Maths Preliminaries

March 18, 2018

Introduction

- This review serves two purposes:
  - Recap relevant maths contents that you may have learned a
  -

    Machine Learning.
- Contents
  - Linear Algebra

Note

- You've probability learned Linear Algebra from matrix/system of linear equations, etc. We will review key concepts in LA

  - Here we emphasize more on intuitions; We deliberately skip many concepts and present some conte

- It is a great exercise for you to view related math models/operations in this perspective e!

A Common Trick in Maths I

Question

Calculate 2^{10}, 2^{-1}, $2^{\ln 5}$ and 2^{4-3i}?

- Properties:
  - $f_a(n) = f_a(n-1) * a$, for $n \geq 1$;
  - $f(u) * f(v) = f(u + v)$.
  - $f(x) = y \Leftrightarrow \ln(y) = x \ln(a) \Leftrightarrow f(x) = \exp\{x \ln a\}$.
  - $e^{ix} = cos(x) + i \cdot sin(x)$.
- The trick:
- Same in Linear algebra

Objects and Their Representations

Goal

- We need to study the objects
- On one side:
  - A good representation helps (a lot)!
-
  -

Basic Concepts I

Algebra

- a set of objects
-

  -
  -
- constraints:
  - Closed for both operations
  - Some nice properties of these operations

    - Communicative: $\mathbf{a} + \mathbf{b} = \mathbf{b}$
    - Associative: $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$.
    - Distributive: $\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

Basic Concepts II

Think: *What about substraction and division?*

Tips

Alwa ra on

Poly

Why these constraints are natural and useful?

Basic Concepts III

Representation matters?

Consider even geometric vectors: $\mathbf{c} = \mathbf{a} + \mathbf{b}$

What if we represent vectors by a column of their coordinates?

Wha

Notes

- Informally, the objects we are concerned w are (column) vectors.
- The set of all n-dimensional real vectors is called $\mathbb{R}^n$.

(Column) Vector

Vector

-

  -

Operations

- Addition: $\mathbf{v}_1 + \mathbf{v}_2 =$
- (Scalar) Multiplication: $\lambda \mathbf{v}_1 =$

Linearity I

Linear Combination: Generalization of Univariate Linear Functions

- Let $\lambda_i \in \mathbb{R}$, given a set of k vectors $\mathbf{v}_i$ ($i \in [k]$), a linear combination of them is
  $\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \ldots + \lambda_k \mathbf{v}_k = \sum_{i \in [k]} \lambda_i \mathbf{v}_i$
  -

$$\mathbf{V} = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v \\ | & | & & | \end{bmatrix}$$

2

- Span: All linear combination of a set of vector of them.
- Basis: The minimal set of vectors whose span is exactly the whole $\mathbb{R}^n$.

Linearity II

- Benefit: every vector has a unique decomposition into basis.

Think: *Why uniqueness is desirable?*

Exa

-
- Span of $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}2\\3\end{bmatrix}$ is $\mathbb{R}^2$. But one o *t*.

Think: *Wh*

  - Decompose $\begin{bmatrix}4\\6\end{bmatrix}$

Linearity III

Exercises

- What are the (natural) basis of all (univariate) Polynomials of degrees up to d?
- Decompose $3x^2 + 4x - 8$ into *the* linear combination of 2,
  2
  ):
- The "same" polynomial is mapped to two di
  under two different bases. **Think**

Matrix I

Linear Transformation

- is a "nice" linear function that maps a vector in $\mathbb{R}^n$ to another vector in $\mathbb{R}^m$.

3

- The general form:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \implies \begin{aligned} y_2 &= M_{21}x_1 + M_{22}x_2 \\ y_3 &= M_{31}x_1 + M_{32}x_2 \end{aligned}$$

Matrix II

Nonexample

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \xrightarrow{f} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \implies \begin{aligned} y_1 &= \alpha x_1^2 + \beta x_2 \\ y_2 &= \gamma x_1^2 + \theta x_1 + \tau x_2 \end{aligned}$$

x_2

Why Only Linear Transformation?

