Welcome to Fall 2025!! :)

A little about me
- grew up in Colorado (dry + mountains!)
- attended Middlebury College
- grad school at NYU
- research in explainable AI

Here to connect with students
- lunch!
- math game nights!
- class meal!

Why are you here?

#1 Learning
- Hard, requires struggling
- Practice is key (HW + class)

#2 Support
- Notes
- Participation form
- Feedback

8/26/25

CSCI 145 Data Mining
(really Machine Learning)

www.rtealwitter.com/datamining2025

Explain why:

Prereqs are Calculus, linear algebra, probability, python.
Foundation for fun math!

Resources Typed notes!
Very helpful to review before, and after.
These slides made the night before

Devices Very distracting, please do not use!

Print these slides if you want help taking notes!

Office Hours when2meet

Grades

P = # points

A if $P \geq 93$
A- if $93 > P \geq 90$
B+ if $90 > P \geq 87$
⋮

(10) Participation

Check in! and feedback!

(10) Problem Sets

main chance to learn
no correctness grading
(disincentivize LLM)

(20) Quizzes

regular check of understanding

(40) Exams: 10/21, 11/25

check understanding

(20) Project

Apply what you learned!

( ) Extra Credit

Improve note resource
↳ first person only
↳ week of or later

Late Policy

24 hr notice to avoid stress of last minute communication

Communication

Discord!

(Please be patient with me)

LLMs

← absolutely *insane*, we'll learn about them!

You may use for *verifiable* info
e.g., documentation, lines of code

Please do not use in lieu of learning

Accommodations

Contact me!

Linear Algebra Review

What <u>is</u> machine learning? Maths

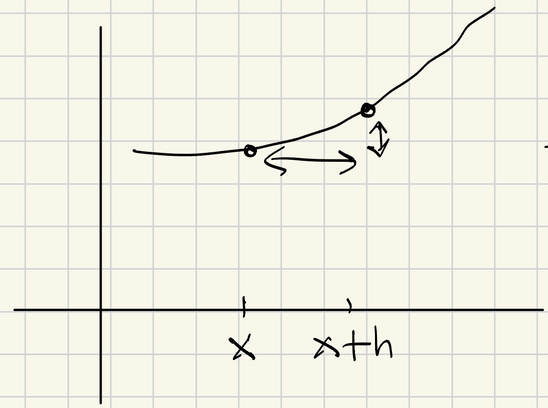
$f: \mathbb{R} \to \mathbb{R}$

$f(x)$	$f'(x)$
x^2	$2x$
x^a	ax^{a-1}
$ax+b$	a
$\ln(x)$	$\frac{1}{x}$
e^x	e^x

"partial"

$$f'(x) = \frac{\partial}{\partial x} f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

d reserved for dimension



$\frac{\text{rise}}{\text{run}}$

Chain Rule:

$g: \mathbb{R} \to \mathbb{R}$

$$\frac{\partial}{\partial x} g(f(x)) = g'(f(x))\, f'(x)$$

Product Rule:

$$\frac{\partial}{\partial x} [g(x) f(x)] = g(x) f'(x) + g'(x) f(x)$$

Gradients

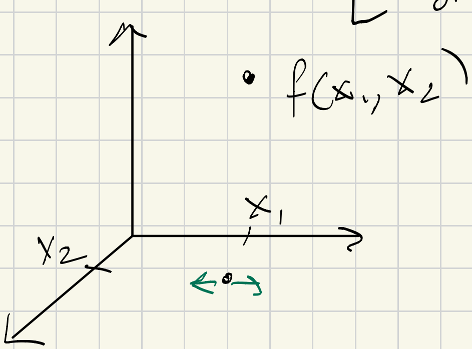
$f : \mathbb{R}^d \to \mathbb{R}$

$f(x_1, x_2, \dots, x_d)$

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d$$

$$\frac{\partial f(\underline{x})}{\partial x_i} = \lim_{h \to 0} \frac{f(\underline{x} + e_i h) - f(\underline{x})}{h}$$

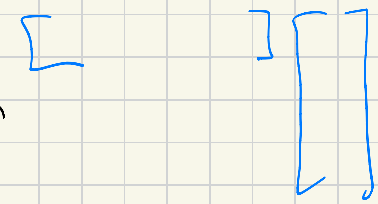
$$\nabla_{\underline{x}} f = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_d} \end{bmatrix}$$



