
Table des matières

- I) Introduction
- II) Modèle Tm
 - 1) Modèle de transformation
 - 2) Les modèles de survie basés sur SVM
 - a) Modèle de survie existant basés sur SVM
 - b) Nouveaux modèles sur SVM
 - 3) Mesures et performances
 - a) Jeux de données
 - b) Mesures de performances
 - c) Expériences
- III) Conclusion

Annexe 1 : Référence

Annexe 2 : Tableau 1

Annexe 3 : Tableau 2

Annexe 4 : Légende des figures

Annexe 5 : Modèle 1

Annexe 6 : Modèle 2

Annexe 7 : Modèle 3

Annexe 8 : Figures

Cigles et Abréviations

SVM : machine à vecteur support

SVM-based : modèle basé sur les contraintes de classement.

Tm : modèle de transformation

SVCR : modèle régression vectorielle de support pour les données censurées :

Résumé

Les machines à vecteurs supports (ou séparateur à vaste marge : SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires (famille d'algorithmes de classement statistique) ; ils calculent la décision par combinaison linéaire des échantillons.

Ils peuvent être utilisés pour résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable.

La résolution de ces deux problèmes passe par la construction d'une famille h qui à un vecteur d'entrée x fait correspondre une sortie y :

Pour le cas, d'une fonction discriminante linéaire, obtenue par combinaison linéaire du vecteur d'entrée $x = (x_1, \dots, x_2)^T$, avec un vecteur de poids $w = (w_1, \dots, w_2)^T$:

$h(x) = w^T x + \varepsilon$, il est alors décidé que x est de classe 1 si $h(x) \geq 0$ et de classe -1 sinon, c'est un classifieur linéaire.

I) INTRODUCTION :

La maladie pulmonaire obstructive chronique est une maladie respiratoire chronique causée principalement par le tabagisme et touchant le plus souvent des personnes âgées de plus de quarante ans. Parce qu'elle est responsable presque partout dans le monde du plus haut taux de réadmission ainsi que de la troisième cause de décès, les chercheurs et cliniciens s'intéressent à prédire le risque auquel les patients font face, de sorte à améliorer les soins et les traitements tout en optimisant les prestations des services de santé. Dans le domaine de l'analyse de survie, on s'intéresse à prédire le temps avant qu'un événement particulier ne survienne, savoir une éventuelle ré-hospitalisation ou un décès dans le cas de la maladie pulmonaire obstructive chronique.

Cet article présente des preuves empiriques de l'utilisation technique d'apprentissage automatique pour l'estimation du non linéarité des modèles de transformation pour l'analyse de survie.

Il faut noter que l'analyse de survie, ou la prédiction de survie intervient dans des domaines variés pour examiner le risque qu'un événement se produise au cours d'une période d'observation fixée. Bien que ce type de prédiction soit surtout utilisé dans le milieu médical, on l'utilise aussi dans beaucoup de domaines en particulier en économie (prédiction de faillite), en mécanique (panne de moteur), en électricité (temps de vie des composants électroniques), en sciences sociales (estimation du temps entre les mariages et les divorces) ainsi que dans divers autres domaines.

La tâche d'estimer un modèle de transformation basé sur un ensemble de données reflétant la survie des patients a été intégrée dans un contexte d'apprentissage et de machine à vecteur de support (SVMs). L'objectif ici, est de reformuler la tâche en tant que problème de classement via l'index de concordance, un problème qui peut être résolu efficacement dans un contexte de risque structure et technique de minimisation et d'optimisation convexe. Pour établir le classement souhaité chaque observation est comparée à son proche voisin. Un avantage principal de cette approche est l'utilisation de noyaux non linéaires implémentant automatiquement les effets non paramétriques des variables. L'objectif de cet article est alors doublé :

- Nous présenterons des données empiriques supplémentaires (matériel supportant cette technique basée sur six cliniques liées au cancer). Et discuterons de la pertinence de la technique actuelle par rapport aux approches classiques pour de telles données
- Et nous étudierons de manière empirique, comment la technique basée sur svm peut être renforcée pour inclure ce qu'on appelle : régressions contraintes interpolant la méthode et la prédiction ponctuelle.

Un inconvénient de la formulation de classement de tels modèles est que le classement est établi sur un choix arbitraire d'observations. On cherche donc à savoir si la performance augmente en divisant les observations en fonction d'intervalles sur les temps d'évènement ; une technique qui peut être mise en œuvre qu'en utilisant des observations factices.

Ainsi l'objectif de ce mémoire est basé sur la prédiction de survie dans un contexte en distinguant deux types de prédictions. Nous comparerons des modèles de survie basée sur svm et des contraintes de régression avec des modèles contenant à la fois des contraintes de classement et de régression. La performance des modèles a été comparée au moyen de trois mesures différentes :

- L'indice de concordance mesurant la capacité discriminante du modèle ;
- La statistique de type log Rank indique si les patients avec un pronostique d'indice inférieur à l'indice de pronostique médian ont une différence significative. La survie que les patients avec un indice de pronostique supérieur à la médiane ;
- Et le taux de risque après normalisation pour limiter l'indice pronostique entre 0 et 1

Les méthodes de support pour l'analyse de survie montrent une amélioration significative des performances pour les méthodes des contraintes de régression par rapport aux méthodes intégrant uniquement les contraintes de classement. De plus, nos expériences ne montrent pas de différence entre les lots de la comparaison de contrainte de classement arbitraires avec l'utilisation d'observation factices.

Résultat : ce travail prouve que les modèles basés sur svm en utilisant des contraintes de régression fonctionnant nettement mieux que svm-based (modèles basés sur les contraintes de classement). Deuxièmement, il a montré que l'estimation de l'indice de pronostique n'est pas améliorée lors de l'identification des comparaisons avec des observations factices.

Notons que nos expériences n'ont pas montré de différence entre les modèles à base de svm et le modèle proportionnel. Cependant, les anciens modèles ont l'avantage d'être extensible vers le modèle non linéaire sans vérification de linéarité vers les variables avant modélisation.

La littérature de survie d'écrit plusieurs modèles pour répondre à ces questions. De nombreuses méthodes courantes, y compris le modèle de risque proportionnel (model cox) et le modèle log-odds peuvent être formulés comme un ensemble de la classe de transformation.

II) Modèle tm

Les tms (mesurent la capacité discriminante) consistent en une séquence :

- D'assemblage d'un index pronostique basé sur les Covariables, et
- D'associer ce pronostique à index aux temps d'événement observés. En général, on ne cherche qu'à trouver une représentation explicite de la 1ère étape.

Le modèle cox standard par exemple évite cette seconde étape en supposant que le risque (le risque instantanée) d'observer l'évènement sachant qu'il ne s'est produit au paravent.

D'autres modèles supposent une forme fixe pour la fonction de transformation h ; le modèle de temps de défaillance accéléré est un exemple qui suppose que la fonction de transformation $h(y)$, avec y le résultat à l'étude, équivaut à la fonction logarithmique et à la probabilité proportionnelle au modèle prend $h(y) = \log y$

L'approche que nous adoptons dans les modèles de survie basés sur le vecteur de support machine (svm) diffère des approches adoptées dans les modèles statistiques. Bien qu'il puisse bien être classé en tant que tm, les modèles basés sur svm n'assument pas les fonctions sous adjacentes pour lesquelles les paramètres doivent être estimés. Le risque empirique est classé sur deux cas de façon erronée en ce qui concerne leur temps d'échec, est minimisé.

Le problème de la survie a donc été reformulé en un problème de classement. Pour réduire la charge de calcul, un système simplifié de la version comparant chaque observation uniquement avec son plus proche voisin au lieu de toutes les autres observations a été proposée en (10). Un cadre plus théorique était fourni dans (8). Nous ferons référence au modèle de survie proposé dans ce travail comme modèle 1.

Dans ce travail, nous nous demandons si l'inclusion de contrainte de régression peut améliorer les performances. Par conséquent, la performance du *modèle 1* est comparée à celle du *modèle 2*, y compris les contraintes de classement et de régression.

Deuxièmement nous rechercherons si l'inclusion d'observations factices pour résoudre le problème du choix arbitraire des contraintes de classement dans le modèle, améliore aussi considérablement la performance. Les deux modèles proposés ci-dessus sont comparés à la survie des modèles ne comprenant que des contraintes de classement et incluant uniquement des contraintes de régression. La table modèle traitée en 1) donne un aperçu des différents modèles traités dans ce travail, leurs contraintes, le nombre de paramètre de réglage dans le cas d'un noyau linéaire et comment les contraintes de classement sont définies.

Cet article est organisé comme suit : la section 1 donne un aperçu des modèles de transformation dans l'analyse de survie. La section 2 commence par un résumé des modèles de survie basés sur svm, suivies de l'introduction de deux nouveaux, proposés par les auteurs. La section 3 illustre les 6 jeux de données d'une clinique et les mesures de performance utilisés dans les expériences. En plus des méthodes mentionnées précédemment, les expériences incluant la performance du model cox en comparaison.

1) **Modèles de transformations**

Les tms d'écrivent le type de modèle dans lesquelles une transformation de la variable de résultat (au lieu de la variable de résultats elle-même) est estimée en fonction de la Covariante.

Initialement, les variables ont été introduites dans les problèmes de régression où l'hypothèse de la normalité sur la distribution des erreurs et la variance constante n'étaient pas satisfaites. Un modèle de régression standard, par exemple qui tente de modéliser le résultat y est une combinaison linéaire des Covariante :

$$y = w^T x + \varepsilon \quad (1)$$

Où w est un vecteur de coefficient et ε est la variable d'erreur. Dans le cas où y est normalement pas distribuée, la régression peut être améliorée en transformant la variable dépendante.

L'idée générale des modèles de transformation peut être résumée comme suit :

Considérons un modèle

$$h(y) = w^T x + \varepsilon, \quad (2)$$

où w représente les paramètres de régression et ε représente la variable avec la densité $f_\varepsilon(\cdot)$.

Notons bien :

Un tm spécifie qu'une fonction monotone h de la réponse variable est liée linéairement à la regression de variable.

La généralisation vers des fonctions non linéaires $\varphi(x) = w^T x$, φ la carte de caractéristiques, des variables conduit à :

$$h(x) = \varphi(x) + \varepsilon \quad (3)$$

Le Weibull (modèle de régression) et le temps de défaillance accélèrent ce type (14). Cependant, le tm le plus couramment utilisé est le risque proportionnel (model cox). Dans ce modèle, le risque au temps t , est défini comme suit :

$$\lambda(t/x) = \lim_{\Delta t \rightarrow 0} \frac{P(t < y < t + \Delta t / x, y \geq t)}{\Delta t} \quad (4)$$

Est modélisé comme le produit d'un risque de base non défini $\lambda_0(t)$ multiplié par l'exponentielle d'une combinaison linéaire des Covariante

$$\lambda(x) = \lambda_0 e^{w^T x} \quad (5)$$

Ceci conduit à estimer la fonction survie $s(x, t) = P(y > t/x)$ comme

$$s(x, t) = s_0(t) e^{w^T x} \quad (6)$$

Prendre $\ln(-\ln(\cdot))$ les deux conduisent à :

$$\ln(-\ln(s(x, t))) = \ln(-\ln(s_0(t))) - w^T \quad (7)$$

Ecrit sous la forme

$$\varepsilon = h(y) - u(x). \quad (8)$$

La fonction précédente $h(\cdot)$ a été corrigée, ce qui n'est pas le cas dans le modèle cox.

