

Machine à Vecteur de Support

SÉPARATEUR À VASTE MARGE OU SUPPORT VECTOR MACHINE (SVM)

Le SVM, introduction

2

Rafik LACHAAL

- ▶ Le **SVM** est un classificateur (donc apprentissage supervisé pour prédire une variable catégorielle).
- ▶ Il à été introduit dans les années 60 et repris dans les années 90.
- ▶ Ce n'est que récemment que cet algorithme est devenue populaire car :
 - Il a une forte capacité à travailler sur des données avec un **grand nombre de variables**.
 - Il à un **faible nombre d'hyperparamètres**.
 - Selon les données, il peut être aussi performant ou même plus qu'un réseaux de neurones.
 - il peut être très puissant pour résoudre de nombreux problèmes.

Principe de fonctionnement

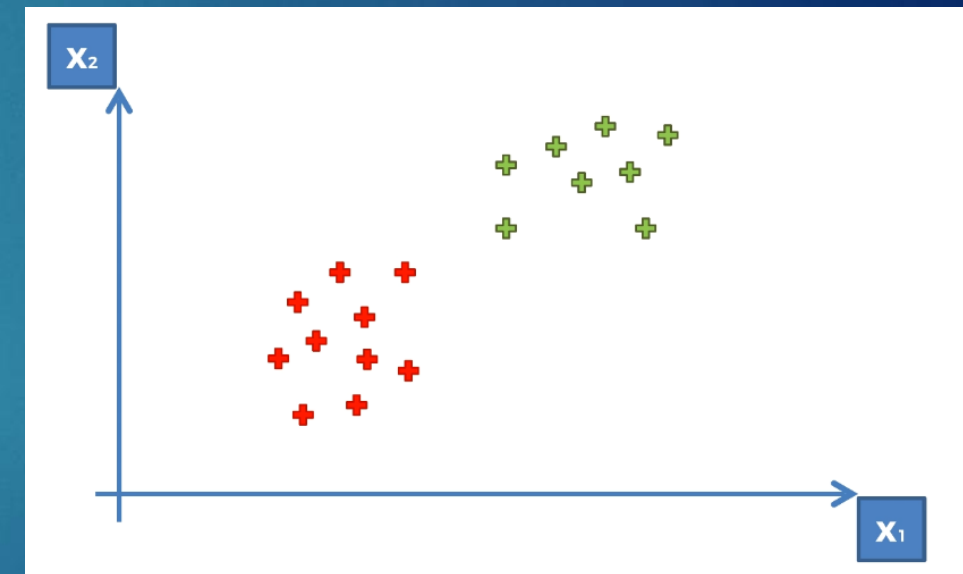
3

Rafik LACHAAL

Comme tout classificateur on cherche à faire la distinction entre deux classes (ou plus dans le cas multinomial).

Ici on a l'exemple d'une variable y (rouge ou vert) que l'on cherche à expliquer en fonction des variables X_1 et X_2 .

- **Le but est de trouver la frontière de séparation** qui nous permettra de différencier au mieux nos deux groupes.
- la frontière de séparation sera donc **une ligne droite** dont il faudra déterminer **l'équation**.



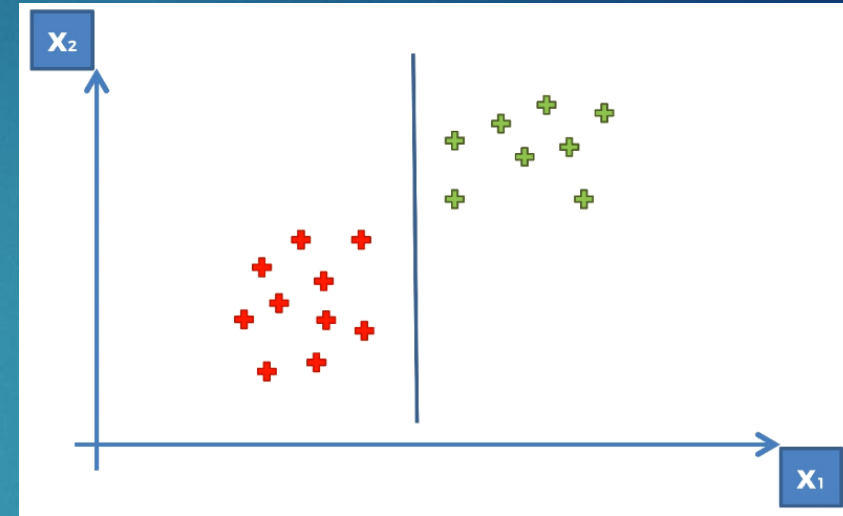
Comme pour la régression logistique le **SVM** est un **modèle linéaire**.

Comment séparer les points ?

