Lecture 1

Brief History

Knowledge Discovery & Data Mining

DM — Exploratory methods + Unsupervised

supervised Learning

ML — reinforcement learning
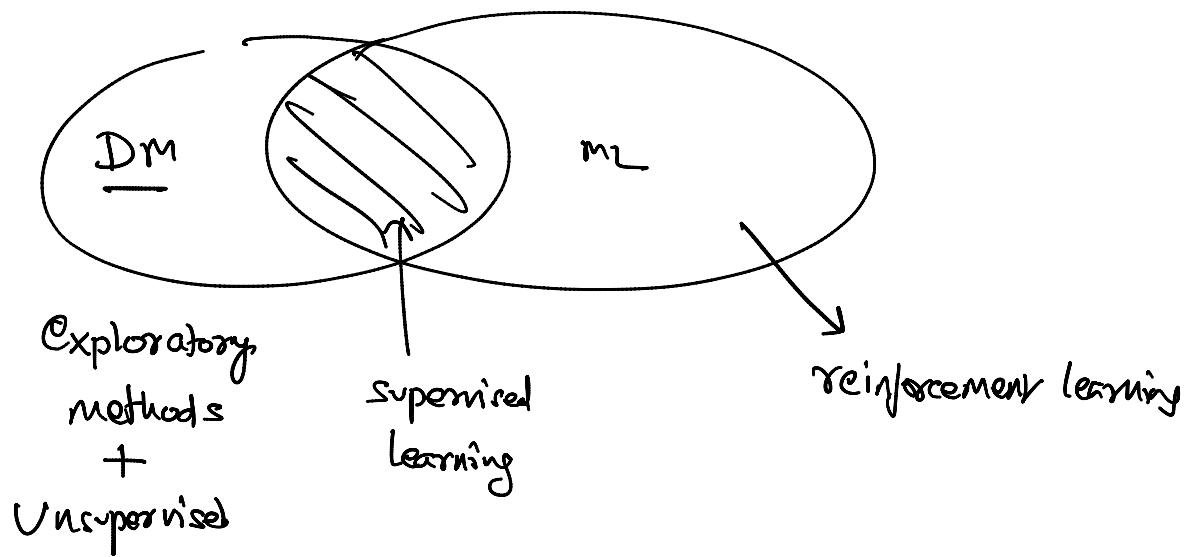
Data ?

↳ Tabular → each instance is a measurement of a set of variables

↓

Spatio-temporal

↳ Text → categorical / symbolic

↳ Images

Data Matrix

$$D =$$

$$d$$
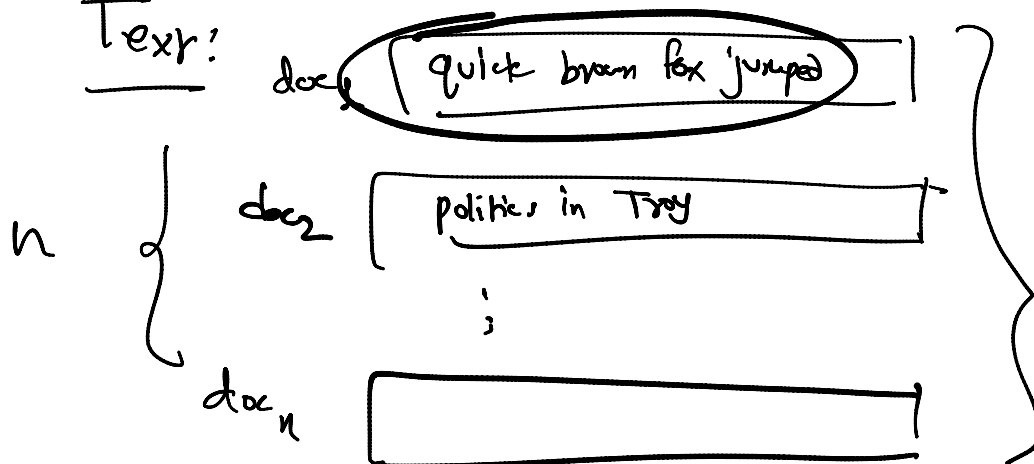
$$X_1 \; X_2 \; \cdots \; X_d$$

$$\underbrace{X_1, X_2, \ldots, X_d}$$

Variables or attributes or features

$n \equiv$ # of examples / instances / objects

$d \equiv$ # of attributes

$$D \in \mathbb{R}^{n \times d}$$

Text:

doc_1 — (quick brown fox jumped)

$n \left\{ \begin{array}{l} doc_2 \text{ — politics in Troy} \\ \vdots \\ doc_n \end{array} \right.$

$$=$$

V

Vocab of tokens

the

One-hot representation for "the"

$|V|$: Size of vocab $\equiv$ 50k

doc$_1$

| quick | brown | fox | jumped |

50k $\equiv |V|$

doc$_1$

for text

$D =$

doc$_1$ | quick | brown | ... | $|V|$

n

d
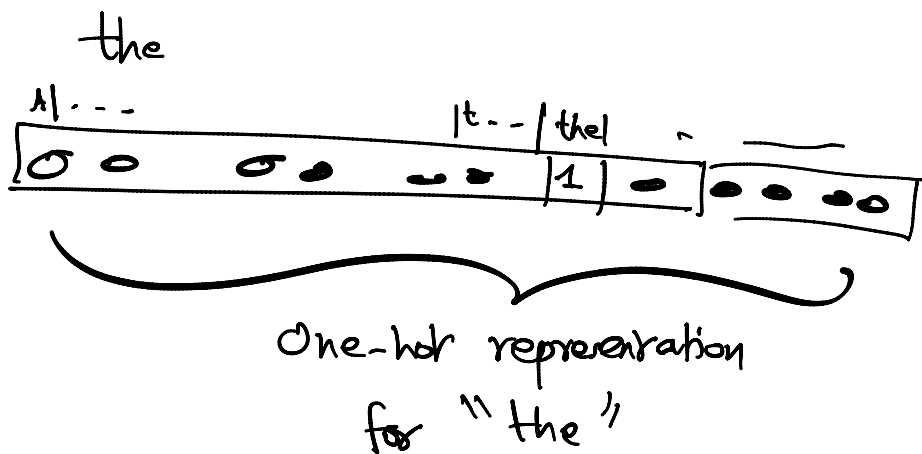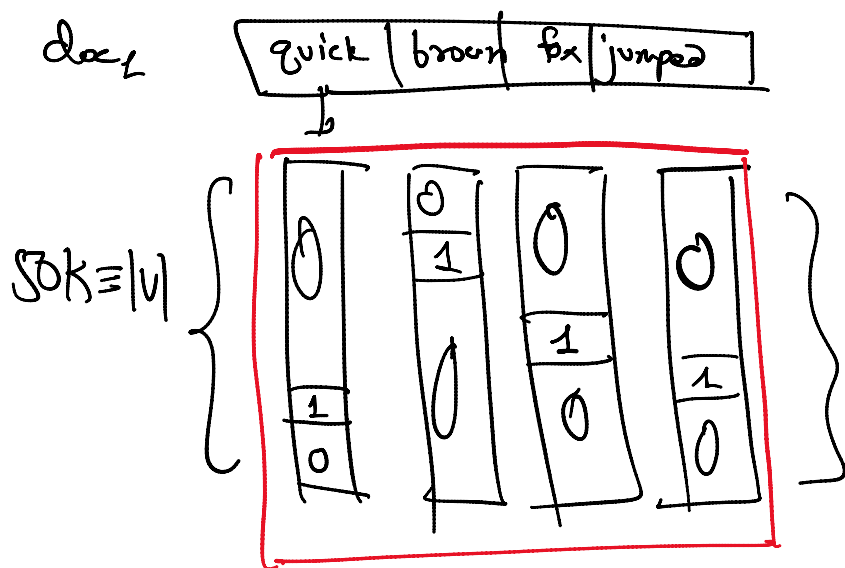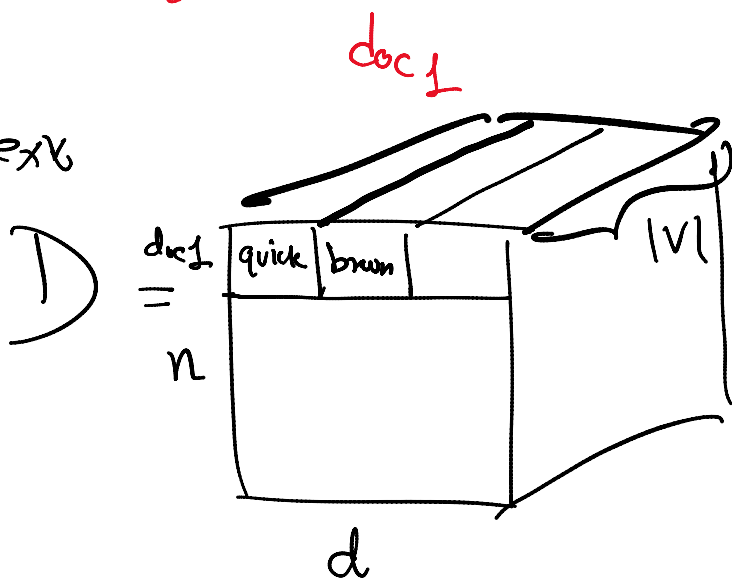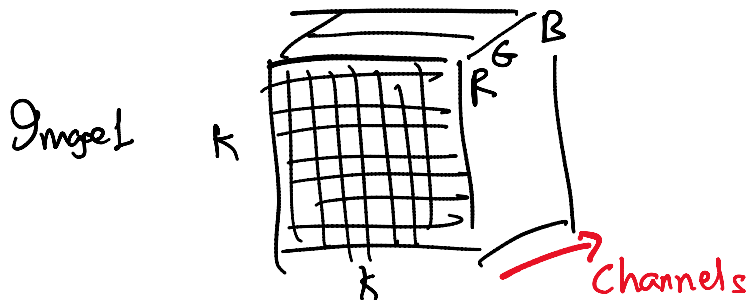
tensor

