Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#421 Vectors & Matrices

été 2020

The Matrix

Data

Provinces et territoires du Canada

Carte	Drapeau	Nom de la province ou du territoire (genre)	Code	Rang Date d'entrée	Population en 2016[4] (% de la population totale)	Superficie (km²)	Terres (km²)	Eaux douces (km²)	% de la superficie totale	Densité de la population	sièges (%)	Capitale	Ville la plus peuplée
		Alberta (f)	AB	10 1er septembre 1905	4 067 175 (11,57 %)	661 848	642 317	19 531	6,6 %	6,15	34 (9,1 %)	Edmonton	Calgary
		Colombie-Britannique (f)	BC	7 20 juillet 1871	4 648 055 (13,22 %)	944 735	925 186	19 549	9,5 %	4,92	42 (11,7 %)	Victoria	Vancouver
		Île-du-Prince-Édouard (f)	PE	8 1er juillet 1873	142 907 (0,41 %)	5 660	5 660	0	0,1 %	25,25	4 (1,3 %)	Charlottetown	Charlottetown
		Manitoba (m)	MB	5 15 juillet 1870	1 278 365 (3,64 %)	647 797	553 556	94 241	6,5 %	1,97	14 (4,5 %)	Winnipeg	Winnipeg
		Nouveau-Brunswick (m)	NB	1 1er juillet 1867	747 101 (2,13 %)	72 908	71 450	1 458	0,7 %	10,25	10 (3,2 %)	Fredericton	Moncton
		Nouvelle-Écosse (f)	NS	1 1er juillet 1867	923 598 (2,63 %)	55 284	53 338	1 946	0,6 %	16,71	11 (3,6 %)	Halifax	Halifax
		Ontario (m)	ON	1 1er juillet 1867	13 448 494 (38,26 %)	1 076 395	917 741	158 654	10,8 %	12,49	121 (34,4 %)	Toronto	Toronto
		Québec (m)	QC	1 1er juillet 1867	8 164 361 (23,23 %)	1 542 056	1 365 128	176 928	15,4 %	5,30	78 (24,4 %)	Québec	Montréal
		Saskatchewan (f)	SK	10 1er septembre 1905	1 098 352 (3,13 %)	651 036	591 670	59 366	6,5 %	1,69	14 (4,5 %)	Regina	Saskatoon
		Terre-Neuve-et-Labrador (m)	NL	12 31 mars 1949	519 716 (1,49 %)	405 212	373 872	31 340	4,1 %	1,28	7 (2,3 %)	Saint-Jean de Terre-Neuve	Saint-Jean de Terre-Neuve

via wikipedia

Vectors

$$\vec{\boldsymbol{u}} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \boldsymbol{u} \in \mathbb{R}^n$$

$$\boldsymbol{u} = (u_1 \cdots u_n)^\top$$

Example :

$\vec{\boldsymbol{u}}_j = (178, 185, 162, 170, \cdots, 169)^\top \in \mathbb{R}^{200}$

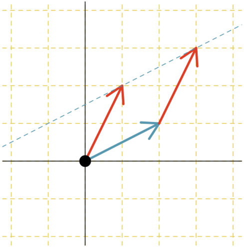
(vector of observations, $n = 200 =$ sample size)

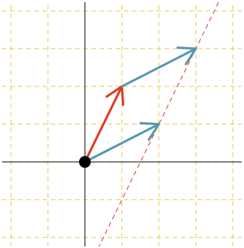
$\vec{\boldsymbol{v}}_i = (178, 74, 1, 0, 14321, 1)^\top \in \mathbb{R}^p$

(vector of individuals, $p = 6 =$ features)

```python
> u = [178, 185, 162, 170]
```

```r
> u = c(178, 185, 162, 170)
```





Scalar Product & Euclidean Distamce

$\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = (u_1 + v_1, \cdots, u_n + v_n)^\top \in \mathbb{R}^n$
Note that $\vec{\boldsymbol{u}} + \vec{\boldsymbol{v}} = \vec{\boldsymbol{v}} + \vec{\boldsymbol{u}}$

Given $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$,

$$\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \rangle = \boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^\top \boldsymbol{v} = \sum_{i=1}^n u_i v_i = u_1 v_1 + \cdots + u_n v_n$$

Example Let $\mathbf{1} = (1, 1, \cdots, 1)^\top$, then $\overline{u} = \frac{1}{n}\langle \vec{\mathbf{1}}, \vec{\boldsymbol{u}} \rangle = \frac{1}{n}\mathbf{1}^\top \boldsymbol{u}$
Example $\boldsymbol{p}$ a probability vector, $\boldsymbol{x}$ a vector of outcome,
$\mathbb{E}(X) = \langle \vec{\boldsymbol{p}}, \vec{\boldsymbol{x}} \rangle$