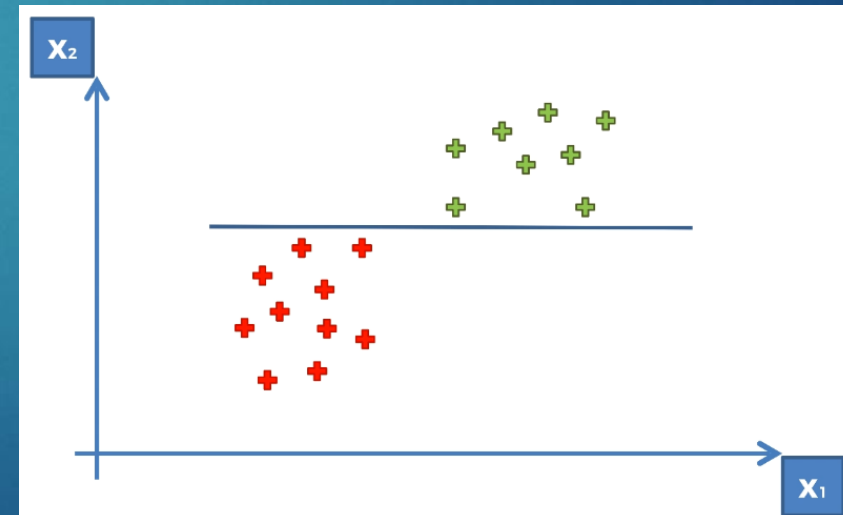
4

Rafik LACHAAL

- Séparation vertical :
Tous ce qui est à gauche de la droite est classifié en rouge et tout ce qui est à droite est classifier vert.



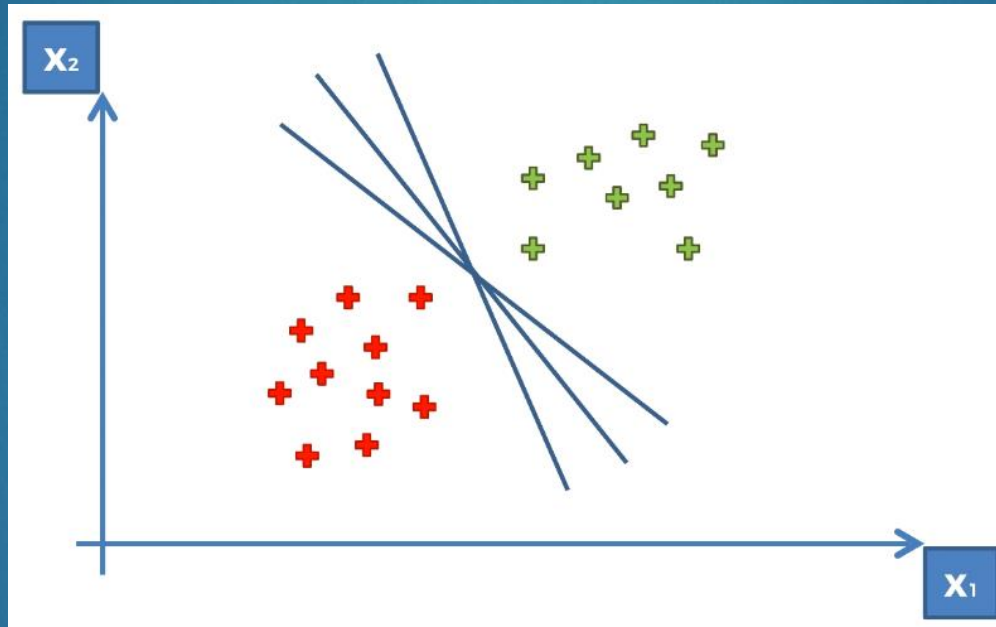
- Séparation horizontal :
Tous ce qui est en bas de la droite est classifié en rouge et tout ce qui est en haut est classifier vert.



Comment séparer les points ?

5

Au final il y a aussi une infinité de droite diagonal qui pourrait très bien séparer notre classe rouge de la classe verte.



Mais alors quel est la meilleur droite qui va le mieux séparer nos classes rouge et verte de sorte que si l'on est une nouvelle observation elle soit attribué intelligemment à l'une de ces deux classes ?

