VECTOR AND MATRIX NORMS

Francois Role - francois.role@parisdescartes.fr

September 2017

1 Vector norms

1.1 Definition

Let E be a vector space on a field K (e.g. R or C). An application $\|.\| \rightarrow R$ is called "norm on E" if: * for every element x of E : * $||x|| \geq 0$ with $\|x\| = 0 \Leftrightarrow x = 0$ (nonnegative and definite) * for every element x of E and every element α of K: * $|\alpha|\|x\| = \|\alpha x\|$ (homogeneous) * for every element x and y of E: * $\|x+y\| \leq \|x\| + \|y\|$

1.2 Usual Vector Norms: p-norms

$\|x\|_p = \left(\sum_{i=1}^m |x_i|^p \right)^{1/p}$

The p-norm we will mostly use is the 2-norm (euclidean norm):

$\|x\|_2 = \left(\sum_{i=1}^m |x_i|^2 \right)^{1/2}$

The two other most common p-norm are:

$$||x||_1 = \sum_{i=1}^{m} |x_i|$$

$$\|x\|_\infty = \max_{i=1 \ldots m} |x_i|$$

Exercise

(a) Compute the distance between the vectors $\vec{x} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} 2 \\ 8 \\ 9 \end{bmatrix}$.

(b)Verify the triangle inequality for $\vec{x}$ and $\vec{y}$.

(c) Prove that $\|x\|_\infty$ is a norm.

1.3 Unit Ball of a Norm

Given a norm $\|.\|$ on a space E, the set B of vectors with a norm less than or equal to one is called the **unit ball of the norm**$\|.\|$:

$$B = \{x \in E \mid \|x\| \leq 1\}$$

2 Matrix Norms

An application from $\mathbb{R}^{m\times n}$ to $\mathbb{R}$ is called a "matrix norm" (denoted $X :\rightarrow \|X\|$) if: * for every matrix $X \in \mathbb{R}^{m\times n}$ * $\|X\| \geq 0$ and $\|X\| = 0 \Leftrightarrow X = 0$ * for every element x of E and every element α of K: * $|\alpha|\|X\| = \|\alpha X\|$ * for every element $X, Y \in R^{m\times n}$: * $\|X+Y\| \leq \|X\| + \|Y\|$

2.1 The Frobenius norm

Given a matrix X of size $m \times n$ the Frobenius norm of X is defined as:

$$\|X\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j}^2} = \sqrt{Tr(X^T X)} \tag{1}$$

$$\tag{2}$$

In the lecture on SVD we will look at another way of computing the Frobenius norm.

2.2 Induced Matrix Norms

Le X be a $m \times n$ matrix and let u be a $n \times 1$ vector. If $\|.\|$ is a vector norm, then the **operator norm** of A induced by $\|.\|$ is:

$$\|A\| = sup_{u\neq 0}\frac{\|Au\|}{\|u\|}$$

As a consequence, we have:

$$\|Ax\| \leq \|A\|\|x\|$$

The norm induced by $\|.\|_1$ is defined as:

$$\|X\|_1 = \max_{j=1...n}\sum_{i=1}^{m}|X_{ij}|$$

The norm induced by $\|.\|_\infty$ is defined as:

$$\|X\|_\infty = \max_{i=1...m}\sum_{j=1}^{n}|X_{ij}|$$

There is also a norm induced by $\|.\|_2$. It is called the **spectral norm** of a matrix and we defer its treatment to the chapter on SVD since it corresponds to the maximum singular value of a matrix.

Exercise

(d) Compute the Frobenius norm for the following matrix:

$$\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \\ 9 & 10 & 11 \end{pmatrix}$$

3 Dot product

3.1 Definition

Consider two vectors $\vec{u}$ and $\vec{v}$ whose components are $u_1, u_2, .., u_n$ and $v_1, v_2, .., v_n$, resp. The dot product of $\vec{u}$ and $\vec{v}$ is denoted $\vec{u} \cdot \vec{v}$ and defined as:

$$\vec{u} \cdot \vec{v} = u_1 v_1 + ... + u_n v_n$$

3.2 Properties

1. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
2. $(\vec{u} + \vec{v}).\vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$
3. $k\vec{u} \cdot \vec{v} = k(\vec{u} \cdot \vec{v}) = \vec{u} \cdot k\vec{v}$
4. $\vec{u} \cdot \vec{u} \geq 0$
5. $\vec{u} \cdot \vec{u} = 0 \Leftrightarrow \vec{u} = \vec{0}$

3.3 Relations between Dot Products and Norms

3.3.1 Important equalities and inequalities

1. $\vec{u} \cdot \vec{u} = \|\vec{u}\|^2 \Leftrightarrow \|u\| = \sqrt{u \cdot u}$ (since $\|\vec{u}\| \geq 0$)
2. $\cos \widehat{u, v} = \frac{u \cdot v}{\|u\| \|v\|}$
3. $|u \cdot v| \leq \|u\| \|v\|$ (Cauchy-Schwartz inequality)

