Mathematical Preliminaries

COMP 527 Data Mining and Visualisation

UNIVERSITY OF LIVERPOOL

Linear Algebra

- In Data Mining, we will represent data points using a set of coordinates (corresponding to various attributes/features). This mathematical representation is compact and powerful enough to describe parallel processing methods.
- The branch of mathematics that concerns with such coordinated representations is called **linear algebra**
- Reference: Chapter 02 of the MML book [https://mml-book.github.io/book/chapter02.pdf]

Vectors

- We will denote a vector **x** in the n-dimensional real space by (lowercase bold fonts) $\mathbf{x} \in \mathbb{R}^n$
- We will use column vectors throughout this module (transposed by T when written as row vectors)
- e.g. **x** = $(3.2, -9.1, 0.1)^T$
- A function can be seen as an infinite dimensional vector, where all function values are arranged as elements in the vector!

Matrices

- We obtain matrices by arranging a collection of vectors by columns or rows.
- We use uppercase bold fonts to denote matrices such as $\mathbf{M} \in \mathbb{R}^{n \times m}$
- When n = m we say M is square
- We denote the (i,j) element of **M** by $M_{i,j}$
- If $M_{i,j} = M_{j,i}$ for all i and j, we say **M** is symmetric. Otherwise, **M** is asymmetric
- If all elements in **M** are real numbers, then we call **M** to be a real matrix, otherwise a complex matrix

Vector arithmetic

- Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ their *addition* is given by the vector $\mathbf{z} \in \mathbb{R}^N$ where i-th element z_i is given by $z_i = x_i + y_i$
- Their element-wise product (Hadamard product $\otimes$) is given by $z_i = x_i y_i$
- Their inner-product (dot product) is defined as

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{N} x_i y_i$$

- Their outer-product ($\mathbf{x}\mathbf{y}^\top$) is defined as the matrix $\mathbf{M} \in \mathbb{R}^{N \times N}$ where $M_{i,j} = x_i y_j$

Quiz

- Given $\mathbf{x} = (1,2,3)^T$ and $\mathbf{y} = (3,2,1)^T$
  - Find $\mathbf{x} + \mathbf{y}$
  - Find $\mathbf{x} \otimes \mathbf{y}$
  - Find $\mathbf{x}^T\mathbf{y}$
  - Find $\mathbf{x}\mathbf{y}^T$

Matrix arithmetic

- Matrices of the same shape (number of rows and columns) can be added elementwise
  - **A** + **B** = **C** where $C_{i,j} = A_{i,j} + B_{i,j}$
- Matrices can be multiplied if the number of columns of the first matrix is equal to the number of rows of the second matrix

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

  - **AB** = **C** where $$C_{i,j} = \sum_{k=1}^{m} A_{i,k} B_{k,j}$$

Quiz

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 3 \\ -1 & 0 & 1 \end{pmatrix}$$

- Compute **A**+**B**
- Compute **B**+**A**

- Compute **AB**
- Compute **BA**
- Is matrix product commutative in general?

Transpose and Inverse

- The transpose of a matrix **A** is denoted by $\mathbf{A}^T$ and the (i,j) element of the transpose is $A_{j,i}$
  - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
- The inverse of a square matrix **A** is denoted by $\mathbf{A}^{-1}$ and satisfies $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
  - Here, $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the unit matrix (all diagonal elements are set to 1 and non-diagonal elements are set to 0)

Computing the inverse of a 2x2 matrix

- Compute the inverse of the following matrix

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

Determinant of a matrix

- Determinant of a matrix **A** is denoted by $|\mathbf{A}|$
- For a 2x2 matrix **A** its determinant is given by

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, |\mathbf{A}| = ad - bc$$

Quiz: Matrix inversion

- Write the generalised form for the inverse of a 2x2 matrix using the matrix determinant.

Linear independence

- Let us consider a vector **v** formed as the linearly-weighted sum of a set of vectors $\{\mathbf{x}_1,\ldots,\mathbf{x}_K\}$ with respective coefficients $\lambda_1,\ldots,\lambda_K$ as follows:

$$\mathbf{v} = \lambda_1\mathbf{x}_1 + \ldots + \lambda_K\mathbf{x}_K = \sum_{i=1}^{K} \lambda_i\mathbf{x}_i$$

- **v** is called a linear combination of $\{\mathbf{x}_1,\ldots,\mathbf{x}_K\}$
- The null vector **0** can always be represented as a linear combination of K vectors (Quiz: show this)
- We are interested in cases where we can represent a vector as the linear combination of non-zero coefficients.

Quiz: Linear independence

- Show that **v** cannot be expressed as a linear combination of **a** and **b**, where
  $\mathbf{v} = (1, 2, -3, 4)^T$
  $\mathbf{a} = (1, 1, 0, 2)^T$
  $\mathbf{b} = (-1, -2, 1, 1)^T$

Rank

- The number of linear independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (m≤n) equals the number of linearly independent rows and is called the **rank** of **A** is denoted by rank(**A**)
- rank(**A**) ≤ min(m,n) = m
- If rank(**A**) = m, then **A** is said to be full-rank, otherwise rank deficit.
- Only full-rank square matrices are invertible.

