

CM 4

Méthodes de Simulation Informatique

Amaya Nogales Gómez
amaya.nogales-gomez@univ-cotedazur.fr

Licence 3 Informatique
Université Côte d'Azur

4 mars 2022

Plan du cours

- ① Introduction
 - Préliminaires
 - Python: numpy, pandas
- ② Base de données
 - Generation des données synthétiques
 - Base de données reels
- ③ Analyse descriptive
- ④ Techniques d'apprentissage supervisée
- ⑤ Techniques d'apprentissage non supervisée
- ⑥ Contrôle de connaissances
- ⑦ Techniques de validation
- ⑧ Elements de la méthodologie scientifique
- ⑨ \LaTeX
 - Écriture de textes scientifiques
 - Beamer: présentations et posters scientifiques

Fréquence absolue

La fréquence (ou fréquence absolue) d'un événement x_i est le nombre n_i de fois où l'observation s'est produite dans une expérience.

$$\sum_{i=1}^k n_i = n$$

Fréquence relative

La fréquence relative de x_i , notée f_i , est la proportion d'occurrences observées pour cet événement, c'est-à-dire,

$$f_i = \frac{n_i}{n}, \quad 1 \leq i \leq k$$

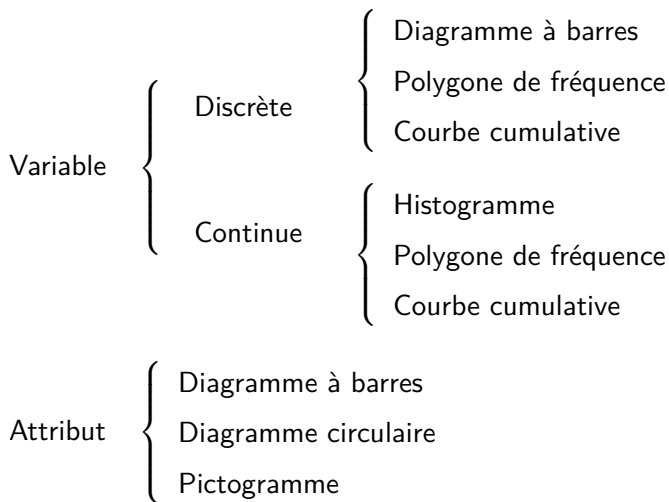
$$\sum_{i=1}^k f_i = 1$$

Distribution des fréquences

Une distribution de fréquence est un tableau (*tableau de fréquences*) ou un graphe (diagramme à barres ou histogramme) qui affiche la fréquence des événements dans un échantillon. Chaque entrée du tableau contient la fréquence des occurrences de valeurs dans un groupe ou un intervalle particulier.

x_i	n_i	N_i	f_i	F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	n_k	N_k	f_k	F_k
	n		1	

Représentations Graphiques



Exemple: variable discrète

Considérons la variable X = “score de risque COMPAS du défendeur”.

x_i	n_i	N_i	f_i	F_i
3	2	2	0.133	0.133
4	4	6	0.266	0.4
5	4	10	0.266	0.666
6	2	12	0.133	0.8
7	3	15	0.2	1
	15		1	

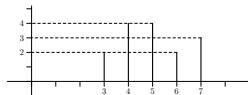
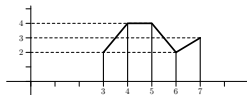
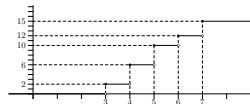


Diagramme à barres



Polygone de fréquences

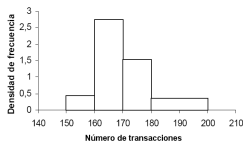


Courbe cumulative

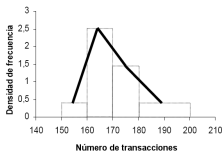
Exemple: variable continue

Considérons la variable X ="Taille en cm", observé chez $n = 50$ suspects.

