Nom et prenom : Anass El Moubaraki
Domaine : Apprentissage statistique.

Quelques stratégies pour améliorer la complexité temporelle de l'algorithme d'évaluation de modèles statistiques prédictives à partir de la mesure Triplet

Abstract :

Ce travail a pour objectif de développer des techniques pour diminuer la complexité d'implémentation d'une mesure de dissimilarité définie sur l'ensemble des matrices à valeurs dans $\mathbb{R}$ à savoir la mesure Triplet. L'utilisation de cette dernière est assez fréquente dans l'apprentissage statistique en général notamment pour la validation de modèles prédictives à partir d'une vérité terrain.

Dans cet article, nous commencerons par présenter brièvement les limites des mesures de dissimilarité classiques comme la ***MSE****, la perte 0-1 pour des problèmes d'apprentissage. Dans un deuxième temps, nous présenterons les avantages de la mesure Triplet vis à vis de l'interprétation de la pertinence des estimateurs et fournirons des méthodes pour implémenter cette dernière en complexité réduite. Nous verrons aussi la nécessité de combiner différentes métriques pour choisir un estimateur pertinent.*

Table des matières

1 Introduction

De nos jours, l'utilisation de l'IA en général et du machine learning en particulier est devenue indispensable pour traiter des données qui affluent à grande vitesse et gagner du temps sur l'étude d'une variable d'intérêt. Cependant ce traitement de données se doit d'être robuste et dans le cas de l'apprentissage supervisé où nous disposons à priori d'une vérité terrain, la validation d'un modèle prédictive se fait par calcul d'une métrique ou une mesure de dissimilarité dans un cadre plus général.

En guise d'illustration, pour des labels que nous présenterons par un vecteur de taille N et que nous notons y et que nous supposons à valeurs dans un espace euclidien une manière de procéder serait de calculer une distance euclidienne entre le vecteur de prédiction $\hat{y}$ et y. Cette dernière est égale à $\frac{1}{N}||y-\hat{y}||^2$ et est appelée **MSE** (mean squared error) dans la littérature. Selon le type du problème a modéliser et la nature des labels (classification binaire, multiclassification, ...), nous trouverons des métriques plus pertinentes que d'autres à utiliser pour valider nos modèles.

Par exemple, pour des problèmes de classification binaire i.e des problèmes où $y \in \{0,1\}^N$ nous pouvons utiliser la perte $1-0$ définie par $l(y,\hat{y}) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{y_i \neq \hat{y}_i}$.

En effet le point commun entre ses mesures (perte $1-0$ et **MSE**) est leur séparabilité par rapport aux données. Autrement dit, le calcul des erreurs de prédiction se fait terme à terme entre l'échantillon de vérité terrain et l'estimée qui lui est associée. Dans ce contexte, nous pourrons nous interroger sur l'existence d'une mesure de dissimilarité plus générale qui vérifie la préservation du rang pour des modèles ou nous cherchons à prédire des matrices modélisant l'interaction entre différents acteurs.

2 Présentation de la distance Triplet à partir d'un exemple

Considérons un problème d'apprentissage statistique où nous désirons déterminer les relations de follow sur un réseau social. Typiquement, nous modélisons le problème par une matrice $\mathbf{D} \in \mathcal{M}_{N \times N}(\{0,1\})$ (N désigne le nombre de personnes qui interagissent) entre elles et qui vérifie :

$$D_{i,j} = \begin{cases} 1 \; si \; i \; follow \; j \\ 0 \; sinon \end{cases}$$

Prenons pour l'instant $N = 5$: Considérons maintenant trois matrices $\mathbf{V}$ (matrice référence qui représente la vérité terrain) et $\mathbf{M}_1$ qui représente les résultats fournis par le modèle 1 et $\mathbf{M}_2$ qui représente les résultats fournis par le modèle 2 . Nous désirons alors valider notre modèle en choisissant le modèle le plus pertinent.

Nous posons

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{M}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{V} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Remarquons que $\frac{1}{N}||\mathbf{M}_2 - \mathbf{V}||_2^2 = \frac{1}{N}||\mathbf{M}_1 - \mathbf{V}||_2^2$ où $||.||_2$ désigne la norme de Frobenius définie sur l'espace des matrices. Nous ne pouvons donc pas conclure à cette étape de validation de modèle.

