Transformations Affines

Cette feuille est centrée autour des applications affines de $\mathbb{R}^n$, plus particulièrement autours des projections et transformations de l'espace affine euclidien. L'objectif est de vous apporter les éléments de langage et les résultats de structure qui vous permettent de les expliciter et les utiliser dans les contextes de ML. Pour rappel la majeure partie des algorithmes de ML sont de nature géométrique.

Nous allons en passant en profiter pour faire quelques rappels de calcul matriiel ainsi que d'algèbre linéaire quand nécessaire.

1 Pivot de Gauss

On se donne une matrice $M \in \mathcal{M}_{m,n}(\mathbb{R})$. Une opération élémentaire sur M est une des deux opérations suivantes :

- pour $i \in \{1, \ldots, m\}$ et $j \in \{1, \ldots, n\}$ interchanger les lignes L_i et L_j
$$L_i \leftrightarrow L_j$$
- pour $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$, $i \neq j$, $\lambda_i \in \mathbb{R}^*$ et $\lambda_i \in \mathbb{R}$, remplacer la ligne L_i par $\lambda_i L_i + \lambda_j L_j$:
$$L_i \leftarrow \lambda_i L_i + \lambda_j L_j.$$

Le pivot de Gauss est un algorithme de transformation d'une matrice M en matrice triangulaire supérieur et éventuellement en matrice identité, dans le but d'inverser une matrice, de calculer un déterminent ou de résoudre un système linéaire. Il est d'utilisation constante en calcul numérique. [1]

1.1 Opérations Élémentaires

On traduit par la suite les opérations élémentaires précédentes à l'aide du produit matriciel.

Question 1-1. Trouver des matrices $P(i, j)$ et $U(i, j, \lambda, \nu)$ dont la multiplications avec M réalise les opérations élémentaires précédentes.

Question 1-2. Trouver les matrices $P_t(i, j)$ et $U_t(i, j, \lambda, \nu)$ qui réalisent les opérations précédentes sur les colonnes.

1.2 Systèmes Linéaires

Question 1-3. Déduire de la section 1.1 que les opérations élémentaires du pivot de Gausse transforme tout système linéaire dont M est une matrice (si appliqué des deux côtés de l'égalité) en un système linéaire ayant les mêmes solutions.

Question 1-4. Comment procéder pour décrire toute les solutions d'un système linéaire ?

1. L'expression de l'inverse d'une matrice inversible à l'aide des cofacteurs est à prohiber ! Elle est théoriquement intéressante et particulièrement esthétique mais numériquement inéfficace.

Question 1-5. Discuter de l'efficacité de l'implémentation de l'algorithme

1.3 Déterminent

Question 1-6. Retrouver les relations standards sur le déterminent d'une matrice qui subit une opération élémentaire à l'aide de la formule

$$\det(AB) = \det(A)\det(B).$$

Question 1-7. Comment procéder pour calculer le déterminent d'une matrice à l'aide du pivot de Gauss ?

1.4 Matrices Équivalentes

Deux matrices A et B sont dites équivalentes s'il existe deux matrices inversibles P et Q telle que

$$A = PBQ.$$

Question 1-8. Justifier le fait que toute matrice est équivalente à une matrice ayant un bloc identité et des zéros partout ailleurs.

2 Applications Affines

Cette section est un kit de survie en milieu hostile. Il s'agit de s'armer des différentes notions de transformations et applications affines dont vous aurez l'usage lors de vos cours de ML. Notre démarche dans la suite est particulièrement pragmatique[2] ; on se limite au cas des applications affines de $\mathbb{R}^n$ dans $\mathbb{R}^m$.

Définition 2.1. Une application $f : \mathbb{R}^n \to \mathbb{R}^m$ est dite affine s'il existe un point $\boldsymbol{a} \in \mathbb{R}^m$ tel que $f - \boldsymbol{a}$ soit une application linéaire.

Cette définition formelle nécessite une réalisation plus concrète afin qu'on puisse l'utiliser. Pour cela on va introduire une généralisation de la notion de bases dans le cas de $\mathbb{R}^n$ qui nous permettra de se placer en un point quelconque pour y travailler.

Définition 2.2. Un repère de l'espace affine $\mathbb{R}^n$ est la donnée d'un point $\boldsymbol{o}$ appelé origine du repère et d'une base $\boldsymbol{v} = (v_1, \ldots, v_n)$ de $\mathbb{R}^n$. On note $(\boldsymbol{o}, \boldsymbol{v})$ une telle donnée.

Le repère canonique de $\mathbb{R}^n$ est le repère $(\underline{0}, e_1, \ldots, e_n)$, que vous aviez eu l'habitude de manipuler dans le secondaire et en physique. Du moins dans les cas de dimensions 1 ou 2.

