Projet de Léo Maury et Slim Kachkachi

**Mandatory Points**

1. Q1 Generate 8 random points of + and 8 random points of - on a 2D plane. The points must be linearly separable.

Les 16 points générés sont représentés dans l’onglet « Q1 » du fichier excel « TP\_Optim\_LM&SK ». Nous avons excentré un point (marqué par un flèche rouge dans le schéma ci-dessous) afin de tester le comportement des « loss » face à des « outliers ». Certains points de même classe ont été « rapprochés sans se toucher » afin de mettre en évidence l’effet de « marge souple » avec la « hinge-loss ».



1. Q2 Implement the model that provides the optimal separating hyperplane for points that are linearly separable.

Les résultats sont présentés dans les onglets “Q2 & Q3” du même fichier Excel.

Juste pour rappel, le « système » à paramétrer dans Excel est :

Pour le minimum de la fonction objective : la fonction « quadratique » est remplacée une fonction linéaire :

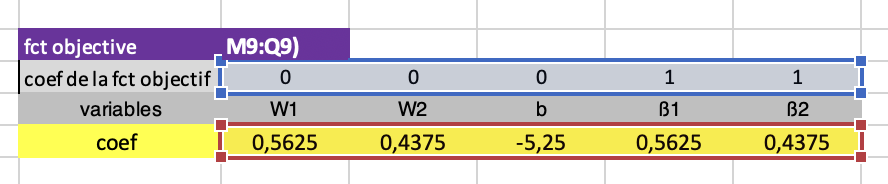
Min remplacé par Min

Et les valeurs absolues sont remplacés par le système suivant :

(

La fonction objective est « paramétrée » dans la cellule violette (col = M, ligne = 9) de l’onglet « Q2 & Q3 »

=SOMMEPROD(M7:Q7;M9:Q9)



Les conditions les s’expriment sous la forme des contraintes suivantes :

|  |  |
| --- | --- |
| ß1 - W1 | >=0 |
| ß1 + W1 | >=0 |
| ß2 - W2 | >=0 |
| ß2 + W2 | >=0 |

Et les conditions «  =   *+*  » sont codées pour chacun des 16 points :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C1 | >=1 | | C9 | >=1 |
| C2 | >=1 | | C10 | >=1 |
| C3 | >=1 | | C11 | >=1 |
| C4 | >=1 | | C12 | >=1 |
| C5 | >=1 | | C13 | >=1 |
| C6 | >=1 | | C14 | >=1 |
| C7 | >=1 | | C15 | >=1 |
| C8 | >=1 | | C16 | >=1 |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

1. Q3 Show the equation of the separating hyperplane.

Les deux graphes suivant représentent l’hyperplan de séparation trouvé avec les marges « -1,+1 » ; avec le zoom à droite qui permet de mieux distinguer les « vecteurs supports ».



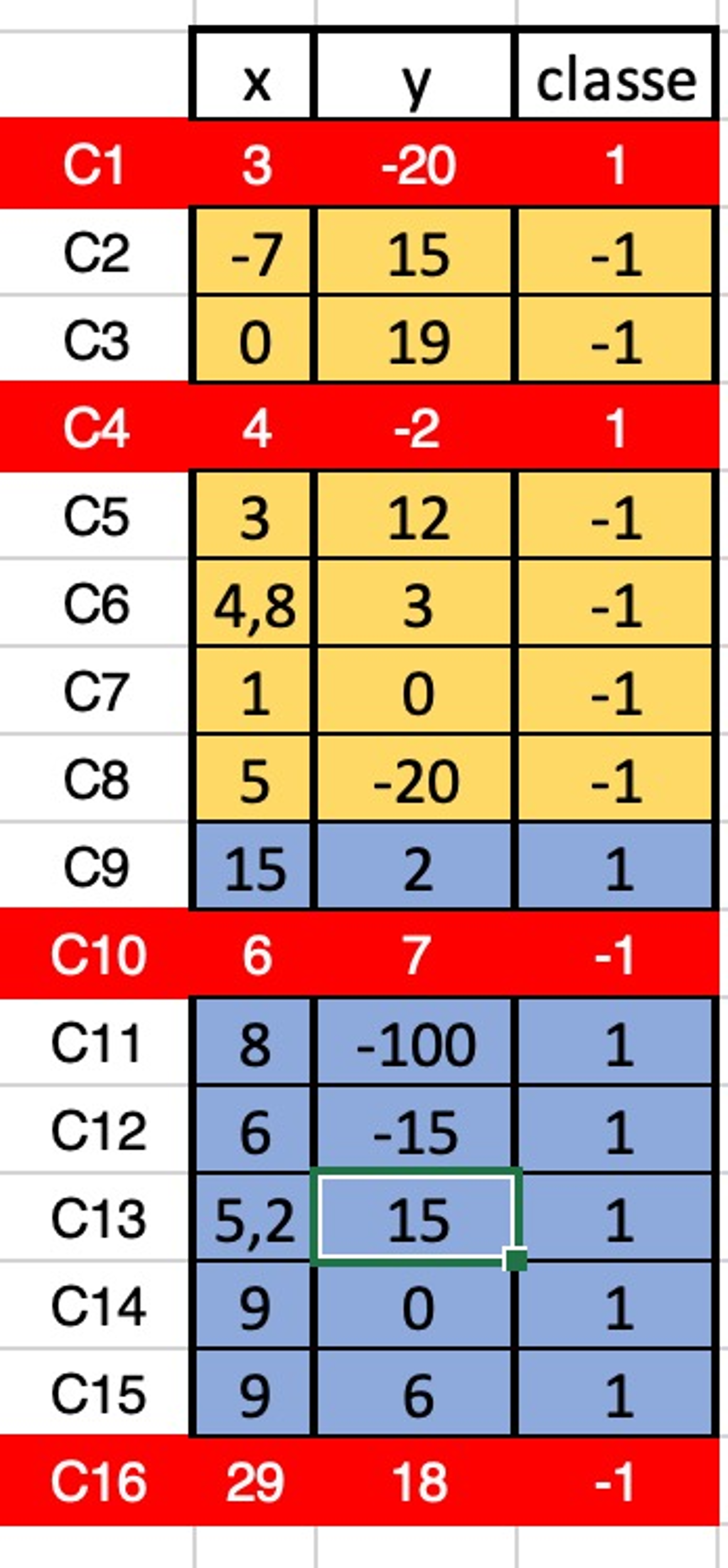


A noter que seuls les vecteurs supports choisis définissent l’hyperplan de séparation. Autrement dit en changeant les autres points, l’hyperplan reste le même (cf. onglet Q2 & Q3 bis) comme l’illustre le graphe suivant (points cerclés ont été modifiés).