Nous introduisons alors une mesure de dissimalirité basée sur un critère préservant le rang dite distance Triplet. Il s'agit en effet de vérifier pour deux matrices $\mathbf{M}$ et $\mathbf{V}$ et pour tout triplet (i, j, k) la proposition suivante :

$$(P) : V_{i,j} < V_{i,k} \; et \; M_{i,j} < M_{i,k} \tag{1}$$

Appliquons cette mesure à notre exemple pour vérifier sa pertinence :

Pour $\mathbf{M}_1$ et $\mathbf{V}$ cette condition est vérifiée pour les triplets (i, j, k) suivants :

$$(1,1,3),\ (1,5,3)$$

Pour $\mathbf{M}_2$ et $\mathbf{V}$ cette condition est vérifiée pour les triplets (i, j, k) suivants :

$$(1,1,3),\ (1,2,3),\ (1,4,3),\ (1,5,3)$$

Nous pouvons conclure que l'utilisation de la mesure Triplet permet de quantifier l'impact d'une erreur de prédiction commise sur un des labels j associé a l'échantillon i (dénoté par $M_{i,j}$) sur les autres labels. Ainsi, en combinant un critère de type **MSE** et la mesure de dissimilarité en question, nous avons une idée sur la répartition des erreurs et nous prenons en compte la dépendance entre différents éléments de sortie représentants un même échantillon de données.

3 Présentation du problème a résoudre

L'application du critère Triplet nécessite une complexité en $\mathcal{O}(N^3)$ puisque nous vérifions une condition logique pour tout triplet (i, j, k) (N est le nombre d'individus présents dans le modèle d'entraînement et qui représente le nombre de lignes et de colonnes de la matrice de vérité terrain). Nous cherchons alors à implémenter ce critère avec une meilleure complexité. Pour se faire nous remarquerons en premier lieu que l'équivalence suivante est vérifiée :

$$(V_{i,j} < V_{i,k} \ et \ M_{i,j} < M_{i,k}) \ ou \ (V_{i,k} < V_{i,j} \ et \ M_{i,k} < M_{i,j}) \Longleftrightarrow (V_{i,j} - V_{i,k})(M_{i,j} - M_{i,k}) > 0 \tag{2}$$

Nous définissons ensuite une fonction $f : [\![1, N]\!]^3 \to \{0, 1\}$ qui vérifie :

$$(\forall (i, j, k) \in [\![1, N]\!]^3) : \ f(i, j, k) = \mathbb{1}_{(V_{i,j} - V_{i,k})(M_{i,j} - M_{i,k}) > 0} \tag{3}$$

Nous cherchons alors $\mathrm{Card}(f^{-1}(\{1\}))$.
Pour se faire nous calculerons dans un premier temps pour tout $i_0 \in [\![1, N]\!]$

$$\mathrm{Card}(\{(i_0, j, k) \ t.q \ f(i_0, j, k) = 1\})$$

Nous déterminerons dans un deuxième temps un score que nous calculerons à partir de la formule suivante :

$$\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i, j, k)$$

Nous détaillerons le calcul pour trois types de problèmes d'apprentissage statistique à savoir la classification binaire, la multiclassification et enfin le cas général (respectivement $\mathbf{D} \in \mathcal{M}_{N \times N}(\{0, 1\})$, $\mathbf{D} \in \mathcal{M}_{N \times N}([\![0, p]\!])$ et $\mathbf{D} \in \mathcal{M}_{N \times N}(\mathbb{R})$)

4 Résolution du problème dans le cas d'une classification binaire

Soient $\mathbf{V} \in \mathcal{M}_{N\times N}(\{0,1\})$ une matrice déterminant la vérité terrain et $\mathbf{M} \in \mathcal{M}_{N\times N}(\{0,1\})$ une matrice représentant les résultats d'un modèle d'apprentissage visant à estimer $\mathbf{V}$.

