatrix}.$$

Matrix Operation

- Matrix addition: $A = A_{m\times n}$, $B = B_{m\times n}$, then $A + B$ is an $m \times n$ matrix with $(A+B)_{i,j} = A_{i,j} + B_{i,j}$. Matrix $A - B$ is similarly defined.
- Matrix multiplication: $A = A_{m\times n}$, $B = B_{n\times p}$, then AB is a $m \times p$ matrix with

$$(AB)_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$

- Let c be a constant, then cA is an $m \times n$ matrix with $(cA)_{i,j} = cA_{i,j}$.
- Let $A = A_{m\times n}$ be a matrix and $x = x_{n\times 1}$ be a vector, then Ax is an $m \times 1$ vector with $(Ax)_i := A_{i,1}x_1 + A_{i,2}x_2 + \cdots + A_{i,n}x_n$.
- $(AB)' = B'A'$.
- For any $A \in \mathbb{R}^{n,m}$, $x \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, then $\langle Ax, c\rangle = \langle x, A'c\rangle$.

Square Matrix (1/3)

- A matrix $A_{m\times n}$ is called a *square matrix* if $m = n$.
- Matrix $I_{m\times m}$ is called *identity matrix*, if $I_{i,j} = 1$ for $i = j$ and $I_{i,j} = 0$ for $i \neq j$. For any $B_{m\times p}$ and $C_{p\times m}$, one has $IB = B$ and $CI = C$.
- A square matrix $A_{m\times m}$ is invertible if there exists some other matrix B such that $AB = I$. Whenever $AB = I$, we always have $BA = I$; vice versa. In this case we write $A^{-1} := B$.
- A square matrix A is called a *projection* matrix if $AA = A$. If this is satisfied, then $A^n = \underbrace{A \cdots A}_{n} = A$.
- The matrix $A_{m\times m}$ is called symmetric, if $A' = A$.
- The matrix $A_{m\times m}$ is called an *orthogonal* (or *orthonormal*) matrix if $A^{-1} = A'$. Therefore $AA' = A'A = I$.

Square Matrix (2/3)

- The determinant of a **square matrix** A: det(A). We have det$(AB) =$ det(A)det(B).
- The matrix A is invertible if and only if det$(A) \neq 0$.
- Eigenvalues: λ is an eigenvalue of A if and only if det$(A - \lambda I) = 0$.
- Real matrix may have complex eigenvalues. Each eigenvalue λ corresponds to at least one eigen vector x, such that

  $$Ax = \lambda x.$$

  When we talk about eigenvector, we mean that it is not zero (meaning at least one coordinate of the vector is not zero).
- When we talk about eigenvalues of A, A must be a square matrix.

- For a square matrix A, the trace of A is defined as $\text{Tr}(A) = \sum_i A_{i,i}$. One has $\text{Tr}(AB) = \text{Tr}(BA)$.
- Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of A, counting multiplicity, then $\det(A) = \prod_{i=1}^m \lambda_i$. Therefore A is invertible if and only if it has no zero eigenvalue. It also holds $\text{Tr}(A) = \sum_{i=1}^m \lambda_i$.
- $\det(A) = \det(A')$, so if $AA' = I$, then $\det(A)$ is either 1, or -1.
- Since $\det(I) = 1 = \det(AA^{-1}) = \det(A)\det(A^{-1})$, one has $\det(A^{-1}) = \frac{1}{\det(A)}$.
- We call A a diagonal matrix if A is a square matrix and $A_{i,j} = 0$ whenever $i \neq j$.

Range, null space, etc.

- The space spanned by a set of vectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ is

$$\{c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m : c_1, \cdots, c_m \in \mathbb{R}\}.$$

- The range, or image of a matrix A, is the space spanned by the vectors which are the columns of A: for $A \in \mathbb{R}^{n\times m}$,

$$\text{range}(A) = \{c_1A_1 + \cdots + c_mA_m : c_1, \cdots, c_m \in \mathbb{R}\} = \{Ac : c \in \mathbb{R}^m\},$$

  where A_i is column i of A.

- The null space of a matrix A, is the set $\{x : Ax = 0\}$.

Question: what are the range and null space of the identity matrix? What about the zero matrix?

Quadratic forms

- Let $A \in \mathbb{R}^{m \times m}$ and $x \in \mathbb{R}^m$. Consider $x'Ax$.
    - First, it is a 1×1 matrix, meaning it is a real number.
    - Second, we have

$$x'Ax = \sum_{i=1}^{m} \sum_{j=1}^{m} A_{i,j} x_i x_j,$$

where each term in the sum is a quadratic form $x_i x_j$. Therefore we call $x'Ax$ a quadratic form.

- A symmetric matrix $A_{m \times m}$ is called positive semidefinite, if

$$x'Ax \geq 0, \qquad \text{for all } x \in \mathbb{R}^m.$$

- A symmetric matrix $A_{m \times m}$ is called positive definite, if

$$x'Ax > 0, \qquad \text{for all } x \in \mathbb{R}^n \text{ such that } x \neq 0.$$

Derivative w.r.t. a vector

- Let $x \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{n \times m}$.

$$D_x \langle Ax, c \rangle = A'c, \quad D_x \langle Ax, Bx \rangle = B'Ax + A'Bx.$$

Linear regression

- The model:

$$y = \beta_0 + \beta_1 z_1 + \cdots + \beta_r z_r + \varepsilon.$$

Here,

- y is called the response variable, or the dependent variable.
- $z_1, z_2, \ldots, z_r$ are called the predictor variables, or the independent variables, or the explanatory variables.
- $\beta_0, \beta_1, \ldots, \beta_r$ are parameters. They are also called the effects.
- ε is called the error term, or the noise.

- Given the dataset (observations) with n samples $\{\boldsymbol{z}_i, y_i\}_{i=1}^n$, where $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \cdots, z_{ir})'$,

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 z_{11} + \cdots + \beta_r z_{1r} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 z_{21} + \cdots + \beta_r z_{2r} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 z_{n1} + \cdots + \beta_r z_{nr} + \varepsilon_n, \end{aligned}$$

estimate $\beta_0, \beta_1, \cdots, \beta_r$.

We write

$$\mathbf{Y} = \begin{bmatrix} 1 & z_{1,1} & z_{1,2} & \cdots & z_{1,r} \\ 1 & z_{2,1} & z_{2,2} & \cdots & z_{2,r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n,1} & z_{n,2} & \cdots & z_{n,r} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \boldsymbol{\varepsilon},$$

or,

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(r+1))}{\mathbf{Z}} \underset{((r+1)\times 1)}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}.$$

Least squares estimation

$$\ell(\boldsymbol{\beta}) = \sum_{j=1}^{n} (y_j - \beta_0 - \beta_1 z_{j,1} - \cdots - \beta_r z_{j,r})^2$$
$$= (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta})$$

When $\mathbf{Z}$ has full rank $r+1$, the least square estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}.$$