Bien que la norme fournisse une relation linéaire entre les explications variables et une fonction de la valeur attendue de la variable dépendante, les méthodologies peuvent être généralement à des relations non linéaires.

Un exemple est l'extension du modèle cox avec des modèles de lissage.

2) Les modèles de survie basés sur le noyau

Cette section commence par une discussion brève des modèles de survie existants basés sur la svm. Dans une deuxième sous-section, deux nouvelles méthodes sont proposées. De plus le résultat de ce type de modèle de survie ne peut généralement pas être interprété comme un échec temps ; nous désignerons le résultat du modèle par l'indice de pronostic $u(x)$ au lieu de la prédiction du modèle. Pour le modèle cox, cela correspond à

$$u(x) = w^T x .$$

a) Modèles de survies existants basés sur svm

Les excellentes performances des machines à vecteurs de support pour la classification et la régression ont conduit à se demander si ce type de modèle pouvait être étendu à un autre problème statistique. Lors de l'analyse des données de survie, on s'intéresse au temps entre un certain point de départ et la survenue d'un événement prédéfini.

Une première approche à laquelle on pourrait penser est d'utiliser des modèles de régression pour modéliser le temps de récurrence. Cependant, les données de survie contiennent généralement des points de données avec des informations incomplètes (information appelées données censurées). Bien qu'il existe différents types de censures ; nous limiterons notre attention à la censure de droite dans ce travail. Les véritables informations censurées sont les informations pour lesquelles l'heure exacte de l'événement n'est pas connue, mais, où une limite inférieure de ce type est connue. Seulement l'incorporation des informations pour lesquelles le temps exact d'échec connu conduirait à une sous-estimation du temps de survie.

Régression de vecteur de support pour les données censurées :

Dans la régression de vecteur de support standard, la prédiction est définie comme une combinaison linéaire d'une transformation des variables $\varphi(x)$, avec $\varphi(.)$ La carte de fonction, plus une constante b ; $\hat{y} = w^T \varphi(x) + b$.

Pour obtenir cette prédiction, il faut calculer les coefficients w et la constante b afin d'obtenir de bonnes propriétés. Les coefficients sont maintenus petits. Les svms sont formulées en tant

que problème d'optimisation, où une fonction de perte doit être modélisée sous certaines contraintes.

Les contraintes stipulent que la prédiction \hat{y} doit être supérieur à $y - \varepsilon$ avec $\varepsilon > 0$. De même, $-\hat{y}$ devrait être plus grand que $-y - \varepsilon^*$, avec $\varepsilon^* > 0$. La fonction de perte identifie les grandes valeurs de ε et ε^* de sorte que les prédictions résultantes \hat{y} seront proches aux valeurs observées y .

Formellement, le problème est formulé comme suit :

$$\min_{w,b,\varepsilon,\varepsilon^*} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*)$$

$$\text{Sous contraintes} \quad \begin{cases} w^T \varphi(x_i) + b \geq y_i - \varepsilon_i & \forall i=1,\dots,n \\ -w^T \varphi(x_i) - b \geq -y_i - \varepsilon_i^* & \forall i=1,\dots,n \\ \varepsilon_i \geq 0 \quad ; \varepsilon_i^* \geq 0 & \text{"} \end{cases} \quad (9)$$

Le paramètre γ est une constante de régularisation positive stricte et ε et ε^* sont

Variables latentes permettant de prévoir les erreurs de données d'apprentissage.

Le résultat prédit pour un nouveau point x^* est alors trouvé comme:

$$\hat{y}(x^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(x_i)^T \varphi(x^*) + b, \quad (10)$$

Avec α_i et α_i^* les multiplicateurs de Lagrange. Un avantage des modèles basés sur svm, est que la carte de caractéristiques (\cdot) n'a pas besoin d'être définie. Selon le théorème (18) de Mercer, $\varphi(x_i)^T \varphi(x_j)$ peut être écrit comme

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (11)$$

À condition que $K(\cdot, \cdot)$ soit un noyau défini positif. Les noyaux souvent utilisés sont

- le noyau linéaire: $K(x, z) = x^T z$,
- le noyau polynomial de degré a : $K(x, z) = (\tau + x^T z)^a$, avec $\tau \geq 0$
- le noyau RBF: $K(x, z) = \exp\left(-\frac{\|x-z\|_2^2}{\sigma^2}\right)$

Plus récemment, un noyau de données cliniques [19] a été proposé comme noyau additif

$K(x, z) = \sum_{p=1}^d K_p(x - z)$ où $K_p(\cdot, \cdot)$ dépend du type de la variable. Pour les variables continues et ordinales, $K_p(\cdot, \cdot)$ est défini comme

$$K_p(x_p, z_p) = \frac{c - |x_p - z_p|}{c} \quad (12)$$

Avec x_p la covariable p^{th} d'observation x , $c = \max x_p - \min x_p$, avec $\min x_p$ et $\max x_p$ la valeur minimale et maximale de la p^{th} covariable dans la formation donnée ensemble de données D . Pour les données qualitatives et binaires, $K_p(\cdot, \cdot)$ est défini comme

$$K_p(x_p, z_p) = \begin{cases} 1, & \text{si } x_p = z_p \\ 0, & \text{si } x_p \neq z_p \end{cases} \quad (13)$$

Le problème de l'utilisation de la régression vectorielle de support pour les données censurées réside dans l'incertitude sur les résultats y . Les premières approches des données censurées en utilisant des stratégies de régression, soit en omettant les observations censurées, ce qui dans des temps d'échec sous-estimés, ou à traiter les observations censurées comme des non événements résultant en des modèles biaisés. Pour utiliser toutes les informations disponibles (les observations censurées donnent des informations sur l'intervalle dans lequel l'événement se produit) Shivaswamy [12] a proposé une approche de régression vectorielle de support pour les données censurées (svcr). Pour les observations non censurées, les mêmes contraintes sont prises en compte modèle de régression svm standard. Pour les observations censurées à droite, il est connu que l'échec ne s'est pas produit avant l'heure de la censure. La première contrainte de

(9) est donc toujours valable. Cependant, la seconde contrainte est trop restrictive pour les observations censurées à droite. Le modèle de régression de vecteur de support pour censurer les données proposées par Shivaswamy peuvent donc être formulées comme suit:

SVCR

$$\min_{w,b,\varepsilon,\varepsilon^*} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*)$$

$$\text{Sous contraintes} \begin{cases} w^T \varphi(x_i) + b \geq y_i - \varepsilon_i & \forall i=1,\dots,n \\ -\delta_i(w^T \varphi(x_i) + b) \geq -\delta_i y_i - \varepsilon_i^* & \forall i=1,\dots,n \\ \varepsilon_i \geq 0 \quad ; \quad \varepsilon_i^* \geq 0 & \text{"} \end{cases} \quad (14)$$

L'index pronostique d'un nouveau point x^* se présente sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n (\alpha_i - \delta_i \alpha_i^*) \varphi(x_i)^T \varphi(x^*) + b, \quad (15)$$

Avec α_i et α_i^* les multiplicateurs de Lagrange.

Une deuxième proposition de régression svm pour les données censurées a été faite par Khan [13] (svmc). La différence entre les deux modèles réside dans la peine appliquée ou perte d'indices pronostiques incorrects (voir Figure 1), ce qui peut être interprété comme les temps d'échec prévus pour les approches de régression. La méthode svcr pénalise les prédictions incorrectes mêmes si la prédiction était supérieure ou inférieure au temps d'échec observé et pénalise les prédictions incorrectes pour le droit censuré données uniquement si la prédiction est inférieure au temps de censure observé. Le raisonnement étant donné qu'il est seulement connu que le temps d'échec est supérieur à celui observé du résultat. De plus, la pénalité pour les mauvaises prédictions est la même, elle consiste en une prévision des temps de défaillance censurés ou observés. Au contraire, svmc applique des pénalités différentes dans les quatre cas possibles: (i) pénalité γ pour les événements avec une survie prévue inférieure à la survie observée; (ii) pénalité γ^* pour les événements avec une survie prévue supérieure à la survie observée, avec $\gamma^* > \gamma$; (iii) pénalité γ_c pour les données censurées à droite avec une survie prévue inférieure au temps de censure observé; et (iv) pénalité γ_c^* pour les données censurées à droite avec survie prévue supérieure au temps de censure observé, avec $\gamma_c^* < \gamma_c$. De plus, ce modèle fournit différents ε -pertes pour les événements et de donnée censuré à droit et pour

des prédictions supérieures et inférieures au temps observé. Par conséquent, là, l'inconvénient majeur de cette dernière méthode est le grand nombre d'hyper-paramètres.

Prise en charge des machines à vecteurs sur la base de contraintes Quand on n'utilise pas les modèles de régression, les problèmes de survie sont souvent traduits en classification des problèmes pour répondre à la question de savoir si le patient survivra à un certain prédéfini temps (par exemple, survivant 5 ans après la chirurgie). Cependant, pour pouvoir inclure autant d'événements que possible, ce délai prédéfini doit être pris très tard.

Au contraire, on souhaiterait un délai précoce, afin de retenir autant de patients que possible, puisque tous les patients censurés avant cette heure seront perdus pour l'analyse. Un autre problème de cette approche est que la validité de la méthode diminue quand de plus en plus de patients sont censurés plus tôt. Pour surmonter les problèmes décrits ci-dessus, il a été proposé de formuler le problème de la survie comme un classement problème dans [9,11] et une simplification informatique a été proposée dans [10].

L'idée de formuler le problème de la survie en tant que problème de classement est que dans les applications cliniques, on s'intéresse souvent à la définition des groupes à risque. L'un n'est pas principalement intéressé par une prédiction de la durée de survie, mais si le patient a un risque élevé ou faible que l'événement se produise, de sorte qu'un traitement approprié peut être donné. Pour atteindre cet objectif, une méthode de classement svm est utilisée, similaire au modèle rankSVM pour l'apprentissage par classement ou par préférence [20]. La méthode proposée par [9] et [11] implique une régularisation comme d'habitude et une pénalisation pour chaque paire comparable de points de données pour lesquels l'ordre dans l'indice de pronostic diffère de l'ordre observé. Un couple de données $\{(x_i, y_i, \delta_i)(x_j, y_j, \delta_j)\}$ est dit comparable chaque fois que l'ordre de leurs temps d'événement est connu (par exemple deux événements, ou un événement et un point de données censuré à droit pour lequel le temps de censure de ce dernier est postérieur au temps d'échec du premier). Plus formellement, une comparabilité indicatrice pour une paire d'observations $\{(x_i, y_i, \delta_i)(x_j, y_j, \delta_j)\}$ est définie comme

$$comp(i, j) = \begin{cases} 1, & \text{si } \delta_i = 1 \text{ et } \delta_j = 1 \\ & \text{si } \delta_i = 1 \text{ et } \delta_j = 0 \text{ et } y_i \leq y_j \\ 0 & \text{sinon} \end{cases} \quad (16)$$

Le modèle est ensuite formulé comme

RANKSVMC

$$\min_{w, \varepsilon} \quad \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \sum_{\substack{j: y_i > y_j \\ \text{comp}(i,j)=1}} \varepsilon_{ij}$$

$$\text{Sous contrainte} \begin{cases} w^T (\varphi(x_i) - \varphi(x_j)) \geq 1 - \varepsilon_{ij}, \\ \forall i = 1, \dots, n; \forall j: y_i > y_j \text{ et } \text{comp}(i, j) = 1 \\ \varepsilon_{ij} \geq 0, \forall i = 1, \dots, n; \forall j: y_i > y_j \text{ et } \text{comp}(i, j) = 1. \end{cases} \quad (17)$$

Notez que l'ajout d'un terme constant n'est pas utile ici car seules les différences dans l'indice de pronostic sont considérés. L'indice de pronostic pour un nouveau point x^* est trouvé comme

$$\hat{u}(x^*) = \sum_{i=1}^n \sum_{\substack{j: y_i > y_j \\ \text{comp}(i,j)=1}} \alpha_{ij} \left(\varphi(x_i) - \varphi(x_j) \right)^T \varphi(x^*), \quad (18)$$

Avec α_{ij} les multiplicateurs de Lagrange.

