Mathématiques : CM

Contents

1 Euclidean space - scalar product

Idea : Let E a vector space of finite dimension

The scalar product is defined as :

$$u \in E, v \in E, < u, v >$$

$$<, > = \begin{cases} E \times E & \rightarrow \mathbb{R} \\ (u, v) & \mapsto < u, v > \end{cases}$$

When u and v are orthogonal ($\perp$) we have the following relation :

$$u \perp v \equiv < u, v >= 0$$

We also have the following relation :

$$||u|| = \sqrt{< u, u >}$$

Standard scalar product of $\mathbb{R}^3$:

$$u = (x_1, y_1, z_1) v = (x_2, y_2, z_2)$$

$$< u, v >= x_1 x_2 + y_1 y_2 + z_1 z_2$$

$$\Longrightarrow \ ||u|| = \sqrt{< u, u >} = \sqrt{x_1^2 + y_1^2 + z_1^2}$$

Definition : symmetric bilinear form

Let a mapping such that :

$$\phi : \begin{cases} E \times E & \rightarrow \mathbb{R} \\ (u, v) & \mapsto \phi(u, v) \end{cases}$$

ϕ is a symmetric bilinear form if :

1. It is bilinear : $\forall\, u_0 \in\, E, v \,\mapsto\, \phi(u_0, v)$ is a linear map
2. It is bilinear : $\forall\, v_0 \in\, E, u \,\mapsto\, \phi(u, v_0)$ is a linear map
3. It is symmetric : $\forall\, (u, v) \in\, E^2, \phi(u, v) = \phi(v, u)$ is a linear map

Explanation :

$$\phi = \begin{cases} E \times E & \to \mathbb{R} \\ (u, v) & \mapsto \phi(u, v) \end{cases}$$

Let $u_0 \in E$:

The mapping : $\begin{cases} E & \to \mathbb{R} \\ v & \mapsto \phi(u_0, v) \end{cases}$ is linear

$$\forall\, (v, v') \in E^2, \forall\alpha \in \mathbb{R}, \phi(u_0, \alpha v + v') = \alpha\phi(u_0, v) + \phi(u_0, v')$$

If $v_0 \in E, \forall(u, u') \in E^2, \forall\alpha \in \mathbb{R}, \phi(\alpha u + u', v_0) = \alpha\phi(u, v_0) + \phi(u', v_0)$

Example : standard scalar product of $\mathbb{R}^3$

$$u = (x_1, y_1, z_1)v = (x_2, y_2, z_2)$$

$$\phi(u, v) = x_1x_2 + y_1y_2 + z_1z_2$$

If u is given (considered as a constant), then the mapping :

$v \,\mapsto\, x_1x_2 + y_1y_2 + z_1z_2$ is a linear map.

Note that the coordinates of u are constant numbers.

Those of v are variables.

for example if $u = (1, 2, 3)$, then $v = (x_2, 2y_2, 3z_2)$

Similarly :, if $v = (x_2, y_2, z_2)$ is given, then the mapping :

$u \mapsto x_1x_2 + y_1y_2 + z_1z_2$ is a linear map .

Note that the coordinates of v are being constant numbers.

Those of u are variables.

Furthermore, $\phi(u, v) = \phi(v, u)$

A symmetric bilinear form ϕ is positive definite if :

$$\forall u \in E, \phi(u, u) \geq 0 \wedge \phi(u, u) = 0 \implies u = 0_E$$

Example 1 : Standard scalar product of $\mathbb{R}^3$

$$\phi(u, u) = x_1^2 + y_1^2 + z_1^2 \geq 0$$

and

$$\phi(u, u) = 0 \implies \begin{cases} x_1^2 = 0 \\ y_1^2 = 0 \\ z_1^2 = 0 \end{cases} \implies u = 0_E$$

The standard scalar product is thus positive definite

Example 2 :

$$\phi(u, v) = x_1x_2 + y_1y_2 - z_1z_2 \text{ is a symmetric bilinear form}$$

If $u = (x_1, y_1, z_1)$ is given (considered as a constant), then the mapping :

$v \mapsto x_1x_2 + y_1y_2 + z_1z_2$ is a linear map.

Note that the coordinates of u are constant numbers.

Those of v are the variables.

Thus the coordinates of u are the coefficient of the linear function.