$(L_{i-1}, L_i]$	n_i	a_i	h_i	N_i
$(150, 160]$	4	10	0.4	4
$(160, 170]$	25	10	2.5	29
$(170, 180]$	14	10	1.4	43
$(180, 200]$	7	20	0.35	50



Histogramme



Polygone de fréquences

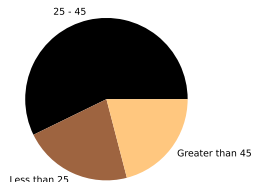
acumulada2.png

Courbe cumulative

Diagramme circulaire

A l'intérieur d'un cercle, chaque catégorie est affectée à un secteur proportionnel à sa fréquence.

Âge du défenseur	n_i	f_i	$f_i \times 360^\circ$
Moins de 25	1347	0.22	79°
Entre 25 et 45	3532	0.57	206°
Plus de 45	1293	0.21	75°
	6172	1	360°



Plan

- Concepts de base
- Maximisation de la marge - Hard-margin
- Cas de la séparation imparfaite - Soft Margin
- Fonctions de perte
- Fonction noyau - Séparation non linéaire
- Calcul des probabilités d'affectation
- Sélection des paramètres et problèmes pratiques
- Extension aux problèmes multi classes
- Pratique des SVM - Logiciels et outils
- Bilan - Avantages et inconvénients
- Références bibliographiques

Apprentissage supervisé: Machines à Vecteurs de Support

- Ω : la population.
- La population est divisée en deux classes, $\{-1, +1\}$.

Apprentissage supervisé: Machines à Vecteurs de Support

- Ω : la population.
- La population est divisée en deux classes, $\{-1, +1\}$.
- Pour chaque objet dans Ω , on a
 - $x = (x^1, \dots, x^d) \in X \subset \mathbb{R}^d$: variables de prediction.
 - $y \in \{-1, +1\}$: étiquettes.

Apprentissage supervisé: Machines à Vecteurs de Support

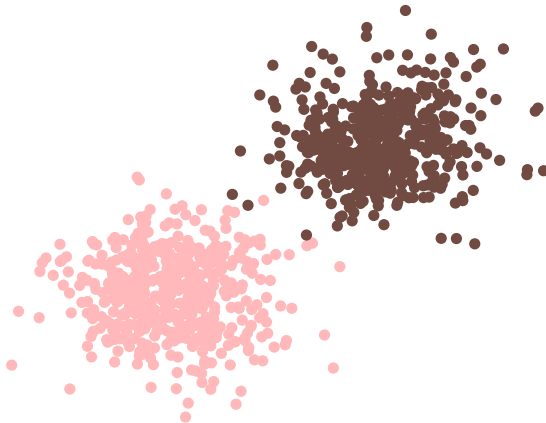
- Ω : la population.
- La population est divisée en deux classes, $\{-1, +1\}$.
- Pour chaque objet dans Ω , on a
 - $x = (x^1, \dots, x^d) \in X \subset \mathbb{R}^d$: variables de prediction.
 - $y \in \{-1, +1\}$: étiquettes.
- L'objectif est de trouver un hyperplan $\omega^\top x + b = 0$ qui vise à séparer, si possible, les deux classes.

Apprentissage supervisé: Machines à Vecteurs de Support

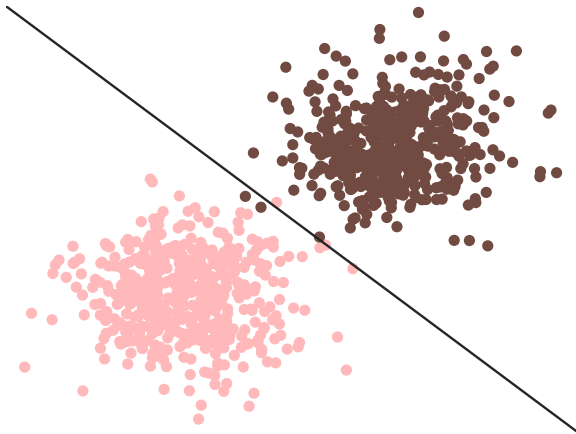
- Ω : la population.
- La population est divisée en deux classes, $\{-1, +1\}$.
- Pour chaque objet dans Ω , on a
 - $x = (x^1, \dots, x^d) \in X \subset \mathbb{R}^d$: variables de prediction.
 - $y \in \{-1, +1\}$: étiquettes.
- L'objectif est de trouver un hyperplan $\omega^\top x + b = 0$ qui vise à séparer, si possible, les deux classes.
- Les objets futurs seront classés comme

$$\begin{aligned} y &= +1 & \text{si} & \quad \omega^\top x + b > 0 \\ y &= -1 & \text{si} & \quad \omega^\top x + b < 0 \end{aligned} \tag{1}$$