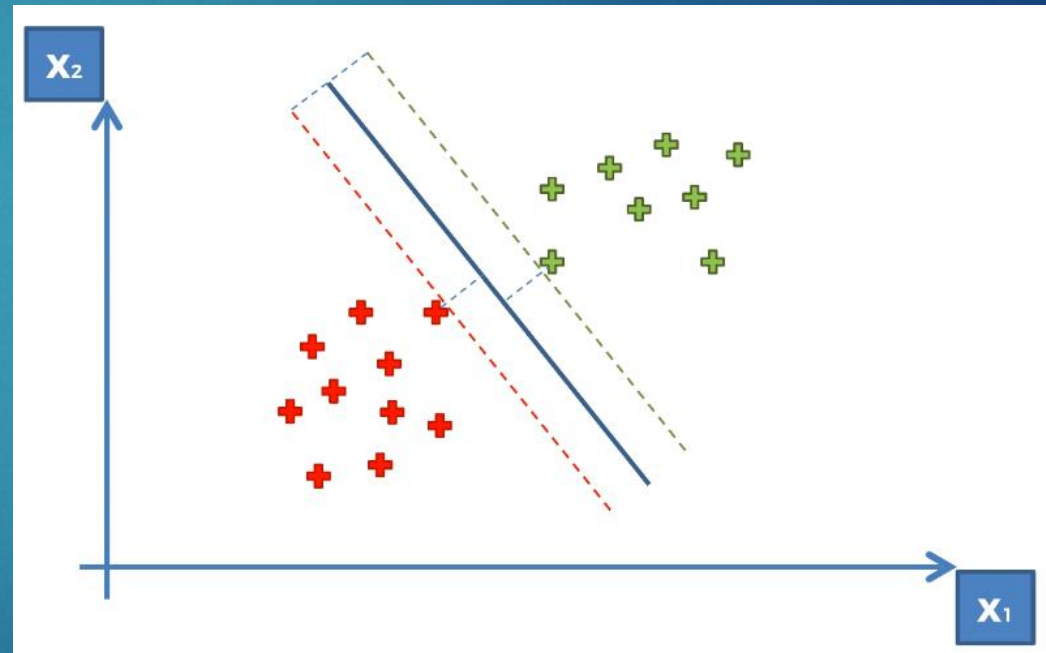
Comment séparer les points ?

6

C'est grâce au **SVM** que l'on va trouver la droite optimal :

Cette droite a été obtenue avec la méthode de **la Marge Maximale** :

- ▶ On prend les deux points rouge et vert qui sont les plus proche l'un de l'autre.
- ▶ La droite optimal sera celle qui :
 - sera équidistante à ces deux points
 - Sera à une distance maximale de ces deux points.



La Marge Maximale

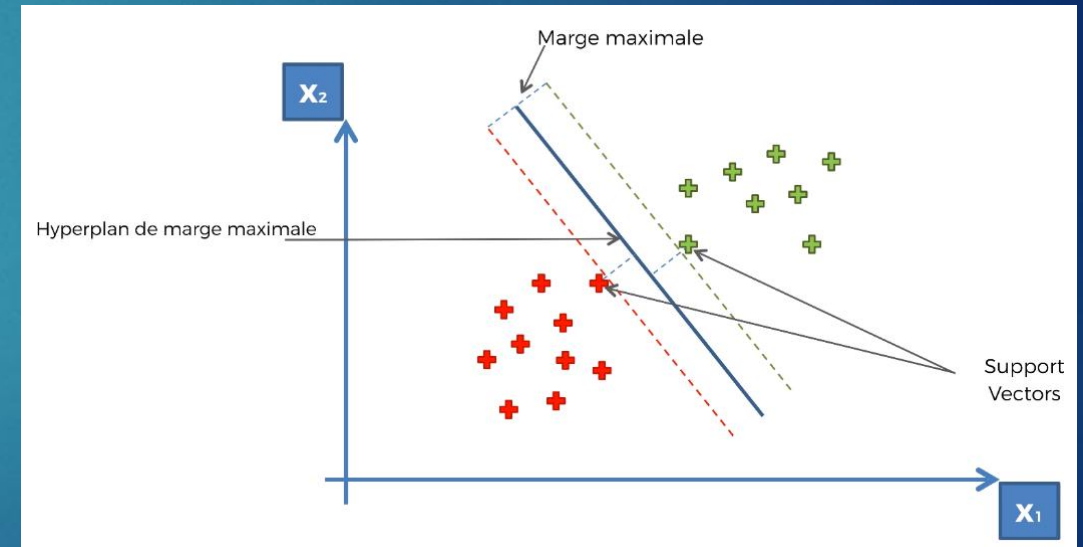
7

Rafik
LACHAAL

- ▶ La somme de la distance de ces deux points rouge et vert est appelé la **Marge Maximale**.
- ▶ Les deux points qui nous ont permis de trouver la droite optimale sont les **Support Vectors**.
 - Remarque : ces deux points suffisent à déterminer la droite optimale. Si on supprime tout les autres points le résultat sera le même.
- ▶ La droite optimale est appelé **Hyperplan de marge maximale**.

Un hyperplan est la forme généralisé d'un plan dans un espace à n dimensions.

 - Un droite dans un espace à deux variables
 - Un plan dans un espace à trois variables
 - Un hyperplan dans un espace à plus de trois variables.



Pourquoi les SVMs sont-ils si spéciaux ?

