



Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas

Modelos não-lineares para Análise de Sobrevivência

Beatriz Nogueira Costa

Orientador: Saul de Castro Leite
Coorientador: Luneque Del Rio de Souza e Silva
Santo André - SP, Agosto de 2023

Beatriz Nogueira Costa

Modelos não-lineares para Análise de Sobrevivência

Dissertação de Graduação apresentada ao Programa de Graduação em Engenharia da Informação (área de concentração: Aprendizado de máquina), como parte dos requisitos necessários para a obtenção do Título de XXX em Engenharia da Informação.

Trabalho aprovado. Santo André - SP, 10 de agosto de 2023:

Saul de Castro Leite
Orientador

**Luneque Del Rio de Souza e Silva
Junior**
Co-Orientador

Professor
Cláudio José Bordin Júnior

Professor
João Ricardo Sato

Santo André - SP
Agosto de 2023

Agradecimentos

Agradeço à todos que me incentivaram, motivaram e não me deixaram desistir da graduação.

Resumo

Neste trabalho estuda-se a eficácia de métodos para o problema de análise de sobrevivência em dados lineares e não-lineares. Para tanto, são estudados métodos estatísticos tradicionais para solução deste problema como o método de Cox e de Kaplan-Meier e Nelson-Aalen. Além disso, são estudados os métodos baseados em aprendizado de máquinas, mais especificamente máquinas de vetores de suporte (SVM), nas formulações clássicas para regressão e ranqueamento. Em particular, as adaptações das formulações para o problema de análise de sobrevivência. Estes métodos podem tratar problemas não-lineares através do "truque do kernel".

São realizados experimentos para avaliar a eficácia de cada método estudado. Utiliza-se o pacote estatístico R, que dispõe da biblioteca virtual *Survival*, de onde são obtidos dados de estudos reais. Também são criados conjuntos de dados artificiais a fim de analisar o resultado dos métodos para dados não-lineares.

Os resultados dos experimentos mostram que, para dados lineares, os métodos clássicos e os métodos baseados em aprendizado de máquina são equivalentes. No entanto, algumas formulações se destacam quando aplicadas em dados não-lineares, indicando melhor eficácia neste cenário.

Palavras-chaves: análise de sobrevivência. máquinas de vetores de suporte.

Lista de ilustrações

Figura 1 – Exemplo de tempo de falha e de censura. Para os pacientes 1 e 3 existe tempo de falha. Já para os pacientes 2 e 4 possuem tempo de censura, sendo que o primeiro estava vivo quando o estudo terminou e, o segundo, abandonou o estudo (JAMES et al., 2013).	4
Figura 2 – Curva de sobrevivência de Kaplan-Meier (linha sólida) e erro padrão (curvas tracejadas) para o conjunto de dados BrainCancer (JAMES et al., 2013).	8
Figura 3 – Lung	22
Figura 4 – Diabetic	22
Figura 5 – Veteran	23
Figura 6 – Dados não lineares (1)	23
Figura 7 – Dados não lineares (2)	24

Lista de tabelas

Tabela 1 – Lung	18
Tabela 2 – Diabetic	19
Tabela 3 – Veteran	19

Sumário

1	INTRODUÇÃO	1
	Introdução	1
1.1	Objetivos	2
2	FUNDAMENTAÇÃO TEÓRICA	3
2.1	Formulação	3
2.2	Métodos Estatísticos Tradicionais	6
2.2.1	Kaplan-Meier e Nelson-Aalen	6
2.2.2	Método de Cox	9
2.3	Abordagem por Máquinas de Vetores de Suporte	10
2.3.1	Máquinas de vetores de suporte para regressão	10
2.3.2	Máquinas de vetores de suporte para regressão para dados censurados	11
2.3.3	Máquinas de vetores de suporte para ranqueamento	12
2.3.4	Máquinas de vetores de suporte para ranqueamento com dados censurados	14
2.3.5	Método combinado	15
3	METODOLOGIA	17
3.1	Métodos Utilizados	17
3.2	Bancos de Dados	17
3.2.1	Dados Reais	17
3.2.2	Dados Artificiais	19
3.3	Métrica de Comparação	20
4	RESULTADOS E DISCUSSÃO	21
	Conclusão e Trabalhos Futuros	25
	REFERÊNCIAS	27

1 Introdução

Atualmente, dados são gerados e coletados com facilidade e este fenômeno tem impulsionado o desenvolvimento de novas ferramentas cada vez mais sofisticadas para tratamento e análise destes dados. A área de aprendizado de máquinas em particular tem um papel importante para compreensão do problema estudado e realização de previsões.

Em aprendizado de máquinas, a análise de dados é geralmente dividida em tarefas, ou problemas, que possuem algoritmos bem definidos para sua solução. Dentre os problemas mais comuns, pode-se destacar o aprendizado supervisionado, não-supervisionado e por reforço. Para o primeiro, em geral, o problema é de classificação ou regressão. Neste caso, um conjunto de dados é conhecido e possui rótulos, também conhecidos, determinados por um especialista. Em um problema de classificação, o rótulo toma valor em um conjunto discreto, já no problema de regressão, pode tomar valor reais (MONARD; BARANAUSKAS, 2003). Por exemplo, é possível pensar no problema de verificar a relação entre expectativa de vida e PIB per capita como um problema de regressão, pois estes valores tomam valores reais, enquanto a verificação de spam em uma caixa de entrada de email é um problema de classificação, pois os valores são discretos (IZBICKI; SANTOS, 2020).

Um problema de aprendizado supervisionado que não recebe tanta atenção como os mencionados no parágrafo anterior é o problema da análise de sobrevivência (JAMES et al., 2013). Análise de sobrevivência teve início como um campo da estatística e possui diversas aplicações. O seu objetivo é analisar e modelar dados onde o evento de interesse é o tempo até a ocorrência de um evento, ou seja, fazer a previsão de eventos que ocorrem no tempo. Por exemplo, em um estudo clínico o evento de interesse pode ser o tempo até o óbito ou tempo até tratamento de um paciente. Muitas vezes esta previsão é feita conhecendo características do indivíduo, que são chamadas de covariáveis. O maior desafio é a existência de eventos com tempo não observado ou tempo não contido no intervalo de monitoramento. Isso é chamado de *censura* e a melhor forma de modelar esses dados é usando técnicas de análise de sobrevivência (LIU, 2012).

Considere um estudo que ocorre em um intervalo de tempo $[0, \tau]$ com um conjunto de covariáveis. O evento de interesse (chamado de "tempo de falha" ou "tempo de sobrevivência") ocorre dentro desse intervalo e é conhecido. Aquele cujo tempo não é conhecido ou está fora do período de estudo é chamado de censurado, sendo essa a principal característica dos dados onde será aplicada a análise de sobrevivência. Nesse caso, tem-se uma estimativa inferior ou superior para este valor, sendo chamada, respectivamente, de censura à direita e à esquerda. O tempo de sobrevivência é independente da censura.

As observações com censura não devem ser descartadas da análise, pois fornecem

informações importantes e, caso omitidas, pode levar à conclusões viciadas ou enviesadas. Por isso, torna-se necessário utilizar a abordagem de análise de sobrevivência para esse conjunto de dados ao invés de técnicas estatísticas clássicas.

Nesse trabalho são estudados métodos estatísticos tradicionalmente aplicados em dados censurados como Kaplan-Meier e Nelson-Aalen e Método de Cox. Também serão apresentadas abordagens através de máquinas de vetores de suporte, mais especificamente usando a técnica de regressão, ranqueamento e a combinação destas.

1.1 Objetivos

O objetivo deste trabalho é comparar os métodos estatísticos tradicionais para solução do problema de análise de sobrevivência com métodos baseados em aprendizado de máquinas, em particular os métodos baseados em kernel, como é o caso da máquina de vetores de suporte. Esta comparação será feita usando dados frequentemente utilizados para o estudo de análise de sobrevivência bem como em dados artificiais não-lineares. Deseja-se verificar se os métodos tradicionais conseguem captar a não-linearidade dos dados.

2 Fundamentação Teórica

Neste capítulo será apresentado formalmente o problema de análise de sobrevivência, bem como alguns dos métodos existentes para solucioná-lo. Inicialmente define-se o problema de análise de sobrevivência e a notação que será empregada. Em seguida, são apresentados alguns métodos clássicos, a saber: Kaplan-Meier, Nelson-Aalen e método de Cox. É apresentado o método tradicional de aprendizado supervisionado - SVM - e, então, sua formulação para um conjunto de dados censurados. Mais especificamente, aborda-se o problema por regressão e ranqueamento. Além disso, apresenta-se um método que combina as duas abordagens estudadas.

