

APRENDIZADO DE MÁQUINA PARA SAÚDE

Aprendizado Clássico: do rótulo à descoberta

Regressão, árvores, ensembles, SVM, clustering e fenotipagem computacional

APOSTILA TÉCNICA · ENCONTRO 06

Prof. Flávio Luiz Seixas
Pós-graduação em Computação

Material acadêmico-didático fundamentado exclusivamente nos PDFs disponíveis no projeto

Sumário

Clique com o botão direito e selecione “Atualizar Campo” se necessário.

Introdução

Algoritmos clássicos oferecem diferentes vieses indutivos para representar relações entre entradas e saídas ou descobrir estrutura sem rótulos. A escolha não deve ser uma disputa abstrata de desempenho: depende da tarefa, da dimensionalidade, do tamanho da amostra, da estabilidade e da necessidade de interpretação. [1]

O encontro compara regressão logística, árvores, ensembles, máquinas de vetores de suporte, agrupamento e redução de dimensionalidade, destacando entradas, saídas, pressupostos e armadilhas de interpretação. [1]



Figura 1 — Árvores e ensembles. Fonte: elaboração própria a partir do material da disciplina, com base em [1] e [2].

1. Supervisionado e não supervisionado

No aprendizado supervisionado, exemplos incluem preditores e uma saída conhecida; o modelo aprende uma função para novos casos. No não supervisionado, não há rótulo-alvo e o algoritmo procura organização nos dados. Essa distinção altera o tipo de afirmação possível: um agrupamento sugere estrutura, mas não demonstra que os grupos são doenças naturais [1].

CONCEITO-CHAVE

No aprendizado supervisionado, exemplos incluem preditores e uma saída conhecida; o modelo aprende uma função para novos casos.

2. Regressão logística e margem

A regressão logística modela a relação entre preditores e a probabilidade de uma classe binária. Coeficientes podem ser transformados em razões de chances, mas essa associação não deve ser interpretada automaticamente como efeito causal. Regularização pode controlar complexidade e instabilidade em conjuntos com muitas variáveis [1].

Máquinas de vetores de suporte procuram uma fronteira com margem entre classes; funções kernel permitem fronteiras não lineares em um espaço transformado. Escala das variáveis, escolha do kernel e parâmetros de penalização influenciam o resultado e sua generalização [2].

3. Árvores e ensembles

Árvores dividem o espaço de preditores por regras sucessivas e produzem classe ou valor nas folhas. Árvores profundas podem ajustar ruído. Random forests combinam árvores construídas com reamostragem e subconjuntos de variáveis, reduzindo variância pela agregação [2].

Boosting constrói modelos em sequência, concentrando etapas posteriores nos erros anteriores. O ganho de flexibilidade amplia a necessidade de regularização, validação e explicações adequadas ao uso clínico [1].

4. Clustering, visualização e fenótipos

Agrupamento depende de representação, distância e critério de formação dos grupos. Mudanças de escala ou escolha de variáveis podem mudar a solução. Reduções como t-SNE e UMAP são úteis para visualizar vizinhanças, mas a geometria bidimensional não deve ser tomada como prova de separação clínica [1].

Fenotipagem computacional usa padrões em dados clínicos para caracterizar grupos ou estados. A interpretação requer validação externa, estabilidade, plausibilidade e relação com decisões ou desfechos; o nome dado ao cluster não substitui evidência de validade [3].

Síntese do encontro

Supervisionado e não supervisionado: no aprendizado supervisionado, exemplos incluem preditores e uma saída conhecida; o modelo aprende uma função para novos casos.
--

Regressão logística e margem: a regressão logística modela a relação entre preditores e a probabilidade de uma classe binária.
--

Árvores e ensembles: árvores dividem o espaço de preditores por regras sucessivas e produzem classe ou valor nas folhas.
--

Clustering, visualização e fenótipos: agrupamento depende de representação, distância e critério de formação dos grupos.
--

Referências

[1] Subramanian, D.; Cohen, T. A. Machine Learning Systems. In: Intelligent Systems in Medicine and Health. Springer, 2022. [Link externo](#)

Arquivo local: pdfs/3-Intelligent Systems/VI-Machine Learning Systems.pdf · páginas PDF utilizadas: 17, 25, 31, 37, 44, 