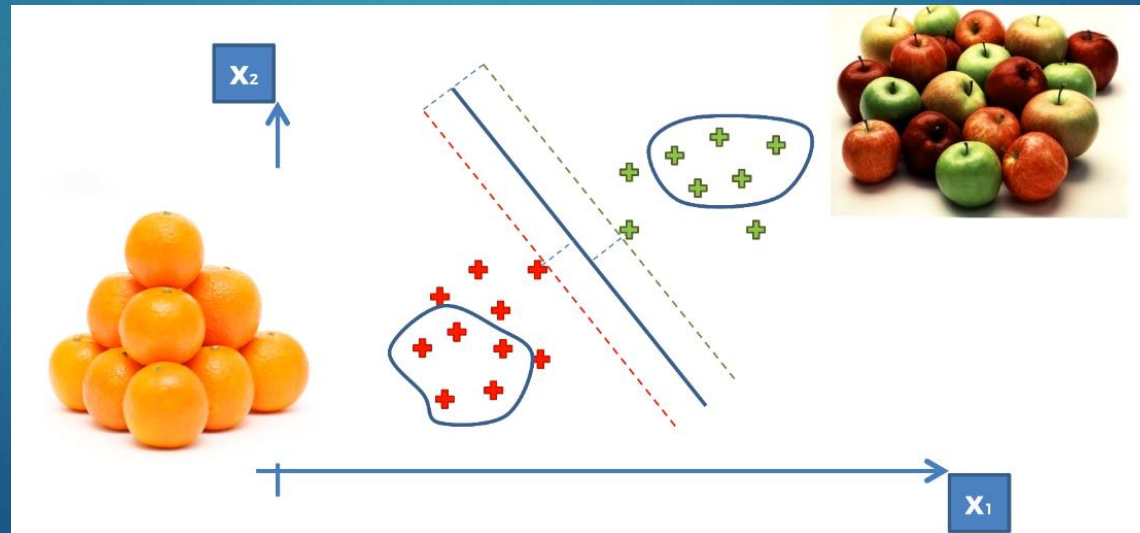
8

Rafik LACHAAL

Dans notre exemple on peut considérer que les points rouges sont des oranges et les points verts des pommes.

Pour différencier les pommes des oranges la plus part des algorithmes de machine learning vont :

- **Regarder l'ensemble des pommes et des oranges** et chercher quel sont les caractéristiques (les variables) qui permettent le plus de les différencier.



Pourquoi les SVMs sont-ils si spéciaux ?

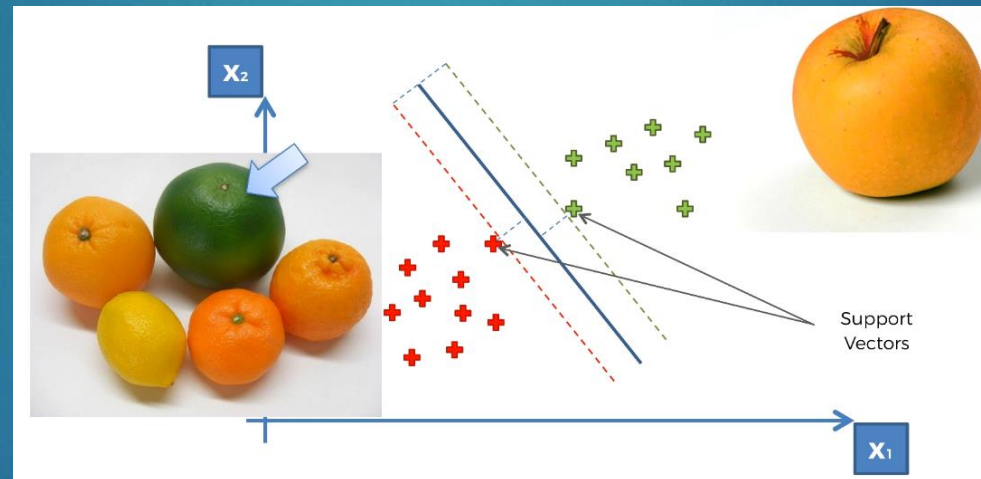
9

Rafik LACHAAL

En revanche le SVM ne vont pas regarder la totalité des observations.

Le SVM va seulement regarder :

- la pommes qui pourra le plus être confondu avec une orange
- l'orange qui pour le plus être confondue avec une pommes



Ces deux fruit vont être nos Support Vectors.

Le SVM va donc regarder seulement le cas extrême de chacune des classes, et c'est ce qui le rend si spéciale.

Pourquoi les SVMs sont-ils si spéciaux ?

10

Rafik LACHAAL

Le SVM va donc regarder seulement le cas extrême de chacune des classes, et c'est ce qui le rend si spéciale.

C'est à partir de ces deux observations qu'il va construire son analyse.

Par conséquent il ne calcule aucune probabilité comme le font la régression logistique et le Naïve Bayes.

Attention !