Similarly, if v is given and constant, $u \mapsto \phi(u, v)$ is linear

$$\implies \phi(u, v) = \phi(v, u)$$

Positive definite ? No

$$\phi(u,u)x_1^2 + y_1^2 - z_1^2$$

If $u = (1,0,0), \phi(u,u) = 1 > 0$

If $u = (0,0,1), \phi(u,u) = -1 < 0$

Example 3 :

$$\phi(u,v) = x_1x_2 + y_1y_2\phi(u,u) = x_1^2 + y_1^2 \geq 0$$

But $\phi(u,u) = 0 \not\Longrightarrow u = 0_E$ as for $u = (0,0,1), \phi(u,u) = 0, u \neq 0_E$

A scalar product is a symmetric bilinear form, which is positive definite.

Then we note $< u,v >$ instead of $\phi(u,v)$

We can define for $u \in E$,

$$||u|| = \sqrt{< u,v >} \implies \sqrt{< u,v >} \in \mathbb{R} \text{ because } < u,v > \geq 0$$

$$||u|| = 0 \implies u = 0_E$$

Theorem : Matrix representation of a symmetric bilinear form :

Let $\mathcal{B} = (e_1, ..., e_n)$ be a basis of E.

u has coordinates : $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ ... \end{pmatrix}$ in $\mathcal{B}$

v has coordinates : $\begin{pmatrix} x_2 \\ y_2 \\ z_2 \\ \dots \end{pmatrix}$ in $\mathcal{B}$

Then : $\exists\, A \in \mathbb{M}_n(\mathbb{R}), \phi(u,v) = {}^{+}X_1 A X_2$ is symmetric

Examples :

Standard scalar product of $\mathbb{R}^3$:

Example 1 :

$\mathcal{B}$ standard basis of $\mathbb{R}^3$

$u = (x_1, y_1, z_1)$ has coordinates $X_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$

$v = (x_2, y_2, z_2)$ has coordinates $X_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$

$${}^{T}X_1 X_2 = (x_1, y_1, y_2) \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

$$= x_1x_2 + y_1y_2 + z_1z_2 \implies A = I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$${}^{T}X_1 A X_2 = x_1x_2 + y_1y_2 + z_1z_2 =< u, v >$$

Example 2 :

$$\phi(u,v) = x_1x_2 + y_1y_2 - z_1z_2$$

$$\text{Then : } A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$^{T}X_1AX_2 = (x_1, y_1, z_1)\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{pmatrix}\begin{pmatrix}x_2\\ y_2\\ z_2\end{pmatrix}$$

$$= (x_1, y_1, z_1)\begin{pmatrix}x_2\\ y_2\\ -z_2\end{pmatrix}$$

$$= x_1x_2 + y_1y_2 - z_1z_2$$

Example 3 :

$$\phi(u, v) = x_1x_2 + y_1y_2$$

$$\text{Then : } A = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}$$

$$^{T}X_1AX_2 = (x_1, y_1, z_1)\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}x_2\\ y_2\\ z_2\end{pmatrix}$$

$$= (x_1, y_1, z_1)\begin{pmatrix}x_2\\ y_2\\ 0\end{pmatrix}$$

$$= x_1x_2 + y_1y_2$$

1.1 Spectral Theorem

Spectral Theorem : all symmetric matrix A is diagonalizable.

We can choose an orthonormal eigenbasis $(\epsilon_1, ..., \epsilon_n)$.

Then the transition matrix P satisfies to :

$$^{T}PP = P \times {}^{T}P = I \Leftrightarrow P^{-1} = {}^{T}P$$

Note : if the eigenbasis is orthonormal, then :

$$< \epsilon_i, \epsilon_j >= \begin{cases} 1 \text{ if } i = j \\ 0 \text{ otherwise} \end{cases}$$

1.2 Theorem

Theorem : Let E = $\mathbb{R}^n$, let A be a symmetric matrix in $\mathcal{M}_n(\mathbb{R})$.

Let us define $\phi(u, v) = {}^{T}X_1 A X_2$.

ϕ is a scalar product $\Leftrightarrow$ All eigenvalues of A are ¿ 0.

Explanation in a an orthogonal eigenbasis $\mathcal{B}'$, $\phi(u, v) = {}^{T}X_1' D X_2'$

Where D is a diagonal matrix.

2 Orthogonal Projection

Let V_0 be the orthogonal projection of V onto F.