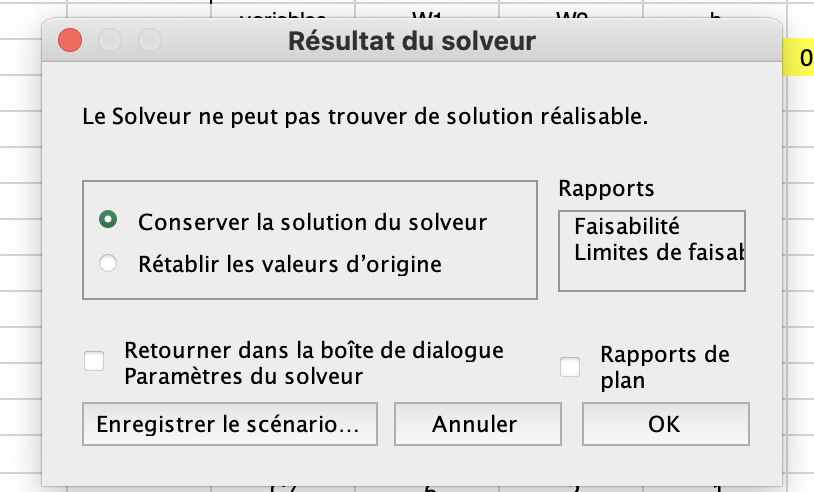


1. Q4 Flip the sign of 4 points (making the new set of points not linearly separable).

Le résultat est présenté dans l’onglet “Q4 ”. Les 4 points ayant changé de classe sont indiqués en rouge.



Avec ce nouveau jeu de données, le solveur ne peut pas trouver de solution comme l’indique le message suivant.

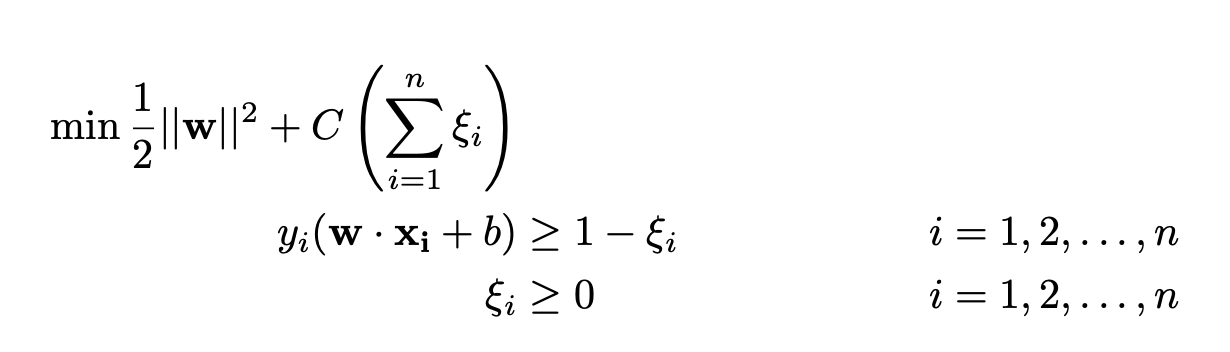


1. Q5 Implement two models (with two of the following types of loss: hinge, ramp or hard) that provide the optimal separating hyperplane for points that are not linearly separable.

Nous allons implémenter les 3 modèles.

**La « hinge loss »** est paramétrée dans l’onglet « Q5\_Hinge\_Loss ».

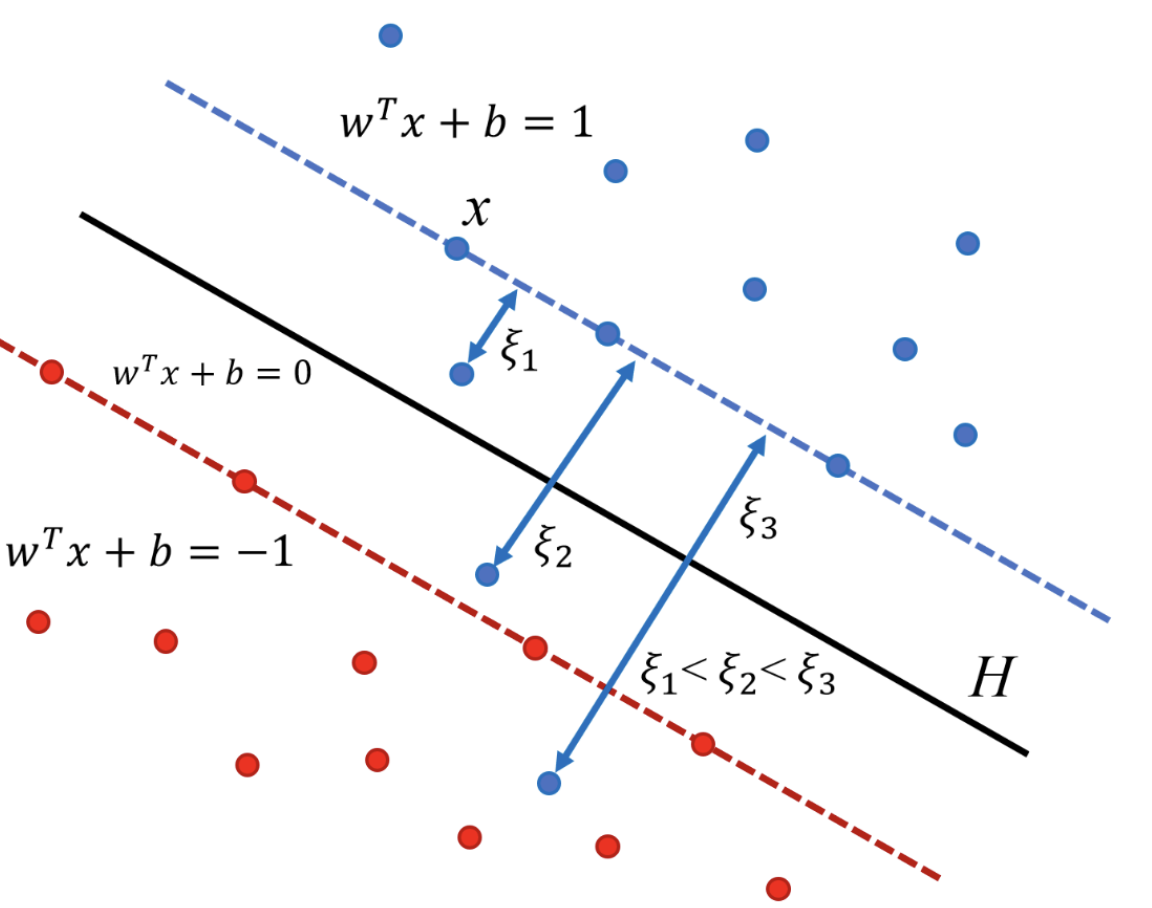
Juste pour rappel, l’hyperplan de séparation est trouvé en résolvant le système d’optimisation suivant **:**



..où est l’erreur par rapport à l’hyperplan de séparation pour un point « i » mal-classé ou localisé dans la marge (sinon égal à zéro pour les points bien classés).

En « interprétant » la fonction objective à optimiser, on « s’aperçoit » que pour un minimum donné, on peut soit jouer sur le terme « min  » soit le terme « C. Notons que réduire  contribue à augmenter la marge (car sa largeur est proportionnelle à ). Nous ferons un essai pour visualiser ce comportement.

On peut anticiper que :



* Plus C est petit, plus les erreurs peuvent être grandes (leur poids relatif dans l’équation sera minoré par C) et donc la marge grandit... ce qui mènera à un modèle de séparation moins complexe.
* Plus C est grand, plus les erreurs doivent être petites pour contribuer à la minimisation de la fonction objective avec une marge réduite et un risque « d’overfitting » (c’est-à-dire un hyperplan de séparation qui va tenir compte des particularités).

