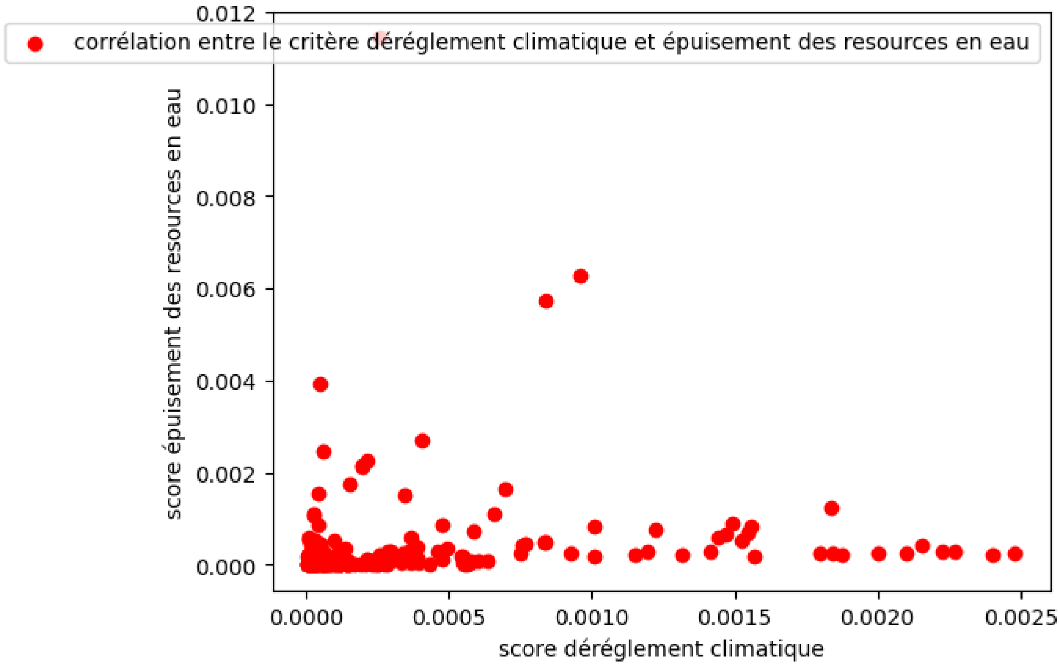
Projet en fondements mathématiques pour l'aide à la décision

Question 3

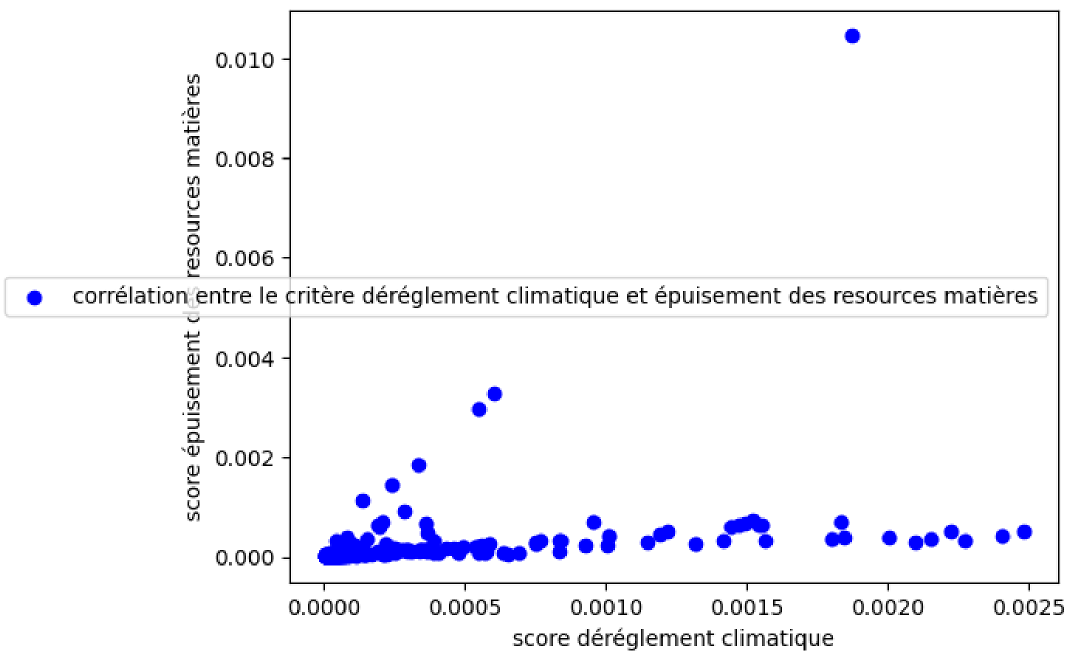
Critères 1 et 14 :



Une régression linéaire de la distribution correspondrait à une droite horizontale. (Autours du score moyen du critère 14).

