AIAA2205

Section 3 Data, Model, Task

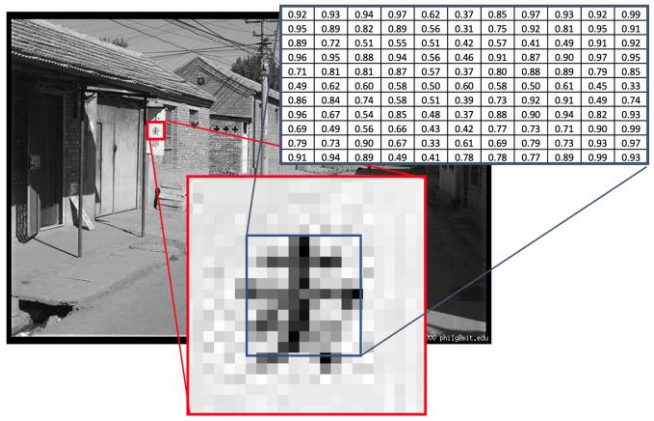
第三节 数据、模型、任务

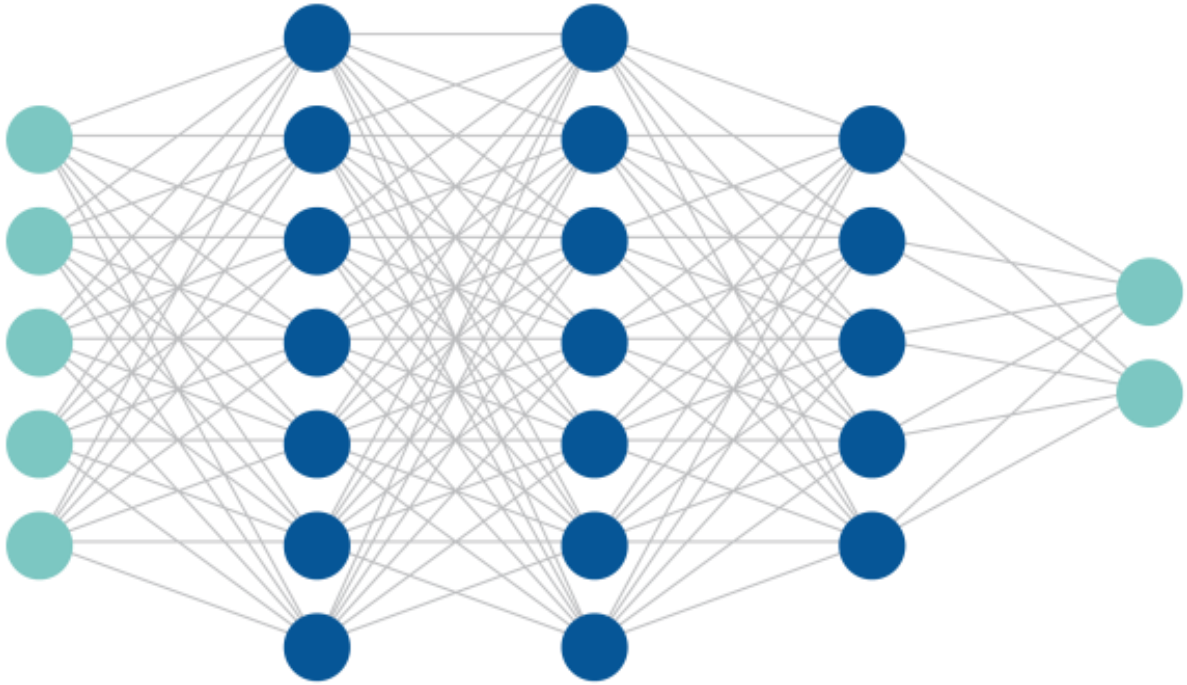
Part (2) The Basic Mathematics
第二部分 基础数学知识

Yutao Yue 岳玉涛

(Declaration: Courtesy to Prof. Yingcong Chen, this part is based on course materials from Prof. Chen)