2.1 Formulação

Considera-se um problema com indivíduos que estão sujeitos a um evento que ocorre aleatoriamente no tempo. Por exemplo, um estudo clínico, que ocorre durante 5 anos, cujo evento é a data de óbito do paciente. Esta data é representada pela variável aleatória T . O evento de interesse é chamado de "tempo de falha" ou "tempo de sobrevivência". Caso o paciente abandone o estudo antes de seu término ou permaneça vivo até o fim do estudo, chamamos o evento de censurado. Em um estudo, é observado o tempo de falha T ou o tempo de censura C (WANG; LI; REDDY, 2019). Ou seja, é a variável aleatória

$$Y = \min(T, C). \quad (2.1)$$

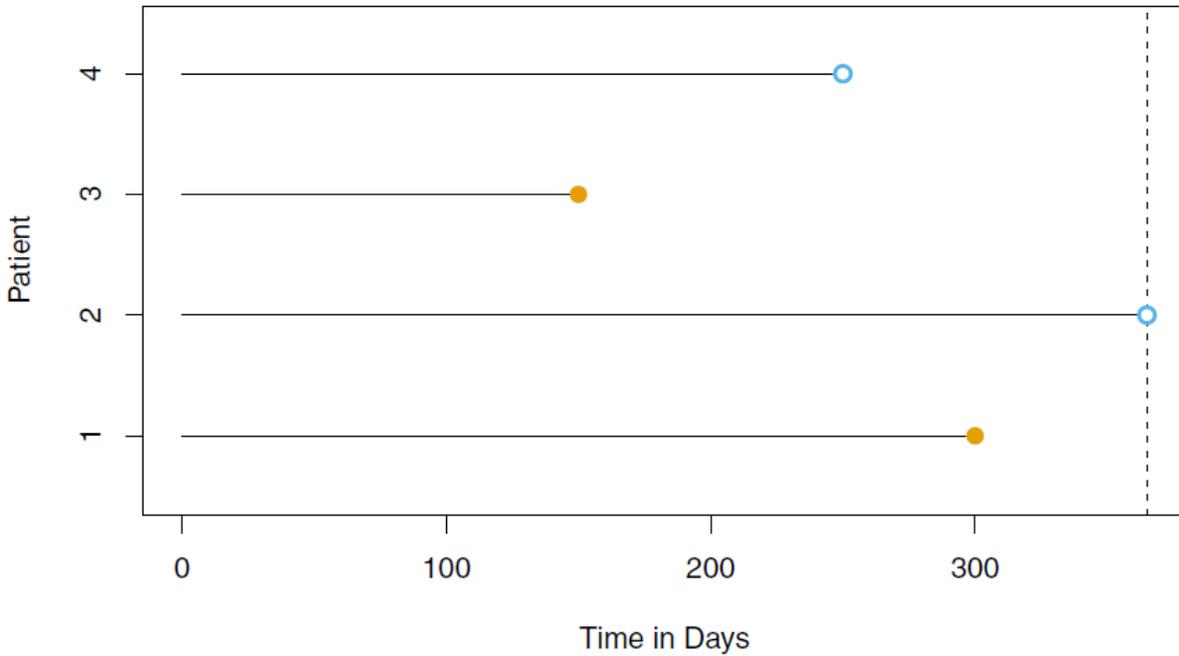
O conjunto de dados dos estudos são caracterizados pelo tempo de falha e censura. Também por conjuntos de covariáveis, ou seja, características que descrevem os indivíduos, como sexo biológico, idade, peso, entre outras. O tempo de falha é constituído de tempo inicial, escala de medida e evento de interesse (falha).

O tempo de início de estudo deve ser definido. Os dados são comparáveis no início do estudo, com exceção das covariáveis. A escala de medida é quase sempre o tempo real. O evento de interesse é o tempo de falha, que deve ser definido de forma clara e precisa.

Este problema pode ser resolvido por técnicas usuais estatísticas, como regressão linear. Porém, como o tempo de falha de alguns indivíduos é incompleto a solução do problema é um pouco mais complicada. Geralmente, esses dados são descartados, porém eles possuem informações valiosas para o estudo e o resultado costuma ser enviesado. A censura pode ser à direita ou à esquerda.

Censura à direita ocorre quando $T \geq Y$, ou seja, o tempo de falha real T é, no

Figura 1 – Exemplo de tempo de falha e de censura. Para os pacientes 1 e 3 existe tempo de falha. Já para os pacientes 2 e 4 possuem tempo de censura, sendo que o primeiro estava vivo quando o estudo terminou e, o segundo, abandonou o estudo (JAMES et al., 2013).



mínimo, o tempo observado Y . Note que $T \geq Y$ é uma consequência de (2.1). Tipicamente o tempo é representado da esquerda para direita, como na Figura 1, por isso o nome *censura à direita*. Censura à esquerda ocorre quando o tempo de falha real T é menor ou igual ao tempo observado Y . Por exemplo, em um estudo de tempo de gravidez são entrevistadas pacientes 250 dias depois da concepção, quando algumas já passaram pelo trabalho de parto. Neste caso, o tempo de gravidez é menor que 250 dias. Assim, o tempo de censura refere-se ao conjunto de dados que não é conhecido o tempo exato do evento, mas o intervalo onde ele cai é conhecido. Neste trabalho, serão estudados casos com censura à direita.

Os dados de sobrevivência de um indivíduo i são representados por (t_i, δ_i, x_i) , onde t_i é o tempo de falha, δ_i indica censura e x_i são as covariáveis (COLOSIMO; GIOLO, 2006), como dado abaixo:

$$\delta_i = \begin{cases} 1 & \text{se } t_i \text{ é um tempo de falha,} \\ 0 & \text{se } t_i \text{ é um tempo censurado.} \end{cases} \quad (2.2)$$

Uma variável importante para descrever estudos de sobrevivência é a função de sobrevivência, que é a probabilidade de uma observação não falhar até o tempo t :

$$S(t) = P(T \geq t). \quad (2.3)$$

Note que a função de distribuição acumulada $F(t) = 1 - S(t)$ é a probabilidade de uma observação não sobreviver até o tempo t . A função de falha ou taxa de falha é formalmente definida como

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad (2.4)$$

onde T é o dado censurado. Considerando Δt extremamente pequeno, podemos definir, informalmente, que

$$\lambda(t) \approx \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \quad (2.5)$$

A taxa de falha é importante, pois está relacionada com a curva de sobrevivência e é a base para modelar conjuntos de dados com censura como uma função de covariáveis.

Para dois eventos A e B , a probabilidade de A dado B é $P(A|B) = P(A \cap B)/P(B)$. Dado que $S(t) = P(T > t)$, temos,

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P((t \leq T < t + \Delta t) \cap (T \geq t)/\Delta t)}{P(T > t)} \quad (2.6)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)/\Delta t}{P(T > t)} \quad (2.7)$$

$$= \frac{f(t)}{S(t)}, \quad (2.8)$$

onde,

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2.9)$$

é a função densidade de probabilidade associada à T , ou seja, é a taxa de falha instantânea no tempo t . Na equação (2.6) usa-se $t < T \leq t + \Delta t$, então $T > t$.

A equação (2.8) mostra uma relação entre a função de falha $\lambda(t)$, a função de sobrevivência $S(t)$ e a função densidade de probabilidade $f(t)$. São formas equivalentes de descrever a distribuição de T . Em particular, a equação (2.8) implica na seguinte forma para $S(t)$: $S(t) = \exp(-\int_0^t \lambda(u) du)$.

A função de taxa de falha acumulada também é útil e de interesse para análise de sobrevivência. É definida por

$$\Lambda(t) = \int_0^t \lambda(u) du. \quad (2.10)$$

O tempo médio de vida pode ser obtido através da função de sobrevivência

$$t_m = \int_0^{\infty} S(t) dt. \quad (2.11)$$

É possível obter o tempo de vida média residual

$$vmr(t) = \frac{\int_0^{\infty} (u-t)f(u) du}{S(t)} = \frac{\int_0^{\infty} S(u) du}{S(t)}, \quad (2.12)$$

onde $f(\cdot)$ é a função de densidade de T . Essa variável significa que para indivíduos com idade t esta quantidade mede o tempo médio restante de vida e é, então, a área sob a curva de sobrevivência à direita do tempo t dividida por $S(t)$.

2.2 Métodos Estatísticos Tradicionais

Este capítulo apresenta as principais técnicas estatísticas para análise de sobrevivência. Serão explicados os métodos de Kaplan-Meier e Nelson-Aalen, taxa de falha e método de Cox.