V_0 is the element of F the closest to V.

2.1 Motivation

Assume you must find $\lambda \in \mathbb{R}$ such that :

$$v = \lambda u$$

The "best solution" is λ such that $v_0 = \lambda u$

Ex: $u = (1, 1), v = (2, 1)$

Solving $v = \lambda u$: $\begin{cases} \lambda = 2 \\ \lambda = 1 \end{cases}$

$$v_0 = \alpha u = (\alpha, \alpha)$$
$$(v - v_0) \perp u$$
$$(2 - \alpha, 1 - \alpha) \perp (1, 1)$$
$$\implies (2 - \alpha) \times 1 + (1 - \alpha) \times 1 = 0$$
$$\implies 3 - 2\alpha = 0 \implies \alpha = \frac{3}{2}$$
$$\implies v_0 = \frac{3}{2}(1, 1) = (\frac{3}{2}, \frac{3}{2})$$

The solution does not have a solution.

However, we can determine v_0.

v_0 represent the solution with the samellest range of error.

2.2 Definition

Let $(E, <, >)$ be an euclidean space.

Let F be a linear subspace of E.

$$F^{\perp} = \{u \in E, \forall\, v \in F, < u, v >= 0\}$$

2.3 Theorem

If F admits a basis $\mathcal{B} = (e_1, e_2, ..., e_n)$, then $F^{\perp} = \mathcal{B}^{\perp}$

$$F^{\perp} = \{u \in E, \begin{cases} < u, e_1 >= 0 \\ < u, e_2 >= 0 \\ ... \\ < u, e_n >= 0 \end{cases} \}$$

2.4 Theorem

If F is finite-dimensional, then $F \oplus F^{\perp} = E$

$$\Leftrightarrow \forall\, u \in E, \exists!(v, w) \in F \times F^{\perp}, u = v + w$$

Then v is the orthogonal projection of u onto F.

2.5 Property

Let $u \in E$ and let $v_0 = P_F(u)$.

P_F is the orthogonal projection onto F.

v_0 is the element of F which minimizes $||u - v||^2$.

It is the solution of the optimization :

$$Min(||u - v||),\ \text{constraint}:\ v \in F$$

2.6 Proof

Let $u = v_0 + w$ where $(v_0, w) \in F \times F^{\perp}$.

Then $\forall\, v \in F$:

$$||u - v||^2 =< u - v, u - v > \implies u - v = (u - v_0) + (v_0 - v)$$

Indeed, $(u - v_0) = w \in F^{\perp}$ and $(v_0 - v) \in F$ as $(v_0, v) \in F^2$

$$\implies ||u - v||^2 - < u - v, u - v >=< (u - v_0) + (v_0 - v), (u - v_0) + (v_0 - v) >$$

$$=< (u - v_0), (u - v_0) > + 2 < (u - v_0), (v_0 - v) > + < (v_0 - v), (v_0 - v) >$$

$$= ||u - v_0||^2 + ||v_0 - v||^2 \geq ||u - v_0||^2$$

$$P_F(u) \in F \Leftrightarrow \exists (\lambda_1, ..., \lambda_n) \in \mathbb{R}^n, u = \lambda_1 e_1 + ... + \lambda_n e_n$$

$$u - P_F(u) \in F^{\perp} \Leftrightarrow \begin{cases} < (u - \lambda_1 e_1 + ... + \lambda_n e_n), e_1 >= 0 \\ ... \\ < (u - \lambda_1 e_1 + ... + \lambda_n e_n), e_n >= 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} < (\lambda_1 < e_1, e_1 > +... + \lambda_n < e_n, e_n >), e_1 >=< u, e_1 > \\ ... \\ < (\lambda_1 < e_1, e_1 > +... + \lambda_n < e_n, e_n >), e_n >=< u, e_n > \end{cases}$$

