Maths Preliminaries

February 18, 202

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: $a + b = b + a$.
        - Associative: $(a + b) + c = a + (b + c)$.
        - Distributive: $\lambda(a + b) = \lambda a + \lambda b$.

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: c = a + 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: $v_1 + v_2 =$
- (Scalar) Multiplication: $\lambda v_1 =$

Linearity I

Linear Combination: Generalization of Univariate Linear Functions

- Let $\lambda_i \in \mathbb{R}$, given a set of k vectors v_i ($i \in [k]$), a linear combination of them is
  $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_k v_k = \;{}_{i\ [k]} \lambda_i v_i$
-

$$V = \begin{vmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_k \\ | & | & & | \end{vmatrix} \qquad {}^{2}$$

- 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: *Who*

  - 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 $3x^2 + 4x - 7$ 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 M:

$$\begin{bmatrix} | & | & & | \\ M_{\bullet 1} & M_{\bullet 2} & \cdots & M_{\bullet n} \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\mathrm{y} = x_1 \mathrm{M}_{\bullet 1} + \ldots + x_n \mathrm{M}_{\bullet n}$$

- Example:

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

Matrix III

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

$$\begin{bmatrix} 1 & 2 \\ -4 & 9 \\ 25 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 10 & 20 \end{bmatrix}$$

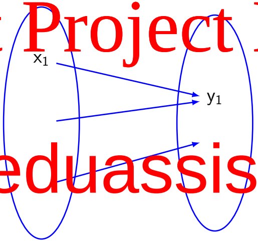
System of Linear Equations I

$$y_1 = M_{11}x_1 + M_{12}x_2$$

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

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

System of Linear Equations II



The above transformation is *injec*

A Matrix Also Specifies a (Generalized) Coordinate System I

Yet another interpretation

- $y = Mx \implies Iy = Mx$
- The vector y wrt standard coordinate system, I, is the same as

Exa

$$
\mathsf{I}: \begin{array}{r} \text{for } 1 \\ \text{for } x \\ \text{for } x^2 \end{array} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies \mathsf{M}: \quad \begin{array}{r} \\ \\ \text{for } 2x^2+5x-4 \end{array} \begin{bmatrix} & & \\ & & \\ 0 & 0 & 2 \end{bmatrix}
$$

$$
\text{Let } \mathsf{x} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \implies \mathsf{Mx} = \mathsf{I} \begin{bmatrix} -7 \\ 13 \\ 6 \end{bmatrix}
$$

Exercise I

- What if 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 $\rightarrow$ scalar: $\langle \mathrm{x}, \mathrm{y} \rangle$.

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

-
  -
  - linearity in the first argument: $\langle a\mathrm{x} + \mathrm{y}, \mathrm{z} \;= a \; \mathrm{x}, \mathrm{z} \; + \; \mathrm{y}, \mathrm{z}$
  - positive definitiveness: $\langle \mathrm{x}, \mathrm{x} \rangle \geq$

- 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.
- $\mathrm{C} = \mathrm{A}^\top \mathrm{B}$: $C_{ij} = \langle A_i, B_j \rangle$
  - Special case: $\mathrm{A}^\top \mathrm{A}$.

Eigenvalues/vectors and Eigen Decomposition

"Eigen" means "characteristic of" (German)

- A (right) eigenvector of a square matrix A is u such that $Au = \lambda u$.
-
-

- We can use all eigenvectors of A to construct a columns). Then $AU = U\Lambda$, or equiv is is the Eigen Decomposition.

  - We can interpret U as a transformation b systems. Note that vectors in 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 n matrix. A is similar to B (denoted
h

-
  - 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 M be $n \times d$ ($n \geq d$).
- Reduced SVD: $\mathrm{M} = \hat{\mathrm{U}}\hat{\boldsymbol{\Sigma}}\mathrm{V}^\top$ exists for any M, such that
  - led
  - *left*
  - $\hat{\mathrm{V}}$ consists of a set of basis vectors v in $\mathbb{R}^d$) ($d \times d$: reduced space)

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

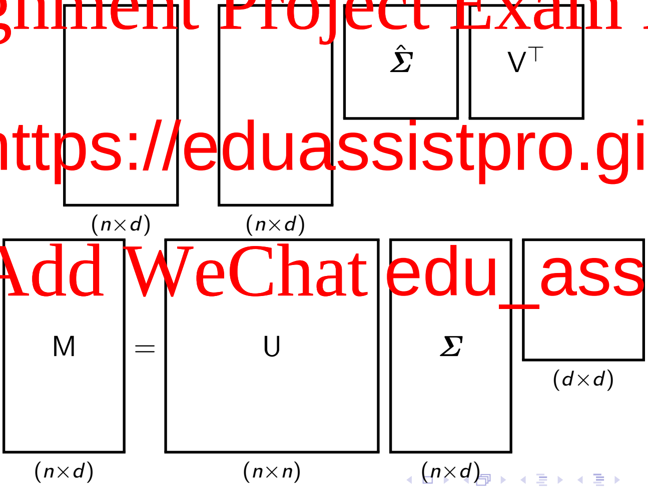
Geometric Illustration of SVD

Geometric Meaning

- $Mv_i = \sigma_i u_i$

Graphical Illustration of SVD I

Figure: Reduced SVD vs Full SVD



Graphical Illustration of SVD II

Mea

-
- Rows of V are the basis of $\mathbb{R}$

SVD Applications I

Relationship between Singular Values and Eigenvalues

- What are the eigenvalues of $M^\top M$?
- Hint: $M = U\Sigma V^\top$ and U and 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