Mathematics for Machine Learning Part II (Analytic Geometry)

K.Bouanane

7 mars 2023

Norms

Norms I

Definition

Let **V** be a vector space. A norm is a function

$$\begin{aligned} \|.\| : \mathbf{V} &\to \mathbb{R} \\ x &\mapsto \|x\| \end{aligned}$$

which associates to each element $x \in \mathbf{V}$, its length$\|x\| \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and $x, y \in \mathbf{V}$ we have

1. $\|\lambda x\| = |\lambda| \cdot \|x\|$
2. $\|x + y\| \leqslant \|x\| + \|y\|$
3. $\|x\| \geqslant 0$ and $\|x\| = 0 \Leftrightarrow x = 0$

Example : In $\mathbb{R}^n$,

Norms II

1. the Manhattan norm, defined by

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

where $|.|$ is the absolute vector. This norm is also called l_1 norm

2. The Euclidean norm :

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^T x}$$

also called l_2 norm

Scalar product

Standard Dot Product

Standard Dot Product I

Defined by

$$x^T \cdot y = \sum_{i=1}^{n} x_i y_i$$

which is a particular case of scalar product

Definition

Let **V**, **W** and **X** be three vector spaces, a bilinear map Ω is a map of $\mathbf{V} \times \mathbf{W} \to \mathbf{X}$ such that

$$\Omega(\lambda x + \psi y, z) = \lambda\Omega(x, z) + \psi\Omega(y, z)$$

$$\Omega(x, \lambda \grave{y} + \psi \grave{z}) = \lambda\Omega(x, \grave{y}) + \psi\Omega(x, \grave{z})$$

Standard Dot Product II

Definition

Let $\mathbf{V}$ be a vector space and $\Omega : \mathbf{V} \times \mathbf{V} \to \mathbb{R}$ a bilinear map, then

1. Ω is said to be symmetric if $\Omega(x, y) = \Omega(y, x)$
2. Ω is said to be positive definite if

$$\forall x \in \mathbf{V} \backslash \{0\} : \Omega(x, x) > 0, \Omega(0, 0) = 0$$

Definition

Let $\mathbf{V}$ be a vector space and $\Omega : \mathbf{V} \times \mathbf{V} \to \mathbb{R}$. Then if Ω is positive definite symmetric it is called the scalar product of $\mathbf{V}$. We then write $< x, y >$ instead of $\Omega(x, y)$

Standard Dot Product III

Example : $\mathbf{V} = \mathbb{R}^2$ we define

$$< x, y >= x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2 x_2 y_2$$

Show that $< x, y >$ is a scalar product

Positive definite matrices

Positive definite matrices I

We consider a vector space **V** and a scalar product $< \cdot, \cdot >$: $\mathbf{V} \times \mathbf{V} \to \mathbb{R}$. Let **B** be a Base of **V** and $x, y \in$ **V**. We then have $x = \sum_{i=1}^{n} \psi_i b_i$ and $y = \sum_{j=1}^{n} \lambda_j b_j$ where $\mathbf{B} = (b_1, b_2, \ldots, b_n)$ and $\psi_i, \lambda_j \in \mathbb{R}, i, j = \overline{1, n}$

We can then write

$$< x, y >= \left\langle \sum_{i=1}^{n} \psi_i b_i, \sum_{j=1}^{n} \lambda_j b_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i < b_i, b_j >= \hat{x}^T \mathbf{A} \hat{y}$$

where $\mathbf{A}_{ij} =< b_i, b_j >$ and $\hat{x}, \hat{y}$ are the coordinates of x and y with respect to the base **B**. This implies that the scalar product

Positive definite matrices II

is uniquely determined through **1.** We can notice that $\mathbf{A}$ is a symmetric matrix and that $\mathbf{A}$ satisfies :

$$\forall x \in \mathbf{V} \backslash \{0\} : x^T \mathbf{A} x > 0$$

We then say that the matrix $\mathbf{A}$ is symmetric positive definite. If $x^T \mathbf{A} x \geqslant 0, \forall x \in \mathbf{V}$ then $\mathbf{A}$ is said to be positive semi-definite symmetric.