Définition 2.3. Soient $(\boldsymbol{o}, \boldsymbol{v})$ et $(\boldsymbol{l}, \boldsymbol{w})$ deux repères respetivement de $\mathbb{R}^n$ et $\mathbb{R}^m$. L'écriture de f dans les repères $(\boldsymbol{o}, \boldsymbol{v})$ et $(\boldsymbol{l}, \boldsymbol{w})$ est la donnée d'une matrice M telle que pour tout $x \in \mathbb{R}^n$,

$$f(x) = M(x - \boldsymbol{o})_{\boldsymbol{v}} + f(\boldsymbol{o}). \tag{1}$$

où $\bullet_{\boldsymbol{v}}$ désigne l'écriture du vecteur $\bullet$ dans la base $\boldsymbol{v}$. Le membre de droite de l'équation 1 est sous-entendu être décrit dans le repère $(\boldsymbol{l}, \boldsymbol{w})$. Cette équation est parfois appelée équation de f dans les repères $(\boldsymbol{o}, \boldsymbol{v})$ et $(\boldsymbol{l}, \boldsymbol{w})$.

2. Elle pourrait heurter certaines âmes sensibles et quelques matheux qu'on garder un peu de sens esthétique.

Exemple 2.1. Une application affine peut être donné par une experssion dans le repère canonique

$$x \mapsto Mx + b$$

où x et b sont des vecteurs respectivement dans $\mathbb{R}^n$ et $\mathbb{R}^m$ et $M \in \mathcal{M}_{m,n}(\mathbb{R})$. Par exemple

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

Ce qui nous donne une fonction de $\mathbb{R}^2$ dans $\mathbb{R}^2$ affine en x et y au sens qu'on entend depuis tout petit.

Dans le but de trouver des expressions agréables ou standardisées d'une application affine on est amené à écrire ces applications affines dans des repères adaptées. Pour pouvoir exprimer un tel contexte il nous faut être en mesure de formaliser la notion de changement de repères.

Définition 2.4. Le changement de repères de $\mathbb{R}^n$ de $(\boldsymbol{o}, \boldsymbol{v})$ vers $(\boldsymbol{l}, \boldsymbol{w})$ est l'application affine donnée dans les repères précédents par

$$x \longmapsto P_{\boldsymbol{v}}^{\boldsymbol{w}}(x - \boldsymbol{o})_{\boldsymbol{v}} + \boldsymbol{l}$$

où $P_{\boldsymbol{v}}^{\boldsymbol{w}}$ est la matrice de passage de la base $\boldsymbol{v}$ vers la base $\boldsymbol{w}$; c'est-à-dire la matrice des vecteurs de $\boldsymbol{v}$ écrits dans la base $\boldsymbol{w}$.

Voici enfin comment formaliser le passage de l'écriture d'une application dans un repère vers un autre.

Proposition 2.1. *Soient $(\boldsymbol{o}, \boldsymbol{v})$ et $(\boldsymbol{o'}, \boldsymbol{v'})$ deux bases de $\mathbb{R}^n$, $(\boldsymbol{l}, \boldsymbol{w})$ et $(\boldsymbol{l'}, \boldsymbol{w'})$ deux bases de $\mathbb{R}^m$. On note M la matrice d'une application affine f dans les repères $(\boldsymbol{o}, \boldsymbol{v})$ et $(\boldsymbol{l}, \boldsymbol{w})$, et M' celle de f dans les repères $(\boldsymbol{o'}, \boldsymbol{v'})$ et $(\boldsymbol{l'}, \boldsymbol{w'})$. Si $P_{\boldsymbol{v'}}^{\boldsymbol{v}}$ désigne la matrice de passage de $\boldsymbol{v}$ vers $\boldsymbol{v'}$ et $P_{\boldsymbol{w'}}^{\boldsymbol{w}}$ la matrice correspondante dans le cas de $\boldsymbol{w}$ et $\boldsymbol{w'}$ alors pour tout $x \in \mathbb{R}^n$*

$$M'(x - \boldsymbol{o'})_{\boldsymbol{v'}} + f(\boldsymbol{o'}) = f(x) = P_{\boldsymbol{w'}}^{\boldsymbol{w}} M \left(P_{\boldsymbol{v'}}^{\boldsymbol{v}}\right)^{-1} (x - \boldsymbol{o})_{\boldsymbol{v}} + f(\boldsymbol{o})$$

Question 2-9. Écrire la fonction du premier exemple dans le repère de $\mathbb{R}^2$ donné au départ et à l'arrivée par $\big((0,0), \{(1,1), (1,-1)\}\big)$.

Remarque 1. Dans les faits, on ne fera pas souvent à faire ce type de changement de repères. À partir des propriétés des applications à l'étude on sera à même de trouver un bon repère dans lesquels les décrire. Il reste que de tels changements de repères doivent être compris et faits quand nécessaire.

