Optimization and Multivariate Calculus

Joe Patten

August 7, 2019

1 Linear Algebra

Trace

The trace of an $n \times n$ matrix, donated tr, is the sum of the (main) diagonal. If $A = \begin{bmatrix} 3 & 7 \\ 2 & 8 \end{bmatrix}$, then $tr(A) = 11$.

Determinants

It is a bit difficult to describe what a determinant is, but this discussion on stack exchange seems to give the most intuitive idea. A determinant can only be computed for a square matrix. The determinant for a matrix, A, can either be denoted as $|A|$ or $det(A)$.

The determinant of a scalar a is just a.

The determinant of a 2×2 matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is:

$$a_{11}a_{22} - a_{21}a_{12}$$

The determinant of a 3×3 matrix $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is:

$$-1^{1+1} \cdot a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + -1^{1+2} \cdot a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + -1^{1+3} \cdot a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Note

We don't have to use the first row to calculate the determinant of a matrix that's bigger than 2×2. For example, if I chose to use the 2nd column, the determinant for the matrix above would now be:

$$-1^{1+2} \cdot a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + -1^{2+2} \cdot a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + -1^{3+2} \cdot a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}$$

If the determinant of a square matrix is nonzero, then that matrix is nonsingular.

Properties

- $|A| = |A^T|$
- $|A||B| = |AB|$

Practice

Use the definition of a determinant for an $n \times n$ matrix to show that the determinant of a 2×2 matrix (which was defined earlier) is equal to $a_{11}a_{22} - a_{21}a_{12}$.

Inverses

An $n \times n$ matrix A is invertible if there exists an $n \times n$ matrix B such that:

$$AB = BA = I_n \tag{1}$$

where I_n is an $n \times n$ identity matrix (described in the special matrices section).

Inverse Properties

1. $(A^{-1})^{-1} = A$
2. $(A^T)^{-1} = (A^{-1})^T$
3. $(cA)^{-1} = c^{-1}A^{-1}$
4. If A, B, and C are invertible $n \times n$ matrices, then $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$
5. $|A^{-1}| = |A|^{-1}$
6. $A^{-1}A = AA^{-1} = I$
7. $A^{-1} = \frac{1}{|A|}adj(A)$

Special Matrices

Square Matrix

The number of rows (n) equals the number of columns (n) for the matrix. The following is an example of a square matrix:

$$\begin{bmatrix} 10 & 5 & 9 \\ 4 & 4 & 3 \\ 6 & 17 & 2 \end{bmatrix} \tag{2}$$

Symmetric Matrix

A symmetric matrix has the following property: $A^T = A$. This means that $a_{ij} = a_{ji}$ for all i, j. Notice that this implies that a symmetric matrix has to be a square matrix ($n \times n$). The following is an example of a symmetric matrix:

$$\begin{bmatrix} 1 & 5 & 6 \\ 5 & 4 & 7 \\ 6 & 7 & 2 \end{bmatrix} \tag{3}$$

Idempotent Matrix

An idempotent matrix (A) has the following property: AA = A

Identity Matrix

An $n \times n$ identity matrix (either donated as I or I_n) has 1's on the diagonal and 0's elsewhere. Example:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

Multiplying any matrix by an identity matrix will return that matrix.

$$AI = A \tag{5}$$

$$IB = B \tag{6}$$

2 Derivatives

Recall from single-variable calculus, the derivative of a function f with respect to x at point x_0 is defined as:

$$\frac{df(x_0)}{dx} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

If this limit exists, then we say that f is differentiable at x_0. We can extend this definition to talk about derivatives of multivariate functions.

2.1 Partial Derivative

Let $f : \mathbb{R}^n \to \mathbb{R}$. The partial derivative of f with respect to variable x_i at $\mathbf{x}^0$ is given by:

$$\frac{\partial f(\mathbf{x}^0)}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_2, ..., x_i + h, ..., x_n) - f(x_1, x_2, ..., x_i, ..., x_n)}{h}$$

Notice that in this definition, the ith variable is affected. To take the partial derivative of variable x_i, we treat all the other variables as constants.

Example

Consider the function: $f(x, y) = 4x^2y^5 + 3x^3y^2 + 6y + 10$.

$\frac{\partial f(x,y)}{\partial x} = 8xy^5 + 9x^2y^2$

$\frac{\partial f(x,y)}{\partial y} = 20x^2y^4 + 6x^3y + 6$

2.2 Jacobian Matrix

We can put all of the partials of the function $F : \mathbb{R}^n \to \mathbb{R}$ at x^* (which we call the derivative of F) in a row vector:

$$DF_{x^*} = \begin{bmatrix} \frac{\partial F(x^*)}{\partial x_1} & \dots & \frac{\partial F(x^*)}{\partial x_n} \end{bmatrix}$$

This can also be referred to as the Jacobian derivative of F.

We can express the derivative in a column vector:

$$\nabla F_{x^*} = \begin{bmatrix} \frac{\partial F(x^*)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x^*)}{\partial x_n} \end{bmatrix}$$

This representation is usually referred to as the gradient vector.

Example

The gradient vector of our previous example would be:

$$\nabla F = \begin{bmatrix} 8xy^5 + 9x^2y^2 \\ 20x^2y^4 + 6x^3y + 6 \end{bmatrix}$$

2.3 Hessian Matrix

Recall that for an function of n variables, there are n partial derivatives. We can take partial derivatives of each partial derivative. The partial derivative of a partial derivative is called the second order partial derivative.

Example

The second order partial derivatives for the example above are defined as:

$\frac{\partial^2 f(x,y)}{\partial x^2} = 8y^5 + 18xy^2$

$\frac{\partial^2 f(x,y)}{\partial y^2} = 80x^2y^3 + 6x^3$

$\frac{\partial^2 f(x,y)}{\partial y \partial x} = 40xy^4 + 18x^2y$

$\frac{\partial^2 f(x,y)}{\partial x \partial y} = 40xy^4 + 18x^2y$

The second order partial derivatives of the form $\frac{\partial^2 f(x,y)}{\partial x \partial y}$ where $x \neq y$ are called the cross partial derivatives. Notice from our example, that $\frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial^2 f(x,y)}{\partial y \partial x}$. This is always the case with cross partials. We see that:

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$$

We can put all of these second order partials into a matrix, which is referred to as the Hessian Matrix:

$$\begin{bmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x^*)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n^2} \end{bmatrix}$$

Exercises

Let A be a convex subset of $\mathbb{R}^n$ where $f : A \to \mathbb{R}$. Let f be concave.

1. Compute the Hessian matrix for the following functions:
   (a) $f(x, y) = 4x^2y - 3xy^3 + 6x$
   (b) $f(x, y) = 3x^2y - 7x\sqrt{y}$
2. Calculate the determinant for the following matrix:

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

3. Let X be an $n \times n$ matrix. Show $X^{-1} = \left(X^T X\right)^{-1} X^T$