2.2.1 Kaplan-Meier e Nelson-Aalen

É difícil estimar $S(t)$ na presença de dados censurados, uma estratégia para contornar esse problema é fazer uma estimativa $\hat{S}(t)$. Seja $t_1 < t_2 < \dots < t_K$ K taxas de falhas entre os dados não censurados e q_k o número de dados que falham no tempo t_k . Para $k = 1, \dots, K$, seja r_k a quantidade de dados sem falha anterior à t_k . Pela lei da probabilidade total (JAMES et al., 2013),

$$P(T > t_k) = P(T > t_k | T > t_{k-1})P(T > t_{k-1}) + P(T > t_k | T \leq t_{k-1})P(T \leq t_{k-1}) \quad (2.13)$$

Como $t_k < t_{k-1}$, então $P(T > t_k | T \leq t_{k-1}) = 0$. Logo,

$$S(t_k) = P(T > t_k) = P(T > t_k | T > t_{k-1})P(T > t_{k-1}) \quad (2.14)$$

$$= P(T > t_k | T > t_{k-1})S(t_{k-1}) \quad (2.15)$$

$$= P(T > t_k | T > t_{k-1}) \times \dots \times P(T > t_1 | T > t_2)P(T > t_1). \quad (2.16)$$

Tomando o estimador

$$\hat{P}(T > t_j | T > t_{j-1}) = (r_j - q_j)/r_j, \quad (2.17)$$

que é a fração do risco no tempo t_j para o evento que durou após t_j . Dessa forma, o estimador de Kaplan-Meier da curva de sobrevivência é

$$\hat{S}(t_k) = \prod_{j=1}^k \frac{r_j - q_j}{r_j}. \quad (2.18)$$

Para os tempos t entre t_k e t_{k+1} , temos $\hat{S}(t) = \hat{S}(t_k)$. Por isso, a curva de sobrevivência de Kaplan-Meier é uma função escada, com degraus nos tempos observados de falha de tamanho $1/n$, onde n é o tamanho da amostra, como mostra a Figura 2.

O estimador de Kaplan-Meier se reduz à função de sobrevivência empírica quando não existem censuras (GOEL; KHANNA; KISHORE, 2010), sendo esta

$$\hat{S}(t) = \frac{\text{número de observações que não falharam até o tempo } t}{\text{número total de observações no estudo}} \quad (2.19)$$

Ele não gera vícios à medida que o tamanho da amostra aumenta, fracamente consistente, converge assintoticamente para um processo gaussiano; é estimador de máxima verossimilhança de $S(t)$ e está disponível em diversos pacotes estatísticos.

O estimador de Kaplan-Meier está sujeito a variações que devem ser descritas em termos de estimações intervalares. Sua variância assintótica é

$$\widehat{\text{var}}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j=1}^k \frac{q_j}{r_j(r_j - q_j)}. \quad (2.20)$$

Esta expressão é conhecida como fórmula de Greenwood (COLOSIMO; GIOLO, 2006).

Para valores de t muito altos, o intervalo de confiança pode apresentar limite inferior negativo ou limite superior maior que 1. Nesses casos, é necessários fazer uma transformação de $S(t)$ para resolver o problema, por exemplo, fazemos $\hat{U}(t) = \log[-\log(\hat{S}(t))]$

$$\widehat{\text{var}}(\hat{U}(t)) = \frac{\sum_{j=1}^k \frac{q_j}{r_j(r_j - q_j)}}{[\sum_{j=1}^k \log(\frac{r_j - q_j}{r_j})]^2}. \quad (2.21)$$

Um intervalo aproximado de $100 \cdot (1 - \alpha)\%$ de confiança para $S(t)$ é dado por

$$\hat{S}(t) \cdot \exp\{\pm z_{\alpha/2} \sqrt{\widehat{\text{var}}(\hat{U}(t))}\}, \quad (2.22)$$

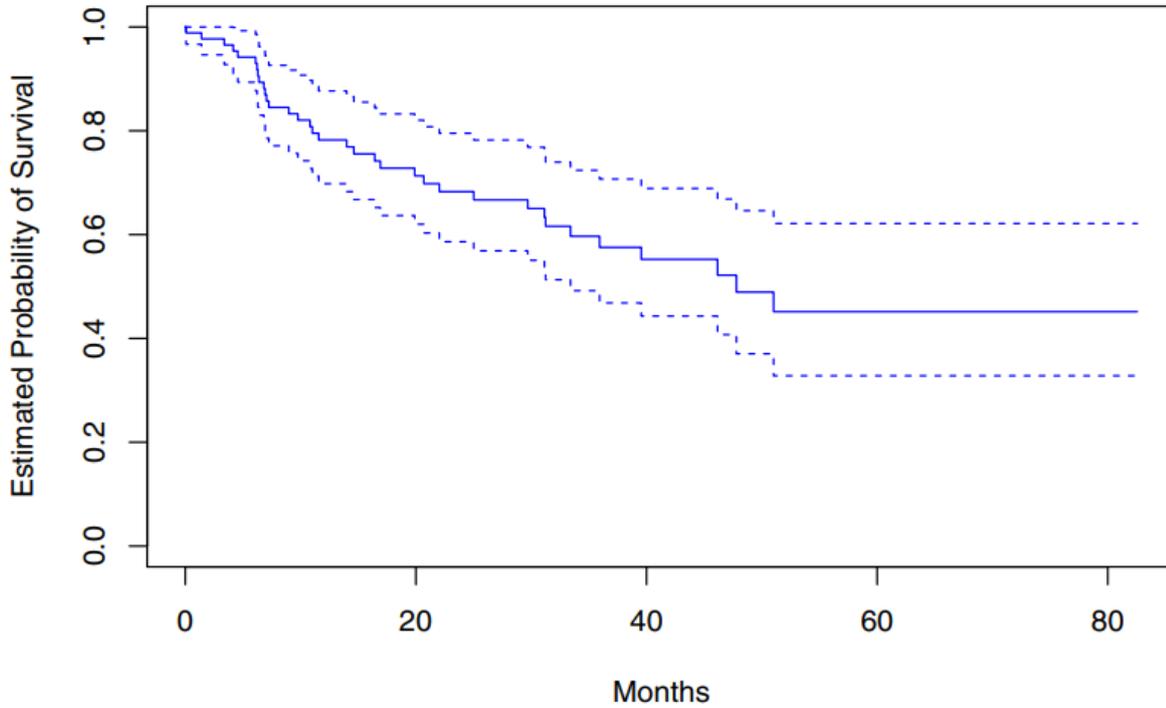
que assume valores no intervalo $[0, 1]$, sendo $\hat{S}(t)$ com distribuição assintótica normal (COLOSIMO; GIOLO, 2006).

O estimador de Nelson-Aalen é mais recente do que o de Kaplan-Meier e baseia-se na expressão

$$S(t) = \exp(-\Lambda(t)) = \exp\left\{-\int_0^t \lambda(u) du\right\}, \quad (2.23)$$

onde $\Lambda(t)$ é a função de risco acumulada. Isto é, o estimador de Nelson-Aalen é dado por $\tilde{S}(t) = \tilde{\Lambda}(t)$, em que

Figura 2 – Curva de sobrevivência de Kaplan-Meier (linha sólida) e erro padrão (curvas tracejadas) para o conjunto de dados BrainCancer (JAMES et al., 2013).



$$\tilde{\Lambda}(t) = \sum_{j=1}^k \frac{q_j}{r_j}. \quad (2.24)$$

A variância do estimador para $\Lambda(t)$, proposta por Aalen (COLOSIMO; GIOLO, 2006), é dada por

$$\widehat{var}(\tilde{\Lambda}(t)) = \sum_{j=1}^k \frac{q_j}{r_j^2}. \quad (2.25)$$

Dessa forma, a variância de $\tilde{\Lambda}(t)$ é dada por

$$\widehat{var}(\tilde{S}(t)) = [\tilde{S}(t)]^2 \sum_{j=1}^k \frac{q_j}{r_j^2}. \quad (2.26)$$

Os estimadores de Kaplan-Meier e Nelson-Aalen apresentam na maioria das vezes estimativas muito próximas para $S(t)$, sendo o último maior ou igual que o primeiro. Ambos não consideram covariáveis e, para contornar o problema, pode-se calcular o tempo de sobrevivência de agrupamentos de indivíduos que compartilham características através de técnicas de estratificação.

2.2.2 Método de Cox

O método de Cox faz uma suposição básica, que é da taxa de falha proporcional, que afirma que

$$\lambda(t|x_i) = \lambda_0(t) \cdot \exp\left(\sum_{j=1}^p x_{ij}\beta_j\right), \quad (2.27)$$

onde $\lambda_0(t) \geq 0$ é uma função não determinada conhecida como parâmetro de falha. Também pode ser escrita de forma genérica como

$$\lambda(t|x_i) = \lambda_0(t) \cdot g\left(\sum_{j=1}^p x_{ij}\beta_j\right). \quad (2.28)$$

A função de falha para um dado com vetor de características x_i é uma função $\lambda_0(t)$ desconhecida multiplicada pelo fator $\exp(\sum_{j=1}^p x_{ij}\beta_j)$, por isso chamamos $\lambda(t|x_i)$ de "falha proporcional". O valor $\exp(\sum_{j=1}^p x_{ij}\beta_j)$ é chamado de risco relativo para o vetor $x_i = (x_{i1}, \dots, x_{ip})^t$, para o vetor de características (covariáveis) $x_i = (0, \dots, 0)^t$

A possibilidade do parâmetro de falha $\lambda_0(t)$ assumir qualquer formato, faz com que a função de falha seja bastante flexível e possa modelar diversos tipos de relação entre as covariáveis e o tempo de sobrevivência. A única suposição feita é de que o crescimento de x_{ij} corresponde a um crescimento de $\lambda(t|x_i)$ por um fator $\exp(\beta_j)$.