Quiz:

- Find the ranks of the following matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix}$$

Matrix trace

- The sum of diagonal elements is called the trace of the matrix. Specifically,

$$\mathrm{tr}(\mathbf{A}) = \sum_i A_{i,i}$$

- Find tr(**A**) for

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

Eigendecomposition

- Let **A** $\in\mathbb{R}^{n\times n}$ be a square matrix. Then $\lambda\in\mathbb{R}$ is an **eigenvalue** of **A** and a nonzero $\mathbf{x}\in\mathbb{R}^n$ is the corresponding **eigenvector** of **A** if
  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$
- We call this the *eigenvalue equation*
- an n-dimensional square matrix has exactly n eigenvectors and we can express **A** using its eigenvectors as follows. This called the **eigendecomposition** of **A**.

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i^\top$$

Quiz:

- Find the eigenvalues and the corresponding eigenvectors of **A**

$$\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$$

Vector Calculus

- This is also known as multivariate calculus, where we have functions of multiple variables (such as the dimensions in a vector) and we must compute partial or total derivatives w.r.t. the variables.
- All what you know from A/L calculus is still valid and can be used to derive the rules in vector calculus starting from the first principles.

Differentiation Rules

Product Rule: $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$

Quotient Rule: $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$

Sum Rule: $(f(x) + g(x))' = f'(x) + g'(x)$

Chain Rule: $\big(g(f(x))\big)' = (g \circ f)'(x) = g'(f(x))f'(x)$

Quiz: Given f(x) = log(x) and g(x) = 2x + 1, compute the four derivatives corresponding to the rules stated above.

Partial derivative

Definition 5.5 (Partial Derivative)**.** For a function $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x} \mapsto f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^n$ of n variables $x_1, \ldots, x_n$ we define the *partial derivatives* as

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= \lim_{h\to 0} \frac{f(x_1 + h, x_2, \ldots, x_n) - f(\boldsymbol{x})}{h} \\ &\vdots \\ \frac{\partial f}{\partial x_n} &= \lim_{h\to 0} \frac{f(x_1, \ldots, x_{n-1}, x_n + h) - f(\boldsymbol{x})}{h}\end{aligned} \tag{5.39}$$

and collect them in the row vector

$$\nabla_{\boldsymbol{x}} f = \operatorname{grad} f = \frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} & \frac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1\times n}\,, \tag{5.40}$$

where n is the number of variables and 1 is the dimension of the image/range/co-domain of f. Here, we defined the column vector $\boldsymbol{x} = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$. The row vector in (5.40) is called the *gradient* of f or the *Jacobian* and is the generalization of the derivative from Section 5.1.

Quiz: For $f(x,y) = (x+2y^3)^2$ compute $\partial f/\partial x$, $\partial f/\partial y$ and $\nabla_{(x,y)} f$

Chain rule for multivariate functions

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ of two variables x_1, x_2. Furthermore, $x_1(t)$ and $x_2(t)$ are themselves functions of t. To compute the gradient of f with respect to t, we need to apply the chain rule (5.48) for multivariate functions as

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1(t)}{\partial t} \\ \frac{\partial x_2(t)}{\partial t} \end{bmatrix} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t} \tag{5.49}$$

where d denotes the gradient and ∂ partial derivatives.

Quiz: Consider $f(x_1,x_2)=x_1^2 + 2x_2$, where $x_1=\sin(t)$ and $x_2=\cos(t)$. Find, df/dt.

Useful identities for computing gradients

$$\frac{\partial}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^\top = \left(\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right)^\top \tag{5.99}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{f}(\boldsymbol{X})) = \operatorname{tr}\left(\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right) \tag{5.100}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \det(\boldsymbol{f}(\boldsymbol{X})) = \det(\boldsymbol{f}(\boldsymbol{X})) \operatorname{tr}\left(\boldsymbol{f}(\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right) \tag{5.101}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^{-1} = -\boldsymbol{f}(\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^{-1} \tag{5.102}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X}^{-1} \boldsymbol{b}}{\partial \boldsymbol{X}} = -(\boldsymbol{X}^{-1})^\top \boldsymbol{a} \boldsymbol{b}^\top (\boldsymbol{X}^{-1})^\top \tag{5.103}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}^\top \tag{5.104}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}^\top \tag{5.105}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{b}}{\partial \boldsymbol{X}} = \boldsymbol{a} \boldsymbol{b}^\top \tag{5.106}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{x}^\top (\boldsymbol{B} + \boldsymbol{B}^\top) \tag{5.107}$$

$$\frac{\partial}{\partial \boldsymbol{s}} (\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s}) = -2(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W} \boldsymbol{A} \quad \text{for symmetric } \boldsymbol{W} \tag{5.108}$$

Note: You do not need to memorise these but must be able to verify these by yourself.