Soit $i_0 \in [\![1, N]\!]$. Nous désirons déterminer

$$\text{Card}(\{(i_0, j, k)\ t.q\ f(i_0, j, k) = \mathbb{1}_{(V_{i_0,j}-V_{i_0,k})(M_{i,j}-M_{i_0,k})>0} = 1\})$$

Remarquons tout d'abord que $f(i, ., .)$ est symétrique. En effet,

$$(\forall (i, j, k) \in [\![1, N]\!]^3) :\ f(i, j, k) = f(i, k, j) \tag{4}$$

Nous allons donc nous intéresser à calculer :

$$\text{Card}(\{(i_0, j, k)\ t.q\ (V_{i_0,j} - V_{i_0,k}) > 0\ et\ (M_{i_0,j} - M_{i_0,k}) > 0\})$$

Proposition 1.

$$(V_{i_0,j}-V_{i_0,k}) > 0\ et\ (M_{i_0,j}-M_{i_0,k}) > 0 \Leftrightarrow (V_{i_0,j} = M_{i_0,j} = 1)\ et\ (V_{i_0,k} = M_{i_0,k} = 0)$$

La preuve est triviale. Il suffit de remarquer que $(V_{i_0,j}, M_{i_0,j}, V_{i_0,k}, M_{i_0,k}) \in \{0,1\}^4$

Nous pouvons exploiter ce résultat en pratique pour déterminer avec une complexité de $\mathcal{O}(N)$ les deux ensembles suivants :

$$A^1_{i_0} = \{(i_0, j)\ t.q\ V_{i_0,j} = M_{i_0,j} = 1\}$$
$$A^0_{i_0} = \{(i_0, k)\ t.q\ V_{i_0,k} = M_{i_0,k} = 0\}$$

Proposition 2.

$$\text{Card}(\{(i_0, j, k)\ t.q\ (V_{i_0,j}-V_{i_0,k}) > 0\ et\ (M_{i_0,j}-M_{i_0,k}) > 0\}) = \text{Card}(A^1_{i_0})\times\text{Card}(A^1_{i_0})$$

Démonstration. Soit $(i_0, j, k) \in \{(i_0, j, k)\ t.q\ (V_{i_0,j} - V_{i_0,k}) > 0\ et\ (M_{i_0,j} - M_{i_0,k}) > 0\}$

$$\begin{aligned}
&(i_0, j, k) \in \{(i_0, j, k)\ t.q\ (V_{i_0,j} - V_{i_0,k}) > 0\ et\ (M_{i_0,j} - M_{i_0,k}) > 0\} \\
&\Leftrightarrow (V_{i_0,j} = M_{i_0,j} = 1)\ et\ (V_{i_0,k} = M_{i_0,k} = 1) \\
&\Leftrightarrow (i_0, j) \in A^1_{i_0}\ et\ (i_0, k) \in A^0_{i_0} \\
&\Leftrightarrow ((i_0, j), (i_0, k)) \in A^1_{i_0} \times A^0_{i_0}
\end{aligned}$$

Nous pouvons alors montrer facilement que l'application

$$\begin{aligned}\Psi : \ \{(i_0, j, k) \ t.q \ (V_{i_0,j} - V_{i_0,k}) > 0 \ et \ (M_{i_0,j} - M_{i_0,k}) > 0\} &\to A^1_{i_0} \times A^0_{i_0} \\ (i_0, j, k) &\to ((i_0, j), (i_0, k))\end{aligned}$$

est bijective. Ainsi,

$$\mathrm{Card}(\{(i_0, j, k) \ t.q \ (V_{i_0,j} - V_{i_0,k}) > 0 \ et \ (M_{i_0,j} - M_{i_0,k}) > 0\}) = \mathrm{Card}(A^0_{i_0}) \times \mathrm{Card}(A^1_{i_0})$$

□

Nous pouvons aussi calculer notre score :

$$\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i, j, k)$$

Proposition 3.

$$\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i, j, k) = \frac{2}{N^3} \sum_{i=1}^{N} \mathrm{Card}(A^0_i) \times \mathrm{Card}(A^1_i)$$

Démonstration.

$$\begin{aligned}\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i, j, k) &= \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{1}_{(V_{i,j} - V_{i,k})(M_{i,j} - M_{i,k}) > 0} \\ &= 2 \times \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{1}_{(V_{i,j} = M_{i,j} = 1)} \mathbb{1}_{(V_{i,k} = M_{i,k} = 0)} \\ &= 2 \times \frac{1}{N^3} \sum_{i=1}^{N} \mathrm{Card}(A^0_i) \times \mathrm{Card}(A^1_i)\end{aligned}$$

Le facteur 2 est utilisé pour prendre en compte la propriété de symétrie de f (voir équation 4) □

Remarque 1.

$$\mathrm{Card}(f^{-1}(\{1\})) = 2 \sum_{i_0=1}^{N} \mathrm{Card}(A^0_{i_0}) \times \mathrm{Card}(A^1_{i_0}) = N^3 \times score$$

En utilisant une telle approche nous avons pu déterminer un score pertinent basé sur la métrique Triplet avec une complexité maximale de $\mathcal{O}(N^2)$. Cela nous permet d'éviter une vérification systématique du critère pour chaque triplet ce qui coûtera un $\mathcal{O}(N^3)$. Nous pouvons utiliser le même raisonnement pour un problème de multiclassification, objet de la prochaine section.