Dans Excel, nous avons traduit le système d’optimisation suivant :

(

avec i

*avec sinon (c'est à dire dans la marge ou mal classé)*

Et l’ensemble des contraintes suivantes :

*+ (que nous appellerons Ci)*

Avec

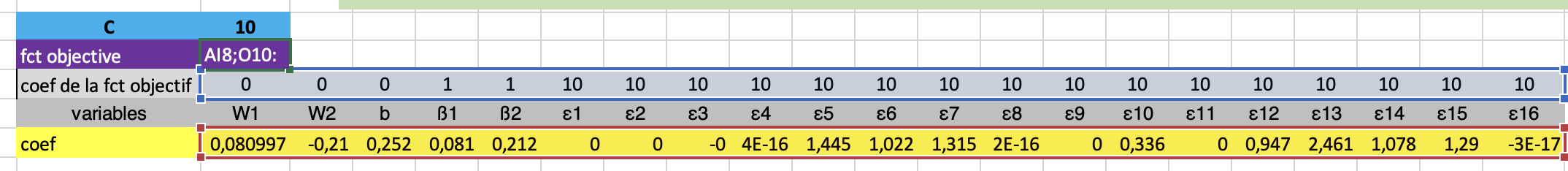
Ce qui se traduit par la déclaration dans le solver des contraintes ci-dessous :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C1 | >=1 | | C9 | >=1 |
| C2 | >=1 | | C10 | >=1 |
| C3 | >=1 | | C11 | >=1 |
| C4 | >=1 | | C12 | >=1 |
| C5 | >=1 | | C13 | >=1 |
| C6 | >=1 | | C14 | >=1 |
| C7 | >=1 | | C15 | >=1 |
| C8 | >=1 | | C16 | >=1 |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ε1 | >=0 | | ε9 | >=0 |
| ε2 | >=0 | | ε10 | >=0 |
| ε3 | >=0 | | ε11 | >=0 |
| ε4 | >=0 | | ε12 | >=0 |
| ε5 | >=0 | | ε13 | >=0 |
| ε6 | >=0 | | ε14 | >=0 |
| ε7 | >=0 | | ε15 | >=0 |
| ε8 | >=0 | | ε16 | >=0 |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

La fonction objective est « codée » dans la cellule violette (col = W, ligne = 7) de l’onglet « Q5\_Hinge\_Loss » :

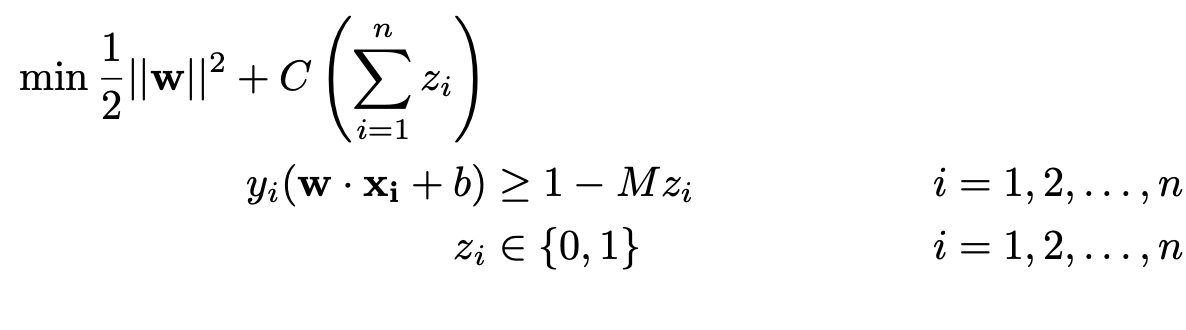
=SOMMEPROD(W8:AQ8;W10:AQ10)



Nous avons testé quatre valeurs de C (0,01,0,1 ;1 ;10). Pour chacune de ces valeurs, le solveur à trouver une solution. L’interprétation de ces résultats est exposée comme demandé dans le chapitre suivant.

**La « Hard Margin loss »** est présentée dans l’onglet Excel « Q5\_HardMargin\_Loss »

L’hyperplan de séparation est trouvé en résolvant le système d’optimisation suivant :



Avec égal 1 si le point est mal classé, égal à 0 sinon

La différence par rapport à la « hinge loss » tient dans la valeur du 2ieme terme « C » où les «  » prennent des valeurs binaires (0 si le point i est bien classé, 1 sinon). Ce qui se traduit une pénalisation « forfaitaire » indépendante de la distance aux hyperplans « +1 » et « -1 » pour tous les points mal-classés.

En « interprétant » la fonction objective à optimiser, on peut dire que :

* Plus C est petit, plus le nombre de points mal classés peut-être élevé (à travers = 1), car leur poids relatif dans l’équation sera minoré par C (ici c’est la quantité de points mal-classés qui joue et non plus les distances par rapport aux hyperplans « +1 » / « -1 ».
* A l’inverse plus C est grand, plus le nombre de points mal classés devra être faible.
* Notons que M, dans l’inégalité joue le rôle de dans l’inégalité pour la « hinge loss » mais cette fois ci avec une valeur constante. On « intuite » que M joue sur la largeur de la marge selon les valeurs de C (à travers une tolérance à l’éloignement des points mal-classés par rapport à leur droite support).

Dans Excel, nous avons traduit le système d’optimisation suivant :

(

et {0,1}

Et l’ensemble des contraintes suivantes

*+*

Avec

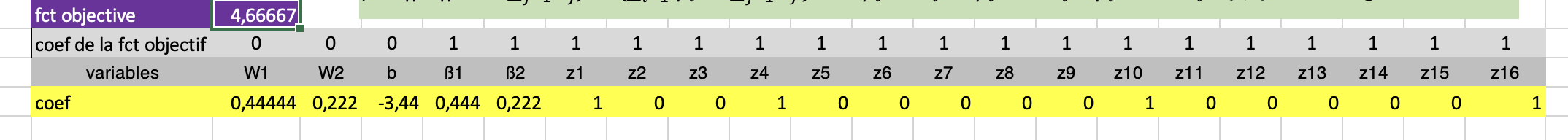
Ce qui se traduit par l’ajout des contraintes suivantes dans le solveur :

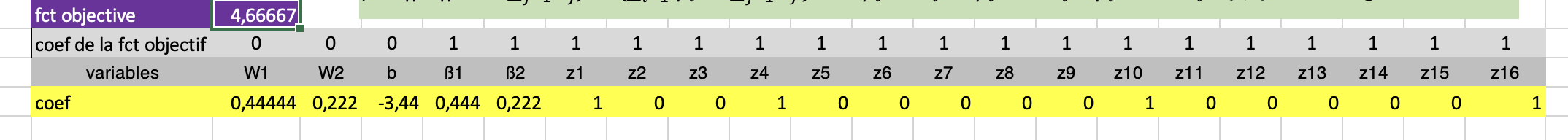
|  |  |  |
| --- | --- | --- |
| z1 | = 1 ou 0 | |
| z2 | = 1 ou 0 | |
| z3 | = 1 ou 0 | |
| z4 | = 1 ou 0 | |
| z5 | = 1 ou 0 | |
| z6 | = 1 ou 0 | |
| z7 | = 1 ou 0 | |
| z8 | = 1 ou 0 | |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C1 | >=1 | | C9 | >=1 |
| C2 | >=1 | | C10 | >=1 |
| C3 | >=1 | | C11 | >=1 |
| C4 | >=1 | | C12 | >=1 |
| C5 | >=1 | | C13 | >=1 |
| C6 | >=1 | | C14 | >=1 |
| C7 | >=1 | | C15 | >=1 |
| C8 | >=1 | | C16 | >=1 |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

La fonction objective est « codée » dans la cellule violette (col = AC, ligne = 6) de l’onglet « Q5\_HardMargin\_Loss ».