A máxima verossimilhança associada com a i -ésima observação é

$$L_i = \begin{cases} f(t_i) & \text{se a } i\text{-ésima observação não é censurada} \\ S(t_i) & \text{se a } i\text{-ésima observação é censurada} \end{cases} \quad (2.29)$$

$$= f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}. \quad (2.30)$$

Isso significa que se $T = t_i$ e a i -ésima observação não é censurada, então L_i é a probabilidade de falha em um intervalo pequeno em torno de t_i . Mas se a i -ésima observação é censurada, então L_i é a probabilidade de sobrevivência até pelo menos t_i . Assumindo que temos n observações independentes e que a probabilidade dos dados é

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i), \quad (2.31)$$

onde a segunda igualdade é derivada de (2.8).

Como $\lambda_0(t)$ é uma função desconhecida, não é possível substituir $\lambda(t|x_i)$ na equação (2.31) para estimar $\beta = (\beta_1, \dots, \beta_p)^t$ por máxima verossimilhança. O método de Cox permite estimar β sem precisar definir $\lambda_0(t)$.

Seja $\delta_i = 0$, ou seja, a i -ésima observação é não censurada e, portanto, t_i é seu tempo de falha. Assim, a função de falha para a i -ésima observação no tempo t_i é $\lambda(t_i|x_i) = \lambda_0(t_i) \cdot \exp(\sum_{j=1}^p x_{ij}\beta_j)$ e a falha total no tempo t_i para as observações em risco de falha é

$$\sum_{i': y_{i'} \geq t_i} \lambda_0(t_i) \cdot \exp\left(\sum_{j=1}^p x_{i'j}\beta_j\right). \quad (2.32)$$

Portanto, a probabilidade de que a i -ésima observação irá falhar no tempo t_i ao invés de uma das outras observações no conjunto de risco é

$$\frac{\lambda_0(t_i) \cdot \exp(\sum_{j=1}^p x_{ij}\beta_j)}{\sum_{i': t_{i'} \geq t_i} \lambda_0(t_i) \cdot \exp(\sum_{j=1}^p x_{i'j}\beta_j)} = \frac{\exp(\sum_{j=1}^p x_{ij}\beta_j)}{\sum_{i': t_{i'} \geq t_i} \exp(\sum_{j=1}^p x_{i'j}\beta_j)}. \quad (2.33)$$

A máxima verossimilhança parcial é o produto dessas probabilidades em todas as observações não censuradas

$$PL(\beta) = \prod_{i: \delta_i=1} \frac{\exp(\sum_{j=1}^p x_{ij}\beta_j)}{\sum_{i': t_{i'} \geq t_i} \exp(\sum_{j=1}^p x_{i'j}\beta_j)}. \quad (2.34)$$

É válida independentemente do valor de $\lambda_0(t)$, tornando o modelo bastante flexível e robusto. Para estimar β basta maximizar (2.34) em relação a β .

A principal vantagem do método de Cox é a consistência dos resultados obtidos. Também permite identificar as covariáveis que tem influência significativa na taxa de risco entre os elementos.

2.3 Abordagem por Máquinas de Vetores de Suporte

Neste capítulo serão apresentados métodos de análise de sobrevivência baseados em aprendizado de máquinas. Aborda-se o método tradicional de máquina de vetores de suporte (SVM, do inglês *support vector machines*) para regressão e para ranqueamento. Em seguida, apresenta-se sua formulação considerando dados censurados. Também são apresentadas duas formulações combinando-se modelos.

2.3.1 Máquinas de vetores de suporte para regressão

A regressão SVM (SVR) é um método para estimar uma função que retorna um número real a partir de um dado de entrada baseado no conjunto de dados de treinamento. A SVR tem as mesmas propriedades que o SVM: maximização da margem e truque do kernel para problemas não-lineares (YU; KIM, 2012).

Dado o conjunto de treinamento $D = (x_i, t_i), \dots, (x_n, t_n)$, o problema consiste em encontrar uma função $f: \mathbb{R}^m \rightarrow \mathbb{R}$ que aproxima t_i para o x_i correspondente. Ou seja,

$$f(x) = w^t x + b, \quad (2.35)$$

onde w é um vetor e b é o viés.

A perda ϵ -insensível, dada por $c(e) = \max(0, |e| - \epsilon)$, onde $e = f(x_i) - t_i$, é minimizada com regularização controlando a qualidade (SHIVASWAMY; CHU; JANSCHKE, 2007). A perda ϵ -insensível é zero desde que a diferença absoluta do valor real e do valor previsto seja menor que ϵ . Há um custo, que cresce linearmente, quando a diferença absoluta é maior que ϵ . A solução do problema linear é

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (2.36a)$$

$$\text{sujeito a } \begin{cases} t_i - w^t x_i - b \leq \epsilon + \xi_i^*, & \xi_i^* \geq 0, & \forall 1 \leq i \leq n, \\ w^t x_i + b - t_i \leq \epsilon + \xi_i, & \xi_i \geq 0, & \forall 1 \leq i \leq n \end{cases}$$

sendo C uma constante de regularização estritamente positiva, ξ_i, ξ_i^* variáveis de folga que permitem erros nas previsões dos dados de treinamento e $\epsilon > 0$ o tamanho da margem.

A formulação dual permite o uso de kernels, simplificando a solução do problema e permitindo a solução de problemas não lineares, mapeando cada x_i para um espaço de Hilbert \mathcal{H} com $\phi : \mathbb{R}^m \rightarrow \mathcal{H}$. O mapeamento pode ser realizado de forma implícita através de $k : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$, tal que $k(x_i, x_j) = \phi(x_i)^t \phi(x_j)$, onde k é um kernel definido positivo. A formulação dual é (SMOLA; SCHÖLKOPF, 2004)

$$\min_{\alpha, \alpha^*} \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) - \sum_{i=1}^n (t_i - \epsilon) \alpha_i + \sum_{i=1}^n (t_i + \epsilon) \alpha_i^*, \quad (2.37a)$$

$$\text{sujeito a } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \quad \forall 1 \leq i \leq n, \quad (2.37b)$$

onde α e α^* são os multiplicadores de Lagrange que satisfazem restrições positivas. Dessa forma, resolvendo a equação acima de forma à obter α e α^* ótimos, podemos reescrever $f(x)$ como

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b. \quad (2.38)$$

2.3.2 Máquinas de vetores de suporte para regressão para dados censurados

Para usar essa formulação para dados censurados há duas opções: desconsiderar os dados censurados, resultando em tempo de falha subestimado, ou tratar os dados

censurados como não-eventos, resultando em um modelo enviesado. Dessa forma, há perda de informação (FOUODO et al., 2018).

O modelo proposto abaixo permite considerar todas as informações do estudo (dados censurados e não-censurados) (BELLE et al., 2011a).

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad (2.39a)$$

$$\text{sujeito a } \begin{cases} t_i - w^t x_i - b \leq \epsilon + \xi_i^*, & \xi_i^* \geq 0, \quad \forall 1 \leq i \leq n \\ \delta_i(w^t x_i + b - t_i) \leq \epsilon + \xi_i, & \xi_i \geq 0, \quad \forall 1 \leq i \leq n, \end{cases}$$

em que $\epsilon \geq 0$ é o raio do tubo insensível, que aparece tomando o valor nulo na formulação apresentada em (BELLE et al., 2011a).

A formulação dual é

$$\min_{\alpha,\alpha^*} \frac{1}{2} \sum_{i,j=1}^n (\alpha_i + \alpha_i^*)(\alpha_j + \alpha_j^*)k(x_i, x_j) - \sum_{i=1}^n (t_i - \epsilon)\alpha_i + \sum_{i=1}^n (t_i + \epsilon)\delta_i\alpha_i^* \quad (2.40a)$$

$$\text{sujeito a } \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \quad \forall 1 \leq i \leq n; \quad (2.40b)$$

Resolvendo a equação acima de forma à obter α e α^* ótimos, podemos reescrever $f(x)$ como

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*)k(x_i, x) + b. \quad (2.41)$$

Esse método penaliza a previsão incorreta se a previsão foi maior ou menor que o tempo de falha observado e, para dados censurados à direita, apenas se a previsão é menor que o tempo de falha observado. Além disso, a penalidade é a mesma para previsões erradas de dados censurados ou tempo de falha observado.

2.3.3 Máquinas de vetores de suporte para ranqueamento

Outra abordagem possível para o problema é vê-lo como um problema de ranqueamento. Dessa forma, o objetivo não é saber o tempo de sobrevivência, mas saber se existe um risco baixo ou alto de um evento acontecer.

A função de ranqueamento (SVM para ranqueamento) se diferencia da função de classificação (SVM para classificação) da seguinte forma

- No problema de classificação o conjunto de treinamento é um conjunto de dados com seus rótulos. Para ranqueamento, é um conjunto de dados ordenados. A condição

“A é preferível à B” é definida como " $A \succ B$ ". O conjunto de treinamento é $R = (x_1, y_1), \dots, (x_n, y_n)$, onde y_i é a posição no ranqueamento de x_i , isto é, $y_i < y_j$ se $x_i \succ x_j$. A classificação retorna uma classe distinta para cada dado, no problema de ranqueamento a saída é uma nota (score) para cada dado, a partir do qual será obtido a ordenação global. A função $F(x_i)$ retorna uma nota tal que $F(x_i) > F(x_j)$ para qualquer $x_i \succ x_j$.

