Derivation of scalar functions with respect to vectors

February 19, 2019

1 Notation

Matrices and vectors are well-known element of linear algebra. They are written by bold fonts.

Row vectors are denoted by as follows:

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_N \end{bmatrix}^T$$

Matrices are usually written by capital letters, e.g.

$$\boldsymbol{B} = \boldsymbol{B}_{M \times N} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1N} \\ b_{21} & b_{22} & \dots & b_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ b_{M1} & b_{M2} & \dots & b_{MN} \end{bmatrix}$$

Matrices can also be composed by column vectors, respectively:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}^1 \\ \boldsymbol{b}^2 \\ \vdots \\ \boldsymbol{b}^M \end{bmatrix}$$

As it is seen above, the upper and lower indices denote the row, and the columns vectors, respectively.

2 Derivatives

The definition of a scalar function J with respect to a vector $\mathbf{x}$ is as follows:

$$\frac{\partial J}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_N} \end{bmatrix}$$

2.1 Notable derivatives

2.1.1 L_2 norm of a vector

$$\frac{\partial \boldsymbol{x^T x}}{\partial \boldsymbol{x}} = 2\boldsymbol{x}$$

Proof:

$$J = \boldsymbol{x}^T \boldsymbol{x} = \sum_{i=1}^{N} x_i^2$$

$$\frac{\partial J}{\partial x_i} = 2x_i^2$$

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_N} \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = 2\boldsymbol{x}$$

2.1.2 Scalar product of two vectors

When the vector $\boldsymbol{x}$, containing of unknown corrdinates, is multiplied by another vector $\boldsymbol{a}$ with the same dimension, the following can be written doe to commutivity:

$$\frac{\partial \boldsymbol{x^T a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a^T x}}{\partial \boldsymbol{x}}$$

The scalas cost function is as follows:

$$J = \boldsymbol{x}^T \boldsymbol{a} = \sum_{i=1}^{N} a_i x_i$$

One element of the derivative vector is

$$\frac{\partial J}{\partial x_i} = a_i.$$

Therefore,

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_N} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \boldsymbol{a}$$

2.1.3 Vector-matrix-vector product

In this case, the scalar function is written as $J = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$,where $\boldsymbol{A}$ is an $N \times N$ matrix.

Thes scalar function can be written by applying double summa as

$$J = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i x_j$$

Its derivative w.r.t. x_k is

$$\frac{\partial J}{\partial x_k} = \sum_{i=1, i \neq k}^{N} a_{ik} x_i + \sum_{i=1, i \neq k}^{N} a_{ki} x_i + 2 a_{kk} x_k =$$

$$a_i^T \boldsymbol{x} + a^i \boldsymbol{x}$$

The whole derivative is given as

$$\frac{\partial \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} a_1^T \\ a_2^T \\ \dots \\ a_N^T \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} a^1 \\ a^2 \\ \dots \\ a^N \end{bmatrix} \boldsymbol{x} = \left(\boldsymbol{A^T} + \boldsymbol{A} \right) \boldsymbol{x}$$

Special case: **B** is symmetric, $\boldsymbol{B}^T = \boldsymbol{B}$. Then

$$\frac{\partial \boldsymbol{x}^T \boldsymbol{B} \boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\boldsymbol{B}^T + \boldsymbol{B} \right) \boldsymbol{x} = 2 \boldsymbol{B} \boldsymbol{x}.$$

3 Inhomogeneous over-detemined liner systems

The task is to solve the linear sisytem of equation

$$\boldsymbol{A} \boldsymbol{x} = \boldsymbol{b},$$

where matrix A is of size $m \times n$. Therefore, vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ have n and m columns and rows, respectively.

If matrix $\boldsymbol{A}$is not a square one, the problem is an estimation if $m > n$. In this case, the difference between the left and the right sides should be minimized. The standard, i.e. least-squares, solution is as follows:

$$\boldsymbol{x} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$$

where $\|\|_2$is the L_2-norm, also called as Euclidean distance. The L_2-norm of a vector $\boldsymbol{v}$ is

$$\|\boldsymbol{v}\|_2 = \sqrt{\boldsymbol{v}^T\boldsymbol{v}}$$

Therefore,

$$J = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) =$$

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{2x^T A^T x + b^T b}$$

The minimal value can be determined by the derivation of the cost function J with repsect to vector $\boldsymbol{x}$

$$\frac{\partial J}{\partial \boldsymbol{x}} = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^T\boldsymbol{b} = 0$$

The solution is obtaned via the so-called normal equation

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}.$$

Thus, it is

$$x = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

4 Case studies

4.1 Line fitting

The theory described above can be applied for line fitting. The standard explicit form, also called as slope-intercept form, of a line in 2D is

$$y = mx + b$$

If at least two points are given, the line can be estimated. Let us assume that there N points are given in 2D: (x_1, y_1), (x_2, y_2), ..., (x_N, y_N). Then the following linear equations can be written:

$$y_1 = mx_1 + b$$

$$y_2 = mx_2 + b$$

$$\vdots$$

$$y_N = mx_N + b$$

The equation can be written in a homogeneous linear system of equaitions

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

The solution is

$$\begin{bmatrix} m \\ b \end{bmatrix} = \left(\begin{bmatrix} x_1 & x_2 & \dots & x_N \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

After elementary modifications,

$$\begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & N \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N x_i y_i \\ \sum_{i=1}^N y_i \end{bmatrix}$$

4.2 Plane fitting

The explicit form of a plane is

$$z = ax + by + c$$

If N pieces of 3D coordinates are given, the i-th point yields a linear equation

$$z_i = ax_i + by_i + c$$

A linear system of equations can be formed from all the data points as

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}$$

The least-squares optimal solution for the plane parameters a, b, and c are given as

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i y_i & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i y_i & \sum_{i=1}^N y_i^2 & \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N y_i & N \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N x_i z_i \\ \sum_{i=1}^N y_i z_i \\ \sum_{i=1}^N z_i \end{bmatrix}$$