



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO, DE CIÊNCIAS EXATAS E EDUCAÇÃO
DEPARTAMENTO DE ENG. DE CONTROLE, AUTOMAÇÃO E COMPUTAÇÃO
CURSO DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Cláudio Lourenço Moreira

Título do trabalho: subtítulo (se houver)

Cláudio Lourenço Moreira

Título do trabalho: subtítulo (se houver)

Trabalho de Conclusão de Curso de Graduação em Engenharia de Controle e Automação do Centro Tecnológico, de Ciências Exatas e Educação da Universidade Federal de Santa Catarina como requisito para a obtenção do título de Engenheiro de Controle e Automação.
Orientador: Prof. Dr. Maiquel de Brito

Ficha de identificação da obra

A ficha de identificação é elaborada pelo próprio autor.

Orientações em:

<http://portalbu.ufsc.br/ficha>

Cláudio Lourenço Moreira

Título do trabalho: subtítulo (se houver)

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de “Engenheiro de Controle e Automação” e aprovado em sua forma final pelo Curso de Graduação em Engenharia de Controle e Automação.

Blumenau, dia de mês de 2024.

Banca Examinadora:

Prof. Dr. Maiquel de Brito
Instituição xxxx

Prof. Segundo, Dr.
Instituição xxxx

Prof. Terceiro, Dr.
Instituição xxxx

Este trabalho é dedicado a todos que me ajudaram,
orientaram e contribuíram para a minha graduação, em
especial meus pais, amigos e professores.

AGRADECIMENTOS

Inserir os agradecimentos aos colaboradores à execução do trabalho.

Texto da Epígrafe. Citação relativa ao tema do trabalho. É opcional. A epígrafe pode também aparecer na abertura de cada seção ou capítulo. Deve ser elaborada de acordo com a NBR 10520. (SOBRENOME do autor da epígrafe, ano)

RESUMO

No resumo são ressaltados o objetivo da pesquisa, o método utilizado, as discussões e os resultados com destaque apenas para os pontos principais. O resumo deve ser significativo, composto de uma sequência de frases concisas, afirmativas, e não de uma enumeração de tópicos. Não deve conter citações. Deve usar o verbo na voz ativa e na terceira pessoa do singular. O texto do resumo deve ser digitado, em um único bloco, sem espaço de parágrafo. O espaçamento entre linhas é simples e o tamanho da fonte é 12. Abaixo do resumo, informar as palavras-chave (palavras ou expressões significativas retiradas do texto) ou, termos retirados de thesaurus da área. Deve conter de 150 a 500 palavras. O resumo é elaborado de acordo com a NBR 6028.

Palavras-chave: palavra-chave 1; palavra-chave 2; palavra-chave 3.

ABSTRACT

Resumo traduzido para outros idiomas, neste caso, inglês. Segue o formato do resumo feito na língua vernácula. As palavras-chave traduzidas, versão em língua estrangeira, são colocadas abaixo do texto precedidas pela expressão “Keywords”, separadas por ponto e vírgula.

Keywords: keyword 1; keyword 2; keyword 3.

LISTA DE FIGURAS

Figura 1 – Relação entre o índice de felicidade e expectativa de vida.	18
Figura 2 – Diferentes correlações entre variáveis.	19
Figura 3 – Exemplo do método dos quadrados ordinários.	21

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LISTA DE SÍMBOLOS

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	AS MUDANÇAS CLIMÁTICAS E AS CATÁSTROFES NATURAIS .	16
2.1.1	Cenário de enchentes no sul do Brasil	16
2.1.2	Dinâmica do Rio Guaíba	16
2.2	APRENDIZADO DE MÁQUINA	17
2.2.1	Categorias de aprendizado de máquina	17
2.3	REGRESSÃO	18
2.4	REGRESSÃO LINEAR	19
2.5	MODELO RIDGE	21
3	SEÇÃO	23
4	CONCLUSÃO	24
	REFERÊNCIAS	25
	APÊNDICE A – Descrição	27
	ANEXO A – Descrição	28

1 INTRODUÇÃO

1.1 OBJETIVOS

Nas seções abaixo estão descritos o objetivo geral e os objetivos específicos deste TCC.

1.1.1 Objetivo Geral

Descrição...

1.1.2 Objetivos Específicos

Descrição...