Apprentissage supervisé



Apprentissage supervisé



Machines à Vecteurs de Support (SVM)

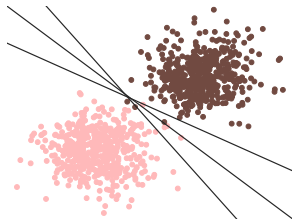
- L'état de l'art en matière de classification supervisée.
- Très bonne précision de classement.
- Pas cher en termes de calcul: formulation de programmation quadratique.
- SVM: Dans de nombreux cas, ils sont compétitifs avec les méthodes de classification existantes.
- Relativement facile à utiliser.
- Techniques du noyau : nombreuses extensions.
- Régression, estimation de densité, PCA noyau, etc.

Hyperplans séparateurs

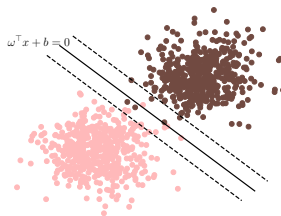
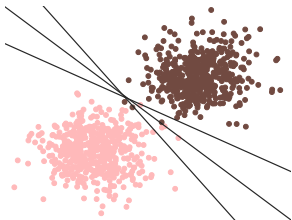
- Hyperplans de séparation possibles infinis.
- Chacun avec des propriétés différentes (métriques).
- Exprimé comme:

$$\omega^T x + b = 0$$

- En géométrie, un hyperplan est un sous-espace dont la dimension est inférieure à celle de son espace ambiant.



Maximisation de la marge



Le SVM vise à trouver la frontière qui maximise la marge entre les classes. La marge est la distance entre l'hyperplan et les données les plus proches. Ces derniers sont appelés vecteurs supports.

Algèbre linéaire d'un hyperplan

- H : hyperplan défini par $\omega^\top x + b = 0$
- Propriétés clés:
 - 1 Pour chaque $x_1, x_2 \in H$
 - $\omega^\top (x_1 - x_2) = 0$ et
 - $\bar{\omega} = \omega / \|\omega\|$ est le vecteur normal à H .
 - 2 Pour chaque $x_0 \in H$, $\omega^\top x_0 = -b$

Distance d'un point à l'hyperplan

- La distance **signée** de tout point x au H est la projection du vecteur v ($x - x_0$), avec x_0 le point d'intersection de H et le vecteur normal) dans le vecteur normal.
- On obtient cette projection via le produit scalaire:

$$\begin{aligned}
 \bar{\omega} \cdot v &= \frac{\omega^\top}{\|\omega\|} \cdot (x - x_0) = \\
 &= \frac{1}{\|\omega\|} (\omega^\top x - \omega^\top x_0) = \\
 &= \frac{1}{\|\omega\|} (\omega^\top x + b)
 \end{aligned}$$

- Distance de l'objet i à l'hyperplan H :

$$d(x_i, H) = \frac{y_i}{\|\omega\|} (\omega^\top x_i + b)$$

Problème d'optimisation

Rappelons que la largeur de la marge est la distance entre la frontière de décision et le point le plus proche.

On veut trouver la marge la plus grande possible (maximisation)

SVM cherche à maximiser, en fonction de ω, b , la quantité :

$$\arg \max_{\omega, b} \left\{ \min_i d(x_i, H) = \frac{1}{\|\omega\|} \min_i y_i (\omega^\top x_i + b) \right\}$$

Approche Hard-Margin

- Échantillon d'apprentissage supposé linéairement séparable, c'est-à-dire l'enveloppe convexe des deux groupes ne sont pas vides et ils ne se chevauchent pas.
- Tous les objets de l'échantillon d'apprentissage doivent être correctement classés!
- L'hyperplan séparateur est celui qui maximise la plus petite distance à une mauvaise classification.