$\vec{\boldsymbol{u}} \perp \vec{\boldsymbol{v}}$ if and only if $\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \rangle = 0$

$$\|\vec{\boldsymbol{u}}\|^2 = \langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{u}} \rangle = \boldsymbol{u} \cdot \boldsymbol{u} = \sum_{i=1}^n u_i^2 = u_1^2 + \cdots + u_n^2$$

Mahalanobis Distance

The standard Euclidean distance is

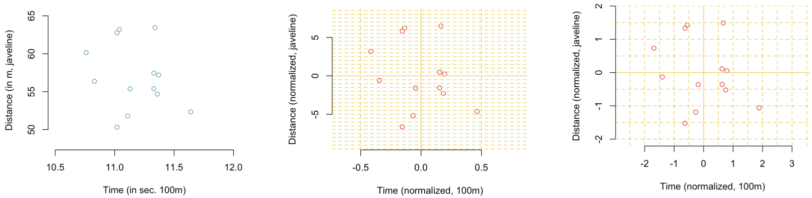
$$\|\vec{\boldsymbol{u}}\|_E^2 = \langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{u}} \rangle = \boldsymbol{u}^\top \boldsymbol{u}$$

Given some diagonal matrix S, and $\boldsymbol{\mu} \in \mathbb{R}^n$,

$$\|\vec{\boldsymbol{u}}\|^2 = (\boldsymbol{u} - \boldsymbol{\mu})^\top S^{-1} (\boldsymbol{u} - \boldsymbol{\mu}) = \|\tilde{\boldsymbol{u}}\|_E^2$$

where $\tilde{\boldsymbol{u}}_i = \dfrac{\boldsymbol{u}_i - \boldsymbol{\mu}_i}{\sqrt{S_{i,i}}}$

(popular in statistics, when $\boldsymbol{u}_i$ is the observation of individual i)



Euclidean Distance

Define the distance $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|$

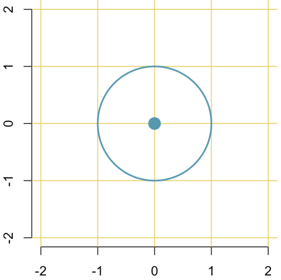
Set of points $\boldsymbol{u}$ such that $d(\boldsymbol{0}, \boldsymbol{v}) = \|\boldsymbol{u}\| = 1$

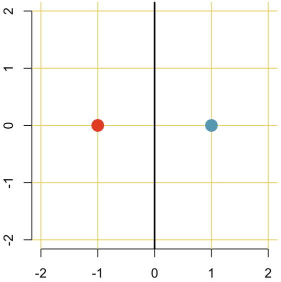
- circle (or sphere in higher dimension)

Set of points $\boldsymbol{x}$ such that $d(\boldsymbol{x}, \boldsymbol{u}) = d(\boldsymbol{x}, \boldsymbol{v})$?

- straight line (or plane in higher dimension)

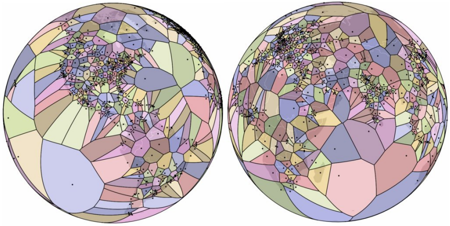
(orthogonal to $\boldsymbol{u} - \boldsymbol{v}$)



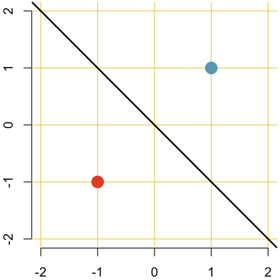


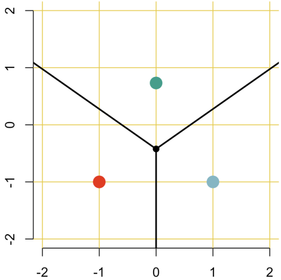
Euclidean Distance & Voronoi Sets

Given a set of points $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots \boldsymbol{u}_n$
Set of points $\boldsymbol{x}$ the closest to $\boldsymbol{u}_i$?
(see also Delaunay triangulation)



(intersections are straight linears or planes)