2 FUNDAMENTAÇÃO TEÓRICA

Explicar brevemente o que será tratado como fundamentação teórica para o entendimento do contexto em que o modelo de aprendizagem de máquina será aplicado.

2.1 AS MUDANÇAS CLIMÁTICAS E AS CATÁSTROFES NATURAIS

As grandes cidades brasileiras enfrentam desafios mais frequentes relacionados às mudanças climáticas, que agravam problemas como enchentes, inundações e deslizamentos. Projeções indicam que, até 2030, a mancha urbana de São Paulo pode aumentar em até 38%, ampliando o risco para mais de 20% das áreas de expansão urbana, que se tornarão suscetíveis a acidentes naturais (Nobre et al., 2011). O estudo também destaca que o aumento na frequência de eventos de chuvas intensas pode dobrar o número de dias com precipitação acima de 10 milímetros, agravando a vulnerabilidade da população, especialmente nas áreas periféricas e de menor infraestrutura.

2.1.1 Cenário de enchentes no sul do Brasil

Com base no histórico das enchentes no Rio Grande do Sul, observa-se que os desastres relacionados ao excesso de chuvas não são um fenômeno recente. Desde 1941, o estado lida com eventos catastróficos, como a enchente que devastou Porto Alegre naquele ano, considerada uma das mais graves da história da cidade. Ao longo das décadas, esses episódios continuaram a ocorrer, expondo a vulnerabilidade da região diante de chuvas intensas e repentinas. A combinação de fatores naturais, como a geografia da região e os ciclos climáticos, aliado às ações humanas nocivas ao meio ambiente, contribui para a repetição e intensificação dessas tragédias (VEJA, 2024).

Em Santa Catarina, estado adjacente ao Rio Grande do Sul, as enchentes também são fenômenos recorrentes que, ao longo dos anos, têm causado impactos sociais, econômicos e ambientais. Um dos eventos mais recentes foi registrado em maio de 2024, quando o estado registrou vários dias com altos índices pluviométricos, levando ao transbordamento de rios, deslizamentos de terra e bloqueios em diversas rodovias (G1, 2024).

2.1.2 Dinâmica do Rio Guaíba

O Rio Guaíba, principal manancial de abastecimento de água para a capital do Rio Grande do Sul e região, é alvo de estudo sobre diversos temas, incluindo sua hidrodinâmica e nível ao longo do ano. No Artigo conduzido pelos pesquisadores (ANDRADE *et al.*, 2017), a variabilidade nas descargas líquidas do Rio Guaíba revelou flutuações significativas nos volumes de descarga, variando de 407 m³/s a 14.270 m³/s, o que indica uma grande influência das condições climáticas sazonais e da vazão dos rios tributários, como o Jacuí, Taquarí, Caí e Sinos. Essas variações extremas foram observadas durante o período de

2014 a 2017 e reforçam a importância de monitorar continuamente o regime de águas do Guaíba para prevenir enchentes e outros desastres associados (ANDRADE *et al.*, 2017).

2.2 APRENDIZADO DE MÁQUINA

Desde que os computadores foram inventados, criou-se o questionamento da possibilidade de fazê-los pensar de modo semelhante ao ser humano. Por meio desse avanço, diversas áreas sofreriam grandes transformações, uma vez que a capacidade da máquina aprender e aprimorar o seu conhecimento sobre determinado assunto traria melhorias e uma maior performance na atividade desejada (CARBONELL; MICHALSKI; MITCHELL, 1983).

Embora os computadores ainda não alcancem o mesmo nível de aprendizado geral do ser humano, nos últimos anos, o aprendizado de máquina (do inglês, *machine learning*, ou ML) se tornou realidade, com aplicações em diversos setores relacionados ou não a tecnologia, agregando valor e conhecimento por meio de dados e informações antes tratados apenas por profissionais da área.

Esse conceito envolve a criação de sistemas que são capazes de aprender a partir de dados, identificando padrões e realizando previsões sem a necessidade de programação explícita. De acordo com (CARBONELL; MICHALSKI; MITCHELL, 1983), o principal objetivo do ML é construir algoritmos que permitam que os computadores adquiram conhecimento e melhorem sua performance de forma autônoma, baseando-se em experiências passadas.

2.2.1 Categorias de aprendizado de máquina

Com pesquisas e algoritmos sendo desenvolvidos para novas aplicações e/ou aprimoramento de implementações existentes, tornou-se necessário criar categorias de ML, a fim de classificar a sua função e estipular em quais cenários o seu uso é adequado.