Hard-margin SVM: marge maximale

- Distance entre $\omega^\top x + b = +1$ et $\omega^\top x + b = -1$:

$$2/\|\omega\| = 2/\sqrt{\omega^\top \omega}$$

- Un problème de programmation quadratique avec linéaire contraintes.

$$\min_{\omega, b} \frac{1}{2} \sum_{j=1}^d \omega_j^2$$

s.t.

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Mais...

$$\min_{\omega, b} \frac{1}{2} \sum_{j=1}^d \omega_j^2$$

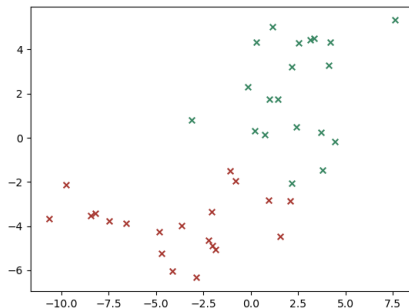
s.t.

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Séparation linéaire



Mais...

$$\min_{\omega, b} \frac{1}{2} \sum_{j=1}^d \omega_j^2$$

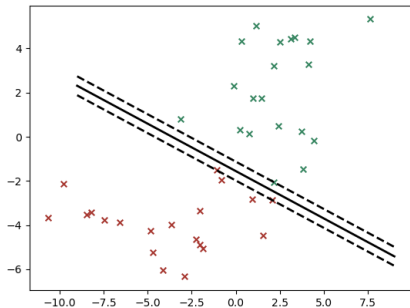
s.t.

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Séparation linéaire



Mais...

$$\min_{\omega, b} \frac{1}{2} \sum_{j=1}^d \omega_j^2$$

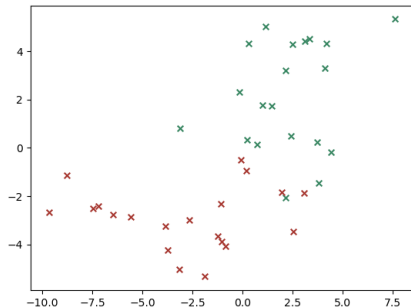
s.t.

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Séparation non linéaire



Mais...

$$\min_{\omega, b} \frac{1}{2} \sum_{j=1}^d \omega_j^2$$

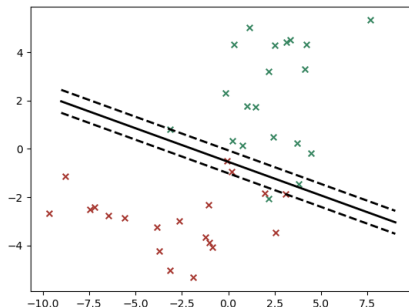
s.t.

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Séparation non linéaire



impossible à résoudre!!

Hard margin SVM: limites

- Les données réelles, très probablement, ne répondront pas aux d'hypothèse séparable linéaire
- La fonction de perte du Hard margin est trop limitante quand il y a des chevauchements de classe
- Hard margin SVM ne pourra pas traiter avec ça
- Solutions possibles:
 - Conservez le SVM à marge ferme mais transformez le données: mappage dans un espace de caractéristiques de dimension supérieure (peut-être infini)

$$\phi(x) = (\phi_1(x), \phi_2(x), \dots)$$

- Assouplir les contraintes (autoriser les erreurs de formation)
- Combinaison des deux

Une solution pour les données non linéairement séparables

- Lorsque les données ne sont pas linéairement séparables, le problème SVM à marge dure est irréalisable.
- Dans l'approche **soft-margin**, les contraintes

$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

sont perturbés.

- Comment? En introduisant des variables auxiliaires ξ_i , rendant le nouveau problème toujours réalisable.

Construire le soft-margin SVM

$$\min_{\omega, b, \xi} \frac{1}{2} \sum_{j=1}^d \omega_j^2 + \frac{C}{n} \sum_{i=1}^n g_i(\xi_i)$$

s.t.

$$y_i(\omega^\top x_i + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, n,$$

- $\xi = (\xi_i) \in \mathbb{R}^n$ est le vecteur des variables d'écart (slack).
- g_i est la fonction de perte (convexe et croissante).
- Plus populaires: *hinge loss*, $g_i(t) = C_i t$ or *hinge loss au carré*, $g_i(t) = C_i t^2$.
- C est un paramètre à régler.

