

Application des Machines à Vecteurs de Support en cardiologie : prévision intelligente des cardiopathies

Depuis mon jeune âge, je fus fasciné par le fonctionnement du cœur, et je le fus encore plus suite à la mort de mon grand-père due à la cardiopathie. De là m'est venue l'idée d'en faire objet de mon étude, et favoriser ainsi la prévention.

Mon étude s'inscrit subtilement dans le thème de cette année: santé et prévention, étant donné que la cardiopathie est une pathologie assez répandue, et de gravité considérable, nécessitant un dépistage efficace et efficient.

Positionnement thématique (ETAPE 1)

MATHEMATIQUES (Mathématiques Appliquées), INFORMATIQUE (Informatique pratique).

Mots-clés (ETAPE 1)

Mots-Clés (en français)	Mots-Clés (en anglais)
<i>Machines à Vecteurs de Support</i>	<i>Support Vector Machines</i>
<i>Hyperplan</i>	<i>Hyperplane</i>
<i>Astuce du noyaux</i>	<i>Kernel trick</i>
<i>Matrice de Confusion</i>	<i>Confusion Matrix</i>
<i>Dimension</i>	<i>Dimension</i>

Bibliographie commentée

Les SVM (Support Vector Machine ou Machine à vecteurs de support) représentent l'ensemble des algorithmes de Machine Learning utilisés pour la classification et la résolution de problèmes de régression. Introduits au 20ème siècle par le mathématicien Vladimir VAPNIK, ces algorithmes ont servi à résoudre des problèmes liés à tant de domaines, et en particulier à la médecine [1]. Dans son état d'introduction, les SVM ont démontré leur efficacité, cependant le processus d'obtention de résultats était laborieux, rempli d'embûches et caractérisé par une infinité de paramètres à considérer avec une technologie encore pas au point, les analystes devaient prendre plusieurs facteurs en considération lors des diagnostics, et une fois l'électrocardiogramme obtenu, son interprétation était douteuse car ce dernier doit être filtré de toutes les perturbations liées aux technologies utilisées et données incomplètes [2]. Toutes les étapes étaient lentes, exécutées manuellement, et donc non fiables.

De nos jours l'accès aux nouvelles technologies a rendu l'utilisation de ces algorithmes bien plus facile et autonome, une vraie automatisation de l'étude a eu lieu. Les chercheurs se contenteront d'établir une base de données qui conditionnera nos résultats après une analyse par des SVM automatisés sous la forme de Machine Learning.

L'intérêt des SVM est de trouver un hyperplan qui sépare un lot d'échantillons positifs et un lot d'échantillons négatifs avec une marge maximale qui les divise rigoureusement. Autrement dit, le but est de trouver l'hyperplan optimal pour distinguer les deux échantillons tout en minimisant la marge d'erreur [3].

Si par exemple nos données sont représentées dans un espace à deux dimensions, le programme essaie de tracer l'hyperplan le plus adapté pour distinguer les deux catégories; ainsi, l'hyperplan est une droite car elle a une dimension en moins que l'espace dans lequel les données ont été représentées.

Cette méthode d'hyperplans n'est plus pratique que dans le cas où les données sont linéairement séparables, en d'autres termes: il est facile de discriminer entre les deux catégories. Qu'en est-il du cas où les données ne sont pas linéairement séparables? Dans cette condition, on recourt à "L'astuce du noyau". Elle consiste principalement à placer les données dans un espace de plus grande dimension où ces dernières seraient linéairement séparables. Grâce à ses paramètres de pénalité, les SVM parviennent à minimiser les erreurs d'entraînement et à maximiser la marge de classification. Par suite, l'ajustement des valeurs de ces paramètres permet l'obtention de l'hyperplan optimal cité précédemment. Ceci est valable à la fois pour les problèmes de classification linéaire et non linéaire [4].

Tout ajustement fait, après avoir trouvé la combinaison adéquate de paramètres, nous évaluons la précision du programme en analysant la matrice de confusion qui est une matrice récapitulant les prédictions faites en se fondant sur les calculs de ces paramètres[5].

Problématique retenue

Dans quelle mesure le machine learning (SVM) peut aider à la prévention des cardiopathies?

Objectifs du TIPE

Mon TIPE vise les objectifs suivants:

- l'explication des fondements mathématiques des SVM
- une simulation informatique: il s'agit du programme (code informatique) à entraîner.
- évaluation de la précision du programme

Références bibliographiques (ETAPE 1)

- [1] ALAA THARWAT : Parameter investigation of support vector machine classifier with kernel functions : <https://doi.org/10.1007/s10115-019-01335-4>
- [2] PIERRE TABOULET : ECG. 2 filtres : <https://www.e-cardiogram.com/filtre-ecg/>
- [3] WANG L : Support vector machines: theory and applications : <https://books.google.co.ma/books?hl=fr&lr=&id=uTzMPJjVjsMC&oi=fnd&pg=PA1&dq=Wang+L>

+(2005)+Support+vector+machines:+theory+and+applications,+vol+177.+Springer,+Berlin&ots
 =GFDGetVJob&sig=bg9IJvJoIxnNci-DYkjkHAYBXkE&redir_esc=y#v=onepage&q&f=false
 [4] MELKI G, CANO A, KECMAN V, VENTURA S :) Multi-target support vector regression via
 correlation regressor chains : *Inf Sci* 415:53–69
 [5] SARANG NARKHEDE : Understanding Confusion Matrix :
<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>