2.7 Finding $P_F(u)$

Give F admits a basis $\mathcal{B_F} = (\epsilon_1, ..., \epsilon_n)$

$$\text{We write } P_F(u) \in F \implies u - P_F(u) \in F^{\perp}$$

$$\implies u = P_F(u) + (u - P_F(u))$$

$$P_F(u) \in F \Leftrightarrow \exists(\alpha_1, ..., \alpha_n) \in \mathbb{R}^n, P_F(u) = \alpha_1\epsilon_1 + ... + \alpha_n\epsilon_n$$

$$u - P_F(u) \in F^{\perp} \Leftrightarrow \begin{cases} < u - P_F(u), \epsilon_1 >= 0 \\ ... \\ < u - P_F(u), \epsilon_n >= 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} < u - \alpha_1\epsilon_1 + ... + \alpha_n\epsilon_n, \epsilon_1 >= 0 \\ ... \\ < u - \alpha_1\epsilon_1 + ... + \alpha_n\epsilon_n, \epsilon_n >= 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} < \alpha_1 < \epsilon_1, \epsilon_1 > +... + \alpha_n < \epsilon_n, \epsilon_n >, \epsilon_1 >=< u, \epsilon_1 > \\ ... \\ < \alpha_1 < \epsilon_1, \epsilon_1 > +... + \alpha_n < \epsilon_n, \epsilon_n >, \epsilon_n >=< u, \epsilon_n > \end{cases}$$

3 Data Analysis

Example of a dataset :

Student	Class	Average	Sport	French	Math	Physics	English
Student1	A1	1	SP1	FR1	MATH1	PHY1	ENG1
Student2	B2	1	SP2	FR2	MATH2	PHY2	ENG2
Student3	C1	1	SP3	FR3	MATH3	PHY3	ENG3
Student4	B2	1	SP4	FR4	MATH4	PHY4	ENG4

Table 1: Table of marks average in a high school.

The question that we adress:

- Consider the student good in maths. Are they good in physics ?

We try to find a relation :

$$PHY = \alpha MAT + \beta + Err_1$$

And compare the error with :

$$PHY = \beta + Err_2$$

The "best" β is the Average of PHY column.

Is $||Err_1|| < ||Err_2||$?

If yes, is $\alpha > 0$?

For the model :

$$PHY = \alpha MAT + \beta + Err_1$$

we look for the best solution of :

$$(S) : \begin{cases} PHY_1 = \alpha MAT_1 + \beta \\ PHY_2 = \alpha MAT_2 + \beta \\ ... \end{cases}$$

$$\Leftrightarrow PHY = \alpha MAT + \beta \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix}$$

Where PHY and MAT are the whole column.

Optimal values : α_0 and β_0 such that $\alpha_0 MAT + \beta_0 \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix}$ is the orthogonal projection of PHY on $Span(Mat, \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix})$

- The students good at maths and sport, do they tend to be even better than those good only at maths ?

$$PHY = \alpha MAT + \beta \begin{pmatrix} 1 \\ 1 \\ ... \\ 1 \end{pmatrix} + \gamma SP + Err_3$$

Find $\alpha_0, \beta_0, \gamma_0$ optimal :

Is $||Err_3|| < ||Err_1||$?

If yes, sign of β ?

Let X be a column of the table.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix}$$

X is an identically distributed random variable and independant.

Empirical expectation of X :

$$\hat{\mathbb{E}}(X) = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{{}^T\begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix} .X}{n}$$

Empirical variance of X :

$$\widehat{\text{Var}}(X) = \hat{\mathbb{E}}((X - \hat{\mathbb{E}}(X))^2)$$

$$= \frac{(x_1 - \hat{\mathbb{E}}(X))^2 + (x_2 - \hat{\mathbb{E}}(X))^2 + ... + (x_n - \hat{\mathbb{E}}(X))^2}{n}$$

Let $X_c = X - \hat{\mathbb{E}}(X) \begin{pmatrix} 1 \\ ... \\ 1 \end{pmatrix}$

$$X_c = \begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) \\ x_2 - \hat{\mathbb{E}}(X) \\ \ldots x_n - \hat{\mathbb{E}}(X) \end{pmatrix}$$

X_c is the centered version of X.

$$\text{Thus } \widehat{\text{Var}} = \frac{(x_1 - \hat{\mathbb{E}}(X))^2 + (x_2 - \hat{\mathbb{E}}(X))^2 + ... + (x_n - \hat{\mathbb{E}}(X))^2}{n}$$

$$= \frac{1}{n} \begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) & x_2 - \hat{\mathbb{E}}(X) & ... & x_n - \hat{\mathbb{E}}(X) \end{pmatrix} \begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) \\ x_2 - \hat{\mathbb{E}}(X) \\ ... \\ x_n - \hat{\mathbb{E}}(X) \end{pmatrix}$$

$$= \frac{{}^T X_c X_c}{n} = \frac{||X_c||^2}{n}$$

Empirical standard deviation of X :

$$\hat{\sigma}(X) = \sqrt{\widehat{\mathrm{Var}}(X)} = \sqrt{\frac{{}^{T}X_cX_c}{n}} = \frac{||X_c||}{\sqrt{n}}$$