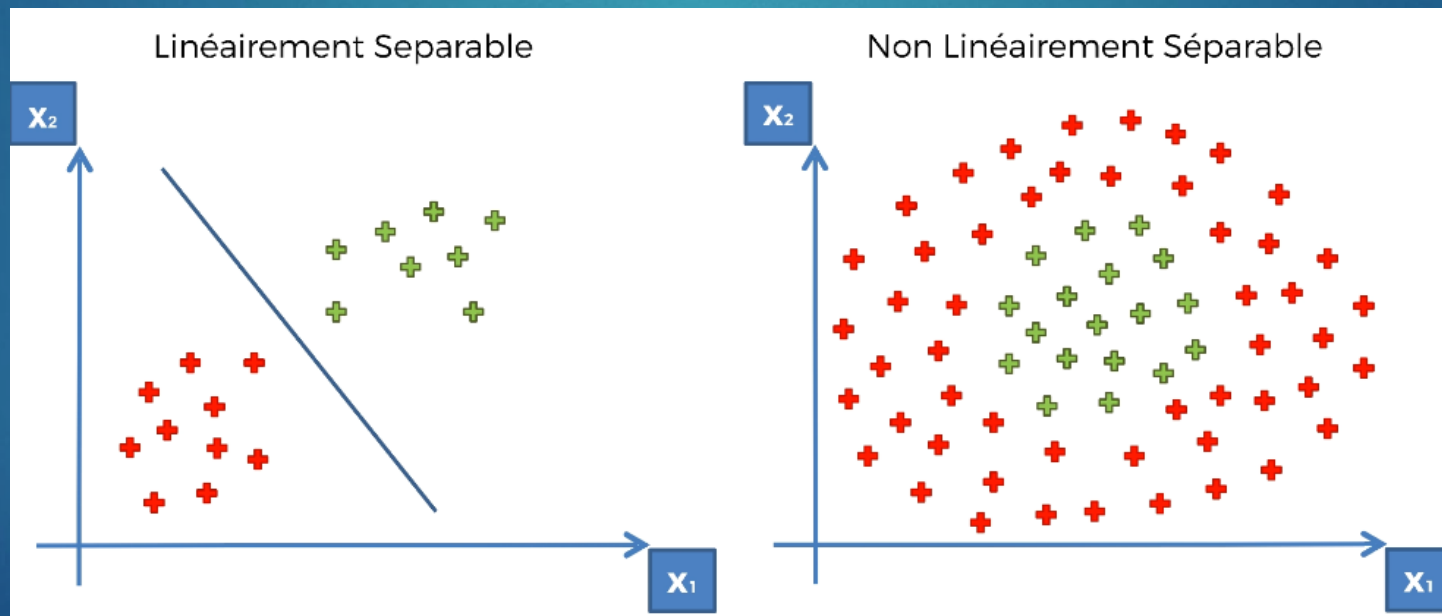
Le SVM est efficace seulement si les classes sont **linéairement séparable**, c'est-à-dire si on peut les séparer en traçant une droite.

Cas des observations non linéairement séparable

11

Rafik LACHAAL

Comment procéder dans les cas où nos observations sont non linéairement séparables ?



Principe de fonctionnement du Kernel SVM

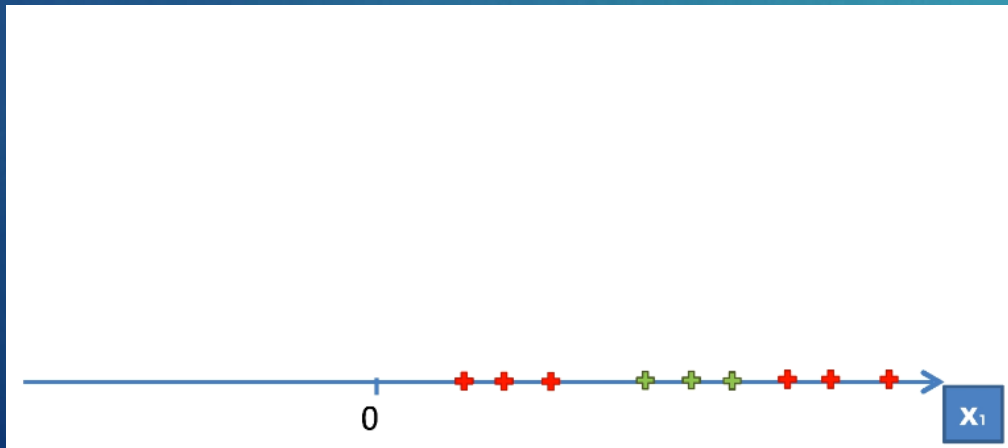
12

Rafik LACHAAL

Commençons simplement :

Ici on a un problème avec une seule variable explicative x_1 et on souhaite expliquer la couleur des points.

Comme c'est un problème à une dimension la solution (l'hyperplan) que l'on obtiendrait avec un SVM serait un point.



Or on voit assez bien qu'il n'est pas possible de séparer les points rouges et verts avec un seul point.