Un inconvénient de la méthode est qu'un problème de programmation quadratique (Qp) de

$O(n^2)$ contraintes doit être résolu. Pour surmonter ce problème, un calcul une approche simplifiée résolvant un Qp de contraintes $O(n)$ a été proposée dans [10]. La réduction a été trouvée en comparant chaque point de données avec son plus proche voisin comparable au lieu de comparer avec tous les points de données comparables. Nous ferons référence à ce modèle simplifié comme RANKSVMC.

Un résultat important sur la manière de relier les modèles de survie statistique standard comme le modèle de risque proportionnel [7] et le modèle de temps de défaillance accéléré [6] avec les modèles de survie à base de svm, se trouve dans les modèles de transformation [8]. Dans le Dans ce dernier travail, un modèle de survie semblable au modèle ranksvmc a été présenté, la différence située dans la partie droite du premier ensemble de contraintes de l'équation (17), où 1 est remplacé par $y_{(i)} - y_{\bar{j}(i)}$, où $\bar{j}(i)$ est défini comme

$$\bar{j}(i) = \arg \max_j j$$

$$\text{Sous contraintes } \begin{cases} \text{comp}(i, j) = 1 \\ y_i > y_j \end{cases} \quad (19)$$

pour une valeur donnée de i . Notez que dans le cas sans censure on a

$\text{comp}(i, j) = 1, \forall i = 1, \dots, n; j \neq i$, tel que $\bar{j}(i) := i - 1$. Puis que ce sera le model à partir duquel nous proposerons de nouveaux modèles, nous le désignerons comme modèle 1. Ce modèle est formulé comme

Modèle 1

$$\min_{w, \varepsilon} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i$$

$$\text{Sous contraintes } \begin{cases} w^T (\varphi(x_i) - \varphi(x_{\bar{j}(i)})) \geq y_{(i)} - y_{\bar{j}(i)} - \varepsilon_i, \quad \forall i = 1, \dots, n \\ \varepsilon_i \geq 0 \end{cases} \quad \forall i = 1, \dots, n \quad (20)$$

L'index pronostique d'un nouveau point x^* se présente sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n \alpha_i \left(\varphi(x_i) - \varphi(x_{\bar{j}(i)}) \right)^T \varphi(x^*) \quad (21)$$

Avec α_i les multiplicateurs de Lagrange

b) Nouveaux modèles basés sur svm

Dans cette section, nous proposons deux méthodes différentes utilisant des modèles de survie svm.

Premièrement, la proposition visant à réduire la complexité de la formulation du classement de [10] être utilisé. Deuxièmement, il sera examiné si la performance du modèle peut être améliorée en remplaçant ce choix arbitraire de comparaisons avec des observations factices. Nous supposons $u = w^T \varphi(x) + b$, sauf indication contraire.

Dans l'analyse de survie, la principale question à laquelle il faut répondre est de ne pas prédire le temps d'échec aussi précisément que possible, mais en définissant des profils de risque en fonction des observations.

C'est exactement ce qui se fait dans le modèle Cox: un indice de pronostique $u(x)$ est estimé et la fonction de lien entre ce prédicteur et la survie réelle de temps est laissée indéterminée. Les modèles basés sur les noyaux pour l'analyse de survie ignorent également $h(\cdot)$, mais aussi laissent l'estimation de l'indice de pronostique libre, résultant en des modèles plus flexibles. L'idée centrale de survie svm dans [8] est qu'un indice de pronostique peut être trouvé en minimisant le risque empirique de mal classer deux observations (voir figure 2).

Au lieu d'essayer d'estimer le temps de survie, on cherche un indice pronostique qui estime l'ordre dans lequel les objets / personnes rechutent. Plus formellement, le modèle devient

$$\min_{w, \varepsilon} \sum_{i=1}^n \varepsilon_i$$

$$\text{Sous contraintes} \begin{cases} u(x_i) - u(x_{j(i)}) + \varepsilon_i \geq \rho_i & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0 & \forall i = 1, \dots, n \end{cases} \quad (22)$$

Où les ρ_i donnent des constantes positives et $u = w^T \varphi(x)$.

Deux autres problèmes doivent être résolus: (i) comment définir ρ_i et (ii) comment inclure la régularisation. On voit que le modèle ranksvmc est un cas particulier de la formulation ci-dessus en prenant $\rho_i = 1$ et en régularisant en utilisant la marge maximale principe [17]. Les méthodes que nous proposons ici vont plus loin dans le travail de [8], en utilisant la constante de Lipchitz au lieu de la marge maximale pour la régularisation fine. Cela résout la question de savoir comment définir ρ , puisque cette méthodologie prend $\rho_i = y_{(i)} - y_{j(i)}$.

Étant donné que le modèle 1 est construit en minimisant le risque empirique lié aux erreurs de classement, la valeur du modèle pronostique n'a qu'une importance relative. Quand on est intéressé par le modèle pronostique pour lequel l'indice pronostique peut être interprété au fur et à mesure, on pourrait inclure des contraintes de régression comme dans [12,13]. Le modèle 1 est donc modifié (modèle 2) comme décrit ci-dessous:

Modèle 2

$$\min_{w, \varepsilon, \xi, \xi^*, b} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i + \mu \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\text{Sous contraintes} \begin{cases} w^T (\varphi(x_i) - \varphi(x_{j(i)})) \geq y_{(i)} - y_{j(i)} - \varepsilon_i, & \forall i = 1, \dots, n \\ w^T \varphi(x_i) + b \geq y_i + \xi_i, & \forall i = 1, \dots, n \\ -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \xi_i^*, & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0, \quad \xi_i \geq 0, \quad \xi_i^* \geq 0 & \forall i = 1, \dots, n \end{cases} \quad (23)$$

Le premier ensemble de contraintes est identique à celui du modèle 1 et constitue la contrainte de classement optimisant l'indice de concordance. Le deuxième ensemble de contraintes indique que l'indice de pronostique devrait être plus grand que le temps d'échec observé. Le troisième ensemble indique que l'indice de pronostique u devrait être inférieur à y . La multiplication avec l'indicateur de censure δ_i , veille à ce que cette contrainte ne soit pas prise en compte pour les données censurées à droite. En prenant les deuxième et troisième séries d'ensemble des contraintes, la valeur de l'index pronostique u est ciblée sur la défaillance de temps. Cependant, pour les données censurées à droite, la troisième contrainte n'a pas besoin d'être rempli. L'indice pronostique $u(x^*)$ pour une nouvelle observation x^* est alors trouvé sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n (\alpha_i + \beta_i - \delta_i (\alpha_i^* + \beta_i^*)) \varphi(x_i)^T \varphi(x^*) + b \quad (24)$$

Où $\alpha_i, \alpha_i^*, \beta_i$ et β_i^* les multiplicateurs de Lagrange.

Les modèles 1 et 2 limitent le nombre de comparaisons ou de classements contraintes à $O(n)$ en comparant chaque observation à sa plus proche (en observés) temps point comparable. Comme ce choix est plutôt arbitraire, nous nous demandons si la performance du modèle peut être améliorée en générant k observations factices avec utilité inconnue v_l (également appelée valeurs de seuil) et échec connu temps z_l , avec $l = 1, \dots, k$. Les valeurs z sont réparties uniformément sur les valeurs observées intervalle de temps. La valeur z minimale est égale à zéro puisqu'il s'agit du plus bas possible résultat, le plus grand z est supérieur au résultat le plus élevé observé. Toute survie et les temps de censure sont donc compris dans l'intervalle $[z_l, z_k]$. L'inclusion du seuil est un moyen élégant de faire un choix non arbitraire de comparaisons.

La différence entre cette approche et celle du modèle 2 est illustrée dans Figure 3). Ne supposez aucune censure pour le moment. Dans le modèle 2, une comparaison sera faite entre chaque observation i et l'observation ayant le temps d'échec le plus élevé pour lequel ce temps d'échec est plus petit que y_i . Dans le modèle 3, ces comparaisons sont rendu moins arbitraires en comparant chaque observation avec une observation fictive.

La modification du modèle 2 en incluant ces seuils conduit à ce qui suit: formulation du problème

modèle 3

$$\min_{w, \varepsilon, \xi, \xi^*, b} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i + \sum_{i=1}^n (\xi_i + \xi_i^*)$$

Sous contraintes

$$\begin{cases} w^T \varphi(x_i) + b - l_i^T v \geq y_{(i)} - l_i^T z - \varepsilon_i, & \forall i = 1, \dots, n \\ \delta_i(w^T \varphi(x_i) + b - r_i^T (v - z) - y_i) \geq -\varepsilon_i^*, & \forall i = 1, \dots, n \\ w^T \varphi(x_i) + b \geq y_i + \xi_i, & \forall i = 1, \dots, n \\ -\delta_i(w^T \varphi(x_i) + b) \geq -\delta_i y_i - \xi_i^*, & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0, \quad \xi_i \geq 0, \quad \xi_i^* \geq 0, \quad v_l - v_{l-1} \geq 0, & \forall i = 1, \dots, n \end{cases}$$

(25)

Dans la formulation du problème ci-dessus, l_i et r_i représentent des vecteurs de longueur k contenant toutes les valeurs zéro, à l'exception de la valeur $\bar{l} = \arg \max_l l$, sous réserve de $z_l < y_i$ pour lequel $l_i(\bar{l})$ est égal à 1 et $\bar{l} = \arg \min_l l$ sous réserve de $z_l < y_i$ pour lequel $r_i(\bar{l})$ est égal à

1. Les premier et deuxième ensembles de contraintes sont similaires à ceux de l'équation

(23), en remplaçant les comparaisons arbitraires avec le plus proche voisin par des comparaisons avec les seuils les plus proches. La dernière série d'équations fait en sorte que les seuils augmentent. L'indice de pronostic $u(x^*)$ pour une nouvelle observation x^* peut alors être trouvé comme :

$$\hat{u}(x^*) = \sum_{i=1}^n (\alpha_i + \beta_i - \delta_i(\alpha_i^* + \beta_i^*)) \varphi(x_i)^T \varphi(x^*) + b, \quad (26)$$

Où

α_i et β_i sont les multiplicateurs de Lagrange.

