Machine Learning Mathematics - Linear Algebra

Machine Learning II

Lecture 2-b

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

- Machine learning combines statistics and computer science fields.
- Statistics, probability, estimation and confidence intervals are some of main topics in machine learning in the statistical part.
- Linear algebra is another mathematical skill which is highly necessary in machine learning.
- Optimization theory and calculus are widely used in machine learning algorithms.

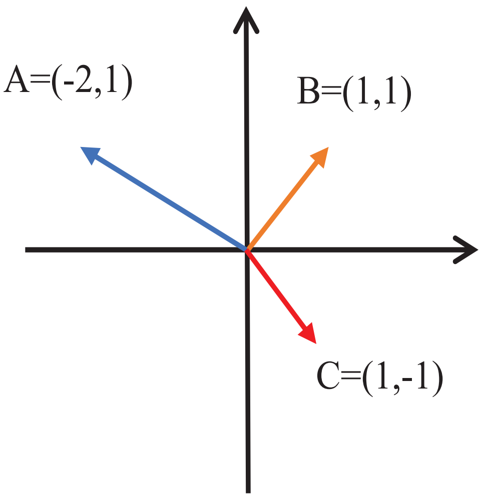
Why we need math?

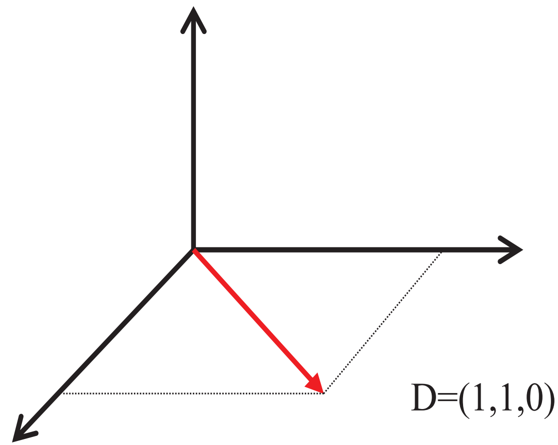
- There are so many machine learning codes out there, and they are fairly simple to run.
- You need to download the packages and library to run the machine learning algorithms.
- However, to get some useful results and meaningful performance, you need to have a good mathematical background.
- After this lecture, you will get a glimpse of what types of mathematical skills you need to practice.
- In this lecture, we go over the main topics and further materials will be provided to you in order to strengthen you mathematical skills.

- Operations on or between vectors and matrices
- Linear regression, dimensionality reduction and solution of linear systems of equations are some application of linear algebra.
- Most common form of data in machine learning is in a vector form. The 2D array where rows represent samples represent columns attributes.
- A vector is a n-tuple of values usually real numbers where n is the the dimension of the vector, n can be any positive number.
- Vector is written in a column form.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

- Vector is a point in space or a line with magnitude and direction.





- Addition of the vectors. $\mathbf{x} = \mathbf{a} + \mathbf{b}$

- Scalar multiplication. $\mathbf{x} = a\mathbf{x}$

- Dot product of vector. $a = \mathbf{x} \,.\, \mathbf{y} = \sum_i^n x_i y_i$ or $a = \mathbf{x} \,.\, \mathbf{y} = ||\mathbf{x}||||\mathbf{y}||cos(\theta)$

- Matrix is mapping. It is an $m \times n$ two dimensional array.
- A vector is the bases set of a matrix. These bases set creates the matrix.
- A vector and matrix can be transposed. Transpose is the swap of rows and columns.
- Matrix is identified by two subscript. First element in subscript shows row number and the second element shows column number.
- A_{21} shows indicates to second row and first column.