Assumindo que R é estritamente ordenado (strict ordering), isso significa que x_i e x_j , em um conjunto D , são $x_i \succ_R x_j$ ou $x_i \prec_R x_j$. Seja R^* o ranqueamento ótimo de dados, onde os dados são ordenados perfeitamente de acordo com a escolha (preferência) do usuário. A função de ranqueamento R^F é tipicamente avaliada por quanto se aproxima de R^* . Assumindo que F é uma função de ranqueamento linear tal que

$$\forall\{(x_i, x_j) : y_i < y_j \in R\} : F(x_i) > F(x_j) \iff w \cdot x_i > w \cdot x_j. \quad (2.42)$$

R é dito linearmente ranqueável se existe uma função F que satisfaz a equação acima para todo $\{(x_i, x_j) : y_i < y_j \in R\}$. O objetivo é encontrar o vetor w tal que $w \cdot x_i > w \cdot x_j$ para a maioria dos pares $\{(x_i, x_j) : y_i < y_j \in R\}$. Assim, o problema pode ser escrito como

$$\min\{L_1(x, \xi_{ij})\} = \frac{1}{2}\|w\|^2 + C \sum \xi_{ij}, \quad (2.43a)$$

$$\text{sujeito a } \begin{cases} \forall\{(x_i, x_j) : y_i < y_j \in R\} : w \cdot x_i \geq w \cdot x_j + 1 - \xi_{ij} \\ \forall(i,j) : \xi_{ij} \geq 0 \end{cases} \quad (2.43b)$$

Através da restrição (2.43b) e minimizando $\sum \xi_{ij}$, a otimização acima satisfaz o problema de ordenamento no conjunto de treinamento R com erro mínimo. Reorganizando (2.43b) tem-se,

$$w \cdot (x_i - x_j) \geq 1 - \xi_{ij}. \quad (2.44)$$

O problema de otimização se torna equivalente ao de SVM para classificação em $(x_i - x_j)$. Assim, é possível utilizar uma implementação de SVM existente para resolver o problema. Os vetores de suporte são obtidos trocando o sinal de \geq por $=$ em (2.44). Usando os multiplicadores de Lagrange, é possível obter a formulação dual

$$\max\{L_2(\alpha)\} = \sum_{ij} \alpha_{ij} - \sum_{ij} \sum_{uv} \alpha_{ij} \alpha_{uv} k(x_i - x_j, x_u - x_v), \quad (2.45a)$$

$$\text{sujeito a } C \geq \alpha \geq 0. \quad (2.45b)$$

Dessa forma, é possível resolver problemas não-lineares. Obtendo o valor ótimo de α , o vetor w pode ser reescrito como

$$w = \sum_{ij} \alpha_{ij} (x_i - x_j). \quad (2.46)$$

2.3.4 Máquinas de vetores de suporte para ranqueamento com dados censurados

Para um conjunto de dados R com dados censurados, um par de dados $\{(x_i, t_i, \delta_i), (x_j, t_j, \delta_j)\}$ é dito comparável quando a ordem de acontecimento dos eventos é conhecida ou então é um evento e um dado censurado, onde o termo da censura é maior que o tempo de falha. Isso pode ser definido como

$$comp(i, j) = \begin{cases} 1 & \text{se } \delta_i = 1 \text{ e } \delta_j = 1, \\ 1 & \text{se } \delta_i = 1 \text{ e } \delta_j = 0 \text{ e } t_i \leq t_j, \\ 0 & \text{caso contrário.} \end{cases} \quad (2.47)$$

O modelo é dado por

$$\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{j: t_i > t_j} \xi_{ij}, \quad (2.48a)$$

$$comp(i, j) = 1, \quad (2.48b)$$

$$\text{sujeito a } \begin{cases} w^t(x_i - x_j) \geq 1 - \xi_{ij} & \forall i = 1, \dots, n; \forall j : t_i > t_j \text{ e } comp(i, j) = 1, \\ \xi_{ij} \geq 0 & \forall i = 1, \dots, n; \forall j : t_i > t_j \text{ e } comp(i, j) = 1. \end{cases} \quad (2.48c)$$

Um problema desta formulação é que ele requer uma quantidade muito grande de restrições, na ordem de n^2 . Desta forma, em (BELLE et al., 2011b) é sugerido um modelo alternativo para solução do problema com dados censurados que é uma adaptação do método SVM para ranqueamento. Este modelo possui somente n restrições e será chamado de modelo 1. A formulação é dada por:

$$\min_{w, \xi} \frac{1}{2} \|w\|^2 + C \sum_{j: t_i > t_j} \xi_{ij}, \quad (2.49a)$$

$$comp(i, j) = 1, \quad (2.49b)$$

$$\text{sujeito a } \begin{cases} w^t(x_i - x_j) \geq t_i - t_{j(i)} - \xi_{ij} & \forall i = 1, \dots, n. \\ \xi_{ij} \geq 0 & \forall i = 1, \dots, n. \end{cases} \quad (2.49c)$$

A formulação dual é

$$\min_{\alpha_i} \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \alpha_i \alpha_k (\phi(x_i) - \phi(x_{j(i)}))^t (\phi(x_k) - \phi(x_{j(k)})) - \sum_{i=0} \alpha_i (t_i - t_{j(i)}), \quad (2.50a)$$

$$\text{sujeito a } 0 \geq \alpha_i \geq C, \forall i = 1, \dots, n, \quad (2.50b)$$

em que $\phi(x)$ é o mapa de características associado ao kernel k , isto é, $k(x_i, x_k) = \phi(x_i)^t \phi(x_k)$.

2.3.5 Método combinado

Também é sugerido um modelo onde combina-se o SVM para ranqueamento e regressão, chamado de modelo 2, sendo sua formulação

$$\min_{w, \xi, \xi^*, b} \frac{1}{2} \|w\|^2 + C \sum_{j: t_i > t_j} \xi_{ij} + \mu \sum_{j: t_i > t_j} \xi_{ij} + \xi_{ij}^*, \quad (2.51a)$$

$$\text{sujeito a } \begin{cases} w^t(x_i - x_{j(i)}) \geq t_i - t_{j(i)} - \xi_i & \forall i = 1, \dots, n, \\ w^t x_i + b \geq t_i - \xi_i & \forall i = 1, \dots, n, \\ \delta_i (w^t x_i + b) \geq -\delta_i t_i - \xi_i^*, \\ \xi_i \geq 0 & \forall i = 1, \dots, n, \\ \xi_i^* \geq 0 & \forall i = 1, \dots, n. \end{cases} \quad (2.51b)$$

A primeira restrição é a mesma que do modelo 1, é restrição de ranqueamento otimizando o índice de concordância. Já a segunda e terceira restrições são iguais às do modelo de regressão. Esse modelo equivale ao modelo 1 quando $\mu = 0$ e o modelo de SVM para regressão quando $C = 0$.

A formulação dual é

$$\begin{aligned}
& \min_{\alpha, \beta, \beta^*} \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \alpha_i \alpha_k (\phi(x_i) - \phi(x_{j(i)}))^t (\phi(x_k) - \phi(x_{j(k)})) \\
& + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n (\beta_i \beta_k + \alpha_i \alpha_k \beta_i^* \beta_k^*) \phi(x_i)^t \phi(x_k) - \sum_{i=1}^n \alpha_i (t_i - t_{i-1}) \\
& + \sum_{i=1}^n \sum_{k=1}^n (\beta_i - \alpha_k \beta_k^*) \alpha_k \phi(x_i)^t (\phi(x_k) - \phi(x_{j(k)})) \\
& - \sum_{i=1}^n \sum_{k=1}^n \beta_i^* \delta_i \beta_i \phi(x_i^t) \phi(x_k) - \sum_{i=1}^n \beta_i y_i + \sum_{i=1}^n \delta_i \beta_i^* t_i,
\end{aligned}$$

$$\text{sujeito a } \begin{cases} 0 \geq \alpha_i \geq \alpha, & \forall i = 1, \dots, n, \\ 0 \geq \beta_i \geq \mu, & \forall i = 1, \dots, n, \\ 0 \geq \beta_i^* \geq \mu, & \forall i = 1, \dots, n. \end{cases} \quad (2.52a)$$

3 Metodologia

Os experimentos realizados visam comparar os resultados obtidos para diferentes métodos de solução do problema de análise de sobrevivência. Para isso, são utilizadas três bases de dados reais disponíveis na biblioteca *survival* do pacote estatístico R. Também são geradas aleatoriamente duas bases de dados artificiais, com dados não lineares. O índice de concordância c-index é utilizado para verificar a assertividade do método e para comparar diferentes métodos.