La formulation du SVM

Séparation non linéaire

$$\min_{\omega, b, \xi} \frac{1}{2} \sum_{j=1}^d \omega_j^2 + \frac{C}{n} \sum_{i=1}^n \xi_i$$

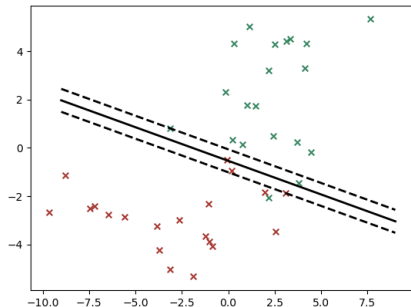
s.t.

$$y_i(\omega^\top x_i + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}$$

$$\xi_i \geq 0 \quad \forall i = 1, \dots, n.$$

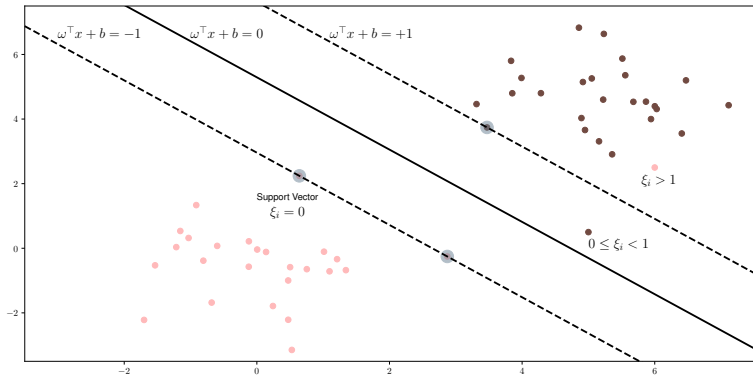


- Un objet i sera correctement classé si $0 \leq \xi_i < 1$
- Mal classé si $\xi_i > 1$.
- Dans le cas $\xi_i = 1$, nous obtenons une égalité (les objets coïncident avec l'hyperplan).
- $\sum_{i=1}^n \xi_i$ est une borne supérieure du nombre de objets mal classés.

Soft Margin SVM

- Soft margin SVM assouplit la contrainte pour permettre aux points d'être à l'intérieur de la marge ou même du mauvais côté de la frontière
- Les limites sont pénalisées par une quantité qui reflète l'étendue de la violation
- Slack variables $\xi_i \geq 0$ pour chaque échantillon à mesurer l'étendue de la violation.

Slack variables



Variables slack

- Pour les points sur ou à l'intérieur de la marge correcte:

$$\xi_i = 0$$

- Pour les autres point :

$$\xi_i = 1 - y_i(\omega^\top x_i + b)$$

- Si un point se trouve dans la frontière de décision:

$$\xi_i = 1$$

- La contrainte hard margin:

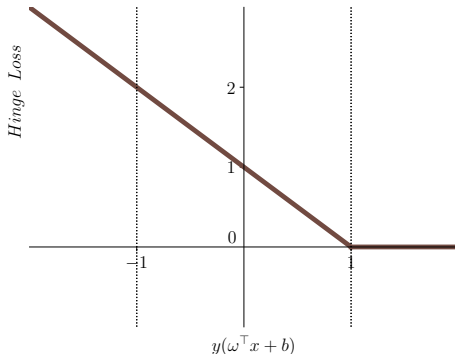
$$y_i(\omega^\top x_i + b) \geq 1 \quad \forall i = 1, \dots, n$$

- devient désormais :

$$y_i(\omega^\top x_i + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n$$

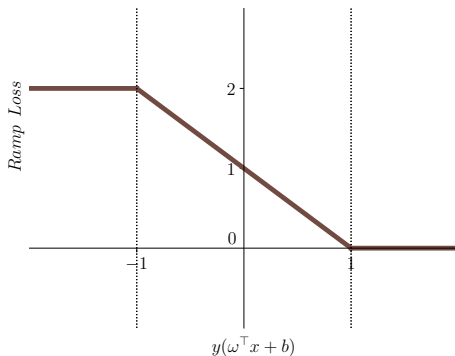
Hinge Loss

$$\ell(x, y) = \max(0, 1 - y(\omega^\top x + b)) = \begin{cases} 0 & \text{if } y(\omega^\top x + b) \geq 1 \\ 1 - y(\omega^\top x + b) & \text{if } y(\omega^\top x + b) \leq 1 \end{cases}$$