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

- $\mathbf{C} = \mathbf{A}+\mathbf{B}$, two matrix with the same dimension can be added ($c_{ij} = a_{ij} + b_{ij}$), the results has the same size.
- $\mathbf{A} = c.\mathbf{B}$, an scalar can be multiplied by each element of a matrix. ($A_{ij} = c + B_{ij}$), the results has the same size.
- Matrix multiplication.
- $\mathbf{A} \,.\, (\mathbf{B} \,.\, \mathbf{C}) = (\mathbf{A} \,.\, \mathbf{B}) \,.\, \mathbf{C}$, associative.
- $\mathbf{A} \,.\, \mathbf{B} \neq \mathbf{B}.\ \mathbf{A}$, commutative.
- $(\mathbf{A} \,.\, \mathbf{B})^T = \mathbf{B}^T.\mathbf{A}^T$, transposition.
- $\mathbf{A}\mathbf{A}^{-1} = I$, the inverse of an n-by-m matrix $\mathbf{A}$ is denoted $\mathbf{A}^{-1}$.

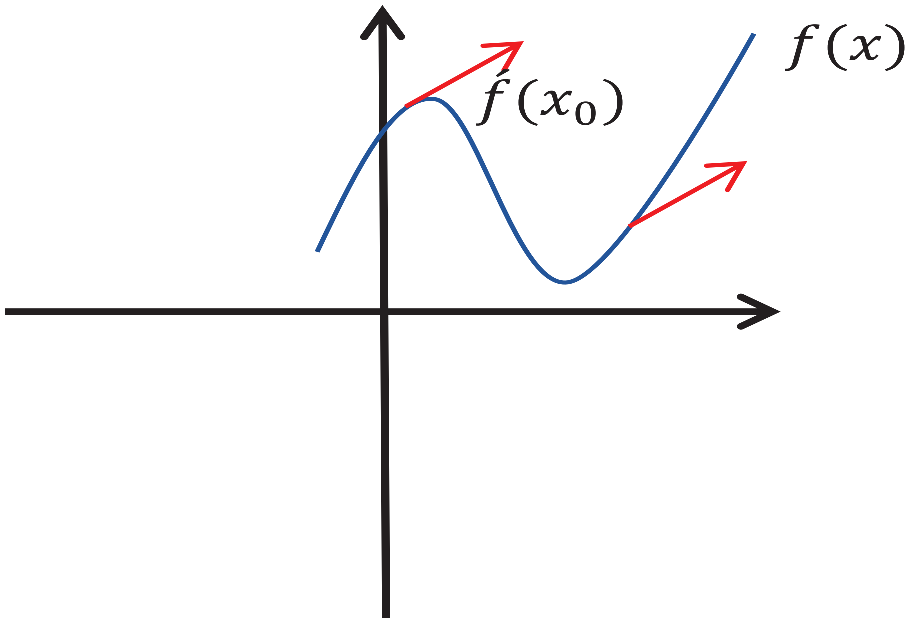
- $\mathbf{C} = \mathbf{A}+\mathbf{B}$, two matrix with the same dimension can be added ($c_{ij} = a_{ij} + b_{ij}$), the results has the same size.
- $\mathbf{A} = c.\mathbf{B}$, an scalar can be multiplied by each element of a matrix. ($A_{ij} = c + B_{ij}$), the results has the same size.
- $\mathbf{A} \,.\, (\mathbf{B} \,.\, \mathbf{C}) = (\mathbf{A} \,.\, \mathbf{B}) \,.\, \mathbf{C}$, associative, $\mathbf{A} \,.\, \mathbf{B} \neq \mathbf{B}.\ \mathbf{A}$, commutative and $(\mathbf{A} \,.\, \mathbf{B})^T = \mathbf{B}^T.\mathbf{A}^T$, transposition.
- In any chain of matrix multiplications, the column dimension of one matrix in the chain must match the row dimension of the following matrix in the chain.
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$, $(k\mathbf{A}^{-1}) = k^{-1}\mathbf{A}^{-1}$, $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$, $(\mathbf{A}\mathbf{B}^{-1}) = \mathbf{B}^{-1}\mathbf{A}^{-1}$