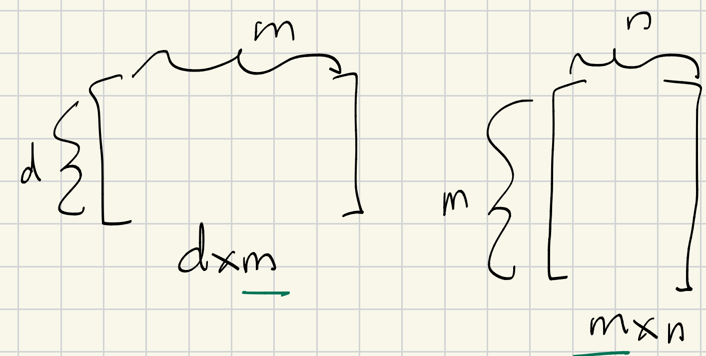
Matrix Multiplication

$\underline{u} \in \mathbb{R}^d$ $\quad \underline{v} \in \mathbb{R}^d$

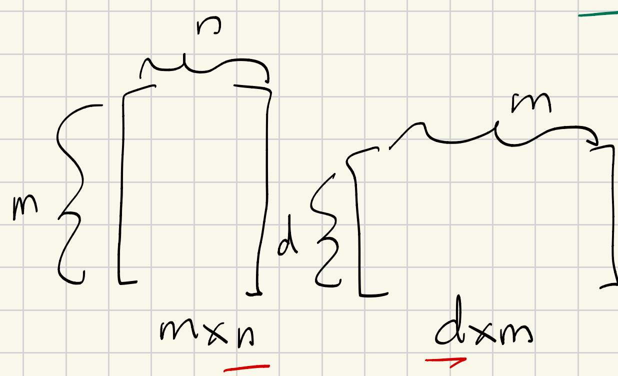
$$\underline{u} \cdot \underline{v} = \sum_{i=1}^{d} u_i v_i = \underbrace{\underline{u}^T}_{1\times d} \underbrace{\underline{v}}_{d \times 1} = \langle \underline{u}, \underline{v} \rangle$$



$A \in \mathbb{R}^{d \times m}$ $\quad B \in \mathbb{R}^{m \times n}$

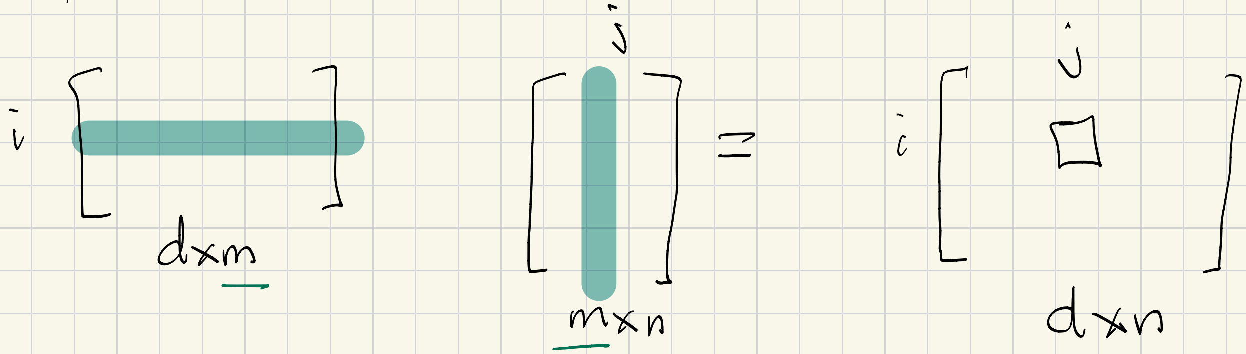


AB is defined!



BA is not!

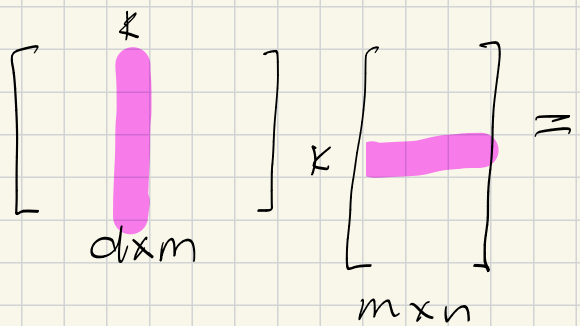
$\underline{A}\,\underline{B} \qquad \underline{A} \in \mathbb{R}^{d \times m} \qquad \underline{B} \in \mathbb{R}^{m \times n}$



$$[AB]_{i,j} = \sum_{k=1}^{m} [A]_{i,k} [B]_{k,j}$$

$$= \underbrace{[\underline{A}]_{i,}}_{1 \times m} \underbrace{[\underline{B}]_{,j}}_{m \times 1}$$

Another perspective...



$$= \sum_{k=1}^{m} \underbrace{[A]_{,k}}_{d \times 1} \underbrace{[B]_{k,}}_{1 \times n} = \underline{A}\,\underline{B}$$

Activity: Confirm that outer product is correct

Eigenvalues & Eigenvectors

"self" in German

Motivation: want basic operations e.g. A^{-1}, A^2, ...