Linear Algebra is everywhere in AI





Data: matrix

Processing: $y = \max(0, W^T x)$

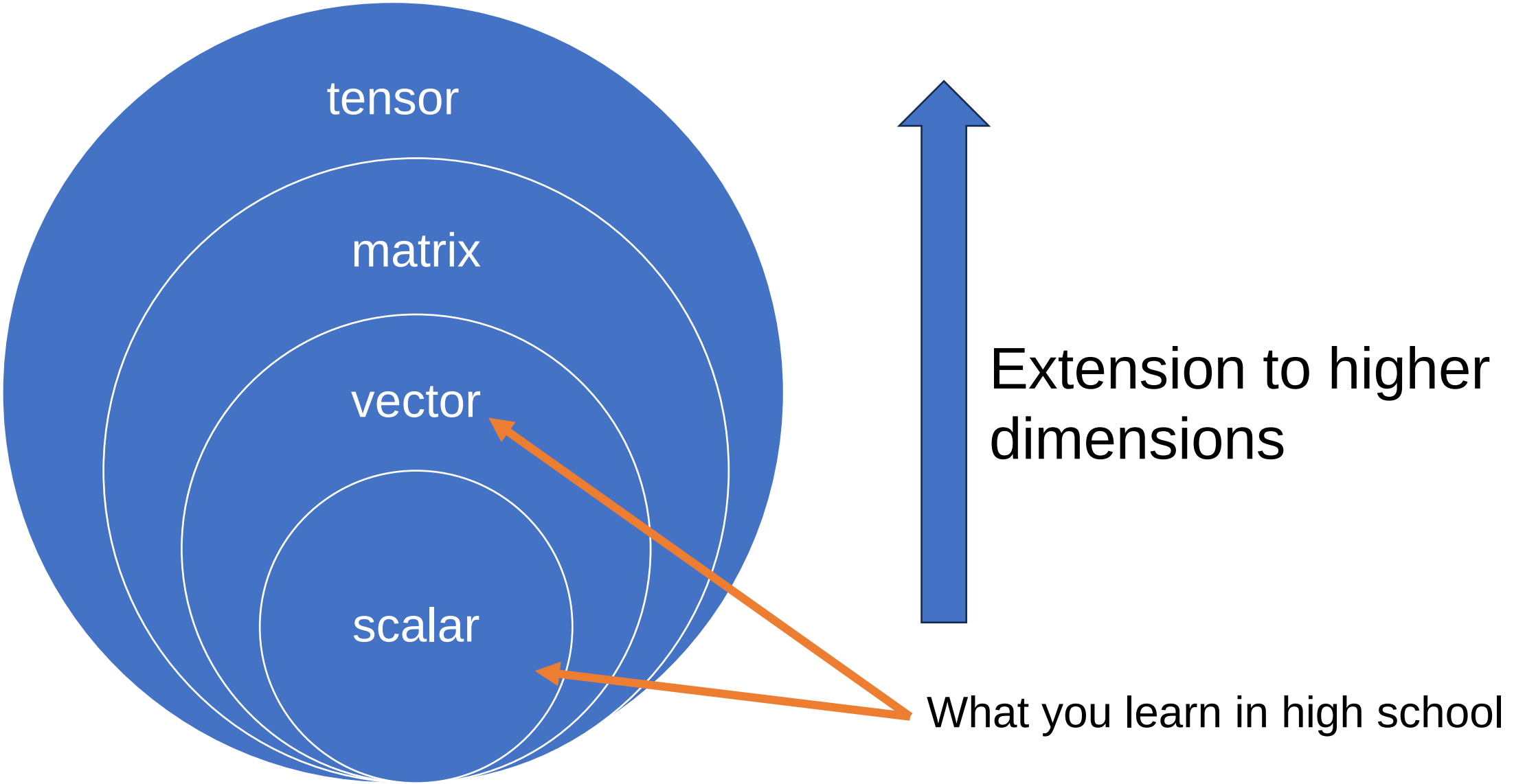
Results Probability

What will this class talk about?