Ramp Loss

$$\ell(x, y) = \begin{cases} 0 & \text{if } y(\omega^\top x + b) \geq 1 \\ 1 - y(\omega^\top x + b) & \text{if } -1 \leq y(\omega^\top x + b) \leq 1 \\ 2 & \text{if } y(\omega^\top x + b) \leq -1 \end{cases}$$



Exemple: données Type A

Les deux classes ont la matrice d'identité comme covariance.

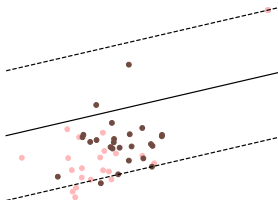
- Classe $+1$: la moyenne est l'origine.
- Classe -1 : la moyenne est $(2/(d), \dots, 2/(d))$.

La base de données d'entraînement est contaminé avec données aberrantes. Les observations aberrantes sont échantillonnées pour la classe $+1$ en utilisant une distribution gaussienne avec une matrice de covariance $0,001$ fois la matrice d'identité et avec une moyenne $(10/(d), \dots, 10/(d))$.

Brooks, J.P. *Support vector machines with the ramp loss and the hard margin loss*.
Operations Research: 59(2), 467-479 (2011)

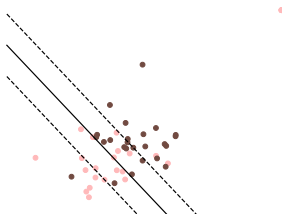
Robustesse aux données aberrantes

SVM avec Hinge Loss



Précision de 44%

SVM avec the Ramp Loss



Précision de 78%

SVM avec la Ramp Loss

$$\min_{\omega, b, \xi, z} \frac{1}{2} \sum_{j=1}^d \omega_j^2 + \frac{C}{n} \left(\sum_{i=1}^n \xi_i + 2 \sum_{i=1}^n (1 - z_i) \right)$$

s.t.

$$(y_i(\omega^\top x_i + b) - 1 + \xi_i) \cdot z_i \geq 0 \quad \forall i = 1, \dots, n$$

$$0 \leq \xi_i \leq 2 \quad \forall i = 1, \dots, n$$

$$z \in \{0, 1\}^n$$

$$\omega \in \mathbb{R}^d$$

$$b \in \mathbb{R}.$$

Multi-class classification

- k classes
- One-against-all: train k binary SVMs
 - 1st class vs. $(2 - k)$ th class
 - 2nd class vs. $(1, 3 - k)$ th class
 - \vdots
- k decision functions

$$(\omega^1)^\top x + b^1$$

$$\vdots$$

$$(\omega^k)^\top x + b^k$$

- Prédiction

$$\arg \max_j (\omega^j)^\top x + b^j$$

- Raison : S'il appartient à la 1re classe, alors nous devons avoir:

$$(\omega^1)^\top x + b^1 \geq 0$$

$$(\omega^2)^\top x + b^2 \leq 0$$

$$\vdots$$

$$(\omega^k)^\top x + b^k \leq 0$$

- Un contre un : entraîner $k(k-1)/2$ SVMs binaire
- Exemple: 4 classes \rightarrow 6 SVMs binaire

$y_i = +1$	$y_i = -1$	Fonctions de décision
Class 1	Class 2	$f^{12}(x) = (\omega^{12})^\top x + b^{12}$
Class 1	Class 3	$f^{13}(x) = (\omega^{13})^\top x + b^{13}$
Class 1	Class 4	$f^{14}(x) = (\omega^{14})^\top x + b^{14}$
Class 2	Class 3	$f^{23}(x) = (\omega^{23})^\top x + b^{23}$
Class 2	Class 4	$f^{24}(x) = (\omega^{24})^\top x + b^{24}$
Class 3	Class 4	$f^{34}(x) = (\omega^{34})^\top x + b^{34}$

- Dans l'ensemble de données de test, nous prédisons toutes les SVM binaires

Classes		gagnant
1	2	1
1	3	1
1	4	1
2	3	2
2	4	4
3	4	3

- Sélectionnez celui avec le plus grand nombre de votes

Classe	1	2	3	4
#votes	3	1	1	1

Probabilité

Définition de probabilité

Une fonction $P : \mathcal{P}(\Omega) \longrightarrow [0, 1]$ on dira qu'il s'agit d'un **probabilité** s'il satisfait

(P1) $0 \leq P(A) \leq 1$, pour tout événement A .