Les deux critères sont indépendants : L'augmentation du score du critère 14 n'est pas associé à une augmentation ou diminution du critère 1.

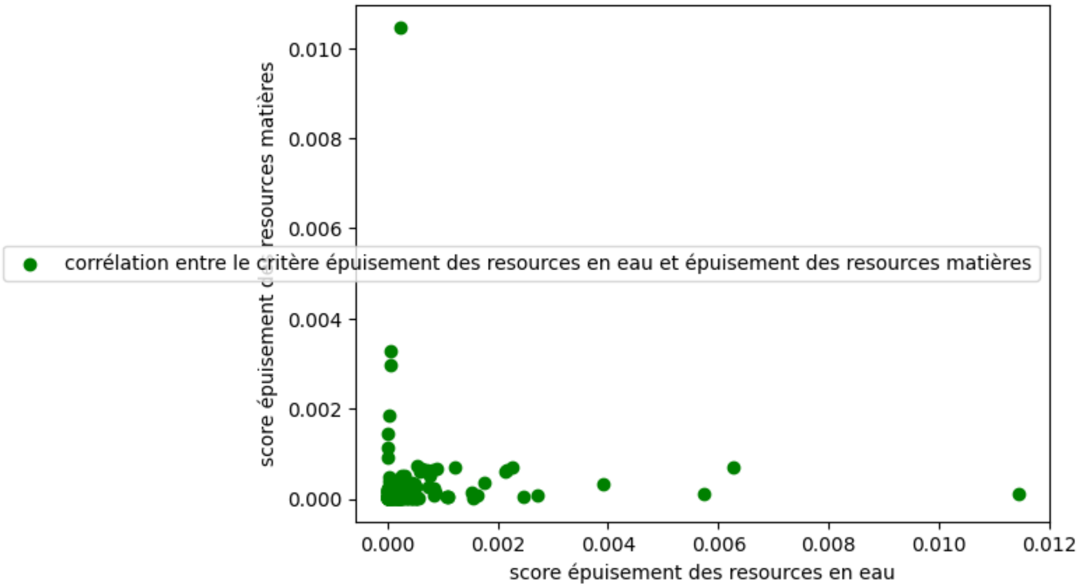
Critères 1 et 16



Une régression linéaire de la distribution correspondrait à une droite horizontale. (Autours du score moyen du critère 16).

Les deux critères sont indépendants : L'augmentation du score du critère 16 n'est pas associé à une augmentation ou diminution du critère 1.

<u>Critères 16 et 14</u>



Les deux critères sont indépendants : Pour des scores faibles du critère 14, on observe des valeurs faibles et importantes pour le score 16. Lorsque le score 14 augmente, le score 16 est distribué de façon aléatoire.

Les deux critères sont indépendants : L'augmentation du score du critère 16 n'est pas associé à une augmentation ou diminution du critère 14.

<u>Conclusion</u> : Les critères 1, 14 et 16 sont mutuellement indépendants.

Question 4

La relation Pareto Dominance est définie sur $A^2 = (\mathbb{R}^{16})^2$.

Dans ce qui suit, je vais noter $X < Y$ pour dire que X Pareto-domine Y.

Symétrie et reflexivité

• Soit (X,Y) tel que $X<Y$, on a $\forall i$, $x_i \leqslant y_i$ et $\exists$ au moins $i_0 \in [\![1,16]\!]$ tq $x_{i_0} < y_{i_0}$.

Ainsi si $X<Y \Rightarrow \neg(Y<X)$, ceci vaut $\forall (X,Y) \in A^2$.

La relation est asymétrique, elle est donc irreflexive.

• On ne peut jamais avoir $X<Y \wedge Y<X$ donc la proposition : $\forall X,Y\ (X<Y \wedge Y<X \Rightarrow X=Y)$ est toujours vraie.