- Very basic concept on linear algebra.

- The simplest AI algorithm: Nearest Neighbor Classification

Linear algebra



Scalars

Basic operations

$$c = a + b$$

$$c = a \cdot b$$

$$c = \sin a$$

Vectors

Collection of scalars

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix}$$

Transpose

$$\boldsymbol{a}^T = [a_1, a_2, \ldots, a_n]$$

Vector norm 范数 (length)

$$\left|\left|\boldsymbol{a}\right|\right|_{\mathbf{1}} = |\boldsymbol{a_1}| + |\boldsymbol{a_2}| + \cdots + |\boldsymbol{a_n}|$$

$$\left|\left|\boldsymbol{a}\right|\right|_{\mathbf{2}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

$$\left|\left|\boldsymbol{a}\right|\right|_{\boldsymbol{p}} = (|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{1/p}$$

Visual understanding of vector norms

Q1: what is L_0 norm?

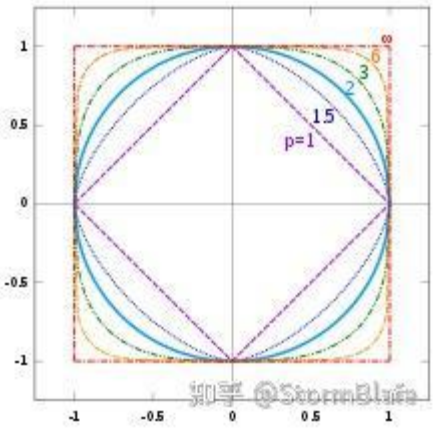
Counting the number of non-zero elements

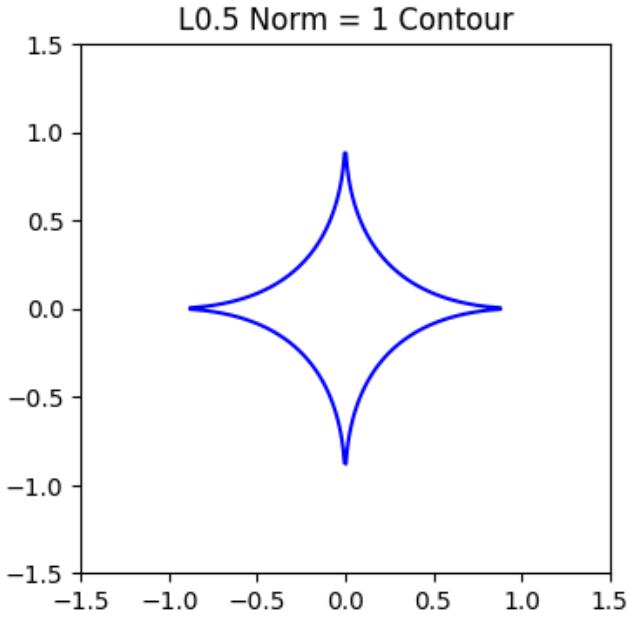
Q2: what is L_∞ norm?

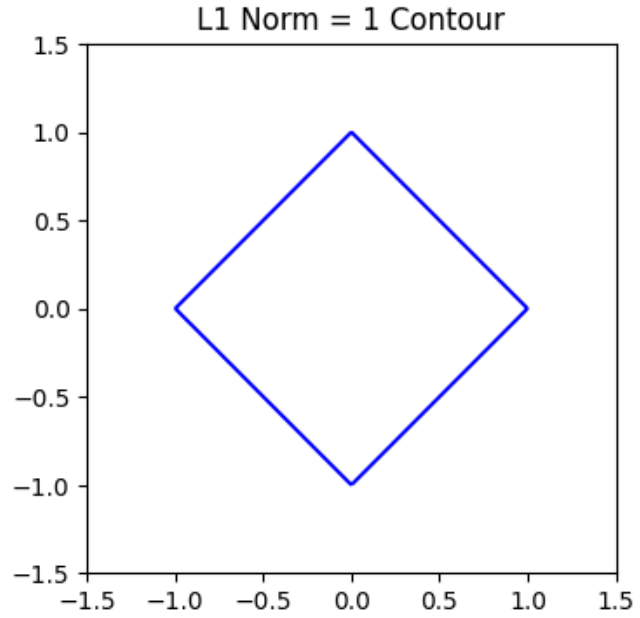
The absolute value of the largest component