Nous nous référons à l'annexe 7 pour plus de détails

3) MESURES ET PERFORMANCES

Cette section décrit les ensembles de données disponibles au public qui sont utilisés dans les expériences de cet article. En outre, trois mesures de performance différentes sont discutées.

a) JEUX DE DONNEES

Nous utiliserons 5 jeux de données disponibles au public pour 6 expériences cliniques. Le premier ensemble de données concerne la prédiction de la rémission complète et du décès de 129 patients avec leucémie [21]. Informations sur le traitement, sexe, âge, score de performance

(Score de Karnofsky), globules blancs, plaquettes, hémoglobine et si le patient reçu une greffe de rein a été donné. Dans une première expérience (LD), l'événement à l'étude est la mort, tandis que dans une deuxième expérience (LCR) l'événement est la rémission complète.

Un deuxième ensemble de données concerne l'essai sur le cancer du poumon administré par le vétéran (VLC) [22,6]. Seulement 9 hommes sur les 137 atteints d'un cancer du poumon avancé inopérable étaient en vie à la fin de l'étude. Les patients ont été randomisés dans un standard au teste de chimiothérapie. Informations sur le score de performance, l'âge, la thérapie, l'histologie de la tumeur, traitement et délai entre le diagnostic et la randomisation était disponible. Le troisième ensemble de données concerne 506 patients atteints de cancer de la prostate (PC) [23]. Les variables utilisées pour cette expérience sont les suivantes: traitement, âge, poids indice, indice de performance, antécédents de maladie cardiovasculaire, taille de la tumeur, un indice combiné du stade du grade histologique et de l'hémoglobine sérique. L'analyse a été réalisée sur les 483 patients avec des informations complètes. 125 (26%) des patients sont décédés à la suite d'un cancer de la prostate au cours de l'étude. Tous les autres patients étaient censurés à droite à la date du dernier suivi. Données sur le cancer du poumon de la clinique Mayo (MLC) [24] contient des informations sur 167 patients atteints d'un cancer du poumon à un stade avancé. L'ensemble de données contient des informations sur l'âge, le sexe, deux scores de performance estimés par le médecin et celui estimé par le patient, le nombre de calories consommées et la perte de poids au cours des six derniers mois. Au cours de la période d'étude, 72% des patients sont décédés. Un dernier ensemble de données contient des informations sur 720 patientes atteintes d'un cancer du sein (étude

allemande sur le cancer du sein) (GBSG) [25,26]. L'information était disponible sur traitement hormonal, statut ménopausique, âge, grade, taille de la tumeur, nombre de ganglions lymphatiques positifs, récepteurs de la progestérone et de l'œstrogène. L'expérience est sur la base des 686 cas contenant des informations complètes. Pendant la période d'étude, 299 patientes ont eu un événement lié au cancer du sein. Tous les jeux de données discutés peut être trouvé sur le Web (<http://lib.stat.cmu.edu/datasets> (consulté: 29 août 2008) et le paquetage R [http://cran.r-project.org/web/packages / survival / index.html](http://cran.r-project.org/web/packages/survival/index.html) (consulté le 20/11/2007)).

b) MESURES DE PERFORMANCE

Afin de comparer les différents modèles, trois mesures de performance sont utilisées dans les expériences:

L'indice de concordance (c-index) L'indice de concordance est le rapport entre le nombre de paires concordantes et le nombre de paires comparables:

$$c - index(u) = \frac{\sum_{i=1}^n \sum_{j \neq i} I[u(x_i) - u(x_j)(y_i - y_j) \geq 0]}{\sum_{i=1}^n \sum_{j \neq i} comp(i, j)} \quad (27)$$

Où I est la fonction indicatrice et comp(i, j) est défini comme précédemment.

Logrank χ^2 statistique : Les cliniciens s'intéressent souvent à la répartition des risques profils aux patients. Dans ce contexte, il est pertinent de tester dans quelle mesure les pays développés l'index pronostique permet de différencier les patients à haut risque et ceux à faible risque.

Les patients sont donc divisés en deux groupes: les groupes à risque faible et élevé si leur indice pronostique est inférieur ou supérieur à la valeur médiane. **Le test logrank est utilisé pour tester les différences significatives de survie entre les deux groupes.** Au lieu de travailler avec la valeur p résultante, nous ferons un rapport sur la Valeur χ^2 du test. Plus cette valeur est élevée, meilleure est la capacité de séparation du modèle.

Taux de risque : estimé par un modèle de variable Cox univarié. Tout produit des indices de pronostique, après normalisation pour limiter sa valeur entre 0 et 1, sera alimenté dans le modèle Cox et le taux de risque résultant sera calculé. Plus le taux de risque est élevé, meilleur est l'indice.

c) EXPERIENCES

Cette sous-section compare les modèles décrits dans la littérature (modèle Cox, ranksvmc, svcr, modèle 1) avec les modèles proposés dans cet article (modèle 2 et modèle 3) sur 6 jeux de données cliniques et 1 expérience artificielle. La validation croisée par 10 avec recuit simulé couplé [27] a été utilisé pour régler les hyper-paramètres.

L'indice de concordance a été utilisé comme critère de sélection du modèle. Tous les modèles ont été implémentés dans matlab3, à l'aide de la boîte à outils d'optimisation mosek4 Matlab.

Données réelles : Les deux modèles proposés (modèle 2 et modèle 3) sont comparés avec le modèle 1, ranksvmc et svcr sur 50 randomisations entre les entraînements (les deux tiers des données) et le groupe d'essais (le tiers restant des données) sur 6 ensembles de données, utilisant un noyau linéaire et un noyau clinique [19]. Les tableaux 2 et 5 résument les résultats. Différences statistiquement significatives entre le modèle 2 et l'autre les modèles ont été calculés en utilisant le test de Wilcoxon. Le modèle cox est testé linéairement dans les covariables (PH-linear) et avec des splines de lissage pénalisées (PH-pspline) [15], à titre de comparaison. On voit que le modèle 1 et le ranksvmc performants considérablement moins que les autres modèles, indépendamment du noyau utilisé. L'ajout des contraintes de régression améliore les performances significativement. Cependant, le nombre de vecteurs de support augmente en même temps.

Aucune différence significative n'est constatée entre le choix arbitraire des comparaisons et l'approche du seuil.

Pour étudier la différence de performance entre les modèles avec et sans contraintes de régression, 1000 échantillons bootstrap d'un jeu d'apprentissage du VLC dataset ont été prises. En utilisant un noyau linéaire, les poids estimés pour chaque covariable et l'indice de concordance résultant sur l'ensemble de tests ont été calculés.

Les figures 4 et 5 montrent la distribution des poids estimés de Karnofsky le score de performance et l'indice de concordance, respectivement. Le modèle seulement pris en compte pour les contraintes de classement entraîne une grande incertitude quant aux estimations de poids et donc de grandes différences dans l'indice de concordance se produisent pour l'ensemble d'entraînement légèrement différent. L'inclusion de contraintes de régression se traduit par plusieurs estimations stables de poids et de l'indice de concordance.

Données artificielles : Dans ce dernier exemple, nous examinons si l'amélioration du modèle 2 par rapport au modèle 1 dépend du pourcentage de censure du problème. Dans cette expérience, le cadre clinique est imité. Donc, 100 jeux de données sont construits avec 10 variables, 100 formations et 500 observations de test.

Toutes les variables étaient normalement distribuées avec une moyenne nulle et une variance égale à 1. Les corrélations entre les variables étaient nulles sauf pour les première et deuxième covariable, les troisième et quatrième, les cinquième et sixième et les septième et huitième, avec un coefficient de corrélation de 0.7, 0.3, -0.7 et

-0.3, respectivement. Le vecteur des coefficients de régression w a été pris comme étant $w = [-0.1, 1, 0.3, 0.4, 0.2, 0.03, 0.02, -0.01, -0.02, 0.01]^T$. Le temps d'échec était

exponentiellement distribué avec le paramètre égal à $w^T x$, avec x le vecteur de la covariable.

Dans un premier exemple, le temps d'échec a été censuré au hasard (imitant des procédures administratives censure); dans un deuxième exemple, le temps de censure a été échantillonné à partir de la même distribution que le temps de survie $F_Y(x)$, l'étendue de la distribution a été modifiée pour créer différents pourcentages de censure ($G_Y(x) = aF_Y(x)$) la figure 6 illustre les résultats. Pour la censure aléatoire et informative, la plus grande différence de performances entre les deux modèles sont notées pour une faible censure pourcentages. Cependant, le modèle 2 fonctionne mieux que le modèle 1 pour tout pourcentage de censure.

III) Conclusions

Ce travail a comparé différentes méthodes de vecteurs de support pour l'analyse de survie. Deux questions ont été posées:

- (i) les contraintes de classement sont-elles suffisantes pour obtenir de bons indices de pronostiques ou l'inclusion de contraintes de régression supplémentaires améliore-t-elle considérablement la performance du modèle? Et
- (ii) L'estimation de l'indice pronostique peut-elle être améliorée en épinglant les comparaisons aux observations factices au lieu de choix arbitraire des observations réelles à comparer. Les expériences révélées une amélioration significative en incluant des contraintes de régression en plus de contraintes de classement. Enfin, l'inclusion des seuils n'améliore pas la performance. Puisqu'un modèle avec des comparaisons avec le voisin le plus proche est plus simple, nous préférons utiliser ce dernier modèle. Comparaison du modèle 2 avec le modèle cox n'a révélé aucune différence significative dans les performances. L'avantage du modèle 2 ci-dessus le modèle Cox réside dans l'extension facile vers les modèles non linéaires sans la nécessité de vérifier les non-linéarités dans les variables avant la modélisation.

Annexe 1

REFERENCES

1. Byar D., Green S., Variables pronostiques pour la survie dans une comparaison randomisée des traitements pour le cancer de la prostate, *Bulletin du Cancer* 67 (1980) 477–490.
2. Cheng S.C., Wei L.J., Ying Z., Analyse des modèles de transformation avec censure données, *Biométrie* 82 (4) (1995) 835–845.
3. Cheng S.C., Wei L.J., Ying Z., Prédire les probabilités de survie avec Semi-paramétrique modèle de transformation, *Journal de l'American Statistical Association* 92 (437) (1997) 227-235.
4. Cox D.R., Modèles de régression et tables de mortalité (avec discussion), *Journal of the Royal Society de statistique, série B* 34 (2) (1972) 187-220.
5. Dabrowska D.M., Doksum K.A., Probabilité partielle dans les modèles de transformation avec données censurées, *Scandinavian Journal de Statistique* 15 (1) (1988) 1–23.
6. Daemen A., De Moor B, Développement d'une fonction de noyau pour les données cliniques, en: *Actes de la 31e conférence internationale annuelle de l'IEEE Engineering Société de médecine et de biologie (EMBS), IEEE, Piscataway, 2009*, p. 5913–5917.
7. Doksum K., Gasko M., Sur une correspondance entre modèles en régression binaire et données de survie, *International Statistical Review* 58 (1990) 243–252.
8. Eiders P.H., Marx B.D., Lissage flexible avec B-splines et pénalités, *Statistique Science* 11 (1996) 89-121.
9. Emerson S.S., Banks P.L.C., *Études de cas en biométrie*, Wiley-Interscience, New York, 1994, ch. L'interprétation d'un essai sur la leucémie s'est arrêtée tôt, p. 275–299.
10. Evers L, Messow C.M., Méthodes du noyau clairsemé pour des données de survie de grande dimension, *Bio-informatiques* 24 (14) (2008) 1632-1638.
11. Herbrich R, Graepel T, Obermayer K, Grandes limites de rangs pour l'ordinal régression, *Les classificateurs des avancées dans les grandes marges* (2000) 115–132.