max doc size
in terms # tokens

$$D \in \mathbb{R}^{n \times d \times |V|}$$

Images

Image 1

k

$k \times k$: pixel grid

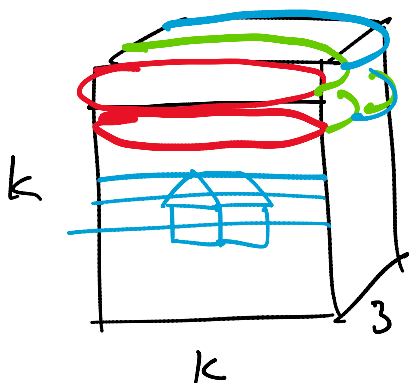$k \times k \times 3$ image

$$\underline{n} \times (k \times k \times 3)$$

# of images

$$D \in \mathbb{R}^{n \times k \times k \times 3}$$

4D tensor

---

k

k

3

$\Rightarrow$

$k \times k \times 3$ Vector

$(\qquad)\frown$

K

$$D \in \mathbb{R}^{n \times d}$$

$$\underline{d = \underline{\underline{k \times 3 \times k}}}$$

$D = $

d

$\vec{A_1} \quad \vec{A_2} \quad \cdots \quad \vec{A_d}$

$\vec{x_1}^T$

$\vec{x_2}^T$

n

$\vec{x_n}^T$

Sample
Size

(data
Size)
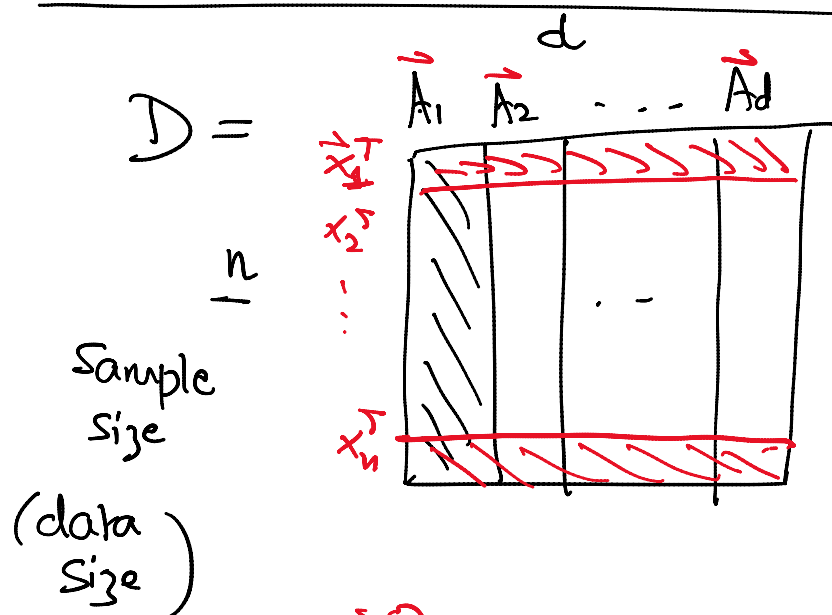
$A_j = j^{th}$ Attribute

$\left(X_j \right)$

$A_j \in \mathbb{R}^n$

all vectors are

<u><u>Column vectors</u></u>

$$\vec{x_i}^{\textcircled{T}} = \left(x_{i1}, x_{i2}, \cdots, x_{id} \right)$$

$$\vec{x_i} = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix}$$

$\vec{x_i} \in \mathbb{R}^d$

feature vector

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

$$\begin{pmatrix} \vdots \\ a_m \end{pmatrix} \qquad \begin{pmatrix} \vdots \\ b_m \end{pmatrix}$$

$$\boxed{\vec{a}^T \vec{b}} \equiv a \cdot b \equiv \langle a, b \rangle \equiv \sum_{i=1}^{M} a_i b_i$$

$$\underbrace{}$$

dot product

inner product

$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

$$\vec{a}^T \vec{b} = 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 2$$

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \Bigg| \quad = 1 + 6 = 7$$

induce a __norm__

$$\|\vec{a}\|^2 = \vec{a}^T \vec{a} = \sum_{i=1}^{m} a_i^{\,\textcircled{2}}$$

L_2 norm

default

$$\|\vec{a}\| = \sqrt{\vec{a}^T \vec{a}} = \sqrt{\sum_{i=1}^{m} a_i^2} \leftarrow$$

Euclidean Norm

(__distance__)

(opposite of similarity)

L_p - Norm

$$\|\vec{a}\|_p = \left(\sum_{i=1}^{m} |a_i|^p \right)^{1/p}$$
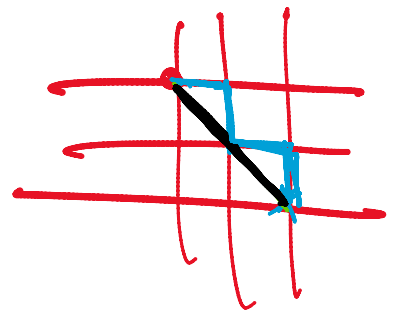
$p = 2 \Rightarrow$ Euclidean Norm

$$\begin{cases} \|\vec{a}\|_1 = \sum_{i=1}^{m} |a_i| = |a_1| + |a_2| + \cdots + |a_n| \quad \longleftarrow \text{ robust} \\ \|\vec{a}\|_2 = \sum_{i=1}^{m} |a_i|^2 = a_1^2 + a_2^2 + \cdots + a_m^2 \end{cases}$$