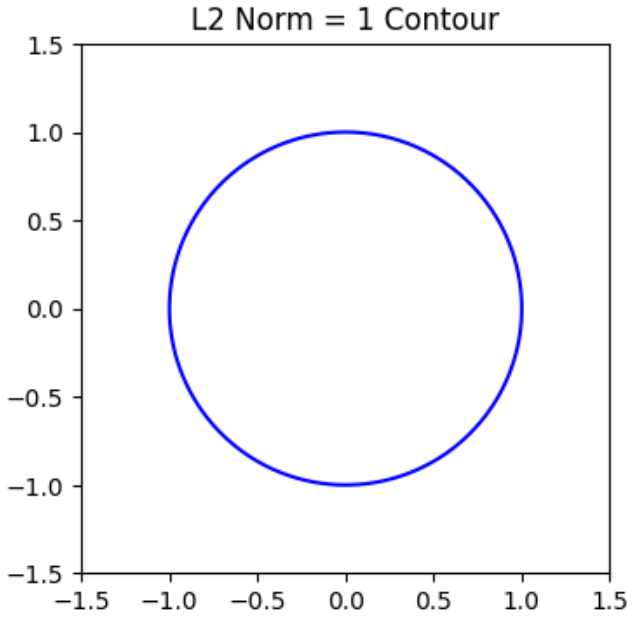
$$\left|\left|\boldsymbol{a}\right|\right|_{\boldsymbol{p}} = (|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{1/p}$$