12. Hurvich C.M., Simonoff J.S., Tsai C.L., Sélection du paramètre de lissage en mode non paramétrique.
Régression utilisant un critère d'information akaike amélioré, Journal de Société royale de statistique: série B 60 (1998) 271–293.
13. Kalbfleisch J.D., Méthodes de vraisemblance et tests non paramétriques, Journal de l'association américaine de statistique 73 (361) (1978) 167–170.
14. Kalbfleisch J.D., Prentice R.L., Analyse statistique des données de temps de défaillance.
Série de Wiley sur les probabilités et les statistiques, New York., 2002.
15. Khan F.M., Zubek V.B., Régression vectorielle de support pour les données censurées (SVRC): un nouvel outil d'analyse de survie, dans: F. Giannotti, D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan, X. Wu (éd.), Actes de la huitième IEEE International Conférence sur le Data Mining (ICDM), société informatique IEEE, Californie, 2008.
16. Lawless J.F., Analyse de la survie et des événements, série de référence de Wiley en bio statistique, Wiley, West Sussex, Angleterre, 2006, ch. Modèles paramétriques en survie analyse, p. 345–355.
17. M. J., Fonctions de type positif et négatif et leur lien avec la théorie d'équations intégrales, Transactions philosophiques de la Royal Society A 209 (1909) 415–446.
18. Prentice R.L., Un modèle log-gamma et son estimation du maximum de vraisemblance, biométrie 61 (3) (1974) 539-544.
19. Sauerbrei W., Royston P., Création de modèles de pronostic et de diagnostique multi-variés: transformation des prédictors en utilisant des polynômes fractionnaires, Journal of the
Société royale de statistique, série A 162 (1) (1999) 71–94.
Méthodes de support pour l'analyse de survie 17
20. Schumacher M., Basert G., H. Bojar, K. Huebner, M. Olschewski, Sauerbrei W., Schmoor C., Beyerle C., Neumann R.L.A., Rauschecker H.F., randomisées 2 x 2 essai évaluant le traitement hormonal et la durée de la chimiothérapie chez les patients non dispositifs atteintes d'un cancer du sein, Journal of Clinical Oncology 12.
21. Shivaswamy P.K., Chu W., Jansche M., Une approche de vecteur de support pour censurer objectifs, dans: Actes de la septième conférence internationale IEEE 2007 sur

- Exploration de données (ICDM), IEEE Computer Society, Californie, 2007, p. 655–660.
22. Therneau T.M. et Grambsch P.M., Modélisation des données de survie: extension de la maladie de Cox modèle, 2e édition, Springer-Verlag, New York, 2000.
 23. Van Belle V., K. Pelckmans, Suykens J.A.K., Van Huffel S., Apprentissage de la transformation
Modèles de classement et d'analyse de survie, Tech. Rep, 09-45, ESAT-SISTA, K.U.Leuven (Louvain, Belgique) (2009).
Méthodes de support pour l'analyse de survie 16
 24. Van Belle V., K. Pelckmans, Suykens J.A.K., Van Huffel S., Machines à vecteurs de support pour l'analyse de survie, dans: E. Ifeakor, A. Anastasiou (Eds.), Actes de la troisième conférence internationale sur l'intelligence numérique en médecine et soins de santé (CIMED), 2007.
 25. Van Belle V., K. Pelckmans, Suykens J.A.K., Van Huffel S., Survival SVM: a Practice Scalable Algorithm., Dans: M. Verleysen (Ed.), Actes du 16ème Symposium européen sur les réseaux de neurones artificiels (ESANN), d-side, Evere, 2008, pp. 89–94.
 26. Vapnik V., Théorie de l'apprentissage statistique, Wiley and Sons, New York, 1998.
 27. . Xavier de Souza S., Suykens J.A.K., Vandewalle J., Bolle D., Couplé simulé annelage, IEEE Transactions on Systems, Man et Cybe

Annexe 2

Tableau 1. Présentation des différentes méthodes de survie basées sur le noyau et de leurs propriétés, dans le cas où un noyau linéaire est utilisé.

model	reference	ranking constraint	regression constraint	# tuning parameters	ranking constraints
RANKSVMC	[11,9,10]	✓		1	nearest neighbor
SVCR	[12]		✓	1	
SVRc	[13]		✓	4	
MODEL 1	[8]	✓		1	nearest neighbor
MODEL 2		✓	✓	2	nearest neighbor
MODEL 3		✓	✓	2	dummy observations

Annexe 3

Tableau 2. Comparaison des différents modèles de survie svm et du modèle cox sur 6 ensembles de données sur le cancer avec variables cliniques, utilisant un noyau linéaire. La médiane performance sur 50 randomisations entre la formation et le test, ainsi que le nombre de vecteurs de support (# Sv) est résumé. Statistique significative les différences entre le modèle 2 (indiqué en gris) et les autres modèles ont été testées en utilisant le test de somme de rang de Wilcoxon. L'utilisation des contraintes de régression améliore la performance de manière significative. Aucune différence significative n'est notée avec le cox modèle. Cependant, les modèles basés sur svm ont l'avantage de ne pas vérifier l'hypothèse de linéarité et de risque proportionnel.

dataset	method	c-index	logrank χ^2	hazard rate	# sv
VLC	MODEL 1	0.58±0.06 ³	2.01±3.72 ³	3.26±4.70 ³	91.0±3.51
	MODEL 2	0.69±0.03	6.27±5.11	9.78±7.46	91.0±1.10
	MODEL 3	0.69±0.03	6.85±5.00	9.85±7.64	90.5±1.16
	RANKSVMC	0.57±0.07 ³	1.60±3.83 ³	2.70±5.36 ³	77.0±6.53³
	SVCR	0.69±0.04	9.88±6.05¹	11.12±9.75	90.0±0.96³
	PH _{pspline}	0.67±0.09 ²	4.52±6.62	8.40±19.14	
LCR	MODEL 1	0.54±0.06 ³	1.04±1.63	1.98±2.85 ²	80.0±8.24 ³
	MODEL 2	0.59±0.04	1.08±1.90	3.56±3.50	83.0±4.72
	MODEL 3	0.58±0.05	1.17±2.08	2.56±3.43	76.5±6.05³
	RANKSVMC	0.52±0.06 ³	0.37±2.17	1.38±2.88 ³	78.0±7.22 ³
	SVCR	0.60±0.04	1.94±1.72¹	4.04±3.25	80.5±3.67 ³
	PH _{pspline}	0.58±0.05	1.39±1.98	3.70±5.83	
LD	MODEL 1	0.66±0.07 ²	3.60±4.35 ³	5.97±12.37 ³	86.0±3.68
	MODEL 2	0.69±0.05	8.50±7.48	17.23±23.33	86.0±2.03
	MODEL 3	0.69±0.05	6.05±6.95	17.47±21.59	86.0±2.24
	RANKSVMC	0.65±0.08 ³	2.68±6.60 ³	5.12±8.95 ³	76.0±5.04³
	SVCR	0.69±0.04	10.7±6.51	13.1±33.54 ³	82.0±4.13³
	PH _{pspline}	0.67±0.05	4.89±5.45 ¹	18.35±62.26	
MLC	MODEL 1	0.61±0.05	3.13±3.74	3.48±7.48	106.5±6.25
	MODEL 2	0.61±0.05	2.53±2.76	3.28±3.83	111.0±8.55
	MODEL 3	0.61±0.05	2.77±2.55	2.96±3.44	111.0±9.75
	RANKSVMC	0.61±0.05	2.55±3.70	3.30±7.42	95.0±4.90³
	SVCR	0.62±0.04	3.70±3.22¹	4.58±5.30¹	96.0±7.84 ³
	PH _{pspline}	0.57±0.04 ³	2.04±2.27	3.08±3.46	
PC	MODEL 1	0.76±0.04 ²	11.22±5.71 ¹	108.39±107.79 ¹	88.5±28.21³
	MODEL 2	0.78±0.03	14.56±6.47	148.08±137.34	112.0±29.54
	MODEL 3	0.78±0.03	14.52±6.86	144.72±163.35	89.5±8.47 ³
	RANKSVMC	0.76±0.03	11.89±6.00	111.26±116.90	264.0±11.38
	SVCR	0.78±0.03	13.15±6.60	150.66±165.24	160.0±62.93 ¹
	PH _{pspline}	0.77±0.03	12.76±5.89	286.6±615.05³	
BC	MODEL 1	0.63±0.04 ³	9.34±8.83 ³	10.10±20.91 ³	443.0±21.24 ³
	MODEL 2	0.68±0.03	22.62±9.09	24.23±13.19	457.0±5.60
	MODEL 3	0.68±0.03	22.42±8.51	24.20±13.30	456.5±22.93 ³
	RANKSVMC	0.62±0.10 ³	8.93±8.83 ³	11.20±15.23 ³	404.0±58.26³
	SVCR	0.68±0.10	19.51±7.89 ¹	18.99±12.02	453.0±75.70 ³
	PH _{pspline}	0.67±0.03	17.36±8.24 ¹	98.93±4052.05³	

¹, $p < 0.05$; ², $p < 0.01$; ³, $p < 0.001$ (Wilcoxon rank sum test)

Annexe 4

Légendes des figures

Figure 1 Fonctions de perte définies par svcr (a-b) [12] et svcr (c-d) [13]. Les fonctions de perte pour les événements sont indiquées en (a) et (c). Les fonctions de perte pour les points de données censurés à droite sont illustrées aux points (b) et (d). Les deux méthodes diffèrent dans la manière dont la fonction de perte est calculée. Un inconvénient majeur de la svcr La méthode est la nécessité de définir 4 hyper paramètres.

Figure 2 Illustration de la recherche d'un indice de pronostic minimisant le risque empirique de mal classer deux observations. La partie supérieure montre le classement de 7 observations en fonction du temps de survie et selon deux modèles pronostiques $u(x)$. La partie inférieure montre le calcul de l'indice de concordance. En supposant que tous les temps de survie sont connus exactement, toutes les paires peuvent être comparées. Une paire est concordante pour un certain modèle à chaque fois le classement dans le temps est égal au classement dans le modèle (pas de croix dans la partie supérieure) partie de la figure). (a) Le premier modèle donne un indice de pronostic qui fait fausse route 3 des 21 paires de points de données possibles. (b) Le deuxième modèle sur au contraire, 10 paires sont mal classées, ce qui donne un indice de concordance beaucoup plus bas.

Figure 3 : Illustration de la différence entre le modèle 2 (a) et le modèle 3 (b).

Dans le modèle 2, chaque observation i est comparée (indiquée par les flèches) à une observation avec un temps d'échec le plus proche de l'échec, mais inférieur à celui-ci heure d'observation i (y_i). Comme il s'agit d'un choix plutôt arbitraire, le modèle 3 comprend des observations factices avec des temps d'échec répartis de manière équidistante sur l'axe du temps (indiqué par les lignes verticales en pointillés). Sur chaque observation, On compare ensuite i avec 2 observations factices: (i) l'observation factice avec le temps de défaillance le plus élevé inférieur à y_i (indiqué par les flèches); et (ii) l'observation fictive avec le temps de défaillance le plus bas supérieur à y_i (non montré).

Figure 4 : Distribution des poids estimés (en utilisant un noyau linéaire) de la

Le score de Karnofsky sur 1000 bootstraps d'un ensemble d'entraînement particulier du

Jeu de données vlc: (a) modèle 1, (b) modèle 2. Le modèle 1 a une large distribution sans estimation claire (estimations avec signes différents). Ajout des contraintes de régression dans le modèle 2 entraînent une relation positive incontestable entre le score de Karnofsky et le temps de survie ($\text{poids} > 0$).