3.1 Métodos Utilizados

Para realização dos experimentos faz-se o uso do pacote estatístico R (versão 3.6.3) (FOUODO et al., 2018), das bibliotecas virtuais *survival* (versão 3.2-7), *tidymodels* (versão 1.0.0), *Hmisc* (versão 4.7-0), *glue* (versão 1.6.2) e *glmnet* (versão 4.1-1) e do *solver Gurobi* (versão 9.5-2).

Foram modelados os métodos de Cox (seção 2.2.2), SVM para regressão (seção 2.3.2) e modelos 1 e 2 com kernel aditivo e radial (seção 2.3.5).

O método de Cox é um método clássico para solução do problema de análise de sobrevivência, aplicado tipicamente em dados lineares. SVM para regressão (SVR) com dados censurados se baseia no método clássico de SVR, porém com algumas considerações adicionais para melhor avaliação da base de dados, conseguindo assim modelar dados censurados. Importante notar que este método pode ser aplicado, sem perda de performance, para dados lineares e não-lineares. O modelo 1 é uma adaptação do método de SVM para ranqueamento, considerando dados censurados. O modelo 2 é uma combinação do modelo 1 e SVR para dados censurados. Dessa forma, espera-se resultados melhores para os demais, excluindo-se o método de Cox.

3.2 Bancos de Dados

3.2.1 Dados Reais

São utilizados os conjuntos de dados *lung*, *diabetic* e *veteran* da biblioteca *survival* do pacote estatístico R.

Lung possui dados sobre pacientes com câncer de pulmão avançado do *North Central Cancer Treatment Group*. Possui os dados do código da instituição, tempo de sobrevivência em dias, status de censura (1 para censura, 2 para morte), idade em anos,

sexo (1 para masculino, 2 para feminino), taxa de performance ECOG (0 para bom, 5 para morte), taxa de performance de Karnofsky (0 para ruim, 100 para bom) dado por um médico e pelo paciente, calorias consumidas, perda de peso nos últimos seis meses. Existem 228 dados, dos quais 63 são censurados. A tabela 1 abaixo mostra um resumo dos dados (THERNEAU THOMAS LUMLEY, 2023).

Tabela 1 – Lung

Statistic	N	Mean	St. Dev.	Min	Max
inst	227	11.088	8.303	1	33
time	228	305.232	210.646	5	1,022
status	228	1.724	0.448	1	2
age	228	62.447	9.073	39	82
sex	228	1.395	0.490	1	2
ph.ecog	227	0.952	0.718	0	3
ph.karno	227	81.938	12.328	50	100
pat.karno	225	79.956	14.623	30	100
meal.cal	181	928.779	402.175	96	2,600
wt.loss	214	9.832	13.140	-24	68

Diabetic possui resultados parciais de um teste de coagulação com laser para o tratamento de retinopatia diabética. Pacientes são escolhidos ao acaso de uma amostra de pacientes com alto risco de retinopatia diabética como definido por *Diabetic Retinopathy Study* (DRS). Em cada paciente, foi escolhido ao acaso qual olho receberia o tratamento. O evento de interesse é o tempo entre o início do tratamento e quando a acuidade visual caiu abaixo de 5/200 após duas visitas seguidas. As visitas ocorriam a cada 3 meses. O tempo de sobrevivência é o tempo que o paciente leva para ficar cego, em meses, menos o tempo mínimo possível para o evento (6, 5 meses).

Possui os dados de tipo de laser (xenon ou argon), idade em que recebeu o diagnóstico, olho (esquerdo ou direito), tratamento (0 sem tratamento, 1 tratamento com laser), grupo de risco (de 6 a 12), tempo para o evento ou última consulta, status (0 para censurado ou 1 para perda visual). Existem 394 dados, dos quais 239 são censurados. A tabela 2 abaixo mostra um resumo dos dados.

Veteran possui dados sobre um estudo de tratamentos para câncer de pulmão, onde há dois possíveis tratamentos e o paciente recebe ao acaso um deles. Possui os dados de tratamento (1 para padrão e 2 para teste), tipo de célula, tempo de sobrevivência, status de censura, taxa de performance de Karnofsky, meses do diagnóstico até início do tratamento, idade, tratamento anterior (0 se não houve, 10 se houve). Existem 137 dados, dos quais 9 são censurados. A tabela 3 abaixo mostra um resumo dos dados.

Tabela 2 – Diabetic

Statistic	N	Mean	St. Dev.	Min	Max
id	394	873.203	495.523	5	1,749
age	394	20.782	14.812	1	58
trt	394	0.500	0.501	0	1
risk	394	9.698	1.475	6	12
time	394	35.579	21.356	0.300	74.970
status	394	0.393	0.489	0	1

Tabela 3 – Veteran

Statistic	N	Mean	St. Dev.	Min	Max
trt	137	1.496	0.502	1	2
time	137	121.628	157.817	1	999
status	137	0.934	0.249	0	1
karno	137	58.569	20.040	10	99
diagtime	137	8.774	10.612	1	87
age	137	58.307	10.542	34	81
prior	137	2.920	4.563	0	10

3.2.2 Dados Artificiais

São gerados dados artificiais para testar o resultado dos modelos com dados não-lineares. A base de dados contém os dados x_1 , x_2 , status de censura e tempo de sobrevivência. x_1 e x_2 são definidos como dois vetores com 500 dados aleatórios de uma distribuição uniforme no intervalo de 0 a 2. O status de censura também é gerado a partir de 500 amostras aleatórias de uma distribuição uniforme no intervalo de 0 a 1. Usa-se os seguintes critérios para transformar os dados gerados em 0 ou 1, onde x é um dado da amostra

$$\text{status}(x) = \begin{cases} 0 & \text{se } x < 0,5 \\ 1 & \text{se } x \geq 0,5. \end{cases} \quad (3.1)$$

Para a primeira base de dados, nao-linear-1, a função de falha é definida como

$$\lambda(t, x) = \lambda \cdot \exp(x_1^2 - 2x_1x_2 + x_2^2), \quad (3.2)$$

λ independe do tempo e é igual a 1. Para a segunda base de dados, nao-linear-2, a função de falha é definida como

$$\lambda(t, x) = \lambda(x_1^2 - 2x_1x_2 + x_2^2), \quad (3.3)$$

λ independe do tempo e é igual a 1. O tempo de sobrevivência real é obtido através da distribuição exponencial com taxa $\lambda(t, x)$. Para obter o tempo de censura, o valor real obtido anterior é multiplicado por um fator aleatório dado por uma distribuição uniforme entre 0,3 e 0,9.

3.3 Métrica de Comparação

Para comparar os resultados obtidos nos experimentos será utilizado o índice de concordância. O índice de concordância é a razão entre o número de pares concordantes e o número de pares comparáveis. É uma métrica bastante comum de ser utilizada em análise de sobrevivência que quantifica a correlação entre o tempo de sobrevivência e a predição dos modelos (ALABDALLAH et al., 2022). Pode ser escrito como

$$c\text{-index}(u) = \frac{\sum_{i=1}^n \sum_{i \neq j} \mathcal{I}[(u(x_i) - u(x_j))(y_i - y_j) \geq 0]}{\sum_{i=1}^n \sum_{i \neq j} \text{comp}(i, j)} \quad (3.4)$$

onde \mathcal{I} é a função indicadora e $\text{comp}(i, j)$ é definido como em 2.3.4. Valores mais próximos de 1 indicam que o modelo é um bom preditor, enquanto valores próximos de 0 sugerem que o modelo apresenta resultados insatisfatórios.

4 Resultados e Discussão

Os resultados são apresentados nos gráficos abaixo para todas as bases de dados testadas. O índice de concordância encontra-se no eixo y e o método testado encontra-se no eixo x . Ao todo foram testados 7 métodos: método de Cox, SVM para regressão com kernel radial e aditivo, modelo 1 com kernel radial e aditivo e modelo 2 com kernel radial e aditivo. Estes são apresentados nessa ordem nos gráficos. No gráfico à esquerda é possível observar o box-plot com os resultados para o conjunto de dados de teste; à direita é possível observar o box-plot com os resultados para o conjunto de dados de treinamento. Deseja-se índice de concordância próximo de 1, pois quanto mais próximo de 1 o resultado estiver, melhor é o resultado do modelo.

Na figura 3 é possível observar que os melhores resultados são obtidos com método de Cox e SVR com kernel radial e aditivo, com índice de concordância próximo de 0,70; já os piores resultados são com o modelo 1 com kernel radial e aditivo, com índice de concordância próximo de 0,3. O modelo 2 com kernel radial apresenta um resultado mediano, com índice de concordância próximo de 0,5.

Na figura 4 é possível observar que o melhor resultado é obtido com o SVR com kernel aditivo; porém o método de Cox e modelo 2 aditivo possuem resultados similares ao primeiro, com índices de concordância próxima de 0,65.

Na figura 5 é possível observar que os melhores resultados são obtidos com método de Cox, SVR com kernel radial e aditivo e modelo 2 com kernel radial e aditivo, com resultados similares e índice de concordância próximo de 0,7; já pior resultado é alcançado com modelo 1 com kernel radial e aditivo, com índice de concordância 0,6 ou menor.