$\underline{A} \in \mathbb{R}^{d \times d}$ and symmetric i.e. $[A]_{i,j} = [A]_{j,i}$

$$\underline{A}\,\underline{v} = \lambda \underline{v}$$

$\underline{v} \in \mathbb{R}^d$ eigenvector $\quad \lambda \in \mathbb{R}$ eigenvalue

r = # eigenvectors

$\underline{v}_1, \ldots, \underline{v}_r$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$

$$\langle \underline{v}_i, \underline{v}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

orthonormal

$$\underline{A} = \underline{V}\,\underline{\Lambda}\,\underline{V}$$

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_r \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_2 \end{bmatrix} \begin{bmatrix} — & \boldsymbol{v}_1^\top & — \\ — & \boldsymbol{v}_2^\top & — \\ & \vdots & \\ — & \boldsymbol{v}_r^\top & — \end{bmatrix}$$

outerproduct!

$$= \sum_{i=1}^{r} \underline{v}_i \lambda_i \underline{v}_i^\top$$

8/28/25

Class Structure

↳ Read typed notes

↳ Optional: print the "slides"

↳ Start thinking about pset

↳ Stop by office hours! :)
(puzzle + Oreo reeses)

Reminders:

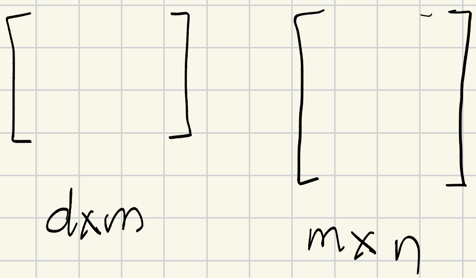
↳ Please no electronic devices

↳ Quiz on Tuesday!

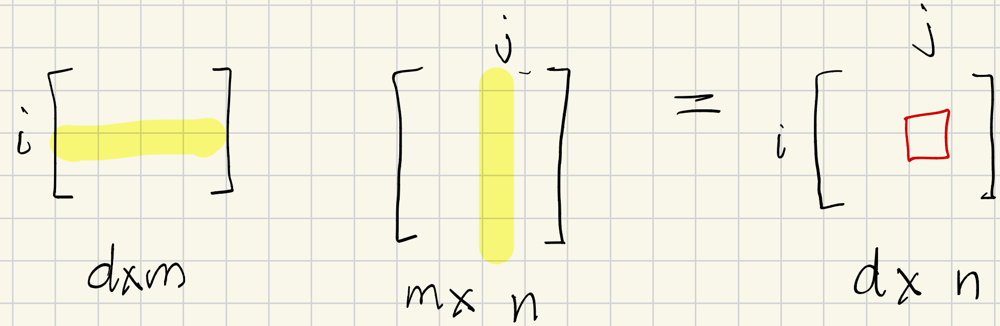
↳ Participation form *every* day

Review

$\underline{A} \in \mathbb{R}^{d\times m}$ $\underline{B} \in \mathbb{R}^{m\times n}$