Os quatro principais tipos de ML são: supervisionado, não supervisionado, semi-supervisionado e reforço (SARAVANAN; SUJATHA, 2018).

- Supervisionado: é o mais comum e envolve a utilização de dados rotulados, no qual o modelo é treinado com entradas e saídas conhecidas para fazer previsões sobre novos dados;
- Não supervisionado: lida com dados não rotulados, onde o sistema busca encontrar padrões ou agrupamentos nos dados;
- Semi supervisionado: combina elementos de ambos os métodos, utilizando uma pequena quantidade de dados rotulados e uma grande quantidade de dados não rotulados, sendo útil em cenários onde a rotulação de dados é cara ou complexa;

- Aprendizado por reforço: se baseia em um sistema de recompensas e punições, onde um agente interage com o ambiente e aprende a otimizar suas ações para alcançar um objetivo a partir de feedbacks recebidos.

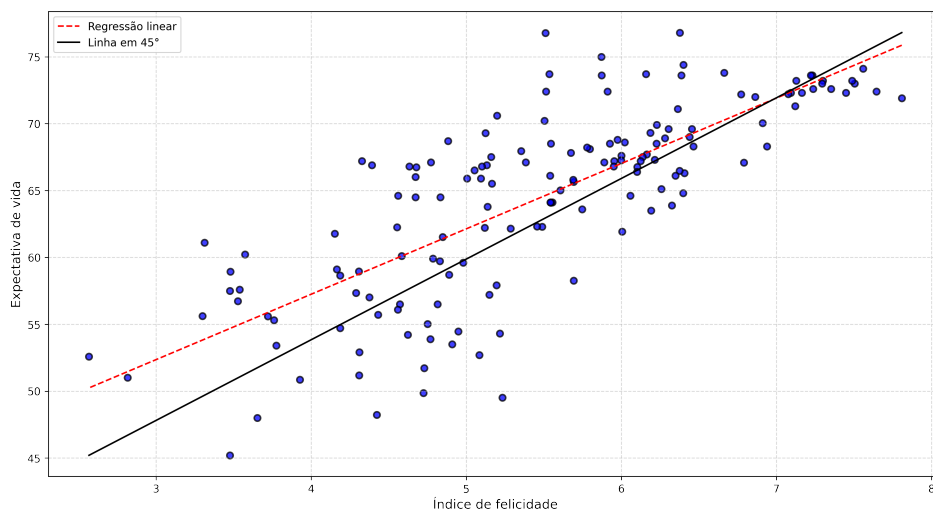
2.3 REGRESSÃO

A partir da necessidade de realizar previsões, visando compreender e estimar a dinâmica dos fenômenos estudados, a regressão se apresenta como uma ferramenta que busca modelar relações entre variáveis dependentes e independentes através de métodos estatísticos (SOTO, 2013).

Em uma equação linear, uma variável independente, comumente representada pela letra x , caracteriza uma grandeza que está sendo manipulada durante um experimento. Dado esse comportamento, a variável x não sofre influência de outras variáveis. A variável independente, normalmente representada pela letra y , caracteriza valores que estão diretamente associados à variável independente. Assim, de forma direta ou indireta, x exerce influência sobre y .

Na figura 1, a fim de exemplificar um caso de regressão, é apresentada a relação entre a expectativa de vida baseada e um índice de felicidade calculado em diversos países obtidos a partir de um levantamento feito por (HELLIWELL *et al.*, 2020). Neste estudo, a variável independente é representada pelo índice de felicidade, enquanto a expectativa de vida representa a variável dependente. Desse modo, uma análise visual do gráfico permite inferir uma tendência de expectativa de vida maior em países com alto índice de felicidade.

Figura 1 – Relação entre o índice de felicidade e expectativa de vida.