Ainsi P.D. est aussi antisymétrique

Transitivité

Si $X<Y$ et $Y<Z$ alors on a $\forall i \in [\![1,16]\!]$:
$x_i \leqslant y_i$ et $y_i \leqslant z_i$ donc $x_i \leqslant z_i$

$\exists\, i_1$, $x_{i_1} < y_{i_1} \leqslant z_{i_1}$ donc $x_{i_1} < z_{i_1}$

Donc $X<Z$
$\Rightarrow$ P.D. est transitive.

Soient $X = (1, 3, 1, \dots, 1) \in \mathbb{R}^{16}$
$Y = (3, 2, 1, \dots, 1) \in \mathbb{R}^{16}$
$Z = (2, 3, 1, \dots, 1) \in \mathbb{R}^{16}$

On a $\neg(X < Y)$ [car $x_2 > y_2$] et $\neg(Y < Z)$ [car $y_1 > z_1$] mais on a $X < Z$ [$\forall i, x_i \leq z_i$].

Ainsi P.D n'est pas négativement transitive.

Soient $X = (1, 2, 3, 1, \dots, 1) \in \mathbb{R}^{16}$
et $Y = (3, 1, 1, 1, \dots, 1) \in \mathbb{R}^{16}$

On a $y_1 > x_1$ donc $\neg Y < X$ et $x_3 > y_3$ donc $\neg X < Y$.

P.D. n'est pas complète.

P.D. est transitive et antisymétrique mais elle n'est pas négativement transitive.

La Pareto Dominance est un ordre strict partiel.

Question 7

Le pourcentage d'ensembles $\{X, Y\}$ de deux alternatives avec d'ensembles $X \in \mathrm{A} \neq \mathrm{Y} \in \mathrm{A}$ tels que X Pareto-domine Y est de 14%.

Le pourcentage d'ensembles $\{X, Y\}$ de deux alternatives avec d'ensembles $X \in \mathrm{A} \neq \mathrm{Y} \in \mathrm{A}$ tels que X Pareto-domine Y sur les critères 1, 14 et 16 est de 27%. On peut voir que ce pourcentage augmente.

Q8 : calcul Taux de substitution

	critère 1	critère 14	critère 16
facteurs de normalisation	7550	11 500	$6,36 \cdot 10^{-2}$
Poids	$21,06 \cdot 10^{-2}$	$8,51 \cdot 10^{-2}$	$7,55 \cdot 10^{-2}$

Taux de substitution : $w_1 / w_{14} = 2,47$

$w_1 / w_{16} = 2,79$

$w_{14} / w_{16} = 1,125$

- une augmentation de 7550 dans le score du critere 1 correspond à une augmentation de $2,47 \times 11500 = 28459$ du score du critère 14
- une augmentation de 7550 dans le score du critere 1 correspond à une augmentation de $2,79 \times 6,36 \cdot 10^{-2} = 0,177$ du score du critère 16
- une augmentation de 11500 dans le score du critere 14 correspond à une augmentation de $1,125 \times 6.36 \cdot 10^{-2} = 0,072$ du score du critère 16