- Simple and nice properties:
  - $(f_1 + f_2)(x) = f_1(x) + f_2(x)$
  - $(\lambda f)(x) = \lambda \cdot f(x)$
  - What about $f(g(x))$?
- Useful

Matrix I

Definition

- A $m \times n$ matrix *corresponds to* a *linear transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$
  -
  -
  - *Transformation* or *Mapping* emphasizes more on the mapping between two sets, rather than the detaile mapping; the latter is more or less the understanding of a *function*. The *morphism* in category theory.

Semantic Interpretation

Matrix II

- Linear combination of columns of $\mathbf{M}$:

$$\begin{bmatrix} | & | & & | \\ M_1 & M_2 & \dots & M_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} | \\ x \\ | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ M_1 & M_2 & \dots & M_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

- Example:

$$\begin{bmatrix} 1 & 2 \\ -4 & 9 \\ 25 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 10 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -4 \\ 25 \end{bmatrix} + 10 \begin{bmatrix} 2 \\ 9 \\ 1 \end{bmatrix} = \begin{bmatrix} 21 \\ 86 \\ 35 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ -4 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 10 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ -4 \end{bmatrix} + 10 \begin{bmatrix} 2 \\ 9 \end{bmatrix} = \begin{bmatrix} 21 \\ 86 \end{bmatrix}$$

Matrix III

Think: *What does* **M** *do for the last example?*

-
-

$$\begin{matrix} 1 & 2 \\ -4 & 9 \end{matrix} \qquad 21 \quad 42$$

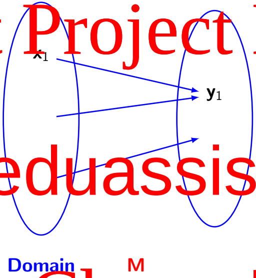
$$\begin{bmatrix} 25 & \\ \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 10 & 20 \end{bmatrix}$$

System of Linear Equations I

$y_1 = M_{11}x_1 + M_{12}x_2$ $\quad y_1 \quad M_{11} \quad M_{12}$ $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

- Interpretation: find a vector in $\mathbb{R}^2$ ... **M**) is exactly the given vector **y**
- How to solve it?

System of Linear Equations II



The above transformation is *injec*

A Matrix Also Specifies a (Generalized) Coordinate System I

Yet a

-
- The vector **y** wrt standard coordinate system, **I**, is the same as **x** wrt the coordinate system defined by M. Think: why columns of M?

A Matrix Also Specifies a (Generalized) Coordinate System II

Example for polynomials

I: f ... fo ... for x^2 0 0 1 ... fo ... -4

Let $\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \implies \mathbf{M}\mathbf{x} = \mathbf{I} \begin{bmatrix} -7 \\ 13 \\ 6 \end{bmatrix}$

Exercise I

- What if $\mathbf{y}$ is given in the above example?
-

0 0 2 0 0 1

- Think about representing polynomials u
  $(x-1)^2$, x^2-1, x^2+1.

Inner Product

THE binary operator – some kind of "similarity"

- Type signature: vector $\times$ vector $\to$ scalar: $\langle \mathbf{x}, \mathbf{y} \rangle$.
  - In $\mathbb{R}^n$, usually called *dot product*: $\mathbf{x} \cdot \mathbf{y} \stackrel{\text{def}}{=} \mathbf{x}^\top \mathbf{y} = \sum_i x_i y_i$.
  - For certain functions, $\langle f, g \rangle = \int_a^b f(t)g(t)\,\mathrm{d}t$. leads to the

-
  - linearity in the first argument: $\langle a\mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = a \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
  - positive definitiveness: $\langle \mathbf{x}, \mathbf{x} \rangle \geq$
- Generalizes many geometric concepts to v (orthogonal), projection, norm
  - $\langle \sin nt, \sin mt \rangle = 0$ within $[-\pi, \pi]$ $(m \neq n)$ $\Rightarrow$ they are orthogonal to each other.
- $\mathbf{C} = \mathbf{A}^\top \mathbf{B}$: $C_{ij} = \langle A_i, B_j \rangle$
  - Special case: $\mathbf{A}^\top \mathbf{A}$.