=SOMMEPROD(AC8:AW8;AC10:AW10))

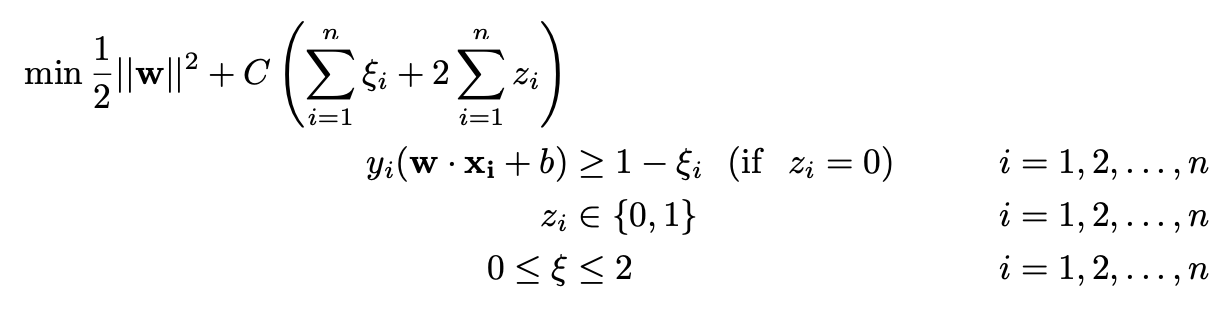
§



Nous avons testé plusieurs valeurs de M & de C (0,5 ;1 ;100 et 1000000). Pour chacune de ces valeurs, le solveur à trouver une solution avec des marges et des positions de l’hyperplan très variables. Les résultats sont présentés et commentés dans le prochain chapitre comme demandé.

**La « Ramp Margin loss »** est présentée dans l’onglet Excel « Q5\_RampMargin\_Loss »

L’hyperplan de séparation est trouvé en résolvant le système d’optimisation suivant :



Avec égal 1 si le point est mal classé, égal à 0 sinon

Dans Excel, nous avons traduit le système d’optimisation suivant :

et {0,1}

et 0 <= <= 2 et M is large value

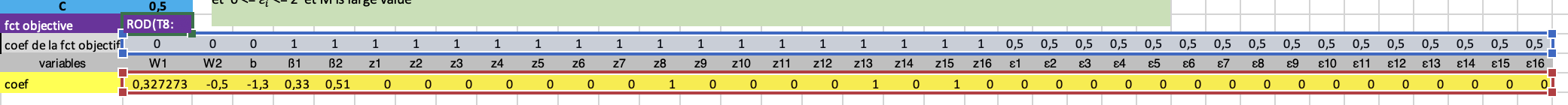
Ce qui se traduit par l’ajout des contraintes dans le solveur :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ε1 | 0 <= 0 & <=2>=0 | | | ε9 | 0 <= 0 & <=2>=0 |
| ε2 | 0 <= 0 & <=2>=0 | | | ε10 | 0 <= 0 & <=2>=0 |
| ε3 | 0 <= 0 & <=2>=0 | | | ε11 | 0 <= 0 & <=2>=0 |
| ε4 | 0 <= 0 & <=2>=0 | | | ε12 | 0 <= 0 & <=2>=0 |
| ε5 | 0 <= 0 & <=2>=0 | | | ε13 | 0 <= 0 & <=2>=0 |
| ε6 | 0 <= 0 & <=2>=0 | | | ε14 | 0 <= 0 & <=2>=0 |
| ε7 | 0 <= 0 & <=2>=0 | | | ε15 | 0 <= 0 & <=2>=0 |
| ε8 | 0 <= 0 & <=2>=0 | | | ε16 | 0 <= 0 & <=2>=0 |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

|  |  |  |  |
| --- | --- | --- | --- |
| C1 | >=1 | C9 | >=1 |
| C2 | >=1 | C10 | >=1 |
| C3 | >=1 | C11 | >=1 |
| C4 | >=1 | C12 | >=1 |
| C5 | >=1 | C13 | >=1 |
| C6 | >=1 | C14 | >=1 |
| C7 | >=1 | C15 | >=1 |
| C8 | >=1 | C16 | >=1 |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |
|  | | | | |  |

|  |  |  |
| --- | --- | --- |
| z1 | = 1 ou 0 | |
| z2 | = 1 ou 0 | |
| z3 | = 1 ou 0 | |
| z4 | = 1 ou 0 | |
| z5 | = 1 ou 0 | |
| z6 | = 1 ou 0 | |
| z7 | = 1 ou 0 | |
| 8 | = 1 ou 0 | |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |
|  | |  |

La fonction objective est « codée » dans la cellule violette (col = T, ligne = 7) de l’onglet « Q5\_Ramp\_Loss ».



=SOMMEPROD(T8:BD8;T10:BD10)

Show the equation of the two obtained separating hyperplanes.

1. Discuss the results obtained at each step with a short report And Optional Points

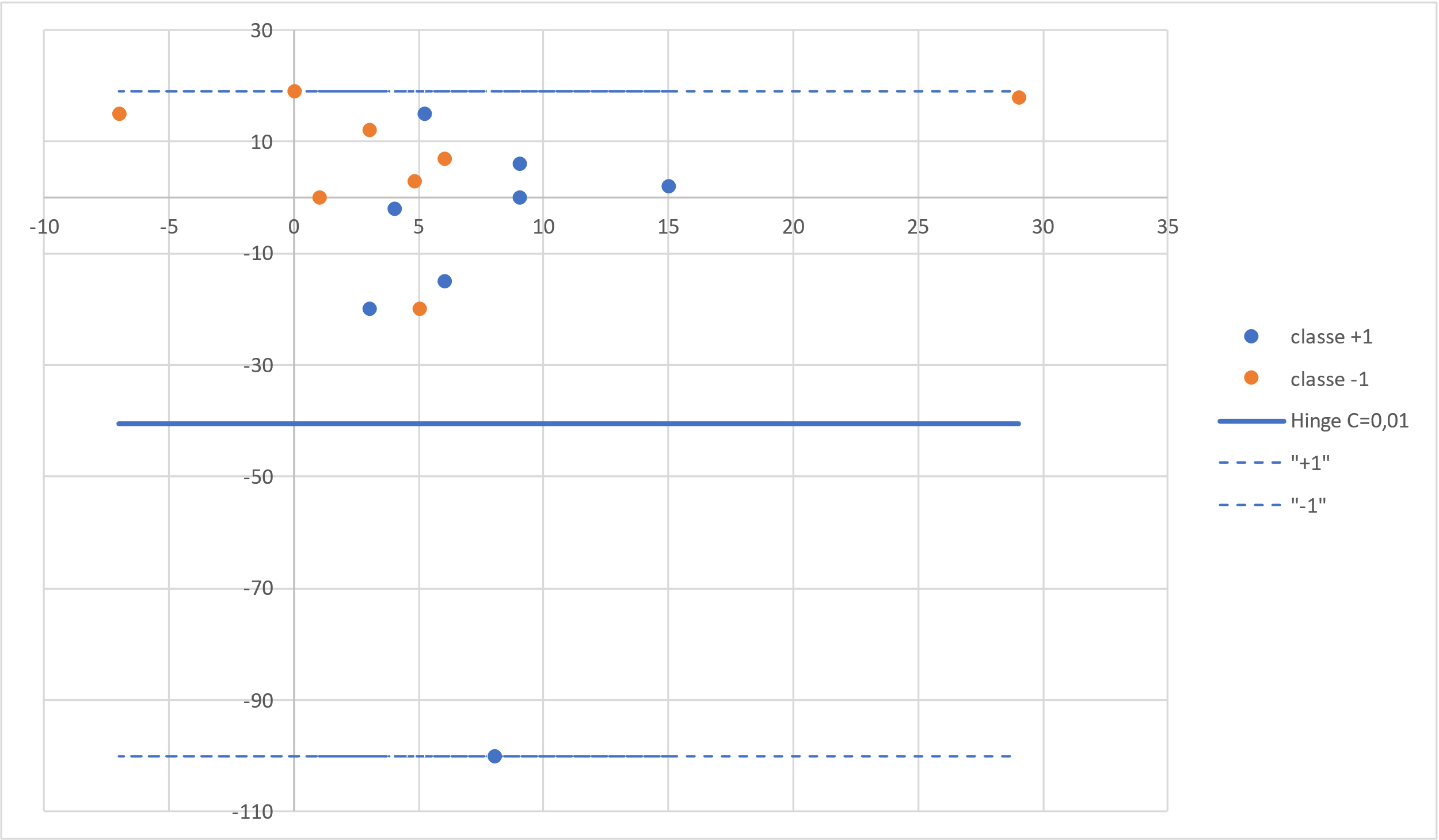
Support your discussion with plots.