Figure 5 : Distribution de l'indice de concordance au test sur 1000 bootstraps d'un ensemble particulier de formation du jeu de données vlc: (a) modèle 1, (b) modèle 2. Le modèle 1 a une large distribution à partir de laquelle on ne peut pas décider si l'indice de pronostic estimé est lié ou inversement au temps d'échec.

L'ajout des contraintes de régression dans le modèle 2 conduit à un rétrécissement de cette Distribution. Il en résulte une relation positive claire entre le pronostique indice et le temps de survie ($\text{c-index} > 0.5$).

Figure 6 : Comparaison entre les modèles modèle 1 (trait continu) et modèle 2 (ligne pointillée) pour la censure aléatoire (a) et informative (b) dans un format à 10 dimensions exemple artificiel. Toutes les variables étaient normalement distribuées avec une moyenne nulle et variance égale à 1. Toutes les variables n'étaient pas corrélées à l'exception de 4 paires de variables avec des corrélations de 0,7, 0,3, -0,7 et -0,3. Le coefficient fixe le vecteur w a été pris. Le temps d'échec a été échantillonné à partir d'un disque exponentiel. la distribution avec un paramètre égal à $w^T x$, x étant le vecteur variable. Le pourcentage d'observations censurées variait de 0 à 100%. Le plus large les différences de performances sont notées pour les faibles pourcentages de censure. Cependant, le modèle 2 semble mieux performer pour tous les pourcentages de censure, tant pour censure aléatoire et informative

Annexe 5

MODÈLE 1

Les méthodes proposées dans cet article sont basées sur le modèle de survie proposé dans [8]. Nous donnons la déviation du modèle dans cette première annexe. La survie modèle a été écrite comme un problème de classement avec des contraintes de classement arbitraires

Modèle 1

$$\min_{w, \varepsilon} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i$$

$$f(x) = \begin{cases} w^T (\varphi(x_i) - \varphi(x_{j(i)})) \geq y_i - y_{j(i)} - \varepsilon_i, & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0 & , \forall i = 1, \dots, n \end{cases} \quad (28)$$

Formulation du lagrangien de (28)

$$\mathcal{L}(w, \varepsilon, \alpha, \beta) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i - \sum_{i=1}^n \alpha_i w^T (\varphi(x_i) - \varphi(x_{j(i)})) - \sum_{i=1}^n \alpha_i (-y_i + y_{j(i)} + \varepsilon_i) - \sum_{i=1}^n \beta_i \varepsilon_i,$$

et résoudre les équations de Karush-Kuhn-Tucker (KKT) [28]

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n \alpha_i (\varphi(x_i) - \varphi(x_{j(i)})) \\ \frac{\partial \mathcal{L}}{\partial \varepsilon_i} = 0 \rightarrow \gamma = \alpha_i + \beta_i & \forall i = 1, \dots, n \\ \alpha_i \geq 0 & \forall i = 1, \dots, n \\ \beta_i \geq 0 & \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i (w^T (\varphi(x_i) - \varphi(x_{j(i)})) - y_i + y_{j(i)} + \varepsilon_i) = 0 \\ \sum_{i=1}^n \beta_i \varepsilon_i = 0 \end{cases} \quad (30)$$

conduit, après l'élimination des variables primales et β_i au double

$$\min_{\alpha_i} \frac{1}{2} \sum_{i=1}^n \sum_{i=1}^n \alpha_i \alpha_k (\varphi(x_i) - \varphi(x_{j(i)}))^T (\varphi(x_k) - \varphi(x_{j(k)})) - \sum_{i=1}^n \alpha_i (y_i - y_{j(i)})$$

Sous réserve de $0 \geq \alpha_i \geq \gamma, \forall i = 1, \dots, n$.

(31)

L'indice de pronostic $u(x^*)$ pour une nouvelle observation x^* peut alors être trouvé sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n \alpha_i (\varphi(x_i) - \varphi(x_{j(i)}))^T \varphi(x^*). \quad (32)$$

Annexe 6

MODÈLE 2

Cette annexe et l'annexe suivante contiennent les dérivés de la nouvelle proposition des modèles. Pour chaque modèle, nous partons de la formulation du modèle donnée dans

Méthodes de support pour l'analyse de survie 24

Section 3, décrivent le lagrangien et dérivent les conditions de KKT pour l'optimalité.

Le modèle 2 a été défini comme

Modèle 2

$$\min_{w, \varepsilon, \xi, \xi^*} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i + \mu \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\begin{cases} w^T (\varphi(x_i) - \varphi(x_{j(i)})) \geq y_{(i)} - y_{j(i)} - \varepsilon_i, & \forall i = 1, \dots, n \\ w^T \varphi(x_i) + b \geq y_i - \xi_i, & \forall i = 1, \dots, n \\ -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \xi_i^*, & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0, \quad \xi_i \geq 0, \quad \xi_i^* \geq 0 & \forall i = 1, \dots, n \end{cases}$$

(33)

Le lagrangien devient

$$\begin{aligned} \mathcal{L}(w, \varepsilon, b, \alpha, \beta) = & \frac{1}{2} w^T w + \gamma \sum_{i=1}^n \varepsilon_i - \mu \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (w^T (\varphi(x_i) - \varphi(x_{j(i)})) - \\ & y_i + y_{j(i)} + \varepsilon_i) - \sum_{i=1}^n \beta_i (w^T \varphi(x_i) + b - y_i + \xi_i) - \sum_{i=1}^n \beta_i^* (-\delta_i (w^T \varphi(x_i) + b - y_i) + \\ & \xi_i^*) - \sum_{i=1}^n \eta_i \varepsilon_i - \sum_{i=1}^n \nu_i \xi_i - \sum_{i=1}^n \nu_i^* \xi_i^*. \end{aligned}$$

(34)

Les conditions KKT pour l'optimalité sont

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n \alpha_i \left(\varphi(x_i) - \varphi(x_{j(i)}) \right) + \sum_{i=1}^n (\beta_i - \delta_i \beta_i^*) \varphi(x_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^n (\beta_i - \delta_i \beta_i^*) \\ \frac{\partial \mathcal{L}}{\partial \varepsilon_i} = 0 \rightarrow \gamma = \alpha_i + \eta_i \quad \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \rightarrow \mu = \beta_i + \nu_i \quad \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \varepsilon_i} = 0 \rightarrow \gamma = \beta_i^* + \nu_i^* \quad \forall i = 1, \dots, n \\ \alpha_i, \beta_i, \beta_i^*, \eta_i, \xi_i, \xi_i^* \geq 0 \quad \forall i = 1, \dots, n \\ \sum_{i=0}^n \alpha_i \left(w^T \left(\varphi(x_i) - \varphi(x_{j(i)}) \right) - y_i + y_{j(i)} + \varepsilon_i \right) = 0 \\ \sum_{i=1}^n \beta_i \varepsilon_i = 0 \\ \sum_{i=1}^n \beta_i^* (-\delta_i (w^T \varphi(x_i) + b - y_i) + \xi_i^*) = 0 \\ \sum_{i=1}^n \eta_i \varepsilon_i = 0 \\ \sum_{i=1}^n \nu_i \xi_i = 0 \\ \sum_{i=1}^n \nu_i^* \xi_i^* = 0 \end{array} \right.$$

Après élimination de toutes les inconnues sauf α_i , β_i et β_i^* la solution est donnée comme suit :

$$\begin{aligned} \min_{\alpha_i, \beta_i, \beta_i^*} & \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \alpha_i \alpha_k \left(\varphi(x_i) - \varphi(x_{j(i)}) \right)^T \left(\varphi(x_k) - \varphi(x_{j(k)}) \right) + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n (\beta_i \beta_k + \\ & \delta_i \delta_k \beta_i^* \beta_k^*) \varphi(x_i)^T \varphi(x_k) - \sum_{i=1}^n \alpha_i (y_i - y_{j(i)}) + \sum_{i=1}^n \sum_{k=1}^n (\beta_i - \delta_i \beta_i^*) \alpha_k \varphi(x_i)^T \left(\varphi(x_k) - \right. \\ & \left. \varphi(x_{j(k)}) \right) - \sum_{i=1}^n \sum_{k=1}^n \beta_i^* \delta_i \beta_k \varphi(x_i)^T \varphi(x_k) - \sum_{i=1}^n \beta_i y_i + \sum_{i=1}^n \delta_i \beta_i^* y_i \end{aligned} \quad (36)$$

$$\text{Avec } \begin{cases} 0 \geq \alpha_i \geq \gamma, \forall i = 1, \dots, n \\ 0 \geq \beta_i \geq \mu, \forall i = 1, \dots, n \\ 0 \geq \beta_i^* \geq \mu, \forall i = 1, \dots, n \end{cases}$$

L'indice de pronostique $\hat{u}(x^*)$ pour une nouvelle observation x^* peut alors être trouvé sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n \left(\alpha_i \left(\varphi(x_i) - \varphi(x_{j(i)}) \right) + (\beta_i - \delta_i \beta_i^*) \varphi(x_i) \right) \varphi_p(x^*)^T + b \quad (37)$$

Annexe 7 :

MODÈLE 3

La formulation du modèle en incluant des observations factices à comparer avec le modèle 2 était la suivante:

Modèle 2

$$\min_{w, \varepsilon, \varepsilon_i^*, \xi, \xi_i^*, b, v} \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*) + \mu \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$\text{Sujet à } \begin{cases} w^T \varphi(x_i) + b - l_i^T v \geq y_{(i)} - l_i^T z - \varepsilon_i, & \forall i = 1, \dots, n \\ -\delta_i (w^T \varphi(x_i) + b - r_i^T (v - z) - y_i) \geq -\varepsilon_i^*, & \forall i = 1, \dots, n \\ w^T \varphi(x_i) + b \geq y_i - \xi_i, & \forall i = 1, \dots, n \\ -\delta_i (w^T \varphi(x_i) + b) \geq -\delta_i y_i - \xi_i^*, & \forall i = 1, \dots, n \\ \varepsilon_i \geq 0, \varepsilon_i^* \geq 0, \xi_i \geq 0, \xi_i^* \geq 0 & \forall i = 1, \dots, n \\ v_l - v_{l-1}, & \forall l = 1, \dots, k \end{cases}$$

(38)

Méthodes de support pour l'analyse de survie 26

Le lagrangien devient

$$\mathcal{L}(w, \varepsilon, \varepsilon_i^*, \xi, \xi_i^*, v, b, \alpha, \alpha_i^*, \beta, \beta_i^*, v, v_i^*, \eta, \eta_i^*, \theta) = \frac{1}{2} w^T w + \gamma \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*) + \mu \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (-w^T \varphi(x_i) + b - l_i^T (v - z) - y_i + \varepsilon_i) - \sum_{i=1}^n \alpha_i^* (-\delta_i (w^T \varphi(x_i) + b - r_i^T (v - z) - y_i) + \varepsilon_i^*) - \sum_{i=1}^n \beta_i (w^T \varphi(x_i) + b - y_i + \xi_i) - \sum_{i=1}^n \beta_i^* (-\delta_i (w^T \varphi(x_i) + b - y_i) + \xi_i^*) - \sum_{l=1}^k \theta_l (v_l - v_{l-1}) - \sum_{i=1}^n \eta_i \varepsilon_i - \sum_{i=1}^n \eta_i^* \varepsilon_i^* - \sum_{i=1}^n \nu_i \xi_i - \sum_{i=1}^n \nu_i^* \xi_i^*$$