We state the theorem : for a finite vector space $\mathbf{V}$ and a basis $\mathbf{B}$ of $\mathbf{V}$. The bilinear map $< \cdot, \cdot >: \mathbf{V} \times \mathbf{V} \to \mathbb{R}$ is a scalar product if there exists a symmetric matrix, positive definite $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $< x, y >= \hat{x}^T \mathbf{A} \hat{y}$ where $\hat{x}, \hat{y}$ are the coordinates of x, y with

Positive definite matrices III

respect to **B** We have the following properties for $\mathbf{A} \in \mathbb{R}^{n \times n}$ symmetric and positive definite (SDF) :

1. The elements of the diagonal of $\mathbf{A}$: a_{ii} are positive, since $a_{ii} = e_i^T \mathbf{A} e_i > 0$, where e_i is the *ith* vector of the standard basis.
2. The kernel of the linear map defined by $\mathbf{A}$ $Ker(\Phi_A) = \{0\}$, because $x^T \mathbf{A} x > 0, \forall x \neq 0$ and therefore $\mathbf{A}x \neq 0$ if $x \neq 0$

Lengths and distance

Lengths and distance I

We can notice that we can define a norm or a length from a scalar product : $\|x\| = \sqrt{< x, y >}$
The converse, however, is not true. We then have the following equation, called the **Cauchy-Schwarz inequation** :

$$| < x, y > | \leqslant \|x\| \cdot \|y\|$$

Lengths and distance II

Definition

We consider a vector space endowed with a scalar product, we then define

$$d(x, y) = \|x - y\| = \sqrt{< x - y, x - y >}$$

called distance between x and y for $x, y \in \mathbf{V}$. If the dot product is the standard dot product, then the distance is called the Euclidean distance

Application :

$$\begin{aligned} d : & \mathbf{V} \times \mathbf{V} \to \mathbb{R} \\ & (x, y) \mapsto d(x, y) \end{aligned}$$

is called metric

Lengths and distance III

A metric satisfies the following criteria :

- d is positive definite,ie : $d(x, y) \geqslant 0, \forall x, y \in \mathbf{V}$ and $d(x, y) = 0 \Leftrightarrow x = y$
- d is symmetric,ie : $d(x, y) = d(y, x) \forall x, y \in \mathbf{V}$
- Triangle inequality :

$$d(x, z) \leqslant d(x, y) + d(y, z)$$

for all $x, y, z \in \mathbf{V}$

Angles and Orthogonality

Angles and Orthogonality I

According to the Cauchy-Schwarz inequality, we have :

$| < x, y > | \leqslant \|x\|\|y\|$

$$\Rightarrow -1 \leqslant \frac{< x, y >}{\|x\| \cdot \|y\|} \leqslant 1$$

So , there exists a unique $\omega \in [0, \pi]$, for which we have :

$$cos\omega = \frac{< x, y >}{\|x\| \cdot \|y\|}$$

ω represents the angle between the vectors x and y

Angles and Orthogonality II

Definition

Two vectors x, y are said to be orthogonal if and only if $< x, y >= 0$. We write $x \perp y$

x and y are orthonormal

Remark

The orthogonality depends on the scalar product. So two vectors can be orthogonal with respect to a dot product and not orthogonal with respect to another

Definition

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be orthogonal if and only if its columns are orthonormal In other words $\mathbf{A}\mathbf{A}^T = \mathbf{I} = \mathbf{A}^T\mathbf{A}$

Which implies that $\mathbf{A}^{-1} = \mathbf{A}^T$

Angles and Orthogonality III

Remark

Orthogonal matrix transformations do not change the length of a vector. for example, for the standard scalar product We have :

$$\|\mathbf{A}x\|^2 = (\mathbf{A}x)^T(\mathbf{A}x) = x^T\mathbf{A}^T\mathbf{A}x = x^Tx = \|x\|$$

Similarly, the angle between two vectors x, y remains unchanged, by a transformation of a matrix A orthogonal :

$$cos\omega = \frac{(\mathbf{A}x)^T(\mathbf{A}y)}{\|\mathbf{A}x\|\|\mathbf{A}y\|} = \frac{x^T\mathbf{A}^T\mathbf{A}x}{\sqrt{x^T\mathbf{A}^T\mathbf{A}xy\mathbf{A}^T\mathbf{A}y}} = \frac{x^Ty}{\|x\|\|y\|}$$

Orthonormal Basis

Orthonormal Basis I

Definition

We consider an n-dimensional vector space and a basis $\mathbf{B} = \{b_1, b_2, \ldots, b_n\}$ of **V**. **B** is said to be orthonormal (ONB) if :

$$< b_i, b_j >= 0 for i \neq j$$

$$< b_i, b_i >= 1$$

for all $i, j = 1, \ldots, n$.

Example :

In $\mathbb{R}^2$, the vectors $b_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, b_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ form an orthonormal basis.

Orthonormal Basis II

Remark

If $\mathbf{B}$ satisfies only condition (1) then it is called orthogonal basis.

Orthogonal Projections

Orthogonal Projections I

Definition

Let $\mathbf{V}$ be a vector space and $\mathbf{U} \subseteq \mathbf{V}$ a vector subspace of $\mathbf{V}$. A linear map $\pi : \mathbf{V} \to \mathbf{U}$ is called a projection if $\pi^2 = \pi \circ \pi = \pi$.