5 Résolution du problème dans le cas d'une multi-classification

Nous considérons un problème de classification où nous disposons de p classes.

Soient $p \in \mathbb{N}^*$, $\mathbf{V} \in \mathcal{M}_{N\times N}(\llbracket 0, p \rrbracket)$ une matrice déterminant la vérité terrain et $\mathbf{M} \in \mathcal{M}_{N\times N}(\llbracket 0, p \rrbracket)$ une matrice représentant les résultats d'un modèle d'apprentissage visant à estimer $\mathbf{V}$.

Soit $i_0 \in \llbracket 1, N \rrbracket$. Nous désirons déterminer

$$a_{i_0} = \mathrm{Card}(\{(i_0, j, k) \ t.q \ f(i_0, j, k) = \mathbb{1}_{(V_{i_0,j} - V_{i_0,k})(M_{i_0,j} - M_{i_0,k}) > 0} = 1\})$$

La propriété de symétrie vérifiée dans l'équation 4 nous permet de restreindre notre recherche au calcul de :

$$\frac{a_{i_0}}{2} = \mathrm{Card}(\{(i_0, j, k) \ tq \ (V_{i_0,j} > V_{i_0,k}) \ et \ (M_{i_0,j} > M_{i_0,k})\})$$

Posons dans ce qui suit :

$$A_{i_0} = \{(i_0, j, k) \ tq \ (V_{i_0,j} > V_{i_0,k}) \ et \ (M_{i_0,j} > M_{i_0,k})\}$$

$$(\forall(\alpha, \beta) \in \llbracket 0, p \rrbracket^2) : A_{i_0}^{\alpha,\beta} = \{(i_0, j) \ t.q \ V_{i_0,j} = \alpha \ et \ M_{i_0,j} = \beta\}$$

Lemme 1.

$$(i_0, j, k) \in A_{i_0} \Leftrightarrow ((i_0, j), (i_0, k)) \in \bigcup_{\alpha=1}^{p} \bigcup_{\beta=1}^{p} \bigcup_{\alpha_1=0}^{\alpha-1} \bigcup_{\beta_1=0}^{\beta-1} A_{i_0}^{\alpha,\beta} \times A_{i_0}^{\alpha_1,\beta_1} \tag{5}$$

Démonstration. Il suffit de raisonner par double inclusion. □

Lemme 2.

$$\mathrm{Card}(A_{i_0}) = \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \sum_{\alpha_1=0}^{\alpha-1} \sum_{\beta_1=0}^{\beta-1} \mathrm{Card}(A_{i_0}^{\alpha,\beta}) \times \mathrm{Card}(A_{i_0}^{\alpha_1,\beta_1}) \tag{6}$$

Démonstration. L'application Φ définie par :

$$\begin{aligned}\Phi : A_{i_0} &\to \bigcup_{\alpha=1}^{p} \bigcup_{\beta=1}^{p} \bigcup_{\alpha_1=0}^{\alpha-1} \bigcup_{\beta_1=0}^{\beta-1} A_{i_0}^{\alpha,\beta} \times A_{i_0}^{\alpha_1,\beta_1} \\ (i_0, j, k) &\to ((i_0, j), (i_0, k))\end{aligned}$$

est bijective. Il suffit d'utiliser le lemme 1 pour vérifier la surjectivité. L'injectivité est donnée par construction. De plus, les ensembles $(A_{i_0}^{\alpha,\beta} \times A_{i_0}^{\alpha_1,\beta_1})_{\alpha,\beta,\alpha_1,\beta_1}$ sont disjoints. $\square$

Proposition 4.

$$\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i,j,k) = \frac{2}{N^3} \sum_{i=1}^{N} \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \sum_{\alpha_1=0}^{\alpha-1} \sum_{\beta_1=0}^{\beta-1} \mathrm{Card}(A_i^{\alpha,\beta}) \times \mathrm{Card}(A_i^{\alpha_1,\beta_1}) \tag{7}$$

Démonstration.