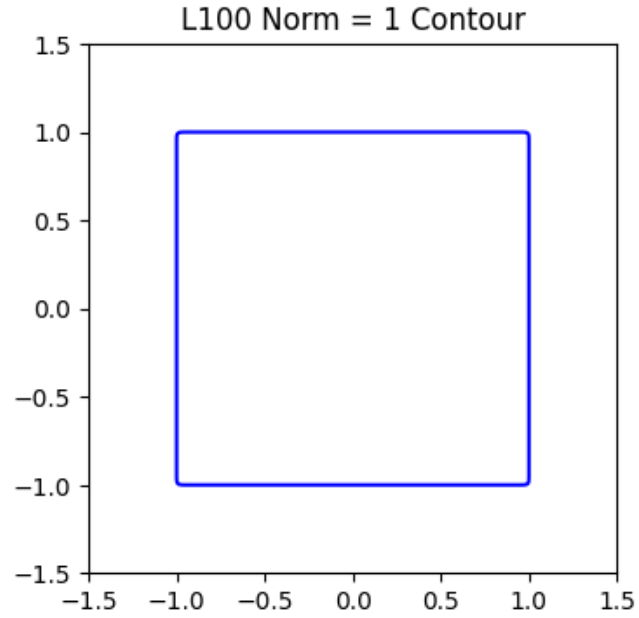
Visual understanding of vector norms









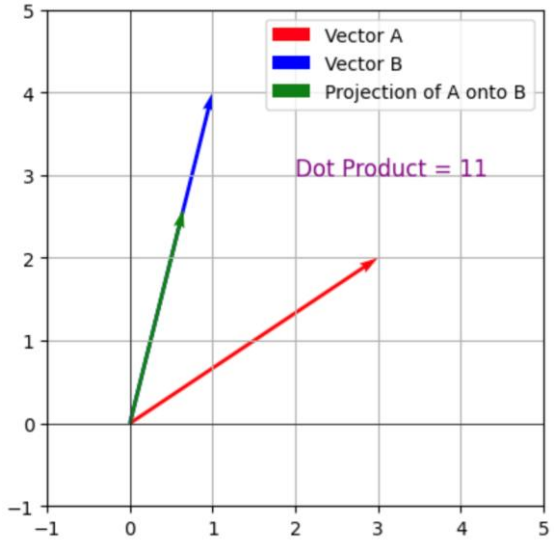


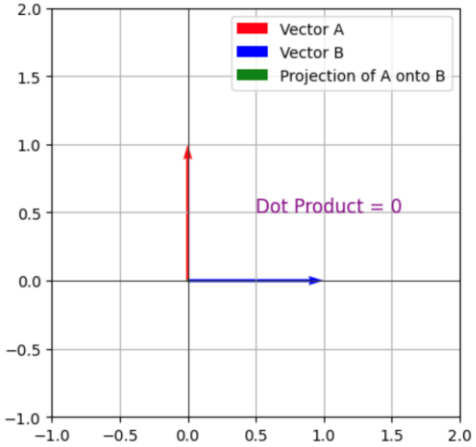
$$\left|\left|\boldsymbol{a}\right|\right|_{\boldsymbol{p}} = (|a_1|^p + |a_2|^p + \cdots + |a_n|^p)^{1/p}$$

Vectors

Dot product: $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = [a_1, a_2, \ldots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i$

$$\boldsymbol{a} \cdot \boldsymbol{b} = \left|\left|\boldsymbol{a}\right|\right|_2 \left|\left|\boldsymbol{b}\right|\right|_2 \cos\theta$$





Orthogonal

Vectors

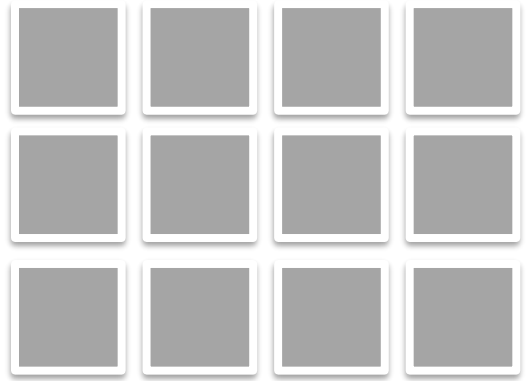
Measuring the similarity between two vectors

$$\left|\left|\boldsymbol{a}-\boldsymbol{b}\right|\right|_1 = |a_1 - b_1| + |a_2 - b_2| + \cdots + |a_n - b_n|$$

$$\left|\left|\boldsymbol{a}-\boldsymbol{b}\right|\right|_2 = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

Cosine similarity $\cos\theta = \dfrac{\boldsymbol{a}\cdot\boldsymbol{b}}{\left|\left|\boldsymbol{a}\right|\right|_2 \left|\left|\boldsymbol{b}\right|\right|_2}$

Matrices



Collection of vectors

- **Simple operations**

$$C = A + B \qquad \text{where } C_{ij} = A_{ij} + B_{ij}$$

$$C = \alpha \cdot B \qquad \text{where } C_{ij} = \alpha B_{ij}$$

$$C = \sin A \qquad \text{where } C_{ij} = \sin A_{ij}$$

Vector is high-dimensional data
Matrix is the mathematical tool to process/manipulate those high-dimensional data