(P2) $P(\Omega) = 1$.

(P3) étant donné A et B , deux événements mutuellement exclusifs, nous avons que $P(A \cup B) = P(A) + P(B)$.

Propriétés

- ① $P(\bar{A}) = 1 - P(A)$. Comme conséquence $P(\emptyset) = 0$.
- ② $P(A) = P(A \cap B) + P(A \cap \bar{B})$, pour toute A et B .
- ③ $A \subseteq B \Rightarrow P(A) \leq P(B)$.
- ④ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ pour toute A et B . Comme conséquence

$$P(A \cup B) \leq P(A) + P(B).$$

- ⑤ Si A_1, A_2, \dots, A_n sont des événements mutuellement exclusifs, c'est-à-dire, si $A_i \cap A_j = \emptyset, \forall i \neq j$, ensuite

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

Calcul des probabilités dans espaces d'échantillonnage fini

Dans cette section, nous verrons comment calculer la probabilité d'un événement lorsque nous avons un espace d'échantillon fini,

$$\Omega = \{w_1, \dots, w_n\},$$

nous connaissons les probabilités de chacun de ses éléments

$$\{P(\{w_i\}), \quad i = 1, \dots, n\}.$$

A noter que les probabilités $\{P(\{w_i\}), \quad i = 1, \dots, n\}$ doivent vérifier:

(a) $0 \leq P(\{w_i\}) \leq 1, \quad i = 1, \dots, n.$

(b) $\sum_{i=1}^n P(\{w_i\}) = 1$, car $\Omega = \bigcup_{i=1}^n \{w_i\}$, avec $\{w_i\}$ exclusive, et donc

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^n \{w_i\}\right) = \sum_{i=1}^n P(\{w_i\}).$$

Dans les conditions ci-dessus, pour tout événement A , la **probabilité de A** est donné par:

$$P(A) = \sum_{w_i \in A} P(\{w_i\}) .$$

Cas particulier de grande importance

Lorsque tous les éléments de l'espace d'échantillonnage sont `textbf` équiprobable, c'est-à-dire lorsque

$$P(\{w_i\}) = \frac{1}{n} , \quad i = 1, \dots, n,$$

on a que

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{cas favorables}}{\text{cas possibles}} ,$$

ce qu'on appelle **Règle de Laplace**.

Probabilité conditionnelle

Définition de la probabilité conditionnelle

étant donné un événement B avec $P(B) > 0$, on définit la **probabilité de un événement A conditionnel à B** , tel que

$$P(A/B) = \frac{P(A \cap B)}{P(B)} .$$

Probabilité conditionnelle

Propriétés de la probabilité conditionnelle

❶ Donné un événement B avec $P(B) > 0$, $P(\cdot/B)$ est une probabilité:

(P1) $0 \leq P(A/B) \leq 1$

(P2) $P(\Omega/B) = 1$

(P3) Si $A \cap C = \emptyset$, ensuite $P(A \cup C/B) = P(A/B) + P(C/B)$.

❷ De la définition de la probabilité conditionnelle, il s'ensuit que

$$P(A \cap B) = P(A/B)P(B).$$

❸ En général, il faut (**règle de multiplication**):

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Qualité d'un classificateur

- Précision
- Sensibilité (Taux Vrai Positif)
- Spécificité (Taux Vrai Négatif)
- Parcimonie (vs. complexité): $\frac{\#(w_j=0, j=1, \dots, d)}{d} \cdot 100$