Discuss how the solution changes when the weight C changes.

**Concernant la « Hinge loss** »

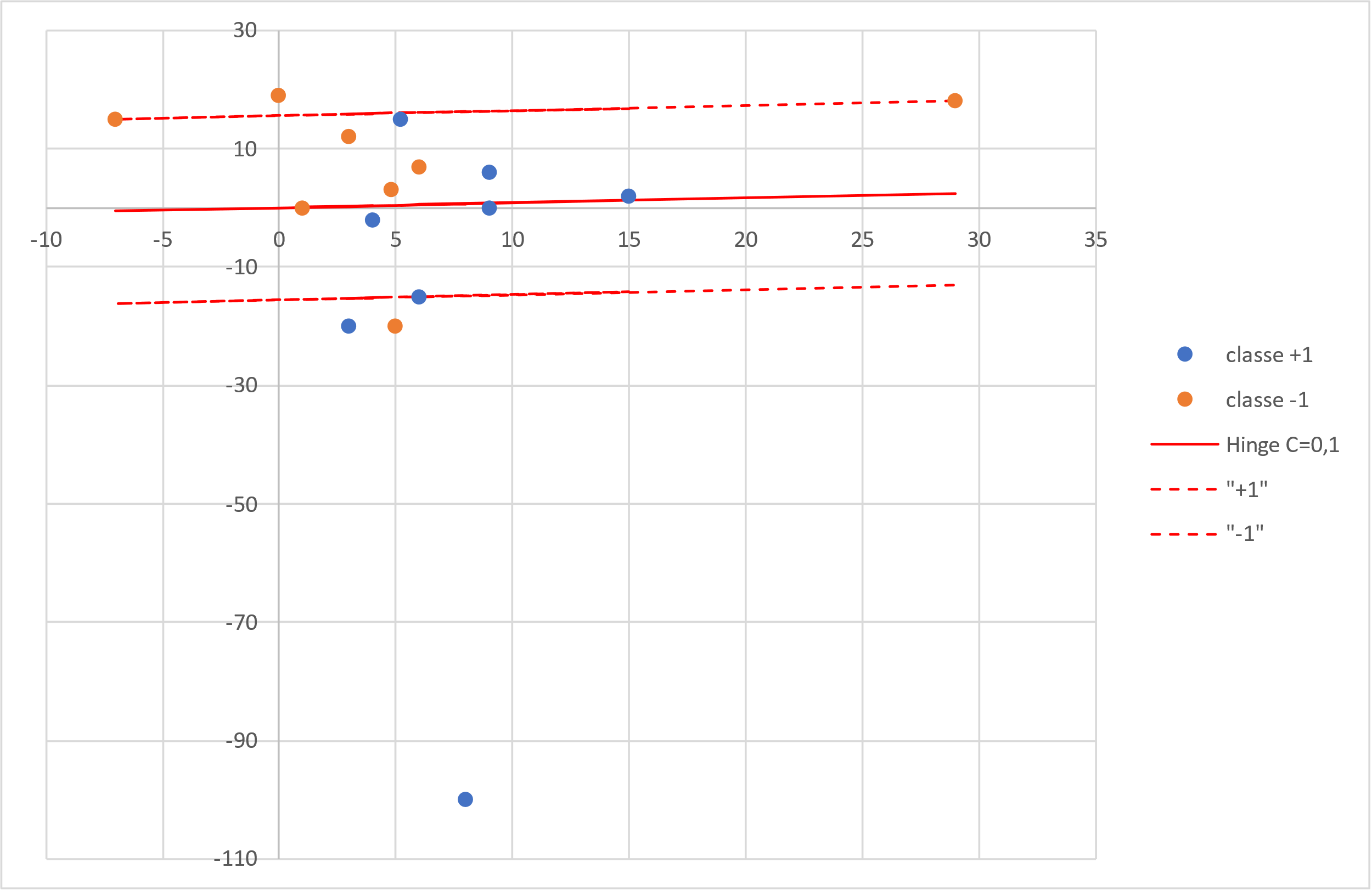
Les graphes suivant représentent les droites de séparation trouvée avec la « Hinge Loss » pour quatre valeurs de C (0,01 ; 0,1 ;1 ;10).

Avec C =0,01, le modèle abouti à un classement où 13 points sont mal-classés (7 de la classe « +1 » et 6 de la classe « -1 »). « L’accuracy » est particulièrement mauvaise (3/16). Pour rappel une valeur de C faible correspond à une large marge (ce qui est bien le cas ici) avec un modèle peu complexe et un risque « d’underfitting » (modèle trop rigide).