(39)



Les conditions KKT pour l'optimalité sont

$$\left\{ \begin{array}{ll} \frac{\partial \mathcal{L}}{\partial w} = 0 \rightarrow w = \sum_{i=1}^n (\alpha_i + \beta_i - \delta_i(\alpha_i^* + \beta_i^*)) \varphi(x_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^n (\alpha_i + \beta_i - \delta_i(\alpha_i^* + \beta_i^*)) \\ \frac{\partial \mathcal{L}}{\partial \varepsilon_i} = 0 \rightarrow \gamma = \alpha_i + \eta_i & \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \varepsilon_i^*} = 0 \rightarrow \gamma = \alpha_i^* + \eta_i^* & \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0 \rightarrow \mu = \beta_i + \nu_i & \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \xi_i^*} = 0 \rightarrow \mu = \beta_i^* + \nu_i^* & \forall i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial v_l} = 0 \rightarrow \sum_{i=1}^n (\alpha_i l_i^T - \delta_i \alpha_i^* r_i^T - \theta_l + \theta_{l-1}) = 0, & \forall l = 1, \dots, k \\ \alpha_i, \alpha_i^*, \beta_i, \beta_i^*, \eta_i, \eta_i^*, \xi_i, \xi_i^* \geq 0 & \forall i = 1, \dots, n \\ \theta_l \geq 0 & \forall l = 1, \dots, k \end{array} \right. \quad (40)$$

$$\begin{aligned} \sum_{i=1}^n \alpha_i (w^T (\varphi(x_i) + b - l_i^T (v - z) - y_i + \varepsilon_i)) &= 0 \\ \sum_{i=1}^n \alpha_i^* (-\delta_i (w^T \varphi(x_i) + b - r_i^T (v - z) - y_i) + \varepsilon_i^*) &= 0 \\ \sum_{i=1}^n \beta_i (w^T \varphi(x_i) + b - y_i + \xi_i) &= 0 \\ \sum_{i=1}^n \beta_i^* (-\delta_i (w^T \varphi(x_i) - b + \delta_i y_i) + \xi_i^*) &= 0 \\ \sum_{i=1}^n \eta_i \varepsilon_i &= 0 \\ \sum_{i=1}^n \eta_i^* \varepsilon_i^* &= 0 \\ \sum_{i=1}^n \nu_i \xi_i &= 0 \\ \sum_{i=1}^n \nu_i^* \xi_i^* &= 0 \\ \sum_{l=1}^k \theta_l (v_l - v_{l-1}) &= 0 \end{aligned}$$

Méthodes de support pour l'analyse de survie 27 avec l_i , l et r_i , l le i ème élément des vecteurs l_i et r_i respectivement. Après élimination de certaines variables, la solution est trouvée comme

$$\begin{aligned} \min_{\alpha_i, \beta_i, \beta_i^*} & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i \alpha_j + \beta_i \beta_j) \varphi(x_i)^T \varphi(x_j) + \\ & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j (\alpha_i \alpha_j + \beta_i^* \beta_j^*) \varphi(x_i)^T \varphi(x_j) + \sum_{i=1}^n \sum_{k=1}^n (\alpha_i \beta_i - \delta_i ((\alpha_i + \beta_i)(\alpha_i^* + \\ & \beta_j^*))) \varphi(x_i)^T \varphi(x_j) - \sum_{i=1}^n \sum_{k=1}^n \delta_i \alpha_i^* \beta_j^* \varphi(x_i)^T \varphi(x_j) - \sum_{i=1}^n (\alpha_i - \delta_i \alpha_i^*) y_i + \sum_{i=1}^n (l_i - \\ & r_i)^T z - \sum_{i=1}^n \beta_i y_i + \sum_{i=1}^n \delta_i \beta_i^* y_i \end{aligned}$$

$$\text{Avec } \left\{ \begin{array}{l} 0 \geq \alpha_i \geq \gamma, \forall i = 1, \dots, n \\ 0 \geq \alpha_i^* \geq \gamma, \forall i = 1, \dots, n \\ 0 \geq \beta_i \geq \mu, \forall i = 1, \dots, n \\ 0 \geq \beta_i^* \geq \mu, \forall i = 1, \dots, n \\ \sum_{i=1}^n (\alpha_i l_i^T - \delta_i \alpha_i^* r_i^T) - \theta_l + \theta_{l-1}, \forall l = 1, \dots, k \end{array} \right.$$

Méthode de vecteur support pour l'analyse de survie : comparaison entre classement et régression approchée

Louis Mamadou Soropogui /201205FTB/option : STAT/DMA/FST/UCAD 2019

L'indice de pronostic $u(x^*)$ pour une nouvelle observation x^* peut alors être trouvé sous la forme

$$\hat{u}(x^*) = \sum_{i=1}^n (\alpha_i + \beta_i - \delta_i(\alpha_i^* + \beta_i^*)) \varphi(x_i)^T \varphi(x^*) + b$$

Annexe 8:

Figure :