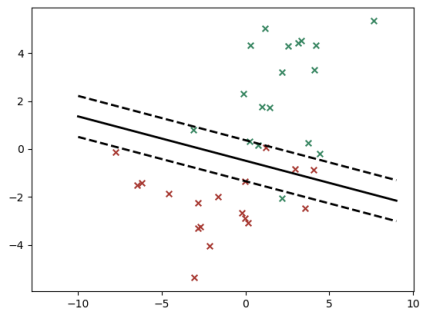
Précision

Étant donné un objet i , il est classé dans la classe positive ou négative selon la valeur de la fonction score, $\text{sign}(\omega^\top x_i + b)$, tandis que pour le cas $\omega^\top x_i + b = 0$, l'objet est classé au hasard. La précision de la classification est définie comme le pourcentage d'objets correctement classés par le classifieur sur une base de données.

$$\text{Précision} = \frac{\text{correct predictions}}{\text{total predictions}} =$$

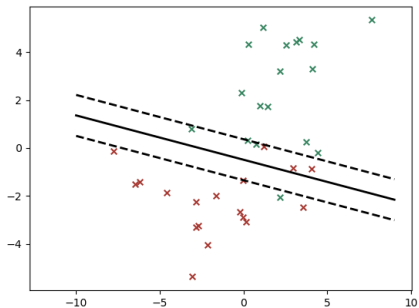
$$= P(\omega^\top x_i + b \geq 0 \wedge y_i = +1) + P(\omega^\top x_i + b < 0 \wedge y_i = -1)$$

Exemple: Précision



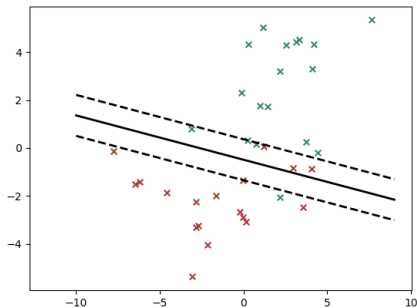
• $|\Omega| = 40$

Exemple: Précision



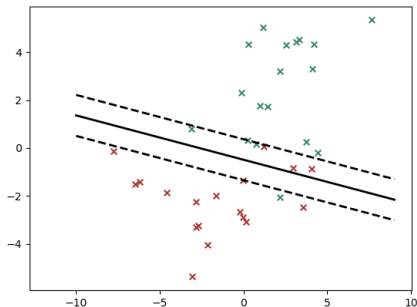
- $|\Omega| = 40$
- $\#\{i, y_i = +1\} = 20$

Exemple: Précision



- $|\Omega| = 40$
- $\#\{i, y_i = +1\} = 20$
- $\#\{i, y_i = -1\} = 20$

Exemple: Précision



- $|\Omega| = 40$
- $\#\{i, y_i = +1\} = 20$
- $\#\{i, y_i = -1\} = 20$
- $\text{Précision} = \frac{19+17}{40} = 0.9$
- 90% d'objets correctement classés.

Matrice de confusion

		Réal	
		Positive	Négative
Predit	Positive	ces valeurs sont nos données positives ET ont également été prédites comme positives	ces valeurs sont nos données négatives ET ont été prédites comme positives
	Négative	ces valeurs sont nos données positives ET ont été prédites comme négatives	ces valeurs sont nos données négatives ET ont également été prédites comme négatives

Sensibilité

La sensibilité est connue sous le nom de taux de vrais positifs. Essentiellement, cela nous informe sur la proportion de cas positifs réels qui ont été prédits comme positifs par notre modèle. C'est le rapport des vrais positifs à tous les positifs.

$$\text{Sensibilité} = \frac{\text{Vrai Positif}}{\text{Vrai Positif} + \text{Faux Negatif}} =$$

Spécificité

La spécificité est connue sous le nom de taux de vrais négatifs. Il nous informe sur la proportion de cas négatifs réels qui ont été prédits comme négatifs par notre modèle. C'est le rapport des vrais négatifs à tous les négatifs.

$$Spécificité = \frac{Vrai\ Négatif}{Vrai\ Négatif + Faux\ Positif} =$$

Bibliography

- Cortes, C., Vapnik, V. *Support-vector networks*. Machine learning, 20(3), 273-297.
- Brooks, J.P. *Support vector machines with the ramp loss and the hard margin loss*. Operations Research: 59(2), 467-479 (2011).