Data Science for Actuaries (ACT6100)

Arthur Charpentier

Rappels # 1 (Vectors, Norms & Inner Product)

automne 2020

https://github.com/freakonometrics/ACT6100/

Norm

A norm $\|\cdot\|$, in $\mathbb{R}^n$, satisfies

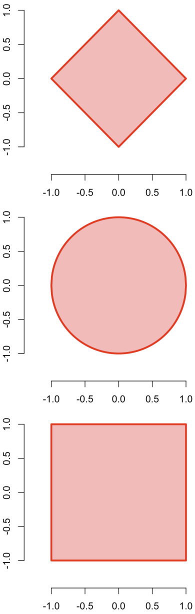
- homogeneity, $\|a\vec{\boldsymbol{u}}\| = |a|\cdot\|\vec{\boldsymbol{u}}\|$, $\forall a$
- triangle inequality, $\|\vec{\boldsymbol{u}}+\vec{\boldsymbol{v}}\| \leq \|\vec{\boldsymbol{u}}\| + \|\vec{\boldsymbol{v}}\|$
- positivity, $\|\vec{\boldsymbol{u}}\| \geq 0$
- definiteness, $\|\vec{\boldsymbol{u}}\| = 0 \iff \vec{\boldsymbol{u}} = \vec{\boldsymbol{0}}$

ℓ_1 norm: $\|\boldsymbol{x}\|_{\ell_1} = |x_1| + \cdots + |x_n|$,
see taxicab geometry

ℓ_2 norm: $\|\boldsymbol{x}\|_{\ell_2} = \sqrt{x_1^2 + \cdots + x_n^2}$,
ℓ_p norm: with $p \geq 1$,

$$\|\boldsymbol{x}\|_{\ell_p} = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

e.g. $\|\boldsymbol{x}\|_{\ell_\infty} = \max\{x_i\}$
Unit balls ($\|\boldsymbol{x}\| \leq 1$) are convex sets



Hilbert Space and Inner Products

An inner product $\langle\cdot,\cdot\rangle$, in $\mathbb{R}^n$, satisfies

- symmetry, $\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}\rangle = \langle\vec{\boldsymbol{v}},\vec{\boldsymbol{u}}\rangle$
- linearity, $\langle a\vec{\boldsymbol{u}} + b\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle = a\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{w}}\rangle + b\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle$
- positivity, $\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{u}}\rangle \geq 0$
- definiteness, $\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{u}}\rangle = 0 \iff \vec{\boldsymbol{u}} = \vec{\boldsymbol{0}}$

Example: On the set of $\mathbb{R}^n$ vectors, $\langle\boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{x}^\top\boldsymbol{y}$

Furthermore

- $\|\boldsymbol{x}\| = \sqrt{\langle\boldsymbol{x},\boldsymbol{x}\rangle}$ defines a norm
- $d(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|$ defines a distance

Example: On the set of $m\times n$ matrices, $\langle\boldsymbol{A},\boldsymbol{B}\rangle = \text{trace}\big(\boldsymbol{A}\boldsymbol{B}^\top\big)$

Example: On the set of random variables, $\langle X,Y\rangle = \mathbb{E}(XY)$

Distance

$d(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|$ defines a distance.

$$d(\boldsymbol{a},\boldsymbol{b}) + d(\boldsymbol{b},\boldsymbol{c}) \geq d(\boldsymbol{a},\boldsymbol{c})$$

Example Let $\boldsymbol{y},\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$. The nearest neighbor of $\boldsymbol{y}$, among the $\boldsymbol{x}_i$'s is

$$\boldsymbol{x}_{i^\star} = \underset{\boldsymbol{x}\in\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}}{\text{argmin}} \left\{d(\boldsymbol{x},\boldsymbol{y})\right\}$$

```
> X = matrix(runif(8*3),8,3)
> dist(X,method = "euclidean")
      1     2     3     4     5     6     7
2 0.832
3 0.698 0.455
4 0.539 1.054 0.716
5 0.834 1.074 0.785 0.921
6 0.755 0.977 0.554 0.354 0.871
7 0.153 0.739 0.691 0.666 0.885 0.849
8 0.734 0.990 0.804 0.977 0.263 0.989 0.752
```

Cauchy-Schwarz Inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \text{ with equality only when } \boldsymbol{x} = \lambda \boldsymbol{y}$$

Application: $x_i \leftarrow x_i - \overline{x}$ and $y_i \leftarrow y_i - \overline{y}$,

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^n x_i y_i$$

$$\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} \cdot \sqrt{\boldsymbol{y}^\top \boldsymbol{y}} = \sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i}$$

$$\text{corr}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\displaystyle\sum_{i=1}^n x_i y_i}{\sqrt{\displaystyle\sum_{i=1}^n x_i^2} \cdot \sqrt{\displaystyle\sum_{i=1}^n y_i^2}} = \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|} \in [-1, +1]$$

and $\text{corr}(\boldsymbol{x}, \boldsymbol{y}) = \pm 1$ only when $\boldsymbol{x} = \lambda \boldsymbol{y}$.