$$\begin{aligned}\frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} f(i,j,k) &= \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{1}_{(V_{i,j}-V_{i,k})(M_{i,j}-M_{i,k})>0} \\ &= \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{1}_{(V_{i,j}>V_{i,k})} \mathbb{1}_{(M_{i,j}>M_{i,k})} + \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \mathbb{1}_{(V_{i,j}<V_{i,k})} \mathbb{1}_{(M_{i,j}<M_{i,k})} \\ &= \frac{2}{N^3} \sum_{i=1}^{N} \mathrm{Card}(A_i) \\ &= \frac{2}{N^3} \sum_{i=1}^{N} \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \sum_{\alpha_1=0}^{\alpha-1} \sum_{\beta_1=0}^{\beta-1} \mathrm{Card}(A_i^{\alpha,\beta}) \times \mathrm{Card}(A_i^{\alpha_1,\beta_1})\end{aligned}$$

$\square$

Remarque 2.

$$\mathrm{Card}(f^{-1}(\{1\})) = 2 \sum_{i=1}^{N} \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \sum_{\alpha_1=0}^{\alpha-1} \sum_{\beta_1=0}^{\beta-1} \mathrm{Card}(A_i^{\alpha,\beta}) \times \mathrm{Card}(A_i^{\alpha_1,\beta_1}) = N^3 \times score$$

En pratique, nous fixons un indice de ligne que nous notons i_0. Nous déterminons alors nos ensembles $A_{i_0}^{\alpha,\beta}$ en définissant un vecteur de taille p^2 où chaque élément va contenir la mise a jour de la taille d'un de nos ensembles. Cela impliquerait de vérifier pour tout j la valeur de $M_{i_0,j}$ et $V_{i_0,j}$ puis ajouter 1 à la case correspondante dans notre liste qui joue la rôle d'un compteur. En fin d'itération nous aurons accès à $\text{Card}(A_i^{\alpha,\beta})$ pour tout α et β. Nous pourrons alors calculer $\text{Card}(A_{i_0})$ et notre score global. La complexité de ce calcul est de $\mathcal{O}(p^4N^2)$ ce qui est plus rapide que $\mathcal{O}(N^3)$ pour une large gamme de valeurs de p. Jusqu'à présent nous avons pu déterminer un score basé sur la distance Triplet implémentable plus rapidement qu'une approche systématique. Malheureusement, une telle méthode n'est pas applicable si l'espace où se trouvent nos variables est infini. Nous verrons dans la prochaine section une méthode approchée qui nous permettra d'interpréter la distance Triplet.

6 Résolution approchée du problème dans le cas général

Considérons deux matrices $\mathbf{V} \in \mathcal{M}_{N\times N}(\mathbb{R})$ (matrice de vérité terrain) et $\mathbf{M} \in \mathcal{M}_{N\times N}(\mathbb{R})$ un prédicteur de $\mathbf{V}$.

Nous désirons calculer le score suivant inspiré du " dot product" classique qui se fait terme à terme.

$$S(\mathbf{V},\mathbf{M}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}(V_{i,j}-V_{i,k})(M_{i,j}-M_{i,k})}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}(M_{i,j}-M_{i,k})^2}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}(V_{i,j}-V_{i,k})^2}}$$

Nous avons ainsi défini une mesure de corrélation entre les différents éléments $M_{i,j} - M_{i,k}$ et $V_{i,j} - V_{i,k}$ pour tout triplet (i, j, k). Nous pouvons vérifier que ce score est toujours inférieur a 1 grâce a l'inégalité de Cauchy-Scwarz. Notre but est de calculer ce dernier en complexité sous-cubique.
Nous verrons dans ce qui suit qu'on peut améliorer cette borne à $\mathcal{O}(N^2)$. Pour se faire, nous introduisons la fonction suivante :

$$\begin{aligned} L_i :& \mathcal{M}_N(\mathbb{R}) \to \mathcal{M}_{1,N}(\mathbb{R}) \\ & \mathbf{M} \to L_i(M) \ (\textit{ieme ligne de } \mathbf{M}) \end{aligned}$$

Proposition 5.

$$S(\boldsymbol{M},\boldsymbol{V}) = \frac{g(N\,\boldsymbol{V}.\boldsymbol{M}) - \sum_{i=1}^{N} g(L_i(V))g(L_i(M))}{\sqrt{g(N\boldsymbol{M}.\boldsymbol{M}) - \sum_{i=1}^{N}(g(L_i(M)))^2}\sqrt{g(N\,\boldsymbol{V}.\boldsymbol{V}) - \sum_{i=1}^{N}(g(L_i(V)))^2}} \tag{8}$$