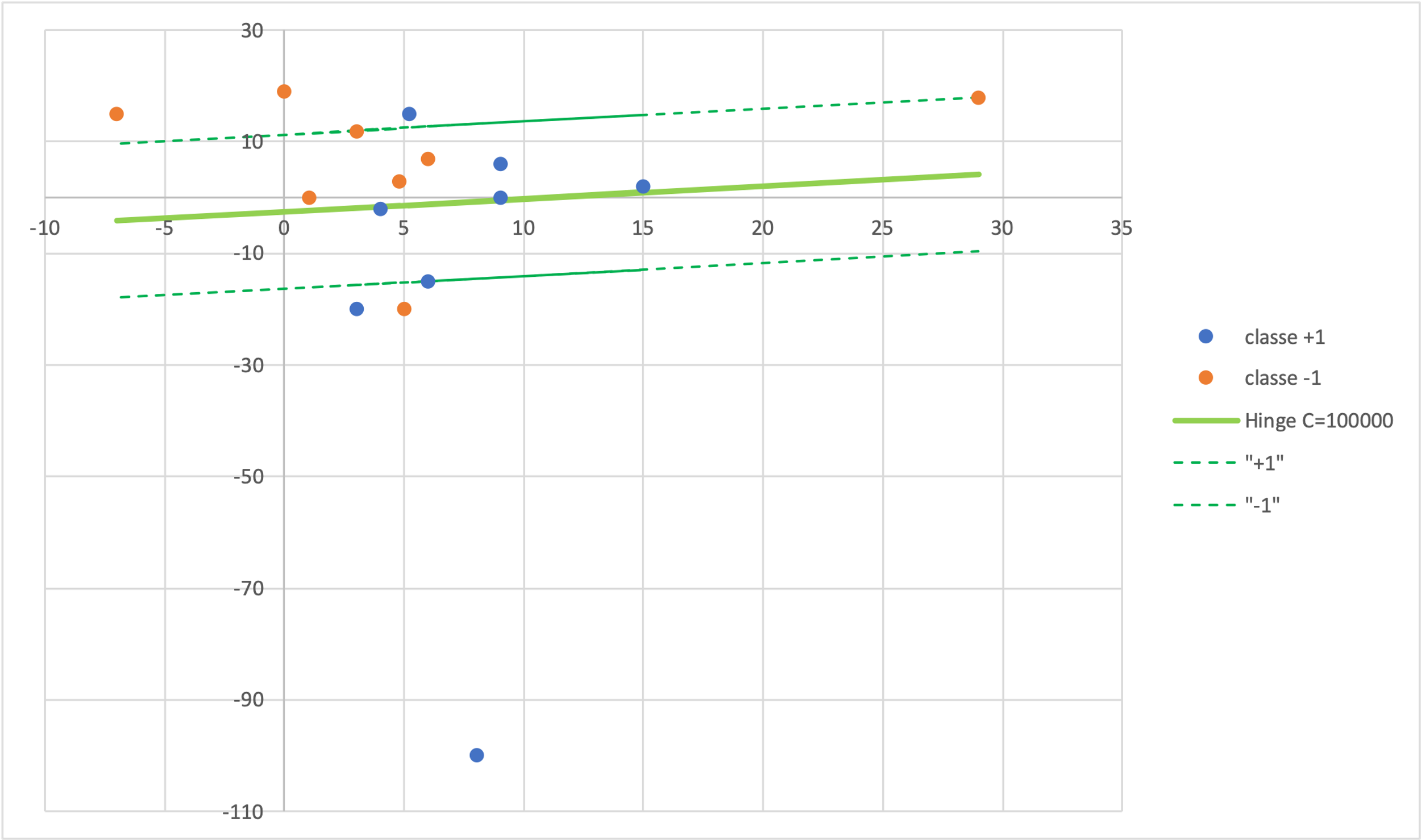


A noter que pour des valeurs de C encore plus petite (0,001), la « solution » fournie par le solver correspond au vecteur …c’est-à-dire pas d’hyperplan. Intuitivement ce résultat parait lié à l’absence de points « vecteurs supports » plus éloignés que ceux utilisés dans le modèle avec C à 0,01.

Avec C=0,1, 9 points sont mal classés (4 de la classe « +1 » et 5 de la classe « -1 »). Ce modèle est clairement moins rigide que le précédent, mais son « accuracy » reste encore assez médiocre (7/16). On constate également que la marge s’est réduite avec l’augmentation de C.



A partir de C=1, 7 points sont mal classés (4 de la classe « 1 » et 3 de la classe « -1 »). Ce modèle « apprend » un peu mieux avec une « accuracy » 9/16. A noter que pour des valeurs de C plus grande, l’hyperplan ne bouge plus. Cela serait-il dû à l’impossibilité de trouver des points supports de la classe « + 1» dans la zone entre les hyperplans  ?

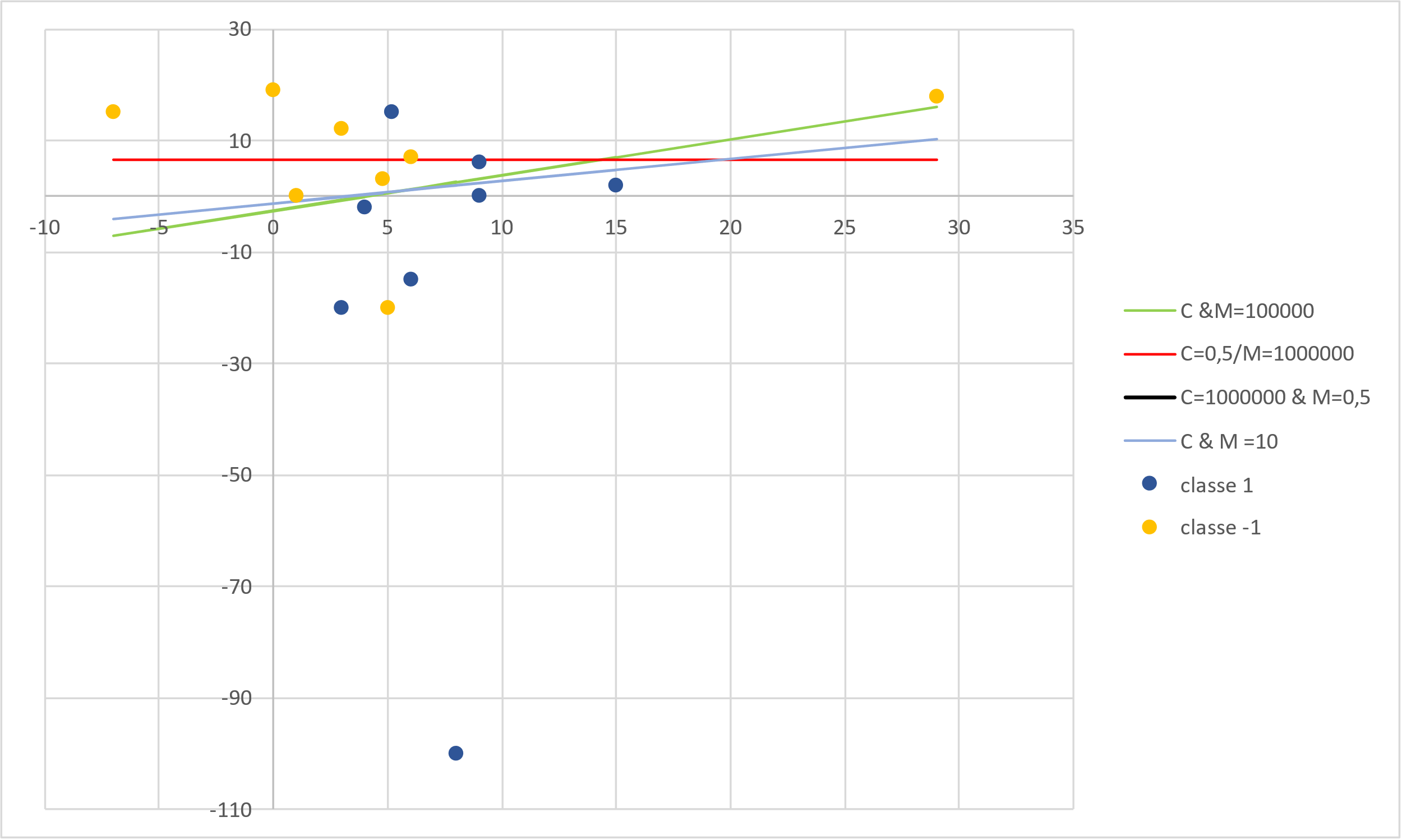


Le dernier point traité sur la « Hinge Loss » concerne l’impact de « l’outlier » sur la position de l’hyperplan (onglet Q5\_HingeLoss\_SansOutlier). En supprimant ce point, on peut constater que la définition de l’hyperplan change considérablement.

