Modern Regression

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

1. Matrices: what you need to know

Outline

Definition 2.1

A two dimensional array of number is called a **matrix***. Typically, if A is a matrix which has 3 rows and 4 columns, we write $A \in \mathbb{R}^{3\times 4}$.*

Proposition 2.1

Let $A \in \mathbb{R}^{m \times n}$ *and* $B \in \mathbb{R}^{p \times q}$. *Assuming the results exist, we have*

- $(A + B)_{i,j} = A_{i,j} + B_{i,j}$.
- $(AB)_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$.
- *The transpose is such that* $(A^T)_{j,i} = A_{i,j}$

Definition 2.2

We say that a square matrix $A \in \mathbb{R}^{n \times n}$ *is invertible if there exists a matrix* $B \in \mathbb{R}^{n \times n}$ *such that*

$$AB = BA = I$$

where I *is a diagonal matrix with only 1's along the diagonal and 0 elsewhere.*

Proposition 2.2

Assuming a square matrix A is invertible, then its inverse is unique.
In this case we say that B is the inverse of A and we may write $B = A^{-1}$.

Definition 2.3

A matrix A is said to be symmetric if and only if $A^T = A$. It is antisymmetric if $A^T = -A$.

Proposition 2.3

Manipulating matrices is not just as trivial as manipulating numbers. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. The following rules apply

- $A + B$ *exists if and only if* $m = p$ *and* $n = q$.
- AB *exists if and only if* $p = n$.
- AB *exists does not imply* BA *exists (let alone* $AB = BA$*)*
- *if the product exists,* $(AB)^T = B^T A^T$.
- *if the inverses exist and the product exists, we have* $(AB)^{-1} = B^{-1}A^{-1}$.

Proposition 2.4

Let A be any square matrix. Then $B = \frac{A+A^T}{2}$ is symmetric.

Definition 2.4

Let A be an $n \times n$ square matrix. Its **trace** *is defined as*

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{i,i}.$$

Definition 2.5

Let $A \in \mathbb{R}^{n \times n}$ *be any square matrix. We define its* **determinant** *recursively via Laplace's formula:*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{i,j} \det(A^{-i,-j}) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det(A^{-i,-j})$$

where the formula holds for any j *in the first expression or any* i *in the second. We have also defined the submatrix* $A^{-i,-j}$ *as the matrix extracted from* A *by deleting row* i *and column* j*.*
The recursion is set by the case $n = 1$ *as* $\det\begin{pmatrix} a \end{pmatrix} = a$.

Proposition 2.5

The determinant of a 2×2 *matrix*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is given by

$$\det(A) = ad - bc.$$

Proposition 2.6

Let $A \in \mathbb{R}^{3\times 3}$*. We may use Sarrus' rule to compute a* 3×3 *determinant:*

$$\begin{aligned}\det(A) = A_{1,1}A_{2,2}A_{3,3} + A_{1,2}A_{2,3}A_{3,1} + A_{1,3}A_{2,1}A_{3,2}\\ - A_{1,3}A_{2,2}A_{3,1} - A_{1,2}A_{2,1}A_{3,3} - A_{1,1}A_{2,3}A_{3,2}.\end{aligned}$$

Proposition 2.7

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$*. If* A *is invertible, its inverse is given by*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Exercise 2.1

Find the inverses of the following matrices:

- $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$
- $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$
- $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

Proposition 2.8

The determinant has some useful properties:

- $\det(I_n) = 1$.
- $\det(A) = \det(A^T)$, *for any square matrix* A.
- *if the product exists for two square matrices* A *and* B, $\det(AB) = \det(A)\det(B)$.
- *if* $c \in \mathbb{R}$ *and* $A \in \mathbb{R}^{n\times n}$, $\det(cA) = c^n \det(A)$.

Definition 2.6

Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *are said to be* **orthogonal** *(or perpendicular) if*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{y}^T \mathbf{x} = 0.$$

Proposition 2.9

Remember that the inner product is linear: for any three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$*,*

$$\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle.$$

Example 2.1

Let $\mathbf{x} = (1, -2, 3)^T$, $\mathbf{y} = (1, 1, 1)^T$ and $\mathbf{z} = (-1, 1, 1)$. Show that $\mathbf{x}$ is orthogonal to $\mathbf{z}$ but not to $\mathbf{y}$.
Is $\mathbf{x}$ orthogonal to the vector $\mathbf{y} + \mathbf{z}$?

Definition 2.7

Let $A \in \mathbb{R}^{n\times n}$. The real-valued function

$$f(x) = x^T Ax$$

is called the **quadratic form** *of the vector x associated to the matrix A.*

- *We say that A is positive definite (resp. semi-definite) if $f(x) > 0$ (resp. ≥ 0) for all non-zero $x \in \mathbb{R}^n$.*
- *We say that A is negative definite (resp. semi-definite) if $f(x) < 0$ (resp. ≤ 0) for all non-zero $x \in \mathbb{R}^n$.*

Definition 2.8

Let $x \in \mathbb{R}^n$ and V a linear subset. The vector

$$y = \operatorname*{argmin}_{v \in V} \|x - v\|_2$$

is called the **projection** *of x onto V.*
We write $y = P_V(x)$.

Definition 2.9

Let $f : \mathbb{R}^d \to \mathbb{R}$ *be a real-valued multivariate and differentiable function. Then its* **gradient** *is defined as the* d *dimensional vector of partial derivatives:*

$$\operatorname{grad}(f)(x) := \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}.$$

Example 2.2

Let f be an affine function, i.e. there exists a vector $a \in \mathbb{R}^d$ and a constant $c \in \mathbb{R}$ such that $f(x) = a^T x + c = \langle x, a \rangle + c$. Then

$$\nabla f(x) = a, \forall x \in \mathbb{R}^n.$$

Definition 2.10

Let $f : \mathbb{R}^d \to \mathbb{R}^n$ *be a vector-valued differentiable function. Its Jacobian is the* $n \times d$ *matrix defined as, for all* $x \in \mathbb{R}^d$

$$J_f(x) = \begin{pmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_n(x)^T \end{pmatrix}$$

Proposition 2.10

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ *and* $g : \mathbb{R}^d \to \mathbb{R}^m$ *be two differentiable functions. The Jacobian of the composite function* $h = f \circ g$ *is obtained via the* **chain rule***:*

$$J_h(\mathbf{a}) = J_f(g(\mathbf{a}))J_g(\mathbf{a}).$$

Example 2.3

Let $A \in \mathbb{R}^{m \times d}$ be a matrix and $b \in \mathbb{R}^m$ a vector. Define $f(\mathbf{x}) = \langle A\mathbf{x}, b \rangle$. The gradient of f is given by

$$\nabla f(\mathbf{a}) = A^T b, \quad \forall \mathbf{a} \in \mathbb{R}^d.$$

Example 2.4

Let $f(x) = x^T Ax$ for $A \in \mathbb{R}^{d\times d}$. Its gradient reads $\nabla f(x) = (A + A^T)x$.

Proposition 2.11 (Derivative of bilinear forms – in a sense)

Let u, v be two functions from $\mathbb{R}^d$ to $\mathbb{R}^m$. Then the gradient of $f(x) = \langle u(x), v(x) \rangle$ is given by

$$\nabla f(x) = J_u(x)^T v(x) + J_v(x)^T u(x).$$

Proposition 2.12

The projection defined is an **orthogonal projection** *in the sense that the vectors* $x - P_V(x)$ *and* $P_V(x)$ *are orthogonal.*