Démonstration.

$$\begin{aligned} g(\mathbf{V}^T\mathbf{M}) &= \sum_{j=1}^{N}\sum_{k=1}^{N}(\sum_{i=1}^{N} V_{i,k}M_{i,j}) \\ &= \sum_{i=1}^{N}(\sum_{k=1}^{N} V_{i,k})(\sum_{j=1}^{N} M_{i,j}) \\ &= \sum_{i=1}^{N} g(L_i(\mathbf{V}))g(L_i(\mathbf{M})) \end{aligned}$$

Nous appliquons le même raisonnement pour $g(\mathbf{M}^T\mathbf{M})$ et $g(\mathbf{V}^T\mathbf{V})$. □

En pratique nous aurons à calculer la somme de chaque ligne de $\mathbf{V}$ et $\mathbf{M}$ avec une complexité de $\mathcal{O}(N^2)$. Nous calculons aussi les produits terme à terme avec la même complexité. Enfin, nous calculons S avec une complexité de $\mathcal{O}(N)$.

Dans cette section, nous avons fourni une métrique qui interprète la mesure Triplet. Cependant, nous obtenons un même score pour une classe d'homothétie de prédicteurs i.e pour une classe d'équivalence associée à la relation suivante. Pour cela, nous commençons par établir un résultat dans la proposition 6

Proposition 6. *Soient $\boldsymbol{M}$, $\boldsymbol{L}$, $\boldsymbol{V}$ trois matrices dans $\mathcal{M}_N(\mathbb{R})$, $\lambda \in \mathbb{R}^*$ et $\gamma \in \mathbb{R}^N$ alors :*

$$\boldsymbol{M} = \lambda \boldsymbol{L} + diag(\gamma)\mathbb{1}_{N,N} \Rightarrow S(\boldsymbol{V}, \boldsymbol{M}) = S(\boldsymbol{V}, \boldsymbol{L}) \tag{9}$$

avec : $diag(\gamma)$ est la matrice diagonale dont l'élément diagonal indexé par i est égal à γ_i et $\mathbb{1}_{N,N}$ est la matrice carrée de taille N dont tous les éléments sont égaux à 1.

Démonstration. Soient $\mathbf{M}$ et $\mathbf{L}$ deux matrices dans $\mathcal{M}_N(\mathbb{R})$, $\lambda \in \mathbb{R}^*$ et $\gamma \in \mathbb{R}^N$ On suppose :

$$(\forall i \in [\![1, N]\!])(\forall j \in [\![1, N]\!]) \; M_{i,j} = \lambda L_{i,j} + \gamma_i$$

alors :

$$(\forall (i,j,k) \in [\![1, N]\!]^3) : M_{i,j} - M_{i,k} = \lambda (L_{i,j} - L_{i,k})$$

En remplaçant dans l'expression du score S nous obtenons le résultat

$\square$

Remarque 3. *L'implication réciproque est vraie si $S(\boldsymbol{M}, \boldsymbol{V}) = 1$ (cas de l'égalité dans l'inégalité de Cauchy-Schwarz). Ainsi si nous disposons d'un prédicteur dont le score s'approche de 1 nous récupérons l'essentiel de l'information sur la ligne de niveau $S(\boldsymbol{V}, .) = S(\boldsymbol{V}, \boldsymbol{M})$ à partir de la classe d'équivalence de $\boldsymbol{M}$ associée à la relation :*

$$\boldsymbol{M} \mathcal{R} \boldsymbol{L} \Leftrightarrow (\exists \lambda \in \mathbb{R}^*)(\exists \gamma \in \mathbb{R}^n) \; \boldsymbol{M} = \lambda \boldsymbol{L} + diag(\gamma)\mathbb{1}_{N,N}$$

Reprenons le raisonnement que nous avons appliqué pour le choix du meilleur candidat parmi les classes d'homothétie. Soit $\mathbf{M}$ un estimateur performant vis à vis du score S. Nous cherchons alors :

$$(\lambda^*, \gamma^*) \in \underset{(\lambda,\gamma) \in \mathbb{R}^* \times \mathbb{R}^N}{\text{Argmin}} ||\mathbf{V} - \lambda \mathbf{M} - diag(\gamma)\mathbb{1}_{N,N}||_2^2$$