Cas des observations non linéairement séparable

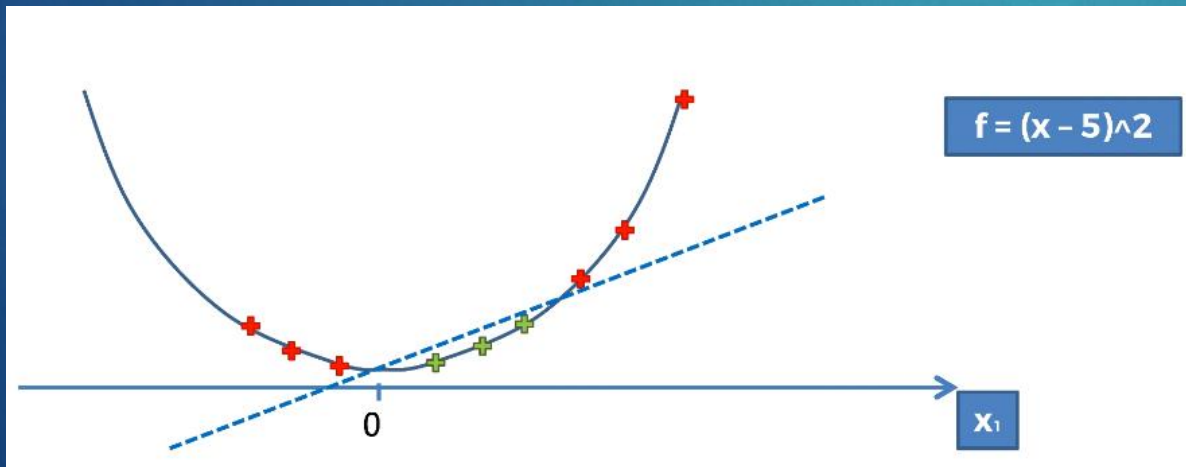
13

Rafik LACHAAL

La solution : le mapping vers une plus grande dimension

On va venir ajouter une variable créée de toute pièce qui va être une fonction de x_1 .

En faisant cela **on passe de une à deux dimensions**.



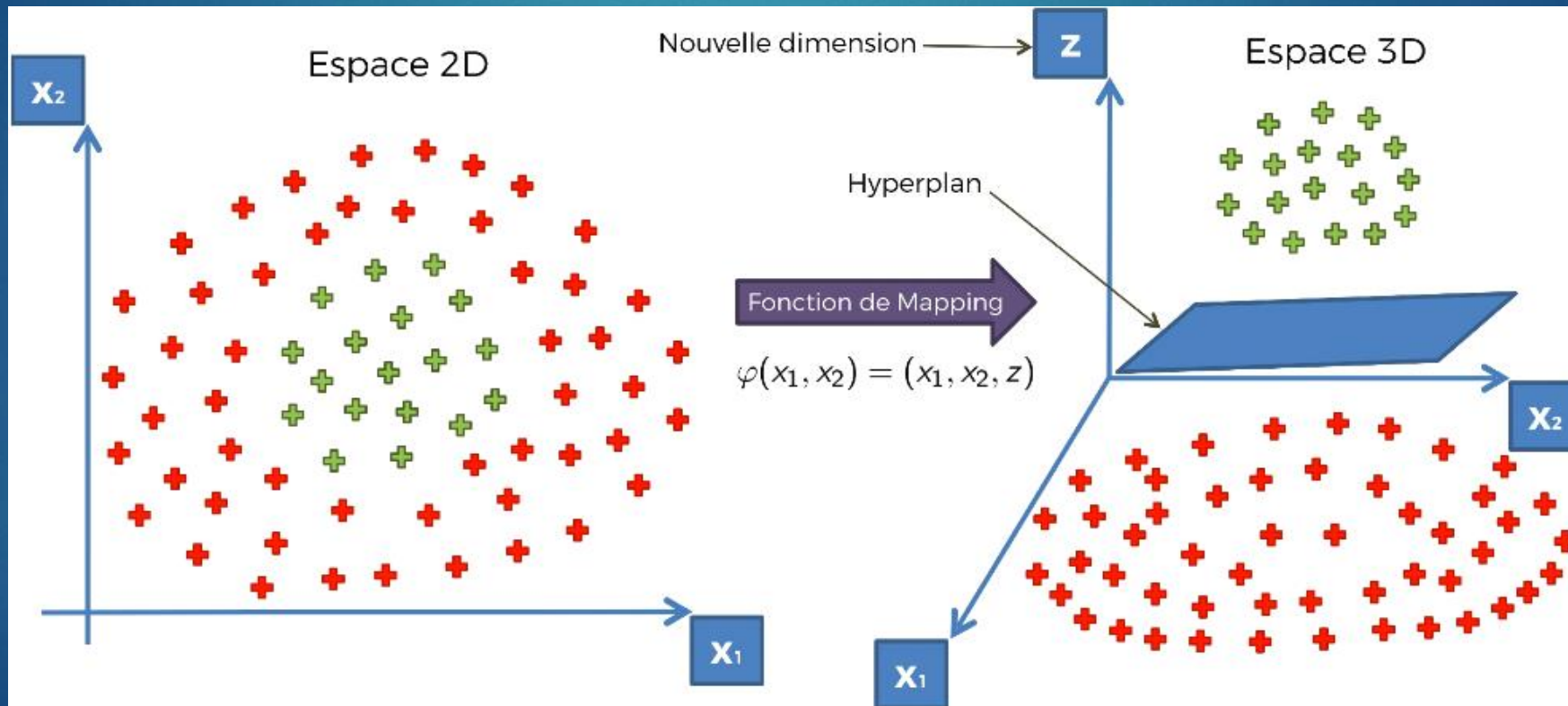
En utilisant la fonction $f = (x - 5)^2$ on peut maintenant tracer une droite qui va séparer nos points.