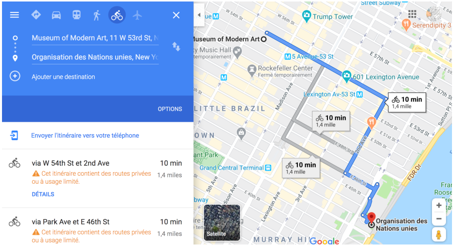


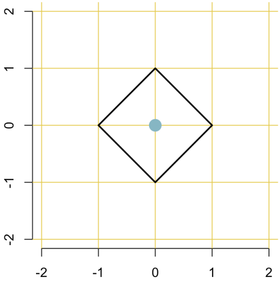
Manhattan (ℓ_1) Distance

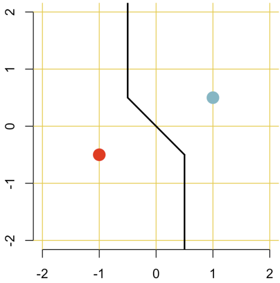
$\|\boldsymbol{u}\| = |u_1| + |u_2|$ is also a norm (Manhattan)

Set of points $\boldsymbol{u}$ such that $d(\mathbf{0}, \boldsymbol{v}) = \|\boldsymbol{u}\| = 1$

- square (or cube in higher dimension)





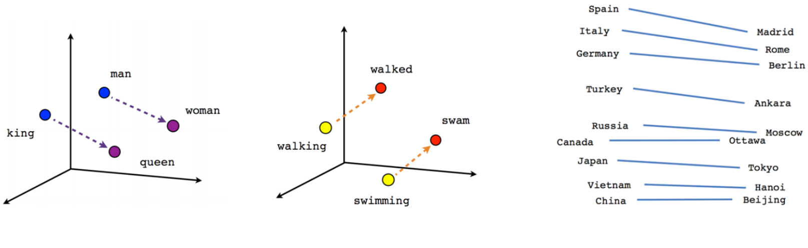


Set of points $\boldsymbol{x}$ such that $d(\boldsymbol{x}, \boldsymbol{u}) = d(\boldsymbol{x}, \boldsymbol{v})$?

- portions of straight lines (or planes)

Word2Vec

Word2vec is one of the most popular models used to create word embeddings (words are represented in a large dimensional space... that can be projected into 2 or 3)

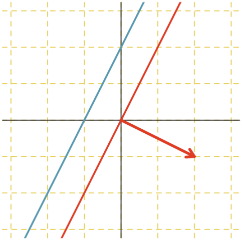


Male-Female   Verb tense   Country-Capital

(source https://www.tensorflow.org)

Linear Function

In dimension d, linear separators are lines (or planes) of the form $\boldsymbol{x} \mapsto \boldsymbol{w}^\top \boldsymbol{x} = \gamma$



Example: $\Delta : y = 2x, \;\; 2x + -1y = 0$

vector $\vec{\boldsymbol{u}} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ is orthgonal to Δ

($\vec{\boldsymbol{u}}$ is a normal vector)

In higher dimension $z = a + bx + cy$ is a plane

Matrices Product

Formally, $AB = C$ with

$$C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^{n} A_{ik}B_{kj}$$

Example:

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$AB = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Hence $AB \neq BA$...

Matrices

Let $\boldsymbol{u} \in \mathbb{R}^n$, then $\boldsymbol{u}^\top \boldsymbol{u} \in \mathbb{R}$ and $\boldsymbol{u}\boldsymbol{u}^\top$ is a $n \times n$ matrix.

Example $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \mathbb{R}^3$

$$\boldsymbol{u}^\top \boldsymbol{u} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1^2 + 2^2 + 3^2 = 14$$

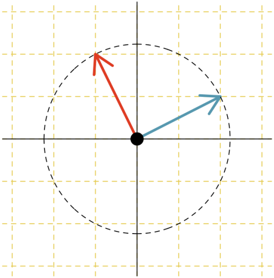
$$\boldsymbol{u}\boldsymbol{u}^\top = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

Matrices & Transformation

Rotation, angle θ, center $\mathbf{0}$:

$$\vec{\boldsymbol{u}} = \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \vec{\boldsymbol{v}} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

or $\vec{\boldsymbol{v}} = \boldsymbol{R}\vec{\boldsymbol{u}}$ with $\boldsymbol{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$



Orthogonal projection, on $\Delta = (\vec{\boldsymbol{\delta}} = (a, b))$,

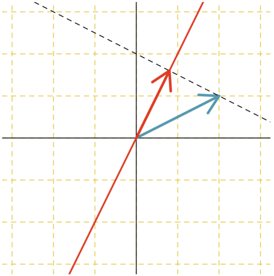
$$\vec{\boldsymbol{u}} = \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \vec{\boldsymbol{v}} = \frac{ax + by}{a^2 + b^2}\begin{pmatrix} a \\ b \end{pmatrix},$$

or $\vec{\boldsymbol{v}} = \boldsymbol{P}\vec{\boldsymbol{u}}$ with $\boldsymbol{P} = \dfrac{1}{a^2 + b^2}\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix}$.