- Assume: **A** is (3x5), **B** is (5x5) and **C** is (3x1).
- $\mathbf{A}\,.\,\mathbf{B}\,.\,\mathbf{A}$, $\mathbf{A}\,.\,\mathbf{B}\,.\,\mathbf{A}^T$, $\mathbf{C}^T.\mathbf{A}\,.\,\mathbf{B}$, $\mathbf{C}.\,\mathbf{A}\,.\,\mathbf{B}$

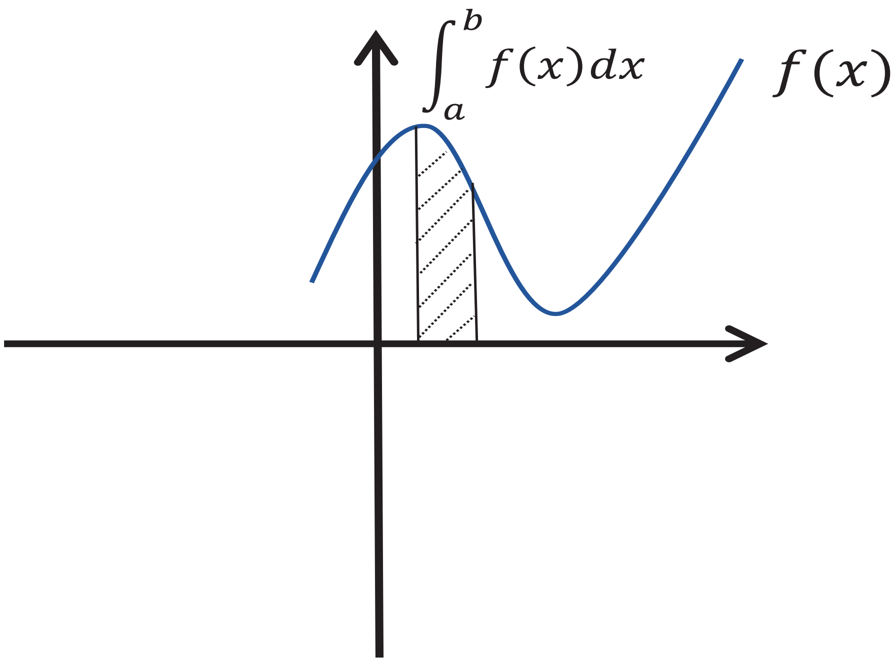
Eigenvalues and Eigenvectors

- Eigenvalues and eigenvectors play a prominent role in the study of ODE and Linear Algebra and in many applications in the physical sciences.
- Let A be an $n \times n$ matrix. The value λ is an eigenvalue of A if there exists a non-zero vector v such that $Av = \lambda v$
- In this case, vector v is called an eigenvector of A corresponding to.

- A derivative, $\frac{d}{dx}f(x)$ is defined as the slope of a function $f(x)$. This is sometimes also written as $f'(x)$.



- Integrals are an inverse operation of the derivative (plus a constant).



- Assume vector $\mathbf{x} = (x_0, x_1, ..., x_{n-1})^T$ and a function $f(x)$

$$\frac{\sigma f(x)}{\sigma x} = \begin{pmatrix} \frac{\sigma f(x)}{\sigma x_0} \\ \frac{\sigma f(x)}{\sigma x_1} \\ \vdots \\ \frac{\sigma f(x)}{\sigma x_{n-1}} \end{pmatrix}$$

- This is called the gradient of the function with respect to vector **x**, written as $\nabla f(x)$ or ∇f

Summary

- Linear algebra is very crucial in machine learning algorithms.
- This lecture just briefly introduces the main and basic concepts of math which we needed for basic machine learning algorithms.
- Check khan academy again if you need to review or refresh your mind.
- We are going to use these mathematical concepts through out the course and implement in into the computer programmes.

- Check my GitHub and answer the exercises.
  - https://github.com/amir-jafari/Machine-Learning/tree/master/Python-Math

Next Lecture

- In the next lecture, we are going to use the power of computer and our math skill to implement the first machine learning algorithm.
- Neural network components will be introduced.
- Neural network architecture will be discussed.
- Check Neural Network Design book.
- Read ahead the Chapter 1 through 4 from the book.