Hyperplans

Dans un espace de dimension n, un hyperplan est un espace de dimension $n-1$

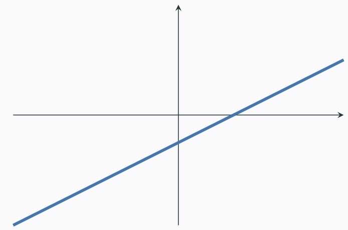
- En dimension 1, les hyperplans sont les points

Hyperplan en dimension 1

Hyperplans

Dans un espace de dimension n, un hyperplan est un espace de dimension $n-1$

- En dimension 2, les hyperplans sont les droites

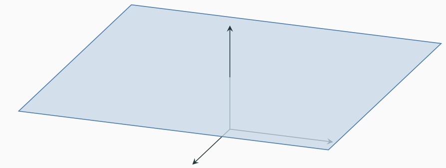


Hyperplan en dimension 2

Hyperplans

Dans un espace de dimension n, un hyperplan est un espace de dimension $n-1$

- En dimension 3, les hyperplans sont les plans



Hyperplan en dimension 3

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 1 : demi-droites

Demi-espaces en dimension 1

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 1 : demi-droites

Demi-espaces en dimension 1

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

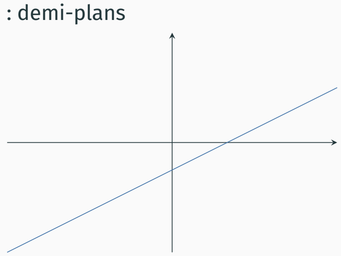
- En dimension 1 : demi-droites

Demi-espaces en dimension 1

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 2 : demi-plans

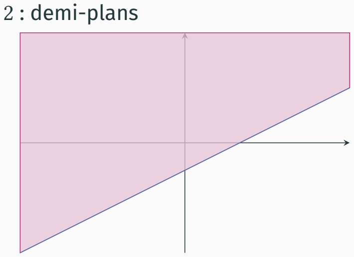


Demi-espaces en dimension 2

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 2 : demi-plans

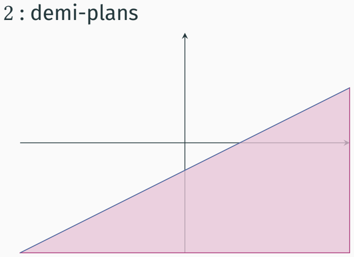


Demi-espaces en dimension 2

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 2 : demi-plans

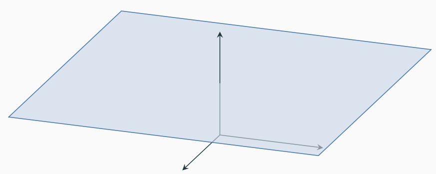


Demi-espaces en dimension 2

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 3 : demi-espaces

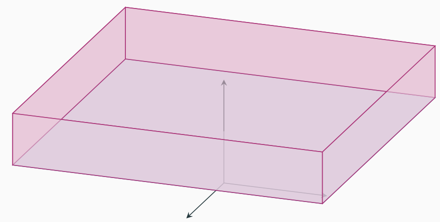


Demi-espaces en dimension 3

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 3 : demi-espaces

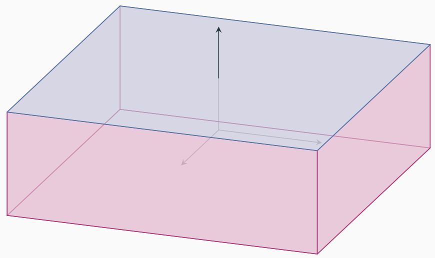


Demi-espaces en dimension 3

Propriétés des hyperplans

Les hyperplans séparent les espaces en demi-espaces

- En dimension 3 : demi-espaces

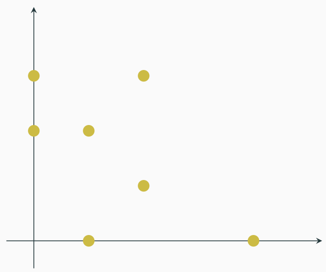


Demi-espaces en dimension 3

Classifieurs linéaires

Classifieurs linéaires

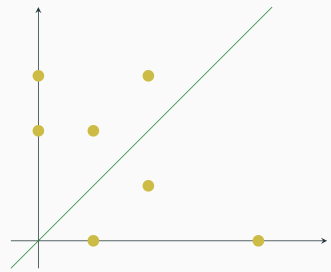
La propriété de séparation des hyperplans permet de construire des classifieurs