(proof : $\vec{\boldsymbol{v}} = \vec{\boldsymbol{\delta}}\lambda = \vec{\boldsymbol{\delta}}\dfrac{\langle\vec{\boldsymbol{\delta}}, \vec{\boldsymbol{u}}\rangle}{\langle\vec{\boldsymbol{\delta}}, \vec{\boldsymbol{\delta}}\rangle} = \vec{\boldsymbol{\delta}}\dfrac{\vec{\boldsymbol{\delta}}^\top\vec{\boldsymbol{\delta}}}{\vec{\boldsymbol{\delta}}^\top\vec{\boldsymbol{\delta}}} = \dfrac{\vec{\boldsymbol{\delta}}\vec{\boldsymbol{\delta}}^\top}{\vec{\boldsymbol{\delta}}^\top\vec{\boldsymbol{\delta}}}\vec{\boldsymbol{u}}$

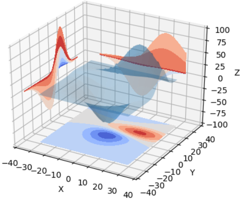
Hence $\boldsymbol{P} = \dfrac{\boldsymbol{\delta}\boldsymbol{\delta}^\top}{\boldsymbol{\delta}^\top\boldsymbol{\delta}}$

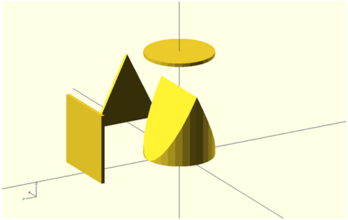
(or more generally $\boldsymbol{P} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^\top\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^\top$).



Projection?

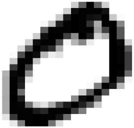
Projection are essential when trying to visualize data in high dimension (≥ 3) (source REAMuseum)

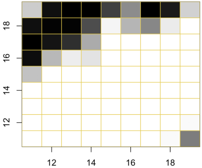


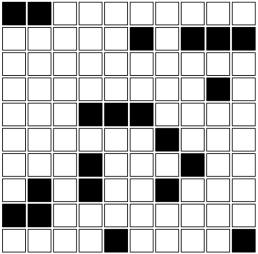


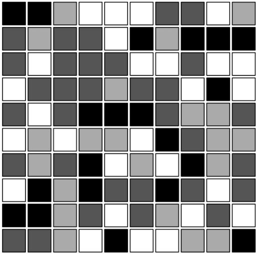
Tensors & Pictures

A $(w \times h)$ picture, with w pixels for the width and h for the height is a "matrix"
Example: black and white picture, matrix M with $M_{i,j} \in [0,1]$ the grey level.





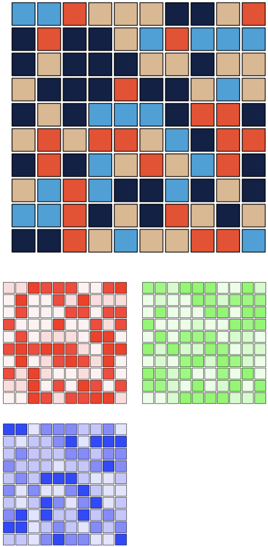




Tensors

A $(w \times h)$ picture, with w pixels for the width and h for the height is a "tensor"
Example: a color picture is a tensor T
$T = (T[,,r], T[,,g], T[,,b])$
(RGB decomposition of a color)
with $T_{i,j,c} \in [0,1]$ the color level.

working with pictures means working with three dimensional matrices



Wrap-up

- a vector is a collector of n numbers, $\boldsymbol{u} = (u_1, \cdots, u_n)^\top$
  $\vec{\boldsymbol{u}}_j = (178, 185, 162, 170, \cdots, 169)^\top$ is a variable
  $\vec{\boldsymbol{v}}_i = (178, 74, 1, 0, 14321, 1)^\top$ is an individual
- $\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \rangle = \boldsymbol{u}^\top \boldsymbol{v} = \displaystyle\sum_{i=1}^{n} u_i v_i = u_1 v_1 + \cdots + u_n v_n$
- $\vec{\boldsymbol{u}} \perp \vec{\boldsymbol{v}}$ if and only if $\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{v}} \rangle = 0$
- $\|\vec{\boldsymbol{u}}\|^2 = \langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{u}} \rangle = \boldsymbol{u}^\top \boldsymbol{u} = \displaystyle\sum_{i=1}^{n} u_i^2 = u_1^2 + \cdots + u_n^2$ (Euclidean)
- data will be related to matrices M
- transformation will be related to matrices T ($\widetilde{M} = TM$)
- orthogonal projection are defined with matrices
- tensors are matrices in higher dimension