Angles

Since $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|$, define θ as the unique number in $[0, \pi]$ that satisfies

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \cos(\theta), \text{ or } \theta = \arccos \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|}$$

Mahalanobis distance

A $n \times n$ symmetric matrix $\boldsymbol{M}$ is positive definite if $\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x} > 0$ for any $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$.

Proposition: If $\boldsymbol{M}$ is a positive definite (symmetric) matrix, then $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{y}$ defines an inner product.

(furthermore, conversely, if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{y}$ defines an inner product, then $\boldsymbol{M}$ is definite positive)

- $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_M = \boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{y}$ defines an inner product,
- $\|\boldsymbol{x}\|_M = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_M}$ defines a norm
- $d_M(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_M$ defines a distance

Given $\boldsymbol{\Sigma}$ some $n \times n$ definite positive matrix, define

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})}$$

Given $\boldsymbol{\mu} \in \mathbb{R}^n$, define the Mahalanobis "norm" as

$$\|\boldsymbol{x}\| = d(\boldsymbol{x}, \boldsymbol{\mu}) = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}$$

If $\boldsymbol{\Sigma}$ is diagonal, it is also called standardized Euclidean distance.

See on the generalised distance in statistics, 1936.

Linear Independence

A collection of k vectors $\{\vec{\boldsymbol{x}}_1, \cdots, \vec{\boldsymbol{x}}_k\}$ (in $\mathbb{R}^n$) are linearly dependent if there are $\alpha_1, \cdots, \alpha_k$ such that

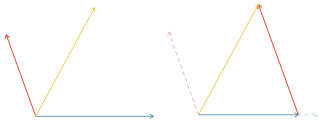
$$\alpha_1\vec{\boldsymbol{x}}_1 + \alpha_2\vec{\boldsymbol{x}}_2 + \cdots + \alpha_k\vec{\boldsymbol{x}}_k = \vec{\boldsymbol{0}}$$

They are linearly independent if they are not linearly dependent, i.e.

$$\alpha_1\vec{\boldsymbol{x}}_1 + \alpha_2\vec{\boldsymbol{x}}_2 + \cdots + \alpha_k\vec{\boldsymbol{x}}_k = \vec{\boldsymbol{0}} \implies \alpha_1 = \cdots = \alpha_k = 0$$

Proposition A collection of linearly independent vectors in $\mathbb{R}^n$ can have, at most, n elements.

$(n+1)$ vectors in $\mathbb{R}^n$ are linearly dependent



Orthonormlization (Gram-Schmidt)

If vectors $\{\vec{\boldsymbol{x}}_1, \cdots, \vec{\boldsymbol{x}}_k\}$ are linearly independent, the Gram-Schmidt algorithm produces an orthonormal collection of vectors $\{\vec{\boldsymbol{u}}_1, \cdots, \vec{\boldsymbol{u}}_k\}$ with the following properties

- every $\vec{\boldsymbol{x}}_j$ is a linear combination of $\{\vec{\boldsymbol{u}}_1, \cdots, \vec{\boldsymbol{u}}_j\}$
- every $\vec{\boldsymbol{u}}_j$ is a linear combination of $\{\vec{\boldsymbol{x}}_1, \cdots, \vec{\boldsymbol{x}}_j\}$

Algorithm 1: Gram-Schmidt

initialization : $\{\vec{\boldsymbol{x}}_1, \cdots, \vec{\boldsymbol{x}}_n\}$;

for t=1,2,...,k **do**

$\quad \vec{\boldsymbol{u}}_t = \vec{\boldsymbol{x}}_t - \big[\langle \vec{\boldsymbol{u}}_1, \vec{\boldsymbol{x}}_t\rangle \vec{\boldsymbol{u}}_1 + \cdots + \langle \vec{\boldsymbol{u}}_{t-1}, \vec{\boldsymbol{x}}_t\rangle \vec{\boldsymbol{u}}_{t-1}\big]$;

$\quad \vec{\boldsymbol{u}}_t = \vec{\boldsymbol{u}}_t / \|\vec{\boldsymbol{u}}_t\|$