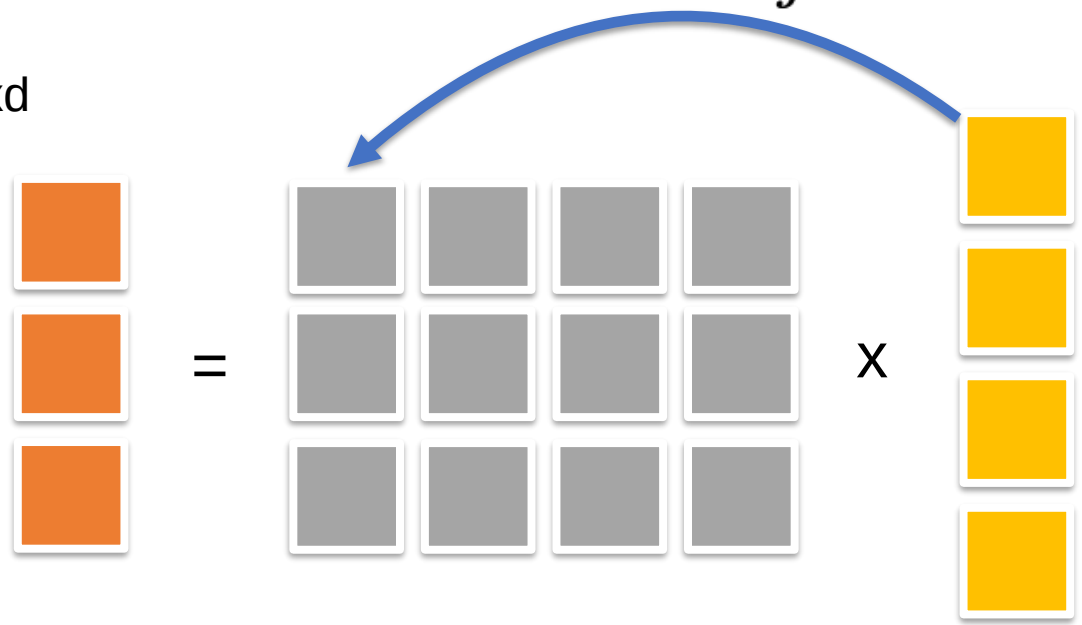
Matrices

- **Multiplications (matrix vector)**

$$c = Ab \text{ where } c_i = \sum_j A_{ij} b_j$$

Shape: Axb matmul b x c→ A x C

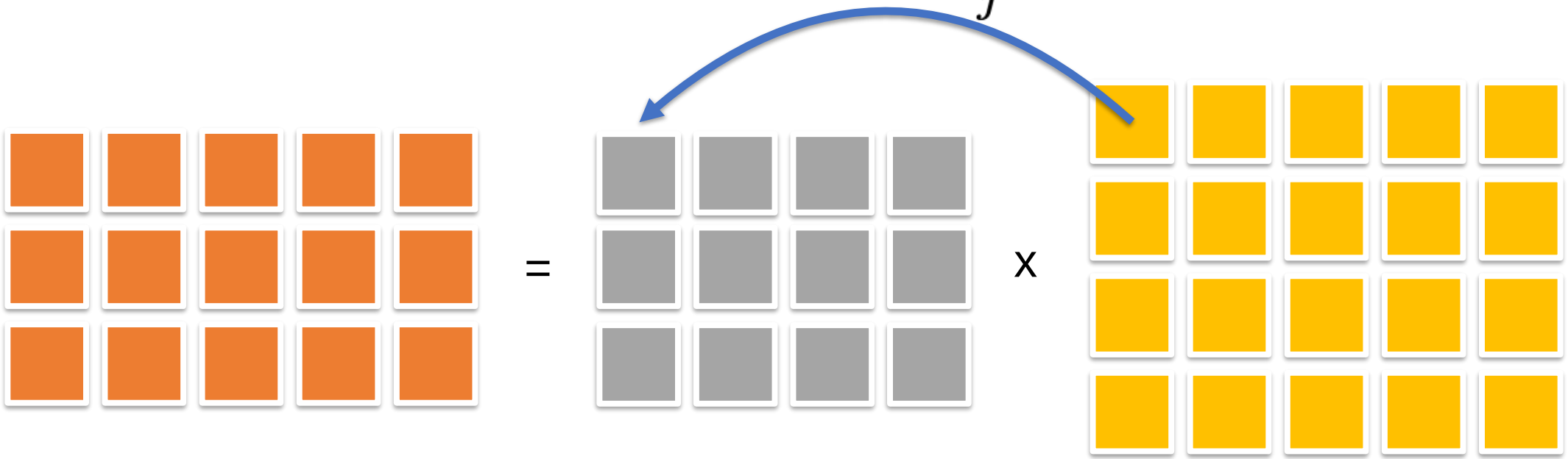
Axbxc matmul cxd -> Axbxd



Matrices

- **Multiplications (matrix matrix)**

$$C = AB \text{ where } C_{ik} = \sum_j A_{ij} B_{jk}$$



Matrices

Matrix norm (simplified version)

Frobenius范数：衡量矩阵作为线性变换的"能量强度"

$$||A||_{Frob} = \left[\sum_{ij} A_{ij}^2\right]^{\frac{1}{2}}$$

$$||A + B||_{Frob} \leq ||A||_{Frob} + ||B||_{Frob}$$

$$||aA||_{Frob} = |a| \cdot ||A||_{Frob}$$

线性代数

居余马

☆ 0 / 1.0 1 comment 平装

this book have watermark of ardvertisement.

Content Type:	书籍	年:	2006
出版社:	清华大学出版社	文件:	PDF, 17.84 MB
IPFS:	CID , CID Blake2b		

线性代数学习指南

居余马，林翠琴

☆ 0 / 0 0 comments 平装

种类:	Mathematics - Mathematical Foundations	Content Type:	书籍
		年:	2003
出版社:	清华大学出版社	语言:	Chinese
页:	345	ISBN 10:	7302065071
ISBN 13:	9787302065074	文件:	PDF, 3.62 MB
IPFS:	CID , CID Blake2b		


大学公共数学系列
高等数学
学习指南
（第二版）
（上册）
唐少锋 程晓风
王孝礼 黄正华 编著
武汉大学出版社

概率论与数理统计（第四版）

盛骤

5.0 / 5.0　　5 comments　　平装

种类:	Mathematics - Probability	Content Type:	书籍
年:	2008	出版:	4
出版社:	北京：高等教育出版社 /	语言:	Chinese
ISBN 10:	7040238969	ISBN 13:	9787040238969
文件:	PDF, 32.78 MB	IPFS:	CID , CID Blake2b

6.050J | Spring 2008 | Undergraduate

Information And Entropy

Course Description

This course explores the ultimate limits to communication and computation, with an emphasis on the physical nature of information and information processing. Topics include: information and computation, digital signals, codes and compression, applications such as biological representations of information, logic …Show more

Course Info

INSTRUCTORS

Prof. Paul Penfield
Prof. Seth Lloyd

DEPARTMENTS

Electrical Engineering and Computer Science
Mechanical Engineering

TOPICS

- Engineering
  - Computer Science
    - Theory of Computation
  - Electrical Engineering
    - Signal Processing
- Mathematics

LEARNING RESOURCE TYPES

Problem Sets with Solutions | Online Textbook | Programming Assignments

stanford.edu/class/ee364a/index.html

EE364a: Convex Optimization I

Professor Stephen Boyd, Stanford University

EE364a is the same as CME364a.

This webpage contains basic course information; up to date and detailed information is on Ed.

Linear Algebra in Pytorch

Setting up your environment

- https://colab.research.google.com
  - Activate the GPU supported runtime
- Pytorch
  - Popular
  - Similar grammar with numpy
  - Install pytorch in colab

    ! pip3 install torch torchvision
- Jupyter notebook
  - Interactive
  - Tip: Don’t write heavy code in jupyter notebook.