Cas des observations non linéairement séparable

14

Rafik LACHAAL

On peut généraliser le mapping vers une plus grande dimension

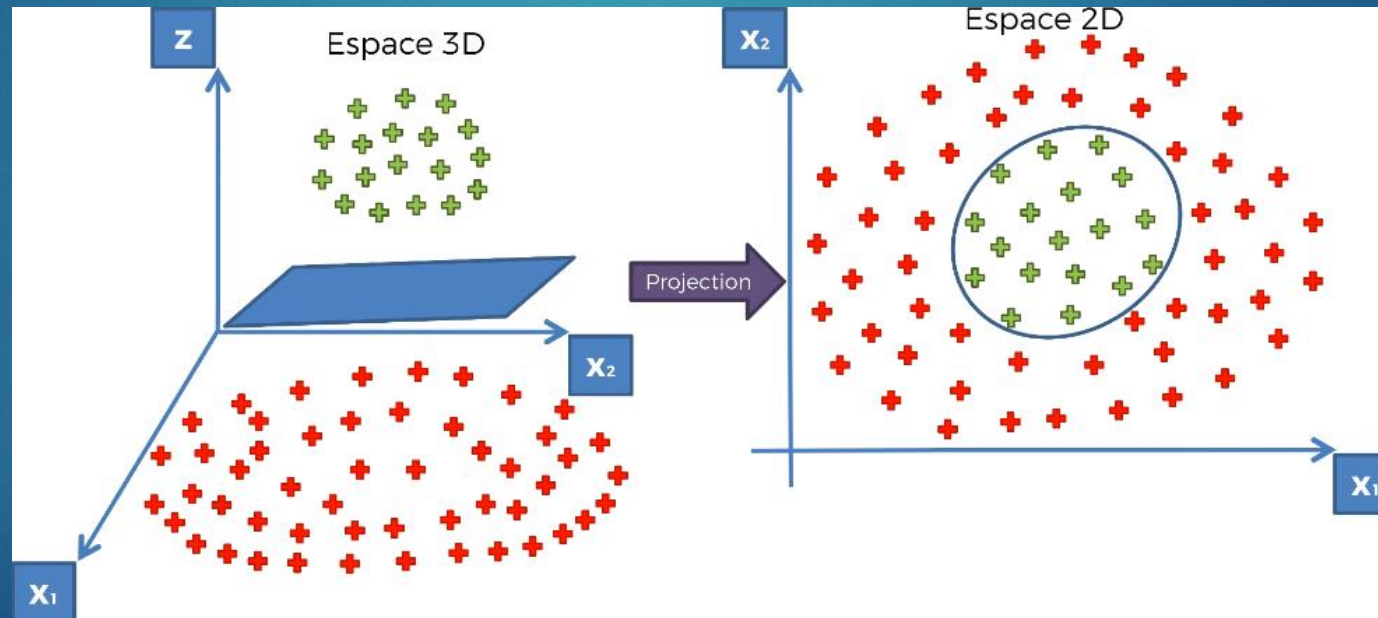


Cas des observations non linéairement séparable

15

Rafik LACHAAL

Une fois l'hyperplan trouvé il ne reste plus qu'à le projeter vers notre espace original à deux dimensions qui devient une courbe élliptiques.



Le Kernel SVM

16

Rafik LACHAAL

Fort heureusement en pratique avec **Sklearn** il n'y pas besoin de créer ces variables supplémentaires ni même d'utiliser le PolynomialFeatures.

Afin de le rendre non linéaire on va venir lui implémenter **un Kernel**.