Para os dados reais testados, é possível notar que, de forma geral, o método clássico de Cox para solução de problemas de análise de sobrevivência possui melhor resultado. O método de SVR, tanto com kernel aditivo, quanto com kernel radial é equiparável ao método de Cox. Já o modelo 1, com kernel aditivo ou radial, de forma geral possui resultado insatisfatório, abaixo dos demais. Por outro lado, o modelo 2 de fato retorna resultados satisfatórios, podendo ser superiores ao método clássico, porém com maior dispersão de resultados. No entanto, é importante observar que nenhum método possui índice de concordância maior que 0,8; os melhores resultados encontram-se entre 0,6 e 0,7.

Na figura 6 é possível observar que o pior resultado alcançado acontece com o método de Cox; já era esperado uma vez que esse método não consegue modelar a não-linearidade dos dados. Os melhores resultados são alcançados com SVR com kernel radial e modelo 2 com kernel radial, com índices de concordância entre 0,6 e 0,7.

Figura 3 – Lung

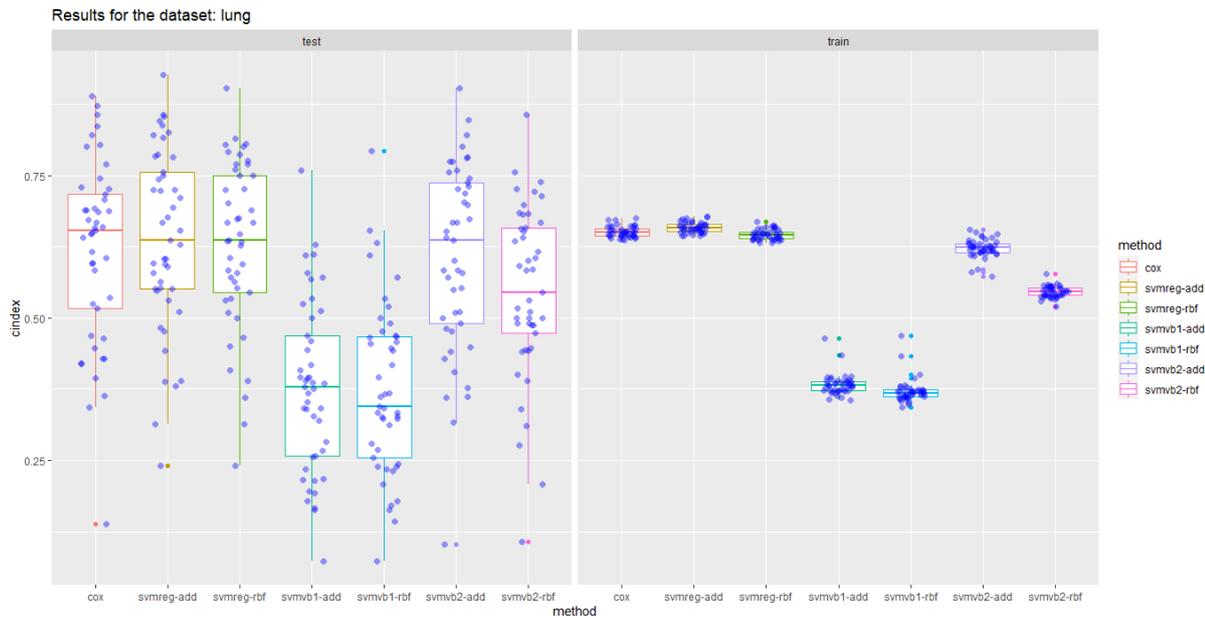
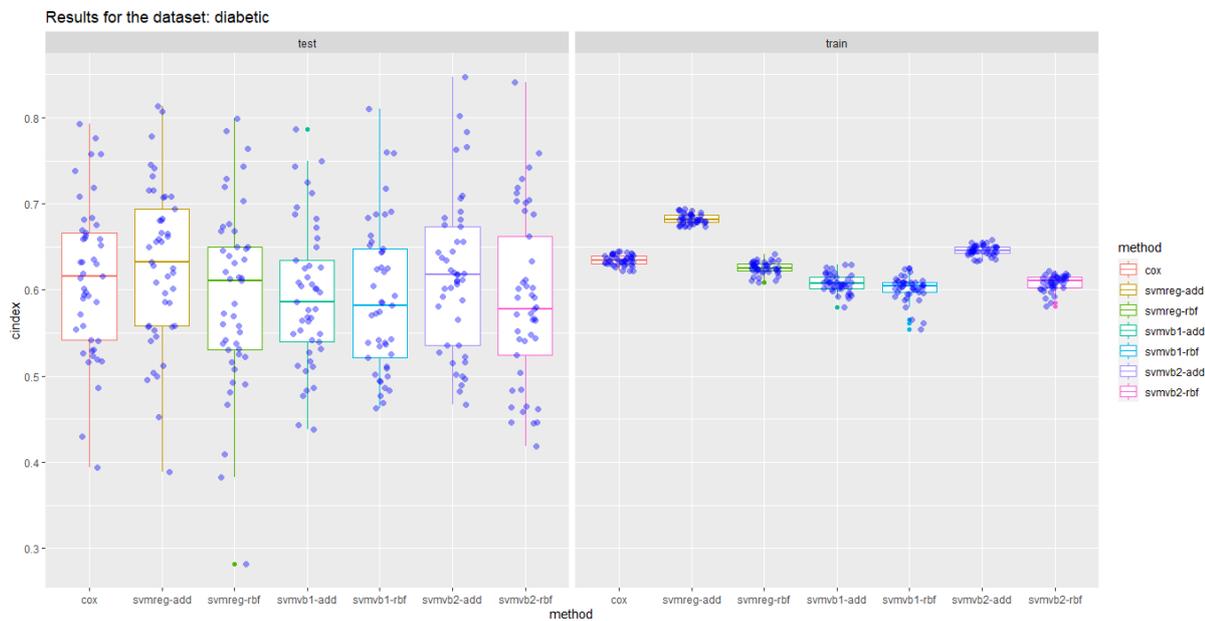


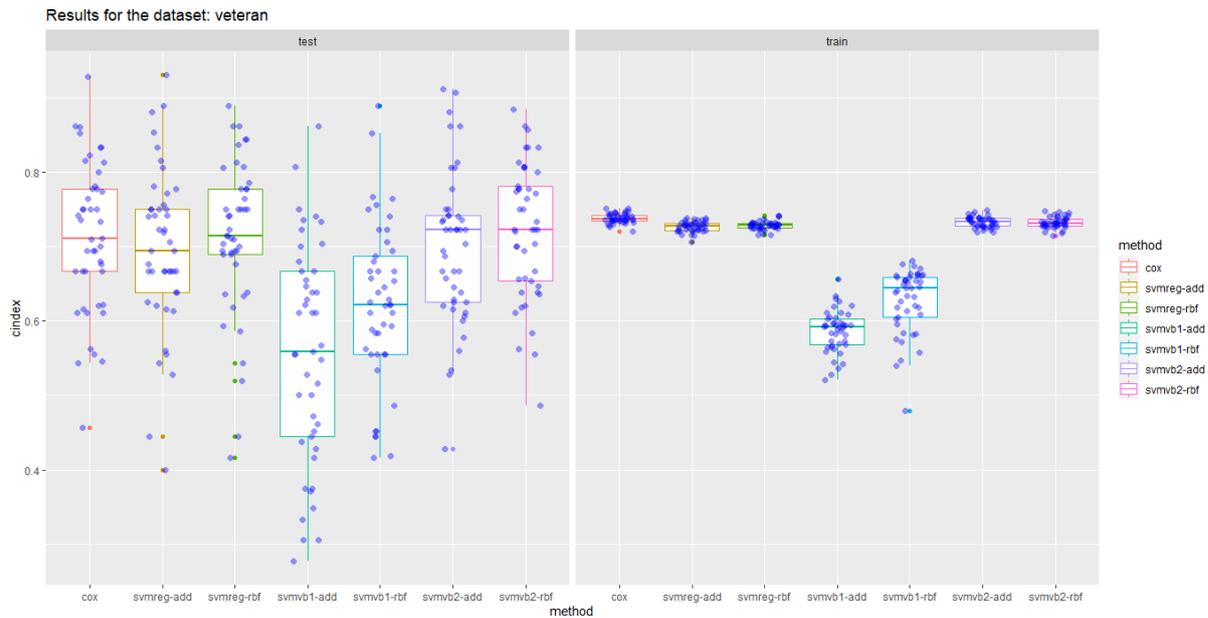
Figura 4 – Diabetic



Na figura 7 é possível observar que os melhores resultados ocorrem com os métodos SVR com kernel radial e modelo 2 com kernel radial, com índices de concordância próximos à 0,8. O pior resultado ocorre com modelo 1 com kernel aditivo, com índice de concordância em torno de 0,5; já os demais métodos possuem grande dispersão de resultados, com índices de concordância variando entre 0,55 e 0,7.