Exemple 2.2. Les applications affines constantes sont celles dont l'expression dans tous les repères a une matrice qui est nulle.

Exemple 2.3. Avec la définition qu'on a donné toute application linéaire va décrire une application affine.

2.1 Translations et Homothéties

On s'attarde un instant sur les plus simples des applications affines non triviales.

Exemple 2.4. La translation de vecteur b est l'application affine donnée dans tous repère $(\boldsymbol{o}, \boldsymbol{v})$ de $\mathbb{R}^n$ par $x \mapsto (x - \boldsymbol{o})_{\boldsymbol{v}} + b$.

Exemple 2.5. L'***homothétie*** de rapport λ et centre $\boldsymbol{o}$ est l'application affine de $\mathbb{R}^n$ dans lui-même dont l'écriture dans le repère $(\boldsymbol{o}, \boldsymbol{v})$ prend la forme

$$x \mapsto \lambda(x - \boldsymbol{o})_{\boldsymbol{v}} + \boldsymbol{o}$$

L'homothétie de rapport 1 est simplement l'identité.

Question 2-10. Donner une particularité que partagent les homothéties et les translations.

Question 2-11. Que donne la composition de deux translations ? Celle de deux homothéties ?

Question 2-12. Dessiner l'image du carré de sommets $(\pm 1, \pm 1)$ dans $\mathbb{R}^2$ par les homothéties suivantes :
- de rapport 2 et de centre l'origine
- de rapport 2 et de centre $(1, 0)$.

2.2 Projections Affines

Les projections affines sont la généralisation des projections linéaires. Pour rappel une projection $p : E \to E$ de l'espace vectoriel E sur lui même est un application linéaire qui satisfait la relation $p^2 = p$. Cette relation garanti le fait que $G = \mathrm{Ker}(p)$ et $F = \mathrm{Ker}(p - \mathrm{id})$ sont des espaces vectoriels supplémentaires dans E. Ainsi, pour tout $x = x_F + x_G$, la projection p est définie par $p(x) = x_F$. L'application p est appelée projection sur F parallèlement à G.

Définition 2.5. Une application affine f de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est une projection affine si sa matrice dans un repère est la matrice d'une projection linéaire.

Une projection affine, dans un repère $(\boldsymbol{o}, \boldsymbol{v})$ de $\mathbb{R}^n$, s'écrit donc sous la forme

$$x \mapsto M(x - \boldsymbol{o})_{\boldsymbol{v}} + \boldsymbol{o}$$

On voit dans cette écriture que $\boldsymbol{o}$ est fixé par la projection. On vient d'écrire la projection de $\boldsymbol{o} + \mathrm{Ker}(M - I_n)$ parallèlement à $\boldsymbol{o} + \mathrm{Ker}(M)$.

Question 2-13. Étudier des exemples de projections affines dans $\mathbb{R}^2$. Dessiner systématiquement les droites affines qui caractérisent cette projection.

3 Transformations Orthogonales

3.1 Aparté sur l'Orthogonalité

On se donne un produit scalaire $\langle \cdot, \cdot \rangle$ sur $\mathbb{R}^n$ [3]. On rappelle que c'est une forme bilinéaire symétrique définie positive. Elle vient avec une norme associée dite norme euclidienne définie par

$$\|x\| = \sqrt{< x, x >}.$$

Définition 3.1. Deux vecteurs x, y in $\mathbb{R}^n$ sont dit orthogonaux pour $\langle \cdot, \cdot \rangle$ si $\langle x, y \rangle = 0$.

Nous avons vu que cette notion correspond dans le cas du produit scalaire usuel au fait que l'angle entre x et y est $\pi/2$ $[\pi]$.

Définition 3.2. Une base $\boldsymbol{v}$ est dite ***orthognale*** si les vecteurs qui la constituent sont orthogonaux deux à deux. Elle est ***orthonormale*** si ses vecteurs sont de norme 1.

L'exemple le plus simple d'une base orthonormée est celui de la base canonique de $\mathbb{R}^n$ pour le produit scalaire usuel.

3. Pensez au produit scalaire usuel si ça vous facilite la vie.

Question 3-14. Donner d'autres bases orthonormales de $\mathbb{R}^3$.

Question 3-15. Comment construire une base orthonormée de $\mathbb{R}^3$ à partir de la base

$$\{(1,0,1),(0,1,1),(1,1,0)\}?$$

Définition 3.3. Étant donné une partie K de $\mathbb{R}^n$ on appelle orthogonal de K pour $\langle\cdot,\cdot\rangle$ le sous-espace vectoriel

$$K^{\perp} = \{x \in \mathbb{R}^n \mid \forall y \in K,\ \langle x, y\rangle = 0\}.$$

Ainsi un hyperplan H de $\mathbb{R}^n$ quand il est donné implicitement est décrit comme l'orthogonal d'un vecteur de $\mathbb{R}^n$ qu'on appelle un vecteur ***normal***.