Fonte: (HELLIWELL *et al.*, 2020)

Embora uma inferência inicial permita constatar uma correlação entre as variáveis

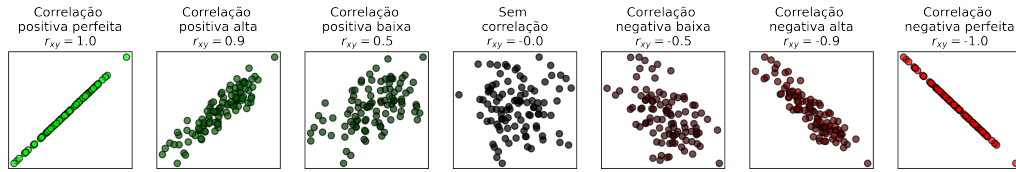
da equação, a criação de um modelo de previsão necessita de métodos que comprovem a correlação pressuposta. Para determinar as relações entre as variáveis dependentes e independentes de um sistema, coeficientes de correlação são calculados, gerando valores que medem e comprovam estatisticamente o grau de correspondência dos fatores estudados. Uma das métricas de correlação mais utilizadas é o coeficiente de Pearson, que mede a associação linear entre duas variáveis (KIRCH, 2008).

Esse coeficiente de correlação pode ser definido pela Equação (1), onde n é o total de amostras, \bar{x} e \bar{y} são as médias aritméticas de ambas as variáveis. Os valores do coeficiente de Pearson variam entre -1 e 1, de tal forma que quanto mais próximos desses extremos, melhor correlacionado estão as variáveis.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

A Figura 2 mostra alguns exemplos com gráficos de dispersão de variáveis com diferentes correlações.

Figura 2 – Diferentes correlações entre variáveis.



Fonte: (HELLIWELL *et al.*, 2020)

Com a aplicação do método de correlação, torna-se possível estimar valores não existentes na amostra ou conjunto de dados, dando forma a uma abordagem preliminar de previsão de dados. Todavia, em casos onde a correlação entre as variáveis é baixa, a previsão de valores passa a ser menos precisa e acurada, sendo necessário outras abordagens mais robustas para realizar as previsões. Em ML, os modelos de regressão podem ser criados a partir de diversos métodos, com ajustes de parâmetros simples ou complexos que auxiliam no ajuste do modelo e seu desempenho na situação aplicada.

2.4 REGRESSÃO LINEAR

A regressão linear é uma técnica de análise de dados que permite estimar valores desconhecidos com base em variáveis conhecidas relacionadas. Esse modelo representa a relação entre uma variável dependente, que se deseja prever, e uma ou mais variáveis independentes através de uma equação linear (SERVICES, 2024).

A aplicação da técnica é relevante devido à sua simplicidade e capacidade de fornecer previsões baseadas em uma fórmula matemática interpretável. Além disso, o

método é base para implementações de algoritmos na área de ciência de dados, como aprendizado de máquina, otimizando o processamento de dados complexos e viabilizando a criação de modelos de previsão.(SERVICES, 2024).

O método de regressão linear é dividido em dois grupos, sendo eles: regressão linear simples (RLS) e regressão linear múltipla (RLM). A RLS tem como objetivo estabelecer uma relação entre duas variáveis através de uma função, cuja definição é dada por:

$$y_i = \alpha + \beta x_i \quad (2)$$

Onde y_i é a variável alvo, enquanto α e β são coeficientes calculados pela regressão, que representam o intercepto no eixo Y e a inclinação da reta, respectivamente.

A RLM, embora seja semelhante à RLS, possui múltiplas variáveis preditoras, sendo definida por:

$$y_i = \alpha + \beta x_{i1} + \beta x_{i2} + \dots + \beta x_{in} \quad (3)$$

Onde y_i é a variável alvo, α permanece sendo o coeficiente de intercepto do eixo Y enquanto β_1 a β_n representam os coeficientes associados à n-ésima variável (SASSI *et al.*, 2012).

O modelo de treinamento utilizando regressão linear consiste em encontrar os valores apropriados para os coeficientes, determinando assim a dinâmica da função a ser estimada (SOCZI, 2024). Isso é feito usando o método dos mínimos quadrados, onde procura-se os valores que minimizam a Soma Residual dos Quadrados (RSS):