$\longleftarrow$ *Manhattan Norm*

Norms are also used for "regularization"

$$\Theta = \text{model } \underline{\text{parameters}}$$

$\hookrightarrow$ learnable

$$\|\Theta\|_2 \quad \text{or} \quad \|\Theta\|_1$$

$\underset{\Theta}{\text{Min}} \quad J \equiv \text{objective} \equiv J(\Theta)$

$\searrow$ regularized objective

$$\boxed{J(\Theta) + c\|\Theta\|_p}$$

Norm is already a <u>distance</u>

$$\vec{a}, \vec{b} \in \mathbb{R}^m$$

$$\| \vec{a} - \vec{b} \| \equiv \text{distance}$$

$$= \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots (a_m - b_m)^2}$$

$$\boxed{\| \vec{a} \| = \| \vec{a} - \vec{0} \|}$$

$$\vec{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^m$$

Zero vector

$$\vec{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^m$$

Ones vector

One-hot vectors $\Big\{$ basis vectors

$$\vec{e}_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \leftarrow i^{th}$$

3D space

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$
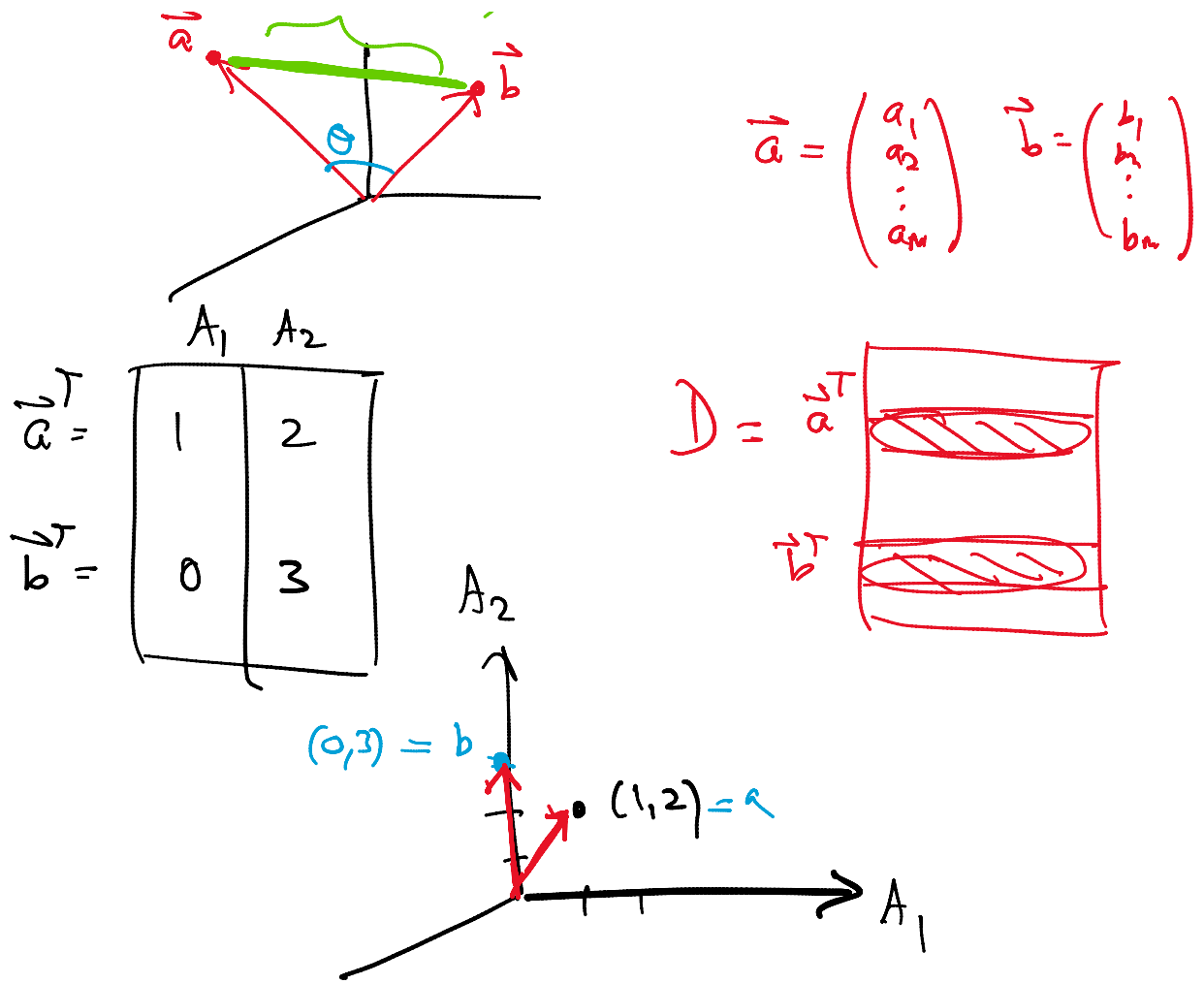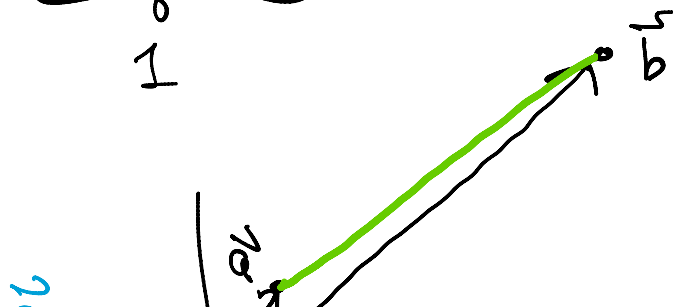
$\underbrace{\hspace{5cm}}$