Exercise

(e) In the previous section, we computed the distance betwen $\vec{x}$ and $\vec{y}$ as np.linalg.norm(x-y) Taking inspiration from equality 1, find an equivalent expression for computing this distance. Verify that you get the same result as for exercise **a** in the previous section.
(f) Verify the Cauchy-Schwartz inequality for $\vec{x}$ and $\vec{y}$ (the same vectors as those in exercise a).
(g) Compute the dot product between the two vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Using equality 2 and computing the norms of the two vectors, deduce the meaning of the obtained value.
(h) Compute the angle between $\vec{x}$ and $\vec{y}$ (the same vectors as those in in exercise a).
(i) Prove that $|\alpha| \|u\| = \|\alpha u\|$ (use the properties: $k\vec{u} \cdot \vec{v} = k(\vec{u} \cdot \vec{v}) = \vec{u} \cdot k\vec{v}$ and $\|u\| = \sqrt{u \cdot u}$.
(j) Use the properties of the dot product to prove the Pythagoras theorem.
(k) Prove the triangular inequality.

4 Dot product and orthogonality

- *Orthogonal vectors*: two vectors u and v are orthogonal iff $u \cdot v = 0$.
- *Orthogonal sets of vectors*: two sets of vectors X and Y are othogonal if every vector in X is orthogonal to every vector in Y.
- *Orthogonal set of vectors*: a set of (nonzero) vectors X is orthogonal if its elemnts are pairwise orthogonal.

- *Orthonormal set of vectors*: a set of vectors X is orthonormal if it is orthogonal and also all its elements are unitary vectors.

Let $B \subseteq \mathbb{R}^m$ be an orthogonal set. if $|B| = m$ then B is is a basis for $\mathbb{R}^m$.

Exercise

(l) What can be said about the set of vectors $\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 8 \end{bmatrix}$

5 Orthogonal Projections

5.1 Projection of a vector on a vector

Let x and u be two vectors. x can be defined as the sum of a vector $x^{||}$ parallel to u and a vector $x^{\perp}$ perpendicular to u:

$$x = x^{||} + x^{\perp}$$

$x^{||} = k\frac{u}{\|u\|}$ with $k = \cos\theta\|x\|$ where θ is the angle between x and u.

$$\begin{aligned} x^{||} &= cos\theta\|x\|\frac{u}{\|u\|} && (3) \\ &= \frac{x.u}{\|x\|\|u\|}\|x\|\frac{u}{\|u\|} && (4) \\ &= \frac{x.u}{\|u\|}\frac{u}{\|u\|} && (5) \\ &= \frac{x.u}{\|u\|^2}u && (6) \\ & && (7) \end{aligned}$$

Note that $x^{||} = (x.u)u$ if u is a unit vector.

5.2 Projection of a vector on a subspace

More generally, if V is a subpace of R^n with an orthonormal basis $u_1, \ldots, u_m$ then:

$$x^{||} = proj_v x = (u_1.x)u_1 + \cdots + (u_m.x)u_m = UU^T x$$

where the u_i are the columns of U. UU^T is called the **projection matrix** onto the subspace$V$.

5.3 Projections and linear applications

Let T be an orthogonal projection onto a subspace V of R.. The set $\{x \mid T(x) = 0\}$ contains elements such that $T(x) = x^{||} = 0$. Since $x = x^{||} + x^{\perp}$, $x = x^{\perp}$. The kernel of T consists of those vectors that are perpendicular to V, and this kernel is called the **orthogonal complement** $V^{\perp}$ of V. Being the kernel of T, $V^{\perp}$ is also a subspace.

6 Cosine and pearson correlation

Let x and y be two vectors. Consider them as two paired samples and compute their sample correlation (covariation divided by the product of the squared variations). We have:

$$\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} = \frac{(x - \bar{x}) \cdot (y - \bar{y})}{\|x - \bar{x}\| \|y - \bar{y}\|} \qquad = \cos(\theta_{x-\bar{x}, y-\bar{y}})) \tag{8}$$

Let now X and Y be two centered variables, the population correlation is:

$$\frac{Cov(X,Y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\Sigma_i x_i y_i}{\sigma_x \sigma_y} \qquad = \frac{\Sigma_i x_i y_i}{\sqrt{n}\sigma_x \sqrt{n}\sigma_y} = \frac{X.Y}{\|X\| \|Y\|} = cos(\theta_{XY}) \tag{9}$$

where we take advantage of the fact that the norm of a centered variable X with n observations is:

$$\|X\| = \sqrt{n}\sigma_x$$

That is, if X and Y are two centered variables then $\rho_{XY} = cos(\theta_{XY})$.

What the above also shows is that the correlation between two variables equals the correlation between the centered versions of these variables. Centering does not change correlation.

Finally, also note that computing the covariance of two standardized variables is equivalent to computing the correlation of the initial variables.

Exercise

(m) Verify that the norm of a centered variable x with n observations is:

$$\|x\| = \sqrt{n}\sigma_x$$

(n) Given the following vectors:
x=np.array([26, 101, 124,205,284], dtype=float) y=np.array([7.5,9.8,16.4,23.3,34.0], dtype=float) compute the Pearson correlation coefficient between the original vectors and then the cosine between centered versions of these vectors. Also compute Pearson correlation between the centered vectors. You should find that the three values are equal.
(o) Verify that computing the covariance of two standardized variables is equivalent to computing the correlation of the initial variables.