Les deux types de Kernel les plus utilisés sont :

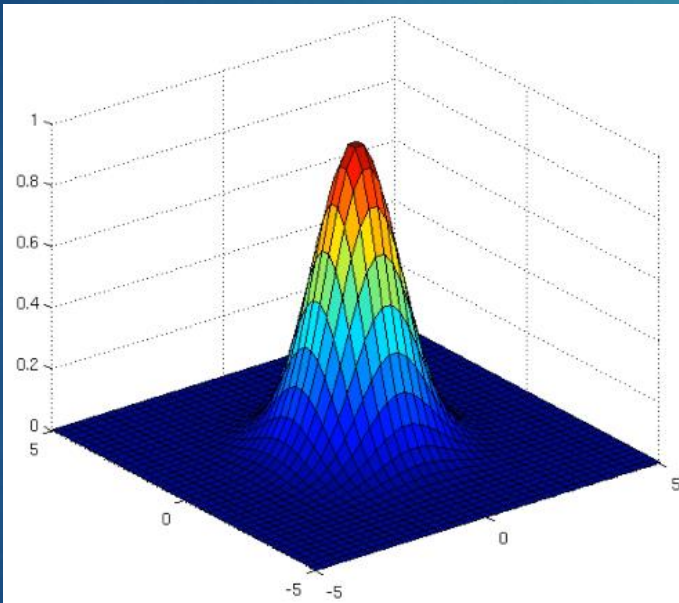
- ▶ Le Kernel polynomial (On choisi le degré du polynôme)
- ▶ Le Kernel radial Gaussien ou RBF Kernel (Aussi appelé méthode des variables de similarités).

Le Kernel SVM

17

Rafik LACHAAL

Autre méthode : le Gaussian RBF Kernel



$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x} - \vec{l}^i\|^2}{2\sigma^2}}$$

- ▶ L est le point de repère (landmark)
- ▶ On peut réécrire la fonction sous la forme :

$$K(x, l) = \exp(-\gamma \|x - l\|^2)$$

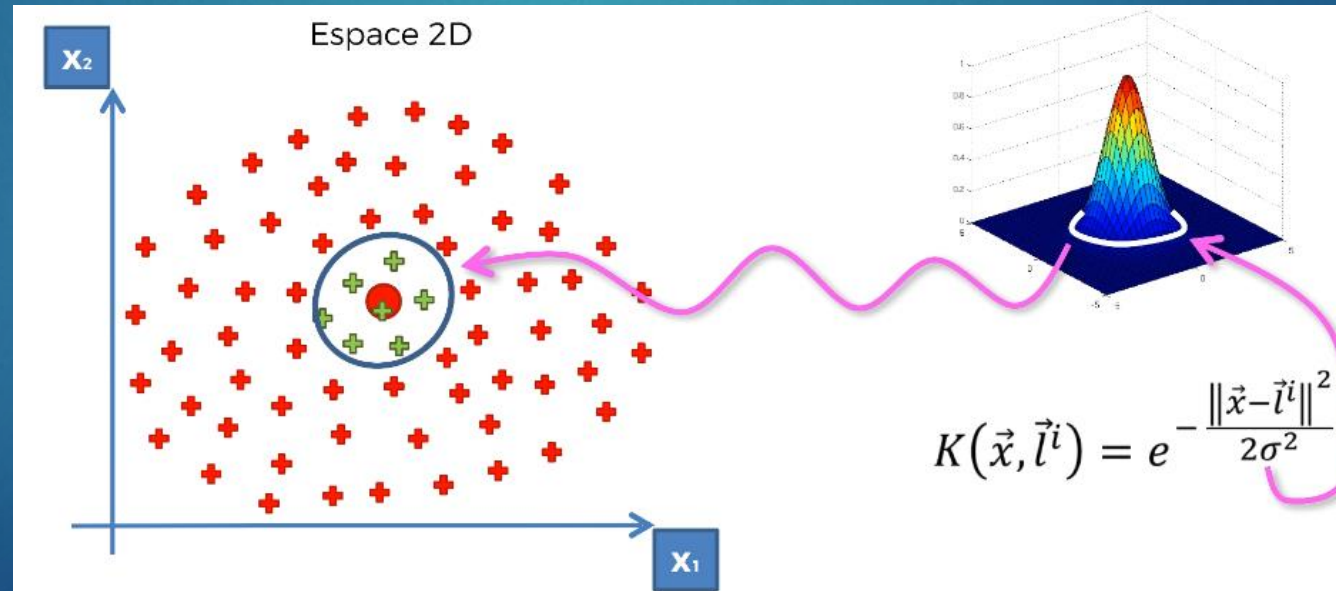
Le Kernel SVM

18

Rafik LACHAAL

Le paramètre gamma :

- Plus γ est faible et plus la circonférence du cercle blanc va se réduire et donc englober moins de points.
- Plus γ est élevé et plus la circonférence du cercle blanc va se s'agrandir et donc englober plus de points.



Le Kernel SVM

19

Rafik LACHAAL

Le paramètre gamma :

- On peut aussi combiner différent Gaussian RBF Kernel pour résoudre des problèmes plus complexes.