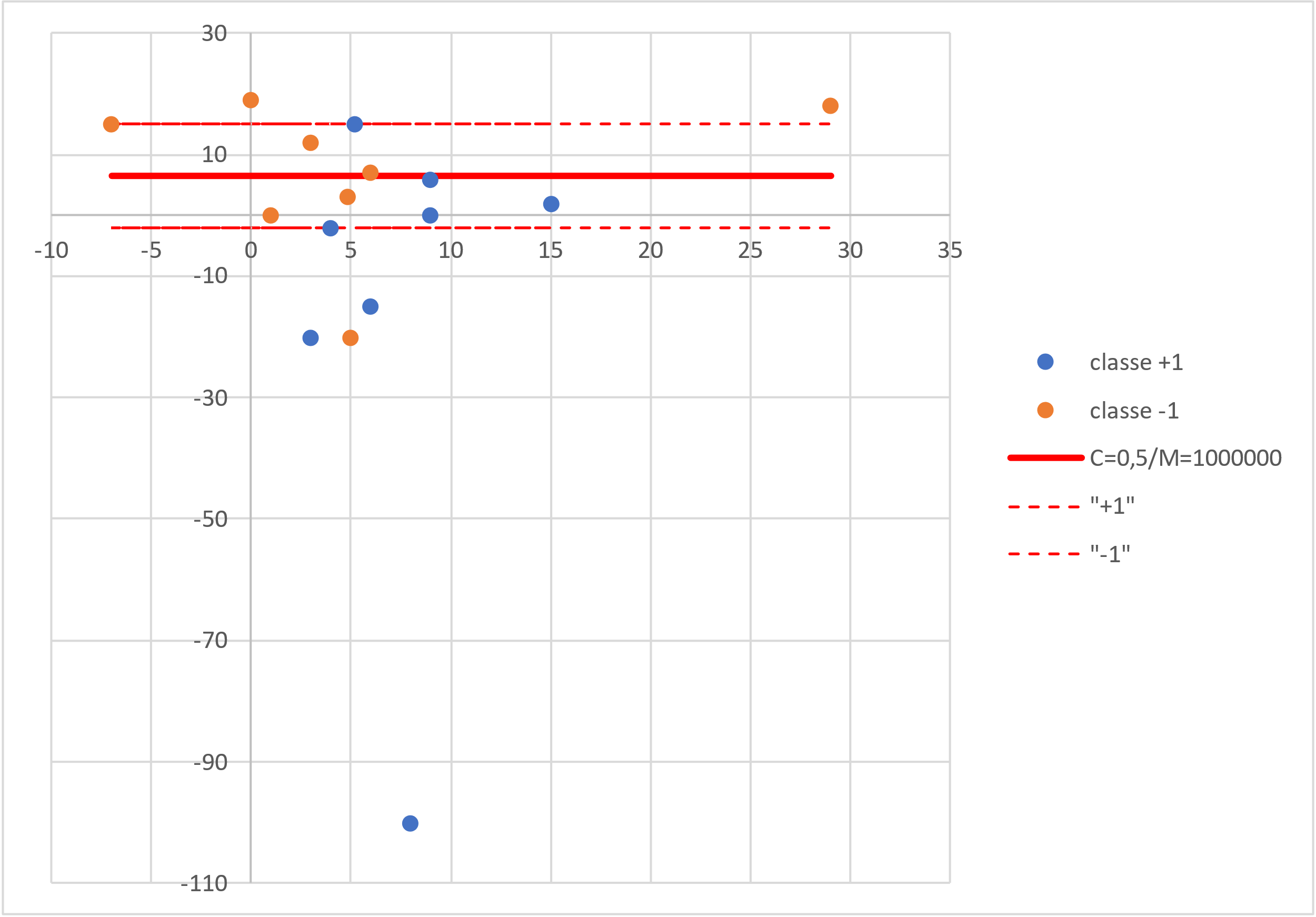


**Concernant la « Hard Margin loss** »

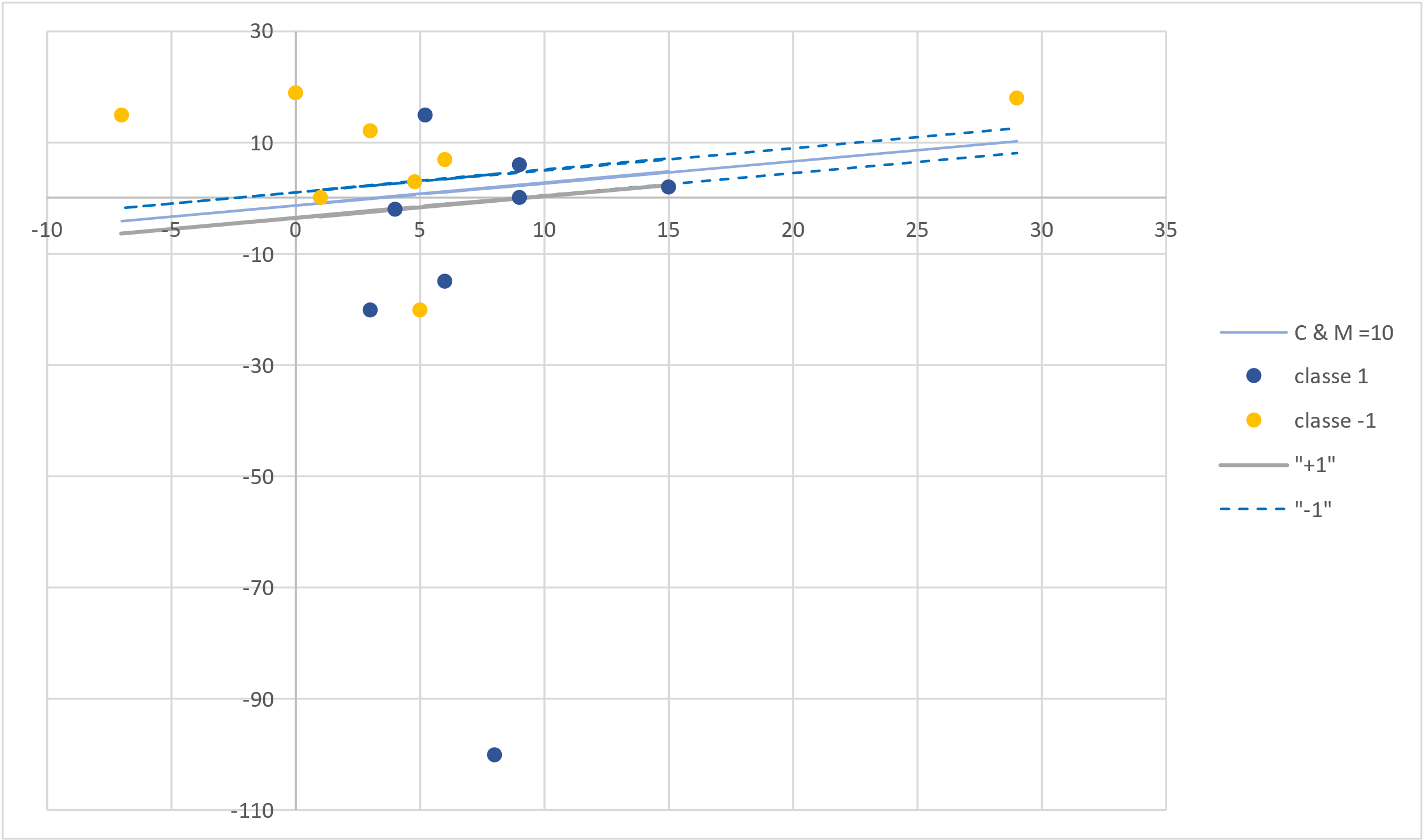
Plusieurs couples de valeurs de C&M ont été testées afin d’évaluer (en grosse maille) le spectre de comportements, avec des résultats très différents, comme le montre la figure suivante sur 4 exemples.