Q9 : Programme linéaire

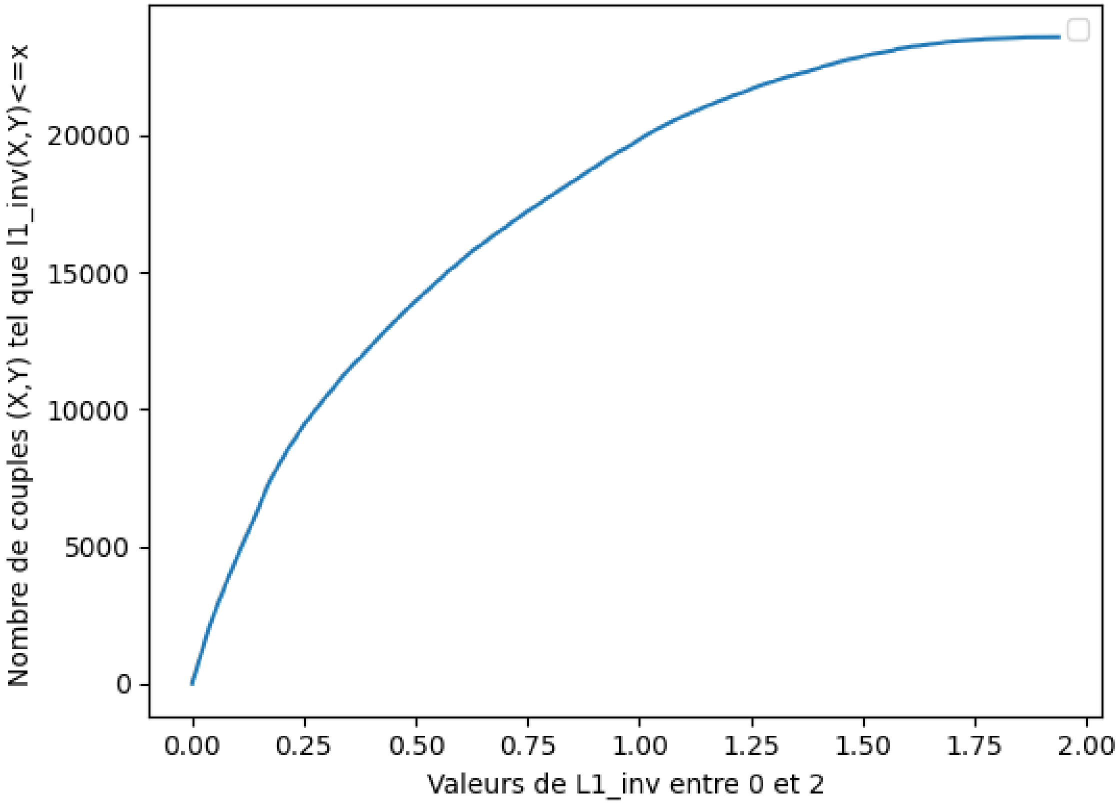
$$z = \min \| w - \hat{w} \|$$

$\Rightarrow$ Modeliser la valeurs absolue

sous contraintes

$$\begin{cases} \bullet\ z = \min \sum_{i=1}^{16} s_i + t_i \\ \bullet\ w_i - \hat{w}_i = s_i - t_i \quad \forall i \in \{1, \dots, 16\} \\ \bullet\ SP_w(X) \geq SP_w(Y) \iff \sum_{i=1}^{16} w_i x_i \geq \sum_{i=1}^{16} w_i y_i \\ \bullet\ SP_{\hat{w}}(X) \leq SP_{\hat{w}}(Y) \\ \bullet\ \sum_{i=1}^{16} w_i = 1 \\ \bullet\ w_i, x_i, y_i, s_i, t_i \geq 0 \end{cases}$$

Question 11



Ce graphique nous indique le nombre de couples pour lesquels il suffit de changer le jeu de poids pour inverser leur comparaison en somme pondérée tout en gardant $||w - \widehat{w}||$ entre 0 et 2.

On peut conclure à partir du graphe qu'il est possible d'inverser la comparaison de plus de 29% des couples (soit plus de 20000 sur 68382) en changeant le jeu des poids, sans dépasser une différence de 2 en norme avec l'ancien jeu de poids.

Ceci indique que la comparaison à partir de la somme pondérée n'est pas fiable pour calculer le score des aliments et les classer, puisque cela dépend du jeu de poids choisi.

Question 12

Nous avons décidé de créer 3 différents types de jeu de poids.

- Premier jeu de poids (w2) :

Une suite croissante et aléatoire telle que $\sum_{i=1}^{16} w_i = 1$.

Ce jeu de poids permet d'amplifier les scores qui sont déjà importants et donc de les prendre plus en considération dans notre comparaison. Ce jeu de poids s'adapte donc à la distribution des scores des critères dans chaque aliment.

- Deuxième jeu de poids (w3):

Une suite décroissante et aléatoire telle que $\sum_{i=1}^{16} w_i = 1$.

Ce jeu de poids permet d'amplifier les scores qui sont moins importants et donc de les prendre autant en considération que les autres critères favorisant ainsi des solutions plus équilibrées.

- Troisième jeu de poids (w4) :

Une suite aléatoire telle que $\sum_{i=1}^{16} w_i = 1$.

<u>Question 13</u>

Moyenne pondérée ordonnée :

Avec la moyenne pondérée ordonnée et en utilisant le jeu de poids w2(croissant), nous nous adaptons à la distribution des scores dans chaque aliment X. Cela permet d’amplifier les scores qui sont déjà grands et de ne pas les camoufler.

Avec la moyenne pondérée et en utilisant le jeu de poids w3(décroissant), nous favorisons des solutions plus équilibrées en donnant à chaque critère évalué une importance égale dans la détermination du score final.