Classification en dimension 2

Classifieurs linéaires

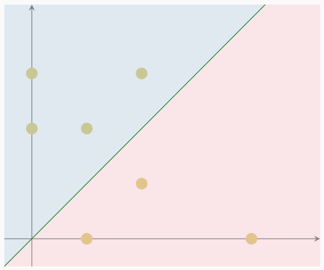
La propriété de séparation des hyperplans permet de construire des classifieurs



Classification en dimension 2

Classifieurs linéaires

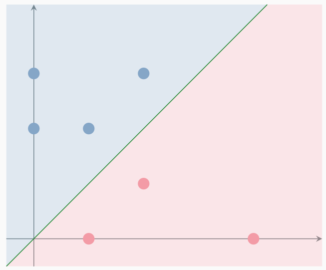
La propriété de séparation des hyperplans permet de construire des classifieurs



Classification en dimension 2

Classifieurs linéaires

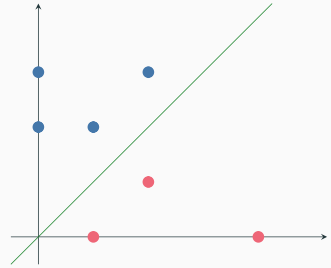
La propriété de séparation des hyperplans permet de construire des classifieurs



Classification en dimension 2

Classifieurs linéaires

La propriété de séparation des hyperplans permet de construire des classifieurs



Classification linéaire en dimension 2

On parle de classifieur linéaire.

Classifieurs linéaires

Le principe des classifieurs linéaire est donc, étant donné un problème à deux classes

- Trouver un hyperplan qui sépare l'espace des attributs en demi-espaces contenant chacun tous les exemples d'entraînement d'une seule classe
- On classe les exemples de test dans la classe correspondant à leur demi-espace

Classifieurs linéaires

Le principe des classifieurs linéaire est donc, étant donné un problème à deux classes

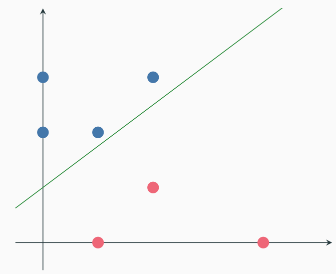
- Trouver un hyperplan qui sépare l'espace des attributs en demi-espaces contenant chacun tous les exemples d'entraînement d'une seule classe
- On classe les exemples de test dans la classe correspondant à leur demi-espace

Cela suppose de résoudre deux problèmes

- Existe-t-il un tel hyperplan ? (on dit dans ce cas que le problème est *linéairement séparable*)
- S'il en existe un, il en existe en général une infinité
  → Lequel choisir ?

Notion de marge

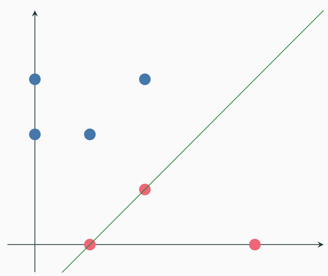
Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



Largeurs de séparation d'hyperplans

Notion de marge

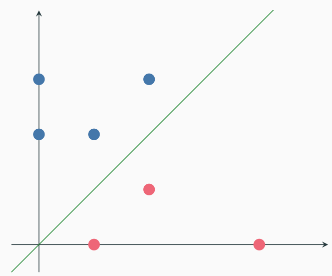
Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



Largeurs de séparation d'hyperplans

Notion de marge

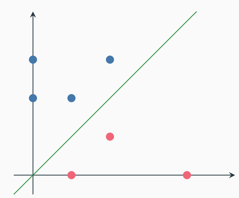
Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



Largeurs de séparation d'hyperplans

Notion de marge

Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



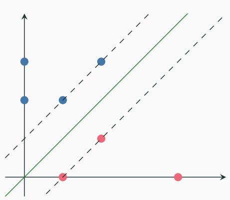
Marge

Définition (marge)
Soit deux classes, C et C', on appelle *marge* d'un hyperplan séparateur H le nombre

$$\mathrm{M}(\mathrm{H}) = \min_{x \in \mathrm{C}} d(x, \mathrm{H}) + \min_{x' \in \mathrm{C}'} d(x', \mathrm{H})$$

