Mathematic for Machine Learning

Created	@Jan 18, 2021 1:39 PM
Created By	K Khanh Vương
Last Edited By	K Khanh Vương
Last Edited Time	@Jan 27, 2021 2:53 PM
Module	Introduction to Machine Learning
Status	
Type	Introduction to Machine Learning

Vector

- We have two main operations on a vectors:
  - **Addition**:

    $$\begin{bmatrix} a_i \\ a_j \end{bmatrix} + \begin{bmatrix} b_i \\ b_j \end{bmatrix} = \begin{bmatrix} a_i + b_i \\ a_j + b_j \end{bmatrix}$$

  - **Multiplied by a scalar number**:

    $$\alpha \begin{bmatrix} a_i \\ a_j \end{bmatrix} = \begin{bmatrix} \alpha.a_i \\ \alpha.a_j \end{bmatrix}$$

- If we want to combine two vectors, we can use `dot product`

  $$\begin{bmatrix} a_i \\ a_j \end{bmatrix} . \begin{bmatrix} b_i \\ b_j \end{bmatrix} = a_i.b_i + a_j.b_j$$

  - `dot product` will return a scalar that indicate:

    $$r.s = \left|r\right| \left|s\right| \cos \theta$$

  - So that, we can use $cos\theta$ to check whether two vectors are pointing in the same direction

- To get the **modulus** of a vector, dotted it with itself and take the square: $\left|a\right| = \sqrt{a.a}$

- If we are facing with changing co-ordinate, think of checking how much one vector is along to other vector:
  - **Scalar Projection**: $\frac{\text{rs}}{\left|r\right|} = \left|s\right| \cos\theta$
  - **Vector Projection**: $r \frac{\text{r.s}}{\left|r\right|\left|r\right|} = r\frac{\text{r.s}}{r.r}$

- Two different vectors can be called **Linear Independence** if we cannot find any combination that change one vector to other vector. Those **Linear Independence Vectors** can be used as the basic vectors (axis) of a co-ordinate system

Matrices

- **Matrices** are the objects that used to rotate, stretch vectors
- If we want to reverse the change, think of **Inverse Matric**. **Inverse Matric** is a matric that undo every change: $A.A^{-}1 \ = \ I$
- The **Identity Matrix** (sometimes ambiguously called a **Unit Matrix**) of size n is the n × n square matrix with ones on the main diagonal and zeros elsewhere. It is denoted by In , or simply by I if the size is immaterial or can be trivially determined by the context.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- **Determinant** is the different in space between before and after changing the space with **Matric**
- A **Matric** only have I**nverse** if the basis vectors describing the matrix are linearly dependent, then the determinant is zero, and that means I can't solve the system of simultaneous equations anymore.
- Note that we must always check for inverse. This allow us to know that this is a transformation you can undo

Multivariate Calculus

Derivate

- The nature of **Derivate**:

$$f_x' \ = \ \frac{\mathrm{d}f}{\mathrm{d}x} \ = \ \lim_{\triangle \ x \to \ 0} \ \frac{f(x + \ \Delta \ x) - f(x)}{\Delta \ x}$$

- **Derivate** has some rule:
  - **Sum Rule**:

$$\frac{\mathrm{d}}{\mathrm{d} \ x} \ (f(x) \ + \ g(x)) \ = \ \frac{\mathrm{d} \ f(x)}{\mathrm{d} \ x} \ + \ \frac{\mathrm{d} \ g(x)}{\mathrm{d} \ x}$$

  - **Power Rule**:

$$f(x) \ = \ ax^b \ \Rightarrow \ f'(x) \ = \ \frac{\mathrm{d} \ f(x)}{\mathrm{d} \ x} \ = \ abx^{b-1}$$

  - **Product Rule:**

$$A(x) \ = \ f(x)g(x) \ \Rightarrow \ A'(x) \ = \ f'(x)g(x) \ + \ f(x)g'(x)$$

  - **Chain Rule:**

$$\frac{\mathrm{d} \ h}{\mathrm{d} \ m} \ = \ \frac{\mathrm{d} \ h}{\mathrm{d} \ p} \ \cdot \ \frac{\mathrm{d} \ p}{\mathrm{d} \ m}$$

  - **Trigonometric Function:**
    - $f(x) \ = \ \sin \ (x)$
    - $f'(x) \ = \ \cos \ (x)$
    - $f^{(2)}(x) = -\sin(x)$
    - $f^{(3)}(x) \ = \ -\cos \ (x)$
    - $f^{(4)}(x) \ = \ \sin \ (x)$
  - **Exponential Function:**

$$f(x) \ = \ e^x \ \Rightarrow \ f'(x) \ = \ \frac{\mathrm{d} \ f}{\mathrm{d} \ x} \ = \ e^x$$

- **Total Derivate**:

$$\frac{\partial f_{(x,y,z)}}{\partial t} = \frac{\partial f}{\partial x} \cdot \frac{d x}{dt} + \frac{\partial f}{\partial y} \cdot \frac{d y}{d t} + \frac{\partial f}{\partial z} \cdot \frac{d z}{d t}$$

Jacobian

- **Jacobian** of a graph:
  - A matric where each entry is the partial derivative of f with respect to each one of those variables in turn.
  - Will return a vector pointing in the direction of steepest slope of this function (it doesn't mean pointing to the highest slope).
  - **Jacobian** vectors are just the gradients
  - The magnitude of **Jacobian** at a point is equal to its local steepness
  - A **Jacobian** vector that equal to

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

indicate that this point is either maximum, minimum or saddle (flat at this point and have no gradient, meaning that

  - Use **Jacobian Matrix** when we want to group many vectors to describe a co-ordiante.

$$J = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right]$$

Hessian

- The square matrix of the quadratic derivative of a function, so it represents the curvature of a multivariable function. The number of rows and columns are equal to the number of variables in the **Jacobian** vector.

$$H = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

- **Determinant**:

$$\left| H \right| = a_{11}.a_{22} - a_{12}.a_{21}$$

- **If the gradient is 0**:
  - If determinant is positive: This point is maximum or minimum.
    - If `a_11` is positive, then we got a minimum.
    - If it is negative, then we got a maximum.
  - If the gradient of this point is 0, but the determinant is negative, then we got a saddle.

Multivariate Chain Rule with Jacobian

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} \frac{\mathrm{d}\, x}{\mathrm{d}\, t} = \begin{bmatrix} \frac{\mathrm{d}\, x}{\mathrm{d}\, t_1} \\ \frac{\mathrm{d}\, x}{\mathrm{d}\, t_2} \\ \frac{\mathrm{d}\, x}{\mathrm{d}\, t_3} \end{bmatrix} \Rightarrow \frac{\mathrm{d}\, f}{\mathrm{d}\, t} = \frac{\partial f}{\partial X} \frac{\mathrm{d}\, X}{\mathrm{d}\, t}$$

Newton-Raphson

- We try a solution and evaluate it and then generate a new guess and then evaluate that again and again, which is called **Iteration**

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Gradient Descent

- **Grad:** Similar to the **Jacobian**, but be written in a column vector
- **Directional Gradient**:
    - Check how much we gone down.
    - **Grad** points up the direction of the steepest descent, perpendicular to the contour lines

$$\bigtriangledown f \ = \ \begin{bmatrix} \frac{\mathrm{d}\ f_{(a,\ b)}}{\mathrm{d}\ x} \\ \frac{\mathrm{d}\ f_{(a,\ b)}}{\mathrm{d}\ y} \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} \ \Rightarrow \ \frac{\partial\ f}{\partial\ x}c \ + \ \frac{\partial\ f}{\partial\ x}d$$