Eigenvalues/vectors and Eigen Decomposition

"Eigen" means "characteristic of" (German)

- A (right) eigenvector of a square matrix $\mathbf{A}$ is $\mathbf{u}$ such that $\mathbf{Au} = \lambda\mathbf{u}$.
-
-
- We can use all eigenvectors of $\mathbf{A}$ to construct a matrix $\mathbf{U}$ (as columns). Then $\mathbf{AU} = \mathbf{U}\Lambda$, or equiv
  This is the Eigen Decomposition.

  - We can interpret $\mathbf{U}$ as a transform coordinate systems. Note that vectors in $\mathbf{U}$ are not necessarily orthogonal.
  - Λ as the scaling on each of the directions in the "new" coordinate system.

Similar Matrices

Similar Matrix

- Let **A** and **B** be two $n \quad n$ matrix. **A** is similar to **B** (denoted such

-
  - Think of **P** as a *change of basis*
  - Relationship with the Eigen decompos

- Similar matrices have the same value wrt ma (e.g., rank, trace, eigenvalues, determin

SVD

Singular Vector Decomposition

- Let $\mathbf{M}$ be $n \times d$ ($n \geq d$).
- Reduced SVD: $\mathbf{M} = \hat{\mathbf{U}}\hat{\Sigma}\mathbf{V}^\top$ exists for any $\mathbf{M}$, such that
  - led

  - left
  - $\hat{\mathbf{V}}$ consists of a set of basis vectors in $\mathbb{R}^d$) ($d \times d$: reduced space)

- Full SVD: $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$.
  - Add the remaining $(n - d)$ basis vectors to $\hat{\mathbf{U}}$ (thus becomes $n \times n$).
  - Add the $n - d$ rows of $\mathbf{0}$ to $\hat{\Lambda}$ (thus becomes $n \times d$).

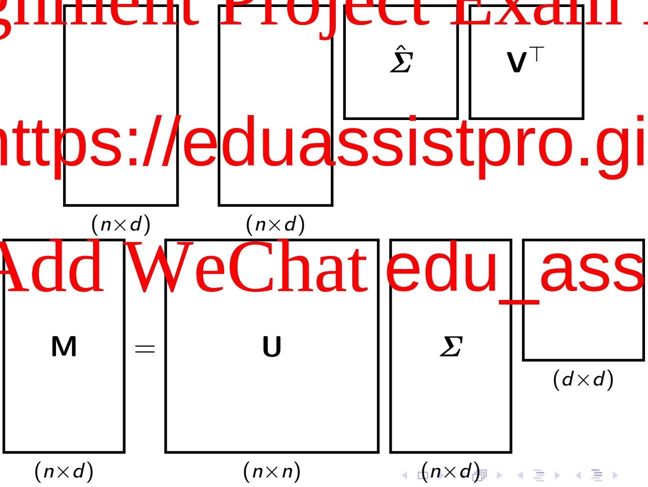
Geometric Illustration of SVD

Geometric Meaning

- $\mathbf{M}v_i = \sigma_i u_i$

Graphical Illustration of SVD I

Figure: Reduced SVD vs Full SVD



Graphical Illustration of SVD II

Mea

-
-

SVD Applications I

Relationship between Singular Values and Eigenvalues

- What are the eigenvalues of $\mathbf{M}^\top \mathbf{M}$?
- Hint: $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$ and $\mathbf{U}$ and $\mathbf{V}$ are unitary (i.e., rotations)

- Related to *PCA (Principle Component Analysis)*

References and Further Reading I

- Gaussian Quadrature: https://www.youtube.com/watch?v=k-yUdqRXijo
- pdf
- Scipy LA tutorial. https://docs.scipy.org/doc/scipy/reference/tutorial/linalg.htm
- We Recommend a Singular Value Decom

  http://www.ams.org/samplings/ vd