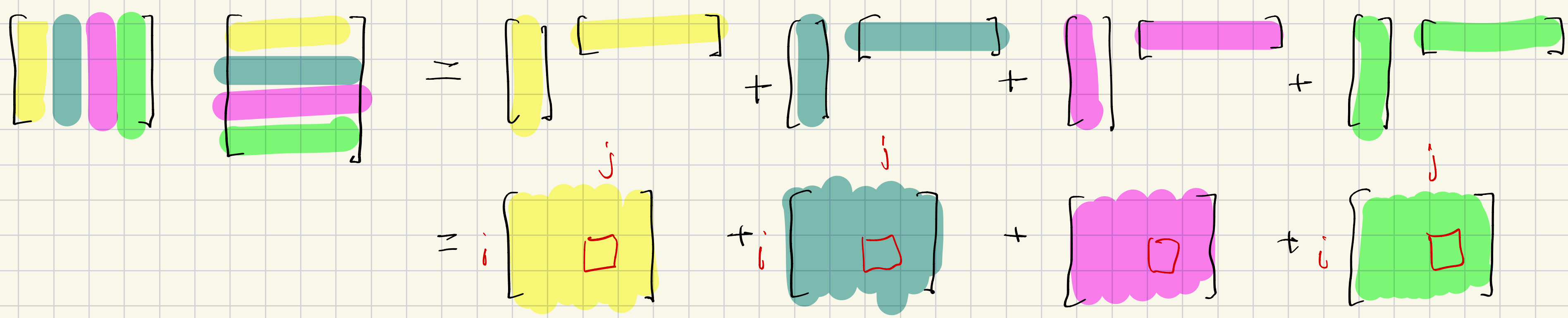


"Normal"



$$[\underline{A}\underline{B}]_{i,j} = [\underline{A}]_{i,:}[\underline{B}]_{:,j} = \sum_{k=1}^{m} [\underline{A}]_{i,k}[\underline{B}]_{k,j}$$

"Outer Product"



$$= [\underline{A}]_{:,1}[\underline{B}]_{1,:} + [\underline{A}]_{:,2}[\underline{B}]_{2,:} + [\underline{A}]_{:,3}[\underline{B}]_{3,:} + [\underline{A}]_{:,4}[\underline{B}]_{4,:}$$

$$[\underline{A}\,\underline{B}]_{i,j} = \left([\underline{A}]_{i,1}[\underline{B}]_{1,j} + \dots\right) = \sum_{k=1}^{m} [\underline{A}]_{i,k}[\underline{B}]_{k,j}$$

Eigenvalues and Eigenvectors

$\underline{A} \in \mathbb{R}^{d \times d}$ diagonalizable matrix

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \in \mathbb{R}$ eigenvalues

$\underline{v}_1, \underline{v}_2, \dots, \underline{v}_r \in \mathbb{R}^d$ right eigenvectors $\qquad \underset{d \times d}{\underline{A}} \; \underset{d \times 1}{\underline{v}_i} = \lambda_i \underset{d \times 1}{\underline{v}_i}$

$\underline{w}_1, \underline{w}_2, \dots, \underline{w}_r \in \mathbb{R}^d$ left eigenvectors $\qquad \underset{1 \times d}{\underline{w}_i^T} \; \underset{d \times d}{\underline{A}} = \lambda_i \underset{1 \times d}{\underline{w}_i^T}$

$$\underline{v}_i \underline{v}_j^T = \underline{w}_i \underline{w}_j^T = \underline{v}_i \underline{w}_j^T = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

$$\underline{A} = \sum_{k=1}^{r} \underline{v}_k \lambda_k \underline{w}_k^T$$

What is ...

$\underline{A} \, \underline{v}_i$? $\qquad \underline{w}_i^T \underline{A}$?

Page Rank

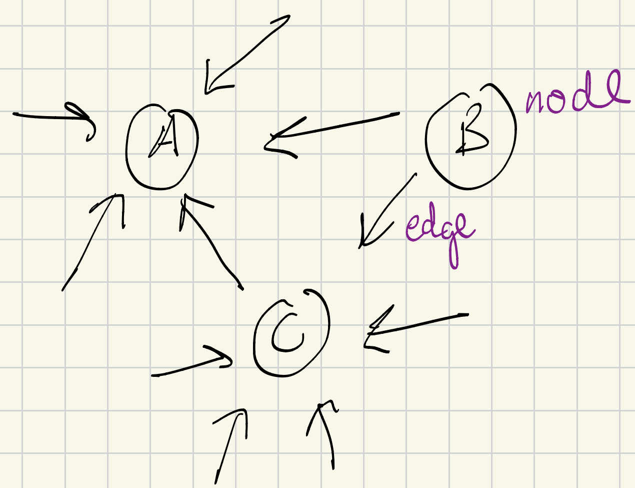
Application of math :)

1. Larry Page
2. wep page

Q: How important is a website?

(!!!) Wikipedia.org
(!!!) Mayo clinic.org
(!) Teals cat.com

How can we $\boxed{\text{algorithmically}}$ answer?