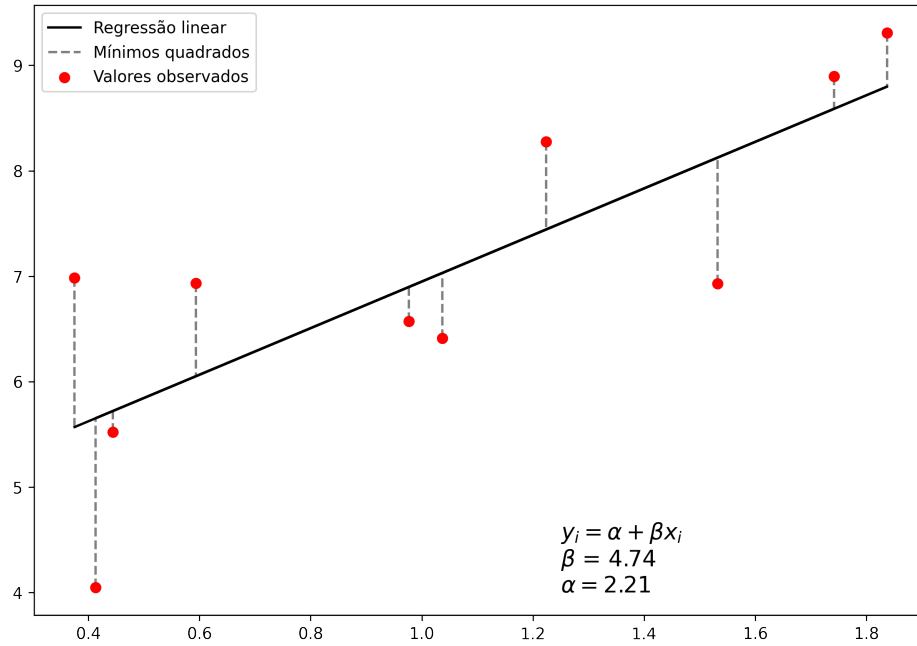
$$RSS = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right) \quad (4)$$

onde:

- y_i é uma variável aleatória e representa o valor da variável resposta (variável dependente) na i-ésima observação
- x_{ij} representa o valor da variável explicativa (variável independente, variável regressora) na i-ésima observação. Nota-se que podem existir múltiplas variáveis independentes para uma variável independente;
- β_0 e β_j são os parâmetros do modelo que serão estimados, e que definem a reta de regressão

Desse modo, a partir de um problema onde uma ou mais entradas geram amostras que resultam em uma saída, torna-se possível estimar uma função que melhor representa seu comportamento, minimizando ao máximo o valor da soma residual dos quadrados entre os pontos amostrais e a curva do modelo.

Figura 3 – Exemplo do método dos quadrados ordinários.



Fonte: (AT, 2024)

2.5 MODELO RIDGE

O modelo Ridge, implementado na biblioteca *scikit-learn* do Python, é uma variação da regressão linear que incorpora um termo de regularização L2 à função de custo, o que ajuda a controlar a complexidade do modelo e prevenir o sobreajuste (*overfitting*). Com essa característica, o método é indicado em casos onde os dados apresentam colinearidade ou onde há muitas variáveis independentes (JOLLY, 2018).

A regressão linear padrão visa minimizar a soma dos erros quadráticos (Erro Quadrático Médio ou MSE) entre as previsões e os valores reais. No entanto, em casos onde o dataset apresenta muitos recursos ou quando os dados apresentam correlações entre as variáveis, o modelo tende a se ajustar demais aos dados de treinamento (*overfitting*). A fim de contornar tal problema, o modelo Ridge adiciona um termo de penalidade à função de custo, que regula o tamanho dos coeficientes do modelo. Esse comportamento é observado na função de custo do modelo, com a adição de um hiperparâmetro λ em relação ao modelo de regressão linear padrão.

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2 + \lambda \sum_{j=1}^n \theta_j^2 \quad (5)$$

onde:

- λ é o hiperparâmetro que controla a intensidade da regularização;
- θ_j são os coeficientes (ou pesos) do modelo;

- Quanto maior o valor de λ , maior será a penalização para grandes coeficientes, resultando em um modelo mais simples.

Sendo assim, a inclusão do termo somado a equação de regressão linear padrão reduz a magnitude dos coeficientes, ajudando a controlar o *overfitting*. Em essência, o modelo Ridge evita que o modelo aprenda padrões específicos do conjunto de treinamento que não se generalizam bem para dados novos.

3 SEÇÃO

Este *template* contém algumas seções criadas na tentativa de facilitar seu uso. No entanto, não há um limite máximo ou mínimo de seção a ser utilizado no trabalho. Cabe a cada autor definir a quantidade que melhor atenda à sua necessidade.