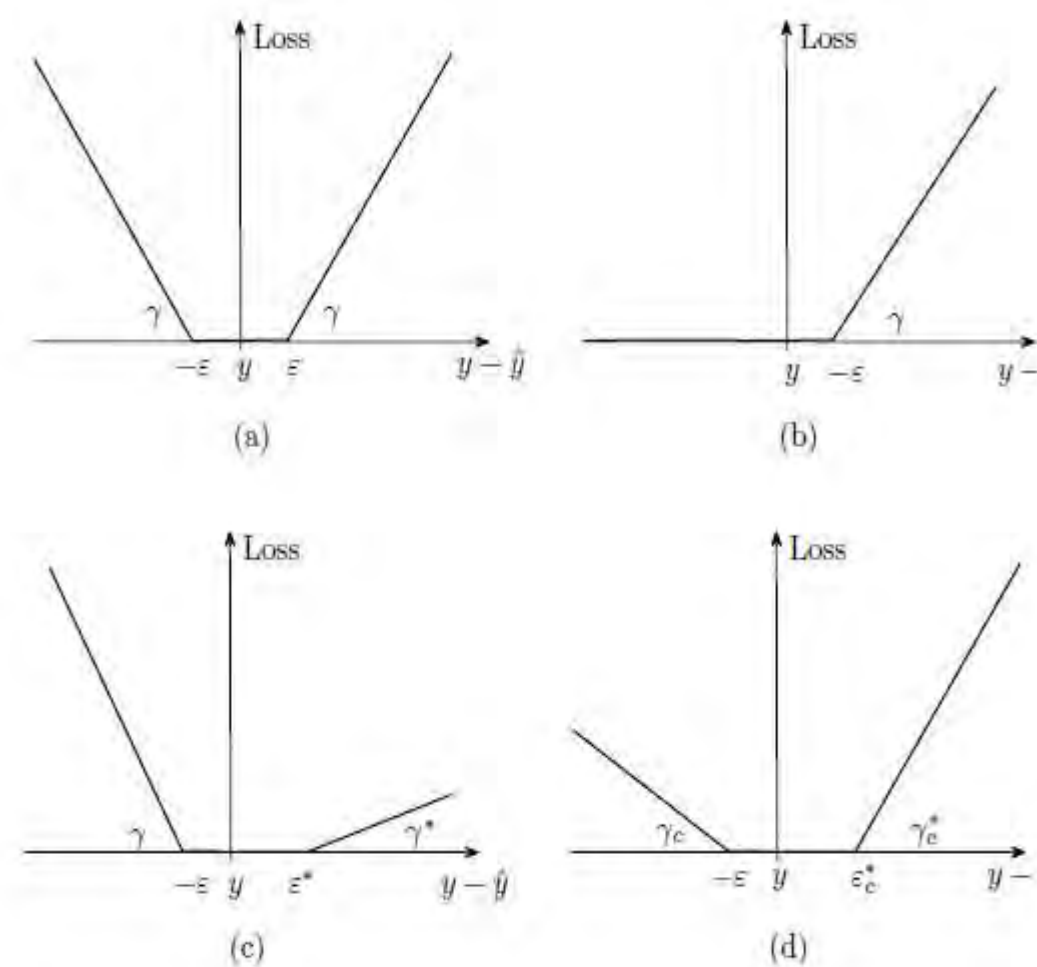


Fig. 1. Van Belle *et al.*

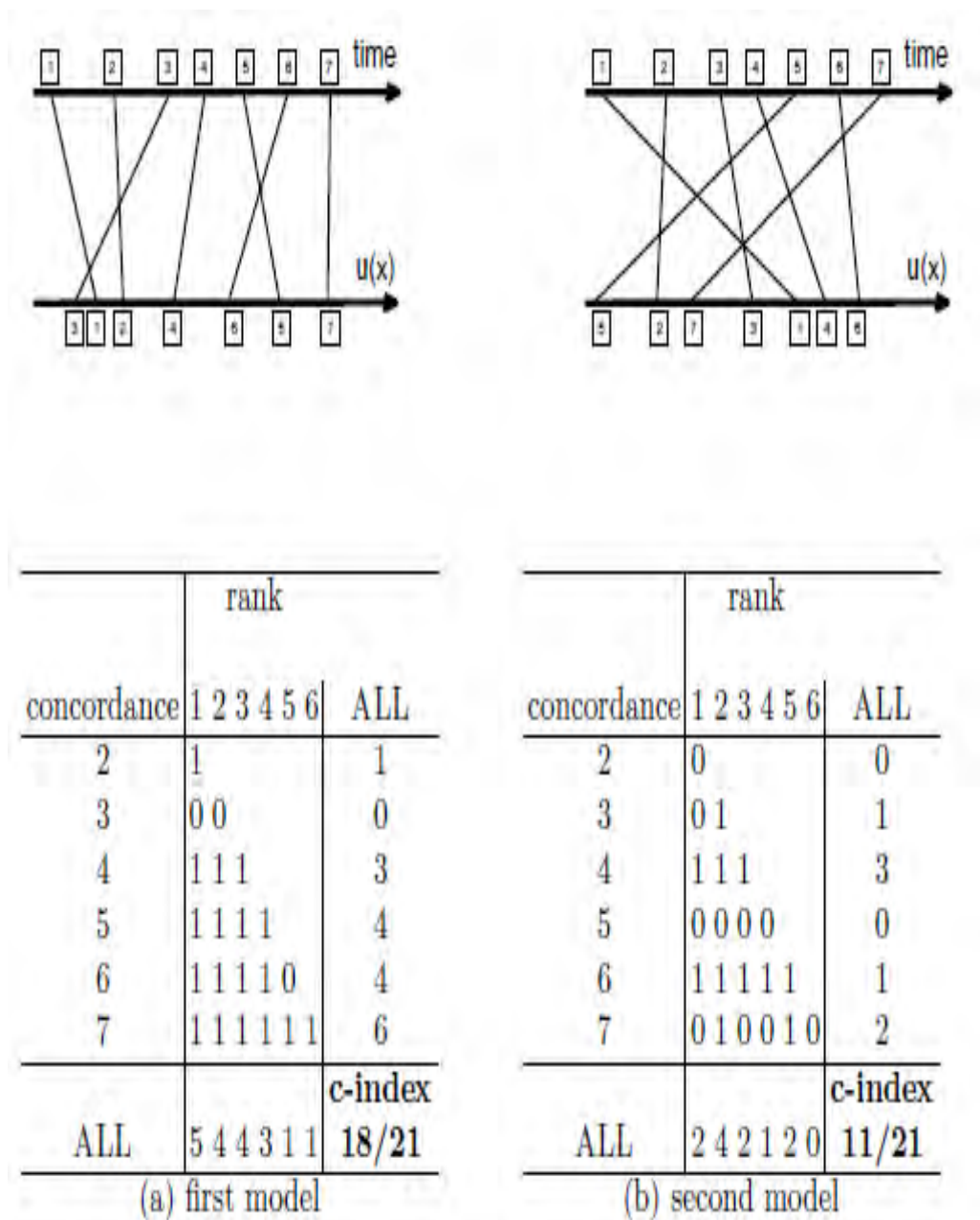


Fig. 2. Van Belle *et al.*

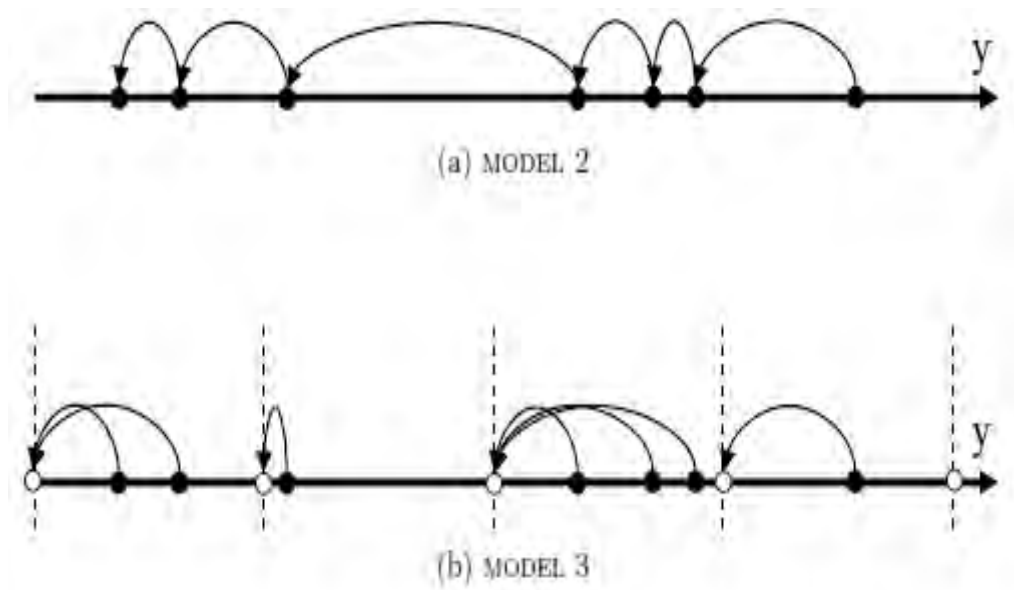


Fig. 3. Van Belle *et al.*

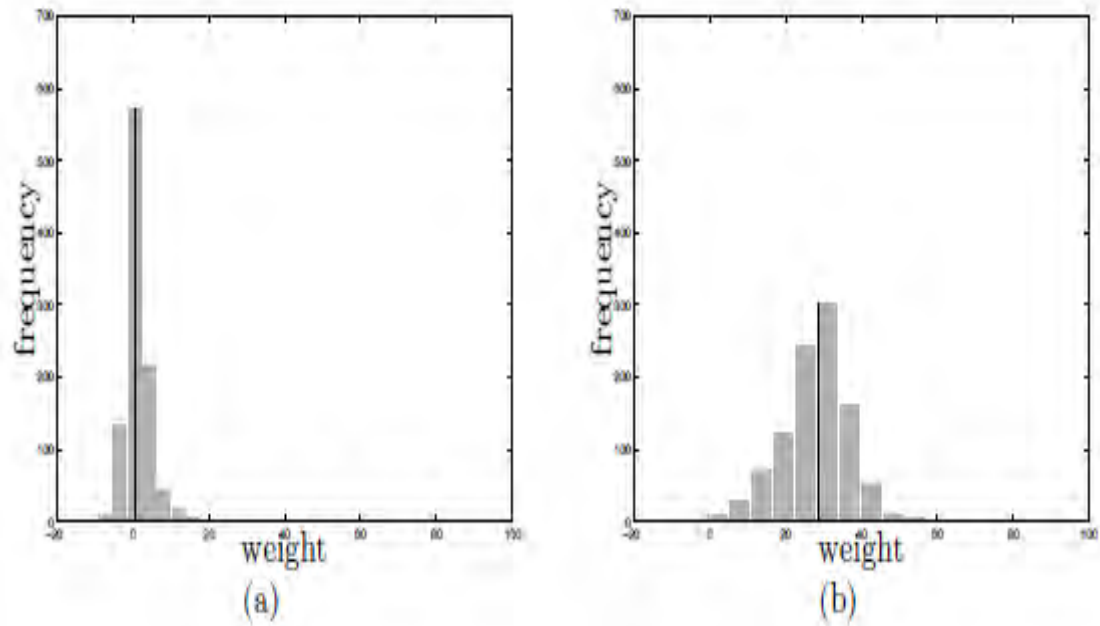


Fig. 4. Van Belle *et al.*

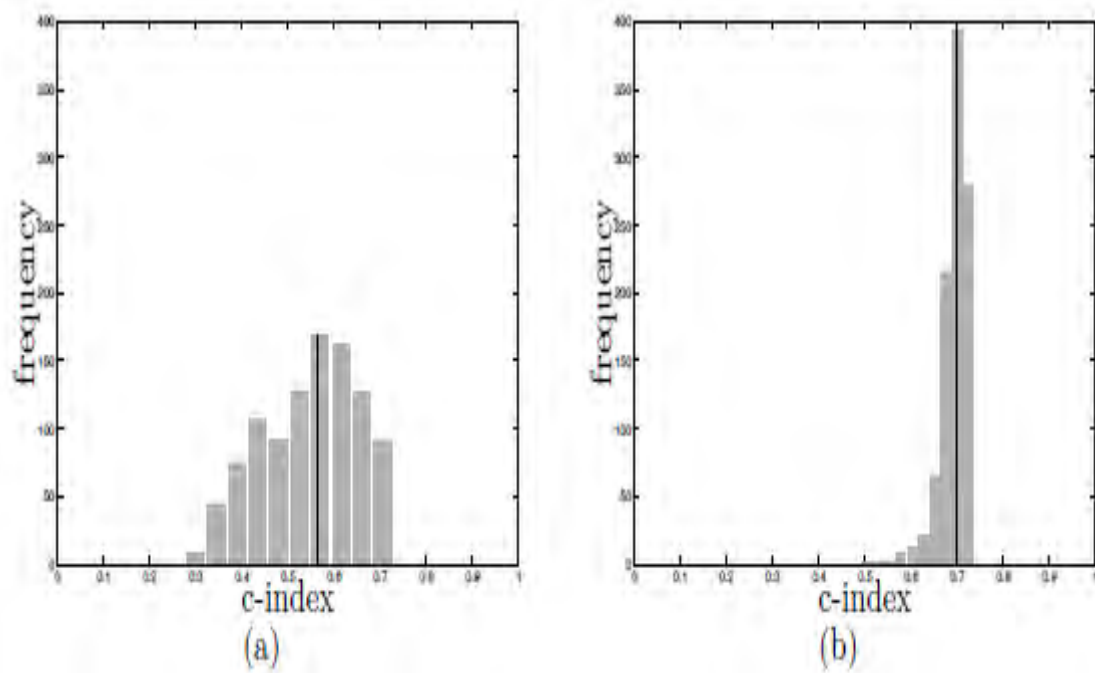


Fig. 5. Van Belle *et al.*

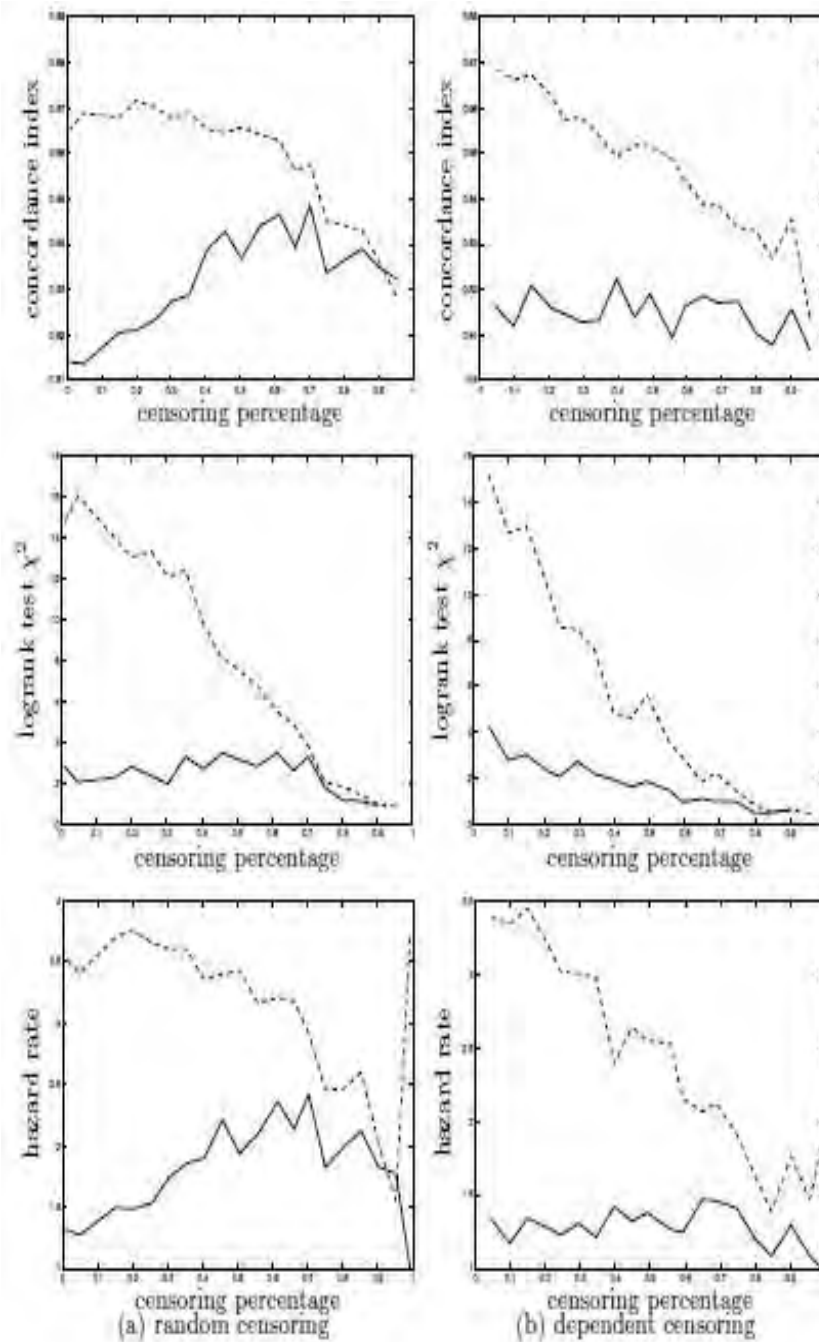


Fig. 6. Van Belle *et al.*