Projection onto One-Dimensional Subspaces (Lines)

Projection onto One-Dimensional Subspaces (Lines) I

Let $\mathbf{U} \subseteq \mathbb{R}^n$ be a 1-dimensional subspace and b a basis of $\mathbf{U}$.
We are looking to project an element $x \in \mathbb{R}^n$ onto $\mathbf{U}$. So we are looking to find a vector $\pi_u(x) \in \mathbf{U}$, the closest to x. $\pi_u(x)$ is called the projection of x onto $\mathbf{U}$. If we want $\pi_u(x)$ to be closest to x, then the distance $\|\pi_u(x) - x\|$ must be minimal.
So the segment $\pi_x(x) - x$ must be orthogonal to $\mathbf{U}$. In particular to b So we have $< \pi_u(x) - x, b >= 0$.
On the other hand ; $\pi_u(x) \in \mathbf{U}$ including $\pi_u(x) = \lambda b$ for $\lambda \in \mathbb{R}$.
Therefore :

$$< \pi_u(x) - x, b >= 0 \Leftrightarrow < x - \lambda b, b >= 0$$

Projection onto One-Dimensional Subspaces (Lines) II

then

$$< x, b > -\lambda < b, b >= 0 \Leftrightarrow \lambda = \frac{< x, b >}{\|b\|^2}$$

for the standard dot product, we get $\lambda = \frac{b^T x}{\|b\|^2}$ if $\|b\| = 1$ then $\lambda = b^T x$.

The length of $\pi_x(x)$ is equal to :

$$\|\pi_u(x)\| = \|\lambda b\| = |\lambda| \cdot \|b\| = \frac{|b^T x|}{\|b\|^2}\|b\|$$

$$= |cost\omega|\|x\|\|b\| \cdot \frac{\|b\|}{\|b\|^2} = |cos^{'}\omega|\|x\|$$

Projection onto One-Dimensional Subspaces (Lines) III

The projection matrix $\mathbf{P}_\pi$ is such that

$$\pi_u(x) = \mathbf{P}_\pi \cdot x = \lambda b = b\lambda = b \cdot \frac{b^T x}{\|b\|^2} = \frac{bb^T}{\|b\|^2} \cdot x$$

by identification, we have :

$$\mathbf{P}_\pi = \frac{bb^T}{\|b\|^2}$$

Example :

Projection onto One-Dimensional Subspaces (Lines) IV

Let $\mathbf{U} = span[\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}]$ We calculate $\mathbf{P}_\pi = \frac{bb^T}{b^Tb} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$ let $x = [1, 1, 1]^T$

$$\pi_u(x) = \mathbf{P}_\pi \cdot x = \frac{1}{9}\begin{bmatrix} 8 \\ 10 \\ 10 \end{bmatrix}$$

Projection on a vector subspace

Projection on a vector subspace I

Let **U** be a vector subspace of $\mathbb{R}^n$ with $dim(\mathbf{U}) = m \geqslant 1$ Let $\mathbf{B} = (b_1, b_2, \dots, b_m)$ be a basis of **U**. We want to determine the orthogonal projection of $x \in \mathbb{R}$ on **U**, that is $\pi_u(x)$. We know that $\pi_u(x) \in \mathbf{U}$ therefore $\pi_u(x) = \sum_{i=1}^{m} \lambda_i b_i$ with $\lambda_i \in \mathbb{R}, i = 1, \dots, m$ To find the orthogonal projection of x onto **U**, we follow these steps :

- Find the $\lambda_1, \dots, \lambda_m$ coordinates of $\pi_u(x)$ with respect to the base **B** :

$$\pi_u(x) = \sum_{i=1}^{m} \lambda_i b_i = \mathbf{B}\lambda$$

Projection on a vector subspace II

with $\mathbf{B} = [b_1, b_2, \ldots, b_m] \in \mathbb{R}^{n \times m}$, $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]^T$ such that $\pi_u(x)$ is closer to $x, \pi_u(x) - x \perp \mathbf{U}$, so x must be orthogonal to $b_1, b_2, \ldots, b_m$ We then have :

$$
\begin{aligned}
< b_1, x - \pi_u(x) >= 0 \Leftrightarrow b_1^T(x_1 - \pi_u(x)) = 0 \\
< b_2, x - \pi_u(x) >= 0 \Leftrightarrow b_2^T(x_2 - \pi_u(x)) = 0 \\
\vdots \\
< b_m, x - \pi_u(x) >= 0 \Leftrightarrow b_m^T(x_m - \pi_u(x)) = 0
\end{aligned}
$$