Question 3-16. Décrire dans $\mathbb{R}^3$ l'orthogonal à la partie $\{(1,1,1),(1,0,1)\}$.

3.2 Transformations vectorielles

La notion de produit scalaire permet à la fois de formaliser la notion d'angle ainsi que celle de distance [4]. Les transformations affines de $\mathbb{R}^n$, c'est-à-dire les applications affines bijectives, qui présèrvent les angles et les distances sont d'un grand intérêt géométrique ; elles ne touchent pas aux dispositions relatives de configurations de parties dans $\mathbb{R}^n$ et donc présèrvent les propriétés géométriques de celles-ci. Elles présèrvent par exemple la perpendicularité ou le parallélisme. Pour l'instant on se limite au cas vectoriel avant d'exposer la situation affine, plus générale.

Avant de poursuivre notre travail on prend un instant pour étudier de plus prêt une écriture matricielle de $\langle\cdot,\cdot\rangle$. Étant donné une base $\boldsymbol{v}$ de $\mathbb{R}^n$, on appelle matrice de $\langle\cdot,\cdot\rangle$ dans la base $\boldsymbol{v}$ la matrice

$$\boldsymbol{A} = \left(\langle v_i, v_j\rangle\right)_{i,j}.$$

Question 3-17. Montrer que, dans la base v, on a pour tout x, $y \in \mathbb{R}^n$, $\langle x, y\rangle = x^T A y$.

Si $\boldsymbol{v}$ est une base orthonormale dans ce cas $A = I$ et on a, pour tout x, $y \in \mathbb{R}^n$ écrits dans cette base

$$\langle x, y\rangle = x^T y.$$

Ce qui nous ramène à l'écriture usuelle du produit scalaire.

Hypothèse 3.1. Quitte à se ramener à une base orthonormée appropriée [5] on peut supposer que notre produit scalaire a la forme usuelle.

Soit M une matrice inversible dans une base orthonormée de $\mathbb{R}^n$. Dire que M préserve le produit scalaire $\langle\cdot,\cdot\rangle$ signifie que pour tout x, $y \in \mathbb{R}^n$ on a

$$(Mx)^T(My) = x^T y.$$

D'où l'on obtient, pour tout x, $y \in \mathbb{R}^n$

$$x^T\left(M^T M - I_n\right) y = 0.$$

En prenant x et y ayant des coefficients nuls sauf un on peut montrer que les coefficients de $M^T M - I_n$ sont tous nuls. Donc

$$M^T M = I_n.$$

4. On étudiera plus en détail la seconde question un peu plus tard.
5. Ça existe toujours. Voir algorithme de Gram-Schmidt.

Proposition 3.2. *Une isomorphisme de $\mathbb{R}^n$ préserve un produit scalaire si et seulement sa matrice dans une base orthonormée satisfait*

$$M^T M = I_n.$$

Une matrice qui satisfait la propriété précédente est dite ***orthogonale****.*

Question 3-18. Quel peut être le déterminant d'une matrice orthogonale ?

Question 3-19. Traduire la proposition précédente en fonction des vecteurs colonnes de la matrice M.

On étudie un premier exemple de matrices orthogonales. On se fixe une base orthonormée de $\mathbb{R}^n$ dans laquelle on travaille. Soit $\theta \in \mathbb{R}$ on note $R(\theta)$ la matrice

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Question 3-20. Montrer que $R(\theta)$ est une matrice orthogonale pour tout θ. Quel est son déterminant ?

Question 3-21. Dessiner l'image du carré de points extrémaux $(\pm 1, \pm 1)$ par $R(\theta)$ pour $\theta = 0$, $\pi/4$.

Un exemple de matrice orthogonale qui n'est pas de déterminant 1 est celui-ci

$$s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Question 3-22. Interpréter géométriquement cette matrice.

Question 3-23. Dessiner l'image du carré de points extrémaux $\{(0,0),(0,1),(1,0),(1,1)\}$ par $R(\theta)$ pour $\theta = 0$, $\pi/4$.

Question 3-24. Déterminer toute les matrices orthogonales de $\mathbb{R}^2$ muni de son produit scalaire usuel. Vous pourrez les étudier suivant leurs spectres.

Question 3-25. Donner des exemples de matrices orthogonales en toute dimension. Essayez d'être le plus général possible.

Question 3-26. Étendre la classification des matrices orthogonales au cas de dimension 3.

3.3 Transformations Affines

TODO