La moyenne géométrique :

Le choix de la moyenne géométrique est pertinent car nous n’avons pas des scores (par critère) nuls ou négatifs. Cette moyenne permet de calculer l’influence globale des critères (grâce au produit) au lieu de sommer les scores individuels et donc de donner autant d’importance à chaque critère dans le calcul de la valeur finale. La MG est aussi moins sensible aux valeurs extrêmes que la moyenne pondérée : Les scores auront des valeurs plus homogènes.

Question 14

d_{kt} est définie sur l'ensemble des relations d'ordre sur $A = \mathbb{R}^{16}$. Notons R_{OA} cet ensemble.

$$d_{KT} : R_{OA}\, R_{OA} \longrightarrow \mathbb{R}_+$$

Nous allons démontrer que d_{KT} est une distance sur R_{OA}^2.

Espace d'arrivée : $\mathbb{R}^+$ ✓

Reste à montrer les propriétés de :

- Symétrie
- Séparation
- Inégalité triangulaire

$\boxed{\text{. Symétrie}}$ $\forall R_1, R_2 \quad d(R_1, R_2) = d(R_2, R_1)$
Vérifiée car la formule est symétique

$\boxed{\text{. Séparation}}$

Soit R_1, R_2 ordres forts sur A tq

$$d_{KT}(R_1, R_2) = 0$$

Or $d_{KT}(R_1, R_2) = \underbrace{\sum_{(X,Y)\in A^2} 0.5\,(X R_1 Y \wedge Y R_2 X) + \dots}$

Somme d'entiers positifs $= 0$
alors tout entier $= 0$

Ainsi $\forall\, X, Y \in A$, on a :

$\neg(X R_1 Y \wedge Y R_2 X) \iff \neg(X R_1 Y) \vee \neg(Y R_2 X)$
(non utilisée)

et

$\neg((X R_1 Y \wedge \neg(X R_2 Y)) \vee (\neg(X R_1 Y) \wedge (X R_2 Y)))$

$\iff \neg(X R_1 Y \wedge \neg(X R_2 Y)) \wedge \neg(\neg(X R_1 Y) \wedge (X R_2 Y))$

$\iff ((\neg X R_1 Y) \vee (X R_2 Y)) \wedge (X R_1 Y \vee \neg(X R_2 Y))$

$\iff \underbrace{(\overbrace{X R_1 Y \wedge X R_2 Y}^{①}) \vee (\overbrace{\neg X R_1 Y \wedge \neg X R_2 Y}^{②})}_{(\mathrm{II})}$

(II) est valable $\forall (X,Y) \in A^2$, donc il faut toujours satisfaire une des propositions ① ou ②.

Soit $(X,Y) \in R1 \iff X R_1 Y$

donc d'après (II), la proposition (1) est vraie (la prop ② ne peut pas être vraie car on a $X R_1 Y$)

donc $X R_2 Y$

donc $(X, Y) \in R_2$ donc $R_1 \subset R_2$

Symétiquement, $(X, Y) \in R_2 \Rightarrow (X, Y) \in R_1$

donc $R_2 \subset R_1$

$\Rightarrow R_1 = R_2$

Ainsi, la propriété de séparation est vérifiée.

Inégalité triangulaire

Soit R_1, R_2, R_3 trois relations d'ordre fort définis sur A^2.

On doit m.q $d_{KT}(R_1, R_3) \leq d_{KT}(R_1, R_2) + d_{KT}(R_2, R_3)$.

Soit $d_{i,j}(R, R')$ l'expression tq

$$d_{KT}(R, R') = \sum_{(i,j) \in A^2} d_{ij}(R, R')$$

On a : $d_{ij}(R, R') = 0 \iff (i R j \wedge i R' j)$

$$d_{ij}(R, R') = 0{,}5 \iff (i R j \wedge \neg(j R' i) \wedge \neg(i R' j))$$

$$d_{ij}(R, R') = 1 \iff (i R j \wedge j R' i)$$

On va démontrer l'inégalité triangulaire sur d_{ij} :

$$\forall i, j \quad d_{ij}(R_1, R_3) \leq d_{ij}(R_1, R_2) + d_{ij}(R_2, R_3)$$

Soit $(i, j) \in A^2$:

1er cas : $d_{ij}(R_1, R_3) = 0$ iméq ✓ car les $d_{ij} \geq 0$.