Para os dados artificiais, é possível notar, de forma geral, que o método de Cox não performa bem, enquanto os resultados para o modelo 2 e SVR, ambos com kernel

Figura 5 – Veteran



radial, são bastante satisfatórios. O modelo 1, com kernels radial retorna resultados pouco superiores ao método de Cox, enquanto com kernel aditivo os resultados são similares.

Figura 6 – Dados não lineares (1)

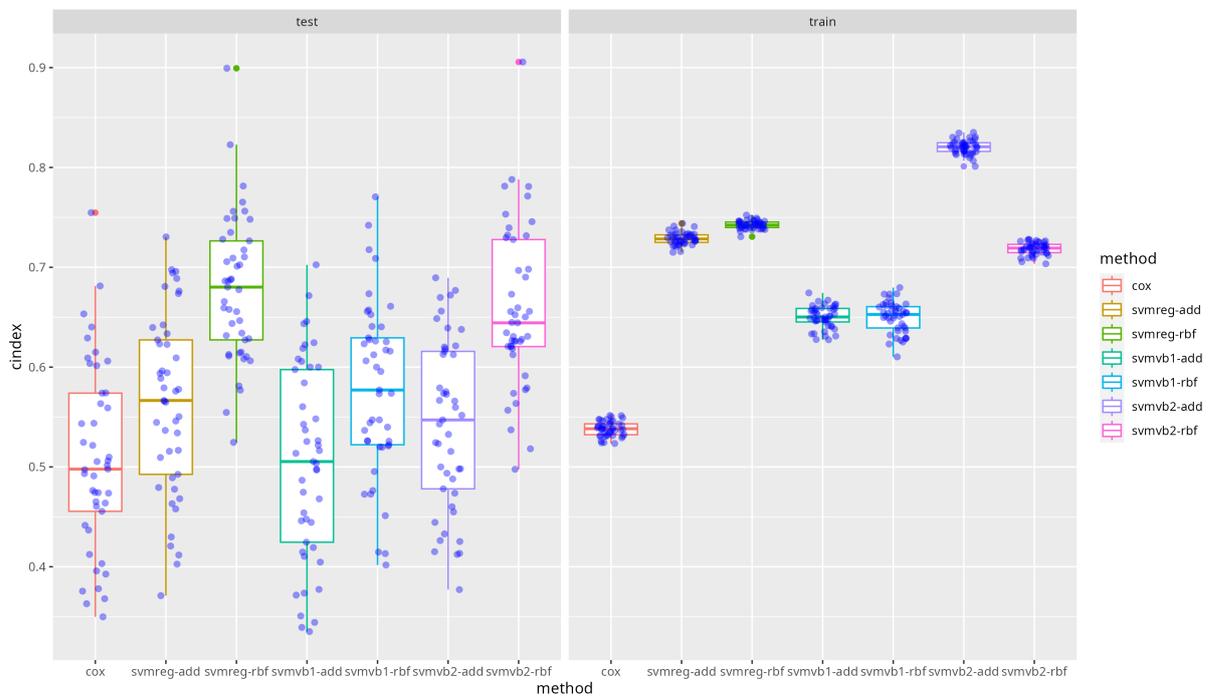
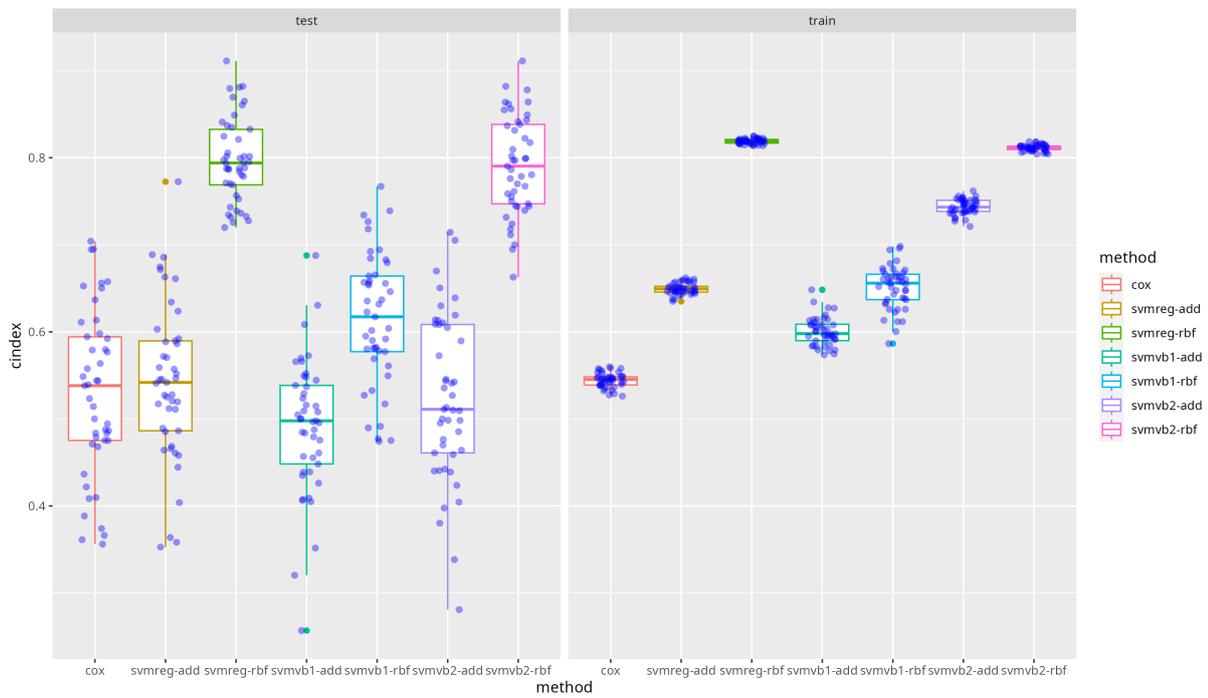


Figura 7 – Dados não lineares (2)



Conclusões e Trabalhos Futuros

Conclusões

Este trabalho mostrou a utilização de máquinas de vetores de suporte na solução do problema de análise de sobrevivência para dados lineares e não-lineares. O método proposto em (BELLE et al., 2011a) mostrou-se eficiente para ambos os casos, em especial o modelo 2, ressaltando-se dos demais. Enquanto o método de Cox apresenta resultados satisfatórios para dados lineares, o mesmo não ocorre com dados não-lineares. Os resultados para o modelo de SVR tanto com kernel aditivo, quanto com kernel radial são inconclusivos.

Trabalhos Futuros

Encontrar outras formas de solução de análise de sobrevivência, em especial para solução de problemas não-lineares. Uma das propostas é abordar o problema usando gradiente estocástico, motivado pelo método do perceptron, que pode simplificar a implementação. Além disso, redes neurais e métodos baseados em árvores também têm sido estudados como abordagens possíveis.

Referências

- ALABDALLAH, A. et al. The concordance index decomposition—a measure for a deeper understanding of survival prediction models. *arXiv preprint arXiv:2203.00144*, 2022. Citado na página 20.
- BELLE, V. V. et al. Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artificial intelligence in medicine*, Elsevier, v. 53, n. 2, p. 107–118, 2011. Citado 2 vezes nas páginas 12 e 25.
- BELLE, V. V. et al. Learning transformation models for ranking and survival analysis. *Journal of machine learning research*, v. 12, n. 3, 2011. Citado na página 14.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Editora Blucher, 2006. Citado 3 vezes nas páginas 4, 7 e 8.
- FOUODO, C. J. et al. Support vector machines for survival analysis with r. *R Journal*, v. 10, n. 1, 2018. Citado 2 vezes nas páginas 12 e 17.
- GOEL, M. K.; KHANNA, P.; KISHORE, J. Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, Wolters Kluwer–Medknow Publications, v. 1, n. 4, p. 274, 2010. Citado na página 7.
- IZBICKI, R.; SANTOS, T. M. dos. *Aprendizado de máquina: uma abordagem estatística*. [S.l.]: Rafael Izbicki, 2020. Citado na página 1.
- JAMES, G. et al. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112. Citado 5 vezes nas páginas 7, 1, 4, 6 e 8.
- LIU, X. *Survival analysis: models and applications*. [S.l.]: John Wiley & Sons, 2012. Citado na página 1.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003. Citado na página 1.
- SHIVASWAMY, P. K.; CHU, W.; JANSCHKE, M. A support vector approach to censored targets. In: IEEE. *Seventh IEEE international conference on data mining (ICDM 2007)*. [S.l.], 2007. p. 655–660. Citado na página 11.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and computing*, Springer, v. 14, p. 199–222, 2004. Citado na página 11.
- THERNEAU THOMAS LUMLEY, A. E. C. C. T. M. *Package ‘survival’*. [S.l.], 2023. R package version 3.5-3. Disponível em: <<https://cran.r-project.org/web/packages/survival/survival.pdf>>. Citado na página 18.
- WANG, P.; LI, Y.; REDDY, C. K. Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 51, n. 6, p. 1–36, 2019. Citado na página 3.

YU, H.; KIM, S. Svm tutorial-classification, regression and ranking. *Handbook of Natural computing*, v. 1, p. 479–506, 2012. Citado na página [10](#).