Notion de marge

Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



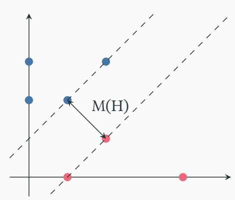
Marge

Définition (marge)
Soit deux classes, C et C', on appelle *marge* d'un hyperplan séparateur H le nombre

$$\mathrm{M}(\mathrm{H}) = \min_{x \in \mathrm{C}} d(x, \mathrm{H}) + \min_{x' \in \mathrm{C}'} d(x', \mathrm{H})$$

Notion de marge

Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



Marge

Définition (marge)
Soit deux classes, C et C', on appelle *marge* d'un hyperplan séparateur H le nombre

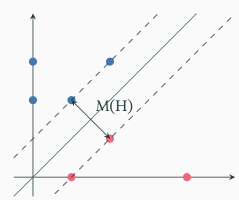
$$\mathrm{M}(\mathrm{H}) = \min_{x \in \mathrm{C}} d(x, \mathrm{H}) + \min_{x' \in \mathrm{C}'} d(x', \mathrm{H})$$

On choisit H pour lequel $\mathrm{M}(\mathrm{H})$ est maximale sur l'ensemble d'entraînement.

→ D'où *séparateur à vaste marge*.

Notion de marge

Intuitivement, pour espérer qu'un classifieur linéaire généralise bien, il faut qu'il sépare les classes le plus largement possible.



Marge

Définition (marge)
Soit deux classes, C et C', on appelle *marge* d'un hyperplan séparateur H le nombre

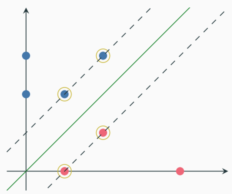
$$\mathrm{M}(\mathrm{H}) = \min_{x \in \mathrm{C}} d(x, \mathrm{H}) + \min_{x' \in \mathrm{C}'} d(x', \mathrm{H})$$

On choisit H pour lequel $\mathrm{M}(\mathrm{H})$ est maximale sur l'ensemble d'entraînement.

→ D'où *séparateur à vaste marge*.

Vecteurs de support

Les points x et x' pour lesquels les minima sont atteints sont appelés vecteurs de support



Vecteurs de support

On peut montrer que H ne dépend que de la position des vecteurs de support.

SVM

Support Vector Machines : *machines à vecteurs de support* ou *Séparateur à Vaste Marge* (V. N. Vapnik et Chervonenkis 1964; Boser, Guyon et V. N. Vapnik 1992; Cortes et V. Vapnik 1995)

Dans Weka classifiers > functions > SMO

SVM

Support Vector Machines : *machines à vecteurs de support* ou *Séparateur à Vaste Marge* (V. N. Vapnik et Chervonenkis 1964; Boser, Guyon et V. N. Vapnik 1992; Cortes et V. Vapnik 1995)

Dans Weka classifiers > functions > SMO

Espace de recherche Ensemble des hyperplans affines de $\mathbb{R}^n$, où n est le nombre d'attributs.

SVM

Support Vector Machines : *machines à vecteurs de support* ou *Séparateur à Vaste Marge* (V. N. Vapnik et Chervonenkis 1964; Boser, Guyon et V. N. Vapnik 1992; Cortes et V. Vapnik 1995)

Dans Weka classifiers > functions > SMO

Espace de recherche Ensemble des hyperplans affines de $\mathbb{R}^n$, où n est le nombre d'attributs.

Technique de recherche

- Traditionnellement : problème d'optimisation quadratique
- Plus récemment : descente de sous-gradient, descente de gradient par coordonnées
  → Linéaire ou quasi-linéaire (au moins en pratique)

Classifieurs linéaires

Raffinements

Problèmes non-séparables

En général, les problèmes réels ne sont pas linéairement séparables. Plusieurs méthodes existent pour y adapter les SVM.

Marge douce Modifier la définition de *marge* pour tenir compte des exemples mal classés

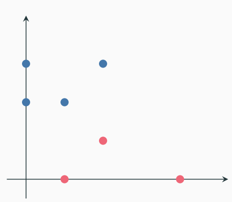
Kernel trick Transformer le problème en un problème équivalent mais linéairement séparable

Marge douce

On remplace la fonction de marge M par une fonction de la forme

$$\mathrm{L(H)} = \mathrm{M(H)} - \mathrm{C(H)}$$

où $\mathrm{C(H)}$ est une pénalité pour les exemples mal classés par H.



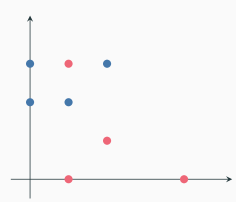
Hyperplan à marge douce maximale

Marge douce

On remplace la fonction de marge M par une fonction de la forme

$$\mathrm{L(H)} = \mathrm{M(H)} - \mathrm{C(H)}$$

où $\mathrm{C(H)}$ est une pénalité pour les exemples mal classés par H.