4 CONCLUSÃO

As conclusões devem responder às questões da pesquisa, em relação aos objetivos e às hipóteses. Devem ser breves, podendo apresentar recomendações e sugestões para trabalhos futuros.

REFERÊNCIAS

- ANDRADE, Mauro M.; SCOTTÁ, Fernando C.; JR., Elírio E. Toldo; WESCHENFELDER, Jair; NUNES, José C. Hidrodinâmica do Rio Guaíba: Resultados Preliminares. *In: XXII Simpósio Brasileiro de Recursos Hídricos*. Porto Alegre: Associação Brasileira de Recursos Hídricos, 2017.
- AT, Data. **Introdução ao Machine Learning: Regressão**. 2024. Disponível em: <https://dataat.github.io/introducao-ao-machine-learning/regress%C3%A3o.html>. Acesso em: 15 out. 2024.
- CARBONELL, Jaime G.; MICHALSKI, Ryszard S.; MITCHELL, Tom M. **Machine Learning: An Artificial Intelligence Approach**. Berlin: Springer, 1983. P. 619.
- G1. **Santa Catarina registra enchente, queda de barreira e dia mais chuvoso de 2024**. 2024. Disponível em: <https://g1.globo.com/sc/santa-catarina/noticia/2024/05/19/sc-enchente-quedas-barreira-dia-mais-chuvoso.ghtml>. Acesso em: 15 out. 2024.
- HELLIWELL, John F.; HUANG, Haifang; WANG, Shun; NORTON, Max. Social Environments for World Happiness. *In: WORLD Happiness Report 2020*. [S.l.: s.n.], 2020.
- JOLLY, Kevin. **Machine Learning with scikit-learn Quick Start Guide: Classification, Regression, and Clustering Techniques in Python**. Birmingham, UK: Packt Publishing, 2018.
- KIRCH, Wilhelm. Pearson's Correlation Coefficient. *In: ENCYCLOPEDIA of Public Health*. Dordrecht: Springer Netherlands, 2008. P. 1090–1091.
- SARAVANAN, R.; SUJATHA, P. A State of Art Techniques on Machine Learning: A Perspective of Supervised Learning Approaches in Data Classification. *In: PROCEEDINGS of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)*. Puducherry: IEEE, 2018.
- SASSI, Cecília P.; PEREZ, Felipe G.; MYAZATO, Letícia; YE, Xiao; FERREIRA-SILVA, Paulo H.; LOUZADA, Francisco. **Modelos de Regressão Linear Múltipla Utilizando os Softwares R e STATISTICA: Uma Aplicação a Dados**

de Conservação de Frutas. São Carlos, SP, Brasil, 2012. Relatórios Técnicos. Disponível em: <http://www.icmc.usp.br>.

SERVICES, Amazon Web. **O que é regressão linear?** 2024. Disponível em: <https://aws.amazon.com/pt/what-is/linear-regression/>. Acesso em: 15 out. 2024.

SOCZI, M. **Ridge Regression Step-by-Step: Introduction with Example.** 2024. Disponível em: <https://medium.com/@msoczi/ridge-regression-step-by-step-introduction-with-example-0d22dddb7d54>. Acesso em: 18 mai. 2024.

SOTO, Timothy. Regression Analysis. *In*: VOLKMAR, Fred R. (Ed.). **Encyclopedia of Autism Spectrum Disorders.** New York, NY: Springer New York, 2013. P. 2538–2538.

VEJA. **De 1941 a 2024: por que as enchentes são um desafio constante no Rio Grande do Sul.** 2024. Disponível em: <https://veja.abril.com.br/ciencia/de-1941-a-2024-porque-as-enchentes-sao-desafio-constante-no-rs>. Acesso em: 15 out. 2024.

APÊNDICE A – Descrição

Textos elaborados pelo autor, a fim de completar a sua argumentação. Deve ser precedido da palavra APÊNDICE, identificada por letras maiúsculas consecutivas, travessão e pelo respectivo título. Utilizam-se letras maiúsculas dobradas quando esgotadas as letras do alfabeto.

ANEXO A – Descrição

São documentos não elaborados pelo autor que servem como fundamentação (mapas, leis, estatutos). Deve ser precedido da palavra ANEXO, identificada por letras maiúsculas consecutivas, travessão e pelo respectivo título. Utilizam-se letras maiúsculas dobradas quando esgotadas as letras do alfabeto.