Standard basis vectors in 3D

X   Y   Z

distance $(L_2) = \| \vec{a} - \vec{b} \|$

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{pmatrix} \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}$$

$$D = \begin{matrix} \vec{a}^T \\ \vec{b}^T \end{matrix}$$

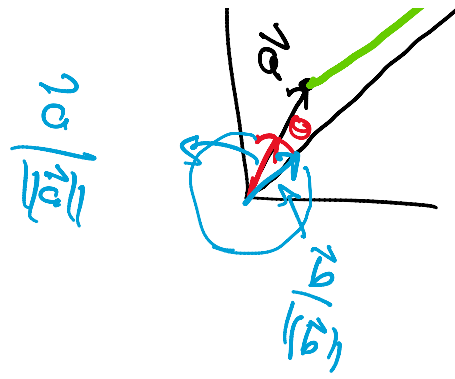|	A_1	A_2
$\vec{a}^T =$	1	2
$\vec{b}^T =$	0	3

$(0,3) = b$

$(1,2) = a$

Angle (as distance)

$$\cos \Theta = \left(\frac{\vec{a}}{\|\vec{a}\|} \right)^T \left(\frac{\vec{b}}{\|\vec{b}\|} \right) = \left(\frac{\vec{a}}{\|\vec{a}\|} \right) \cdot \left(\frac{\vec{b}}{\|\vec{b}\|} \right)$$

$$\frac{\vec{a}}{\|\vec{a}\|} = \text{unit length vector in direction of } \vec{a}$$

$$\underbrace{\qquad}_{1}$$

$$\frac{\vec{a}}{\|\vec{a}\|}$$

$\vec{a}$

θ

$$\frac{\vec{b}}{\|\vec{b}\|}$$

angle is small

$$\|\vec{a} - \vec{b}\|$$
is large!