Important pages have

$\boxed{1.}$ Lots of incoming links

$\boxed{2.}$ Incoming links from important pages

$\underline{p}^{(0)} \in \mathbb{R}^{n = \#\text{ pages}}$ is (initial) importance

$p_i^{(0)} = \frac{1}{n}$ all equally important

We will update $\underline{p}^{(t+1)}$ from $\underline{p}^{(t)}$ to incorporate $\boxed{1.}$ and $\boxed{2.}$

$$p_i^{(t+1)} = \sum_{j=1}^{n} \mathbb{1}[j \text{ links to } i] \frac{p_j^{(t)}}{d_j}$$

normalize for # links from j

$$p_i^{(t+1)} = \sum_{j=1}^{n} \mathbb{1}[j \text{ links to } i] \frac{p_j^{(t)}}{d_j}$$

Inner product structure!

$$= [\underline{A}]_{i,:} \, \underline{p}^{(t)} \quad \text{where} \quad [\underline{A}]_{i,j} = \begin{cases} 1/d_j & \text{if } j \text{ links to } i \\ 0 & \text{else} \end{cases}$$

$$\underset{d \times 1}{\underline{p}^{(t+1)}} = \underset{d \times d}{\underline{A}} \; \underset{d \times 1}{\underline{p}^{(t)}}$$

column-stochastic $\iff$ $\displaystyle\sum_{i=1}^{n} [\underline{A}]_{i,j} = \sum_{i=1}^{n} \mathbb{1}[j \text{ links to } i] \frac{1}{d_j}$

$$= \frac{1}{d_j} \sum_{i=1}^{n} \mathbb{1}[j \text{ links to } i] = \frac{d_j}{d_j} = 1$$

$$\sum_{i=1}^{n} p_i^{(t+1)} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} [A]_{i,j} \, p_j^{(t)} \right) = \sum_{j=1}^{n} p_j^{(t)} \sum_{i=1}^{n} [A]_{i,j} = \sum_{j=1}^{n} p_j^{(t)}$$

$$\implies \sum_{i=1}^{n} p_i^{(0)} = \sum_{i=1}^{n} p_i^{(t)} = 1$$

$\underline{p}^{(t)}$ is a probability distribution! $\quad p_i^{(t)} = \Pr(\text{random user on page } i \text{ at time } t)$

Challenges:

1. Reducible if a cluster doesn't link out or in
2. Periodic if probability oscillates eg. $A \to B \to A \to \dots$

Solution: Jump to random page w.p. $1-\alpha$

$$\underline{p}^{(t+1)} = \alpha \underline{A}\, \underline{p}^{(t)} + (1-\alpha)\, \underline{1}\, \frac{1}{n}$$

Let's make this nicer. Recall $\sum_{i=1}^{n} p_i^{(t)} = 1 = \underline{1}^T \underline{p}^{(t)}$

$$\underline{p}^{(t+1)} = \alpha \underline{A}\, \underline{p}^{(t)} + \frac{(1-\alpha)}{n} \underbrace{\underline{1}}_{d\times 1}\, \underbrace{\underline{1}^T}_{1\times d}\, \underbrace{\underline{p}^{(t)}}_{d\times 1} = \underbrace{\left(\alpha \underline{A} + \frac{1-\alpha}{n} \underline{1}\, \underline{1}^T\right)}_{\text{call this } \underline{M}} \underline{p}^{(t)}$$

$$\underline{p}^{(t+1)} = \underline{M}\, \underline{p}^{(t)} = \underline{M}\, \underline{M}\, \underline{p}^{(t-1)} = \underline{M}^t\, \underline{p}^{(0)}$$

eigendecomposition

$$\underline{M} = \sum_{i=1}^{r} \lambda_i \underline{v}_i \underline{w}_i^T$$

$$\underline{w}_i^T \underline{v}_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$$

$$\underline{M}^2 = \sum_{i=1}^{r} \lambda_i \underline{v}_i \underline{w}_i^T \sum_{j=1}^{r} \lambda_j \underline{v}_j \underline{w}_j^T$$

$$=$$

$$\underline{M}^t =$$

$$\underline{p}^{(t)} = \underline{M}^{t}\,\underline{p}^{(0)} = \sum_{i=1}^{r} \lambda_i^t\, \underline{v}_i\, \underline{w}_i^T\, \underline{p}^{(0)} = \lambda_1^t \sum_{i=1}^{n} \left(\frac{\lambda_i}{\lambda_1}\right)^t \underline{v}_i\, \underline{w}_i^T \underline{p}^{(0)}$$

$$\lim_{t\to\infty} p^{(t)} = \lambda_1^t\, \underline{v}_1 [\underline{w}_1^T \underline{p}^{(0)}] \overset{\lambda_1 = 1}{=} \underline{v}_1\, (\underline{w}_1^T\, \underline{p}^{(0)})$$

Page Rank is just top eigenvector!