Consider two columns $X = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ ... \\ y_n \end{pmatrix}$

Empirical covariance of X and Y :

$$\widehat{\mathrm{Cov}}(X,Y) = \hat{\mathbb{E}}((X - \hat{\mathbb{E}}(X)) \times (Y - \hat{\mathbb{E}}(Y)))$$

$$= \frac{1}{n}(x_1 - \hat{\mathbb{E}}(X))(y_1 - \hat{\mathbb{E}}(Y)) + ... + (x_n - \hat{\mathbb{E}}(X))(y_n - \hat{\mathbb{E}}(Y))$$

$$= \frac{1}{n}\begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) & x_2 - \hat{\mathbb{E}}(X) & ... & x_n - \hat{\mathbb{E}}(X) \end{pmatrix} \begin{pmatrix} y_1 - \hat{\mathbb{E}}(Y) \\ y_2 - \hat{\mathbb{E}}(Y) \\ ... \\ y_n - \hat{\mathbb{E}}(Y) \end{pmatrix}$$

$$= \frac{{}^{T}X_cY_c}{n}$$

$$\frac{1}{n} < X_c, Y_c >$$

Empirical correlation coefficient of X and Y :

$$\hat{c}(X,Y) = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\hat{\sigma}(X)\hat{\sigma}(Y)}$$

Note that $\widehat{\mathrm{Cov}}(X,X) = \widehat{\mathrm{Var}}(X)$ and $\widehat{\mathrm{Cov}}(Y,Y) = \widehat{\mathrm{Var}}(Y)$

Let $D = \text{Dataset} = \begin{pmatrix} X & Y \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ \dots & \dots \\ x_n & y_n \end{pmatrix}$

$D_c = \text{centered data set} = \begin{pmatrix} X_c & Y_c \end{pmatrix}$

The Empirical covariance matrix of D is :

$$\widehat{\mathrm{Cov}}(D) = \begin{pmatrix} \widehat{\mathrm{Var}}(X) & \widehat{\mathrm{Cov}}(X,Y) \\ \widehat{\mathrm{Cov}}(X,Y) & \widehat{\mathrm{Var}}(Y) \end{pmatrix} = \frac{{}^T D_c D_c}{n}$$

$${}^T D_c D_c = \begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) & ... & x_n - \hat{\mathbb{E}}(X) \\ y_1 - \hat{\mathbb{E}}(Y) & ... & y_n - \hat{\mathbb{E}}(Y) \end{pmatrix} \begin{pmatrix} x_1 - \hat{\mathbb{E}}(X) & y_1 - \hat{\mathbb{E}}(Y) \\ ... & ... \\ x_n - \hat{\mathbb{E}}(X) & y_n - \hat{\mathbb{E}}(Y) \end{pmatrix}$$

$$= \begin{pmatrix} {}^T X_c X_c & {}^T X_c Y_c \\ {}^T Y_c X_c & {}^T Y_c Y_c \end{pmatrix}$$

Ex : find $\widehat{\mathrm{Var}}(aX + bY)$ where $aX + bY = D \begin{pmatrix} a \\ b \end{pmatrix}$

$$D \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \cdots & \cdots \\ x_n & y_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + ny_2 \\ \cdots \\ ax_n + by_n \end{pmatrix} = aX + bY$$

$$\widehat{\mathrm{Var}}(X) = {}^T X_c X_c$$

$$\widehat{\mathrm{Var}}(D \begin{pmatrix} a \\ b \end{pmatrix}) = (D_c \begin{pmatrix} a \\ b \end{pmatrix})(D_c \begin{pmatrix} a \\ b \end{pmatrix}) \times \frac{1}{n}$$

$$= \begin{pmatrix} a & b \end{pmatrix} {}^T D_c D_c \begin{pmatrix} a \\ b \end{pmatrix} \times \frac{1}{n}$$

$$= \begin{pmatrix} a & b \end{pmatrix} \widehat{\mathrm{Cov}}(D) \begin{pmatrix} a \\ b \end{pmatrix}$$