Pour résoudre ce problème nous choisissons comme métrique une norme de Frobenius compte tenu de ces propriétés analytiques (classe $\mathcal{C}^\infty$ et convexité). Il suffit donc d'annuler le gradient pour obtenir le minimiseur. Écrivons notre système d'équations aux dérivées partielles à résoudre.
Nous posons :

$$(\forall(\lambda,\gamma)\in\mathbb{R}^*\times\mathbb{R}^N) : h(\lambda,\gamma) = ||\mathbf{V} - \lambda\mathbf{M} - diag(\gamma)\mathbb{1}_{N,N}||_2^2$$

$$\left\{\begin{array}{c}\frac{\partial h}{\partial\lambda}(\lambda^*,\gamma^*) = 2\sum_{i=1}^N\sum_{j=1}^N M_{i,j}(\lambda^* M_{i,j} + \gamma_i^* - V_{i,j}) = 0\\ \frac{\partial h}{\partial\gamma_j}(\lambda^*,\gamma^*) = \sum_{j=1}^N 2(\gamma_i^* + \lambda^* M_{i,j} - V_{i,j}) = 0\end{array}\right.$$

$$\Leftrightarrow\left\{\begin{array}{l}\lambda^* = \frac{tr(\mathbf{M}^T\mathbf{V}) - g(\mathbf{M}^T\gamma^*)}{||M||_2^2}\\ \gamma_i^* = \frac{g(L_i(\mathbf{V})) - \lambda^* g(L_i(\mathbf{M}))}{N}\end{array}\right.$$

En remplaçant γ^* dans la première équation nous obtenons :

$$\lambda^* = \frac{tr(\mathbf{M}^T\mathbf{V}) - \frac{1}{N}\sum_{i=1}^N g(L_i(\mathbf{M}))g(L_i(\mathbf{V})))}{||\mathbf{M}||_2^2 - \frac{1}{N}\sum_{i=1}^N g(L_i(M))^2}$$

Nous pouvons alors calculer les composantes de γ^* en injectant la valeur de λ^* dans la deuxième équation du système. A partir des valeurs obtenues, nous pouvons déduire que $\lambda^*\mathbf{M} + diag(\gamma^*)\mathbb{1}_{N,N}$ est l'estimateur le plus robuste dans cette classe vis à vis du score S et de l'erreur d'estimation. Pour vérifier la pertinence de l'estimateur $\mathbf{M}$ obtenu il faudrait vérifier les deux conditions suivantes :

$$\lambda^* \approx 1$$

$$\gamma^* \approx \mathbf{0}_{\mathbb{R}^N}$$

Remarque 4. *Toutes les constantes à déterminer qui figurent dans les expressions de λ^* et γ^* ont été calculé au préalable pour déterminer le score $S(\boldsymbol{V},\boldsymbol{M})$. Le calcul de λ^* et γ^* se fera alors en complexité linéaire $\mathcal{O}(N)$.*

Nous donnons dans ce qui suit un protocole expérimental qui permet de choisir un estimateur selon sa pertinence vis a vis d'un critère qui préserve le rang et d'une métrique qui mesure l'erreur d'estimation
Considérons alors $\mathbf{M}$ un estimateur et $\mathbf{V}$ une vérité terrain. Nous supposons que les deux matrices sont à valeurs dans $\mathbb{R}$.

Étape 1 :
Nous calculons $S(\mathbf{V},\mathbf{M})$ avec une complexité quadratique

Étape 2 :
Nous calculons λ^* et γ^* avec une complexité linéaire

Étape 3 :
Nous calculons un compromis entre le score $S(\mathbf{V}, \mathbf{M})$ et l'erreur d'estimation évaluée a l'étape 2. Par exemple, nous pouvons calculer : $|\lambda^* - 1|^2$, $||\gamma^*||^2$ et $1 - S(\mathbf{V}, \mathbf{M})$ (dans la perspective de choisir le prédicteur qui minimise ces trois quantités) .

Étape 4 :
Nous répétons les trois étapes précédentes pour un autre predicteur.

Avantages :
Tout le calcul s'effectue en complexité quadratique ($\mathcal{O}(N^2)$). De plus, nous avons établi un lien entre la mesure Triplet approchée par le score S et la norme de Frobenius qui mesure l'erreur terme à terme. En effet, nous avons construit un schéma cohérent pour valider nos modèles à partir du critère composite (Triplet-Frobenius).