Projection on a vector subspace III

So we can write :

$$b_1^T(x - \mathbf{B}\lambda) = 0$$

$$b_2^T(x - \mathbf{B}\lambda) = 0$$

$$\vdots$$

$$b_1^T(x - \mathbf{B}\lambda) = 0$$

Projection on a vector subspace IV

$$\Leftrightarrow \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix} (x - \mathbf{B}\lambda) = 0$$

$$\Leftrightarrow \mathbf{B}^T(x - \mathbf{B}\lambda) = 0 \Leftrightarrow \mathbf{B}^T x = \mathbf{B}^T\mathbf{B}\lambda \Leftrightarrow \lambda = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T x$$

$(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ is called the pseudo inverse matrix of $\mathbf{B}$

- Find the projection $\pi_u(x) \in \mathbf{U}$

$$\pi_u(x) = \mathbf{B}\lambda = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T x$$

- Find the projection matrix $\mathbf{P}_\pi$ It can easily be deduced that $\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$

Projection on a vector subspace V

Example : Let $\mathbf{U} = span[\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}] \subseteq \mathbb{R}^2$ and $x = \begin{bmatrix}0\\0\\0\end{bmatrix} \in \mathbb{R}^3$

We are looking for the projection of x onto $\mathbf{U}$

Gram–Schmidt method

Gram–Schmidt method I

This method is used to transform a basis $(b_1, \ldots, b_n)$ of a sub-vector space or vector space into an orthogonal or orthonormal basis $(u_1, \ldots, u_n)$. The process iteratively constructs the base $(u_1, \ldots, u_n)$ as follows :

$$u_1 = b_1$$
$$u_k = b_k - \pi_{U^{k-1}}(b_k), k = 2, \ldots, n$$

where $\mathbf{U}^{k-1} = span[u_1, u_2, \ldots, u_{k-1}]$ Also constructed the vectors $u_1, u_2, \ldots, u_k$ are orthogonal
If we want to build an orthonormal basis, we just have to divide each vector by its length

Projection on an affine subspace

Projection on an affine subspace I

Let $\mathbf{L} = x_0 + \mathbf{U}$ be an affine space.
To determine the orthogonal projection $\pi_L(x)$ of x on $\mathbf{L}$, we transform the pb into a projection problem on a vector subspace . To do this, simply subtract the point of the support x_0 from x and from $\mathbf{L}$. Thus, the problem is reduced to the projection of vector $x - x_0$ on the vector subspace $\mathbf{U} = \mathbf{L} - x_0$.
We thus obtain the projection $\pi_u(x - x_0)$ which we can translate to obtain from x to $\mathbf{L}$, adding x_0 to $\pi_u(x - x_0)$:

$$\pi_L(x) = x_0 + \pi_U(x - x_0)$$

Rotations

Rotations I

A rotation is a linear map that performs a rotation of a plane by an angle θ, By convention, we rotate in a counterclockwise direction with $\theta > 0$

Rotation in $\mathbb{R}^2$

Rotation in $\mathbb{R}^2$ I

We consider the canonical basis

$$\{e_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} of \mathbb{R}^2$$

We want to perform a rotation in $\mathbb{R}^2$. we therefore want to find the coordinates of $x \in \mathbb{R}^2$ in a new basis, obtained by changing the basis.
We thus define the rotation matrix $\mathbf{R}(\theta)$ which represents the coordinates of the canonical vectors in the new base obtained by transformation

$$\Phi(e_1) = \begin{pmatrix} cos\theta \\ sin\theta \end{pmatrix}, \ \Phi(e_2) = \begin{pmatrix} -sin\theta \\ cos\theta \end{pmatrix},$$

Rotation in $\mathbb{R}^2$ II

The rotation matrix is then written :

$$\mathbf{R}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$$

Rotation in $\mathbb{R}^3$

Rotation in $\mathbb{R}^3$ I

We can rotate a 2-dimensional plane around the third axis. A simple way to describe this transformation is to specify the images of the vectors of the canonical basis and check that these images are indeed orthonormal

1. Around e_1-axe :

$$\mathbf{R}_1(\theta) = [\Phi(e_1), \Phi(e_2), \Phi(e_3)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & -sin\theta \\ 0 & sin\theta & cos\theta \end{bmatrix}$$

Rotation in $\mathbb{R}^3$ II

2. Around e_2-axe :

$$\mathbf{R}_2(\theta) = \begin{bmatrix} cos\theta & 0 & sin\theta \\ 0 & 1 & 0 \\ sin\theta & 0 & cos\theta \end{bmatrix}$$

3. Around e_3-axe :

$$\mathbf{R}_3(\theta) = \begin{bmatrix} cost\theta & -sin\theta & 0 \\ sin\theta & cost\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$