数值计算库

工具	类型	特点	应用场景
NumPy	基础库	多维数组基础操作	通用科学计算
CuPy	GPU加速版	兼容NumPy API，支持CUDA	大规模数据加速处理
JAX	函数转换	自动微分 + XLA编译优化	高性能数值计算

深度学习框架

框架	特性	优势领域
PyTorch	动态图优先，Pythonic	研究/快速原型开发
TensorFlow	静态图优化，生产部署友好	工业级模型部署
Keras	高层API，易用性强	快速模型搭建
MXNet	多语言支持，内存优化	分布式训练

交互式环境

环境	特点	适用场景
Jupyter Lab	增强版Notebook，模块化界面	数据科学全流程开发
Google Colab	云端免费GPU资源	深度学习原型开发
VS Code	本地IDE + Jupyter扩展	工程化开发

领域专用工具

工具	领域	特点
Pandas	数据分析	表格数据处理与清洗
SciPy	科学计算	高级数学函数和算法库
OpenCV	计算机视觉	图像/视频处理基础库
Dask	并行计算	分布式NumPy/Pandas替代方案

典型技术栈组合：

1. **研究导向**：Jupyter + PyTorch + NumPy + Matplotlib
2. **生产导向**：VS Code + TensorFlow + TFX + Docker
3. **数据分析**：Jupyter Lab + Pandas + Seaborn + Scikit-learn
4. **高性能计算**：JAX + Dask + CuPy + MPI

N-dimensional Array Examples

- N-dimensional array, short for ndarray (as in numpy), is the main data structure for machine learning and neural networks

0-d (scalar)

1.0

A class label

1-d (vector)

[1.0, 2.7, 3.4]

A feature vector

2-d (matrix)

[[1.0, 2.7, 3.4]
[5.0, 0.2, 4.6]
[4.3, 8.5, 0.2]]

A example-by-feature matrix

ND Array Examples, cont

3-d

```
[[[0.1, 2.7, 3.4]
  [5.0, 0.2, 4.6]
  [4.3, 8.5, 0.2]]
 [[3.2, 5.7, 3.4]
  [5.4, 6.2, 3.2]
  [4.1, 3.5, 6.2]]]
```

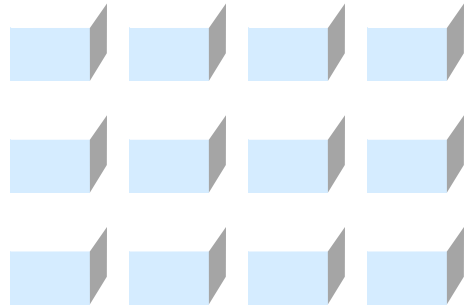
A RGB image (width x height x channels)

4-d

```
[[[[. . .
   . . .
   . . .]]]
```

A batch of RGB images (batch-size x width x height x channels)

5-d



```
[[[[. . .
   . . .
   . . .]]]
```

A batch of videos (batch-size x time x width x height x channels)

Matrix/vector computation vs Loops

```
[27] a = torch.ones(1000)
     start = time.time()
     print('L1 norm of a is {}'.format(a.norm(1)))
     end = time.time()
     print('elapsed time is {}'.format(end - start))

     L1 norm of a is 1000.0
     elapsed time is 0.0008735656738281 25
```

vector computation

```
[28] start = time.time()
     l1 = 0
     for i in range(a.shape[-1]):
       l1 += a[i].abs()
     print('L1 norm of a is {}'.format(l1))
     end = time.time()
     print('elapsed time is {}'.format(end - start))

     L1 norm of a is 1000.0
     elapsed time is 0.01016688346862793
```

loops

In pytorch/tensorflow/numpy, matrix/vector computation is much faster than loops.

Thank you!

Email电邮：yutaoyue@hkust-gz.edu.cn

HKUST-GZ香港科技大学（广州）
（Wechat Subscription公众号）

Yutao Yue岳玉涛
（Personal Wechat个人微信）

Pillars.Creation 创世之柱（Left左: C8HD, Right右: Hubble.Space.Telescope哈勃） | Target: d~5700Ly/D~5Ly(pillar)/VM=+6.4 | Telescope: D=203mm/F=1439mm/T~336min | Time.Loc.: 20230528/CN-SZ-KJY@Bortle9 | Photo by: 馒头星系MattO