Le graphe suivant représente la droite de séparation trouvée pour M=1 000 000 et C=0,5. Ce modèle produit une solution avec 8 points correctement classés (« accuracy » de 1/2). A noter que le même résultat a été obtenu pour des valeurs de M entre 100 et 1 000 000. Avec des valeurs petites pour C et grandes pour M, on s’attendait à un nombre potentiellement élevé de points mal-classés et une marge relativement grande.

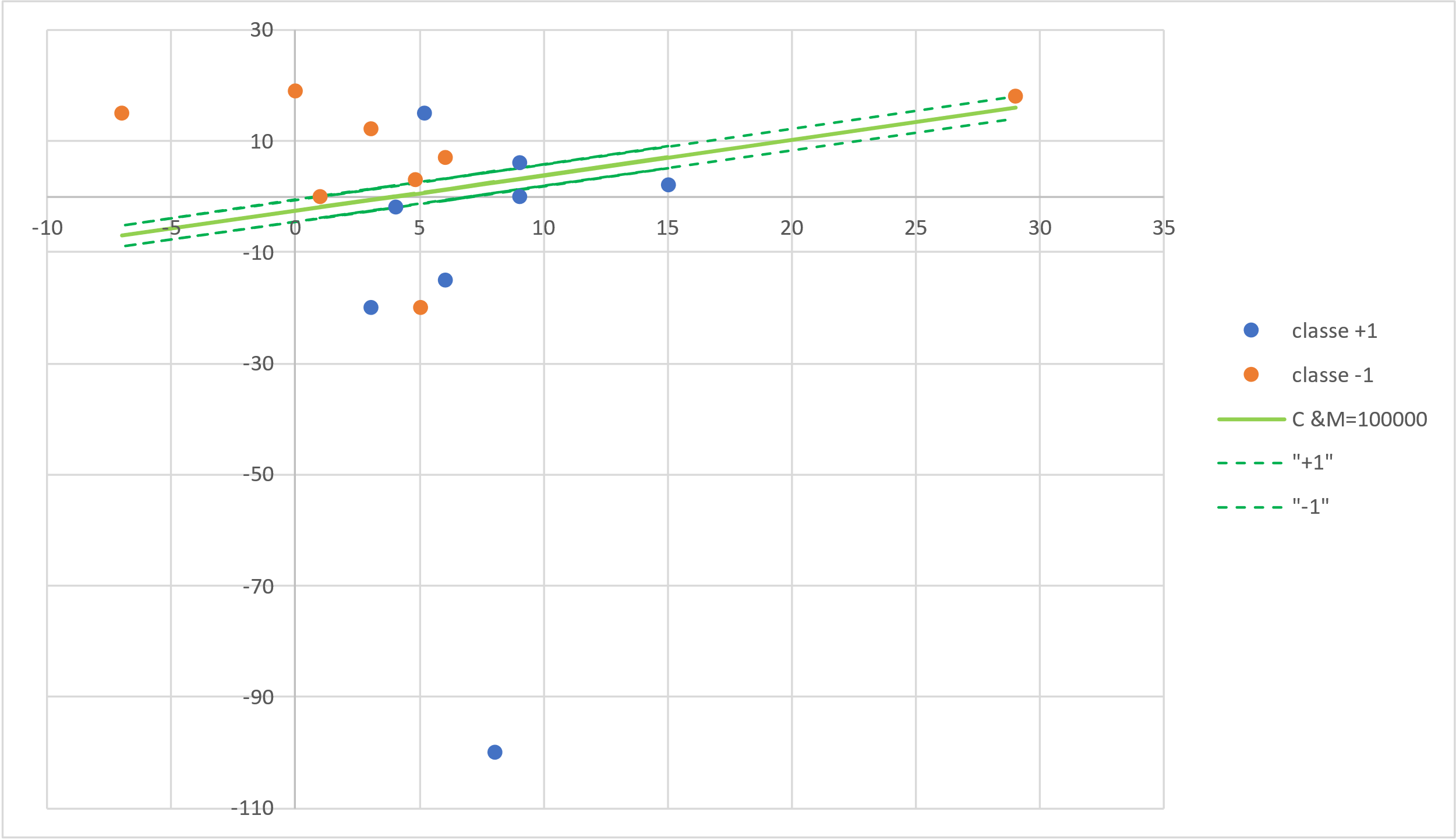


Le graphe suivant représente la droite de séparation trouvée pour M & C =10. Ce modèle produit une bonne « accuracy » de 13/16 (2 points de classe « 1 » et 1 point de classe « -& » mal-classés). La marge est nettement plus étroite que celle du modèle précédent.



Le graphe suivant représente la droite de séparation trouvée pour M & C =1 000 000. Ce modèle produit un modèle avec 13 points correctement classés (« accuracy » de 7/8) avec une marge encore plus étroite.

.



Par curiosité nous avons essayé les valeurs de 1 000 000 pour C et M=0,5. Le solver n’a pas trouvé de solution (pour rappel C très grand impose un nombre de points mal-classés faible).

Et enfin dernier point, nous avons évalué l’intolérance de la « Hard Loss » aux « outliers ». La position de l’hyperplan est modifiée mais semble-t-il dans des proportions plus réduites en comparaison de la « hinge-loss » sans cet outlier.



**Concernant la « Ramp Loss** »

Plusieurs couples de valeurs de C&M similaires aux valeurs testées avec les autres Loss pour se donner des points de comparaisons.

**De façon globale, le comportement de cette loss semble beaucoup « plus cadrée » avec des amplitudes de variations nettement plus réduites.**

Comme le montrent les figurent suivantes, les hyperplans de séparation restent dans des zones très similaires.

Ici l’hyperplan pour les valeurs de C et M égales à 1000000. On voit que l’hyperplan de la « Ramp Loss » est identique à celui de la « Hard Margin Loss ».



Ici l’hyperplan pour les valeurs de C et M égales à 10. On peut constater une légère différence entre les hyperplans de la « Ramp Loss » et de la « Hard Margin Loss ».



Ici l’hyperplan pour les valeurs de C égal à 0,5 et M égal à 1. On peut constater une différence assez nette entre les hyperplans de la « Ramp Loss » et de la « Hard Margin Loss ».