2e cas : $d_{ij}(R_1, R_3) = 0{,}5 \iff i R_1 j \wedge \neg(j R_3 i) \wedge \neg(i R_3 j)$

dans ce cas, $d_{ij}(R_2, R_3) \geq 0{,}5$ car pour être nul, il faut $(j R_3 i)$ or on a $\neg(j R_3 i)$.

donc inégalité ✓

3^e cas : $d_{ij}(R_1,R_3) = 1 \Leftrightarrow i R_1 j \wedge j R_3 i$

↳ sous cas 1 : $d_{ij}(R_1, R_2) = 0$

donc $i R_2 j$ et on a $j R_3 i$

donc $d_{ij}(R_2,R_3) = 1$

inégalité ✓

↳ sous cas 2 : $d_{ij}(R_1, R_2) = 0,5$

$\Leftrightarrow i R_1 j \wedge \neg(j R_2 i) \wedge \neg(i R_2 j)$

On a ainsi :

$j R_3 i \wedge \neg(j R_2 i) \wedge \neg(i R_2 j)$

donc $d_{ij}(R_3, R_2) = 0,5$

inégalité ✓

On a ainsi $\forall (i,j) \in A^2$:

$$d_{ij}(R_1,R_3) \leqslant d_{ij}(R_1,R_2) + d_{ij}(R_2,R_3)$$

En sommant sur tous les couples $(i,j) \in A^2$:

$$d_{KT}(R_1,R_3) \leqslant d_{KT}(R_1,R_2) + d_{KT}(R_2,R_3)$$

Ceci vaut $\forall R_1, R_2, R_3$ ordres forts sur A.

Inégalité triangulaire vérifiée.

$\Rightarrow d_{KT}$ est une distance sur les ordres forts de A.

Question 15

La distance entre MPO et MG permet d'évaluer la corrélation des scores générés par les agrégateurs. Plus la distance est grande, plus les évaluations des agrégateurs sont différentes. Avec le jeu de poids w2 et w3 (suites croissantes et décroissantes de poids aléatoires), nous arrivons à avoir une distante deux fois plus petite qu'avec le jeu de poids initial (noté w1). Ainsi, les jeux de poids w2 et w3 permettent de garder une certaine cohérence même lorsqu'on change d'agrégateur.

Question 16

Grâce à la somme pondérée $SP_{\hat{w}}$, il est possible de ramener l'éco-score à un score entre 0 et 100 grâce à la formule : $score(x) = 100 - \frac{\ln(10x+1)}{\ln(2+\frac{1}{100x^4})} \times 20.$

L'éco-score est utilisé pour catégoriser chaque aliment à partir du score calculé par la formule. Plus l'éco-score est élevé, moins l'aliment aura d'impact sur l'environnement, c'est-à-dire, que sa consommation polluera moins.

La formule logarithmique permet de réduire l'écart entre les produits ayant des éco-scores très différents, ce qui facilite la comparaison entre eux. Elle permet aussi de rendre les scores plus compréhensibles pour les consommateurs en les présentant sous forme de pourcentage plutôt que de valeurs numériques brutes.

Cependant, cette formule peut présenter certains inconvénients. Elle peut masquer des différences importantes entre les produits ayant des éco-scores très proches, car la normalisation logarithmique tend à "compresser" les valeurs basses, les aliments dont le score sera élevé ou moyen seront favorisés par rapport à ceux dont le score sera très faible.

Une alternative serait une normalisation linéaire. Cette dernière a l'avantage d'être facile à comprendre pour les consommateurs tout en gardant les écarts de valeurs de l'éco-score entre les aliments.

Question 17

On remarque que pour notre base de données, il n'existe aucun jeu de poids qui puisse faire changer la catégorie de nos aliments.

Q17 : Programme linéaire

$$z = \min \| w - \hat{w} \|$$

$\Rightarrow$ Modeliser la valeurs absolue

- $z = \min \sum_{i=1}^{16} s_i + t_i$

sous contraintes
- $w_i - \hat{w}_i = s_i - t_i \quad \forall i \in \{1, \dots, 16\}$
- changement de categorie : $SP_w(X) \geq 0,28 \iff \sum_{i=1}^{16} w_i x_i \geq 0,28$
- $\sum_{i=1}^{16} w_i = 1$
- $w_i, x_i, s_i, t_i \geq 0$