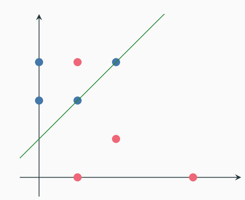
Hyperplan à marge douce maximale

Marge douce

On remplace la fonction de marge M par une fonction de la forme

$$\mathrm{L(H)} = \mathrm{M(H)} - \mathrm{C(H)}$$

où $\mathrm{C(H)}$ est une pénalité pour les exemples mal classés par H.



Hyperplan à marge douce maximale

On choisit comme précédemment l'hyperplan H qui maximise $\mathrm{L(H)}$.

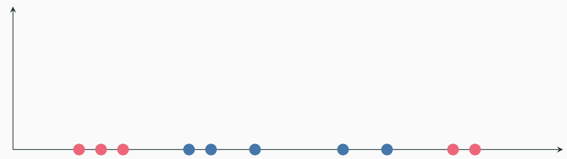
Kernel trick

L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.

Hyperplan séparateur en dimension supérieure

Kernel trick

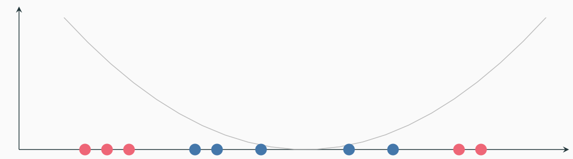
L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.



Hyperplan séparateur en dimension supérieure

Kernel trick

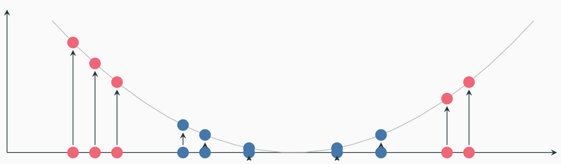
L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.



Hyperplan séparateur en dimension supérieure

Kernel trick

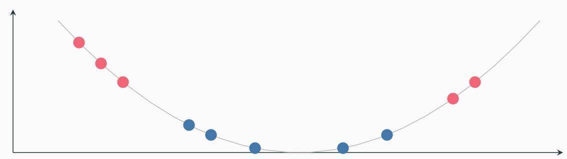
L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.



Hyperplan séparateur en dimension supérieure

Kernel trick

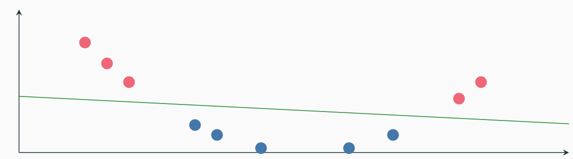
L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.



Hyperplan séparateur en dimension supérieure

Kernel trick

L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.

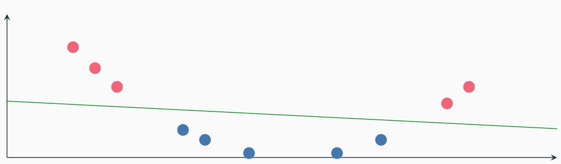


Hyperplan séparateur en dimension supérieure

- Les *noyaux* en question sont des fonctions qui permettent d'éviter d'avoir à calculer explicitement les transformations pour passer en dimension supérieure.

Kernel trick

L'*astuce du noyau* consiste à transformer un problème non-linéairement séparable en un problème linéairement séparable dans un espace de dimension supérieure.



Hyperplan séparateur en dimension supérieure

- En réalité, le problème en dimension supérieure est souvent lui aussi non-linéairement séparable, et on combine en général le *kernel trick* avec une marge douce.

Propriétés

Les SVM sont des modèles très efficaces, qui jusqu'à récemment étaient systématiquement les plus performants.

- Une fois l'apprentissage fait, la classification est très rapide
- Les modèles sont peu gourmands en mémoire
- Les modèles marchent mieux que les autres modèles présentés (en général)

Propriétés

Les SVM sont des modèles très efficaces, qui jusqu'à récemment étaient systématiquement les plus performants.

- Une fois l'apprentissage fait, la classification est très rapide
- Les modèles sont peu gourmands en mémoire
- Les modèles marchent mieux que les autres modèles présentés (en général)

Avec quelques inconvénients

- Les modèles sont difficilement interprétables
- Le choix d'un noyau tient plus de l'artisanat que de la science
- Les algorithmes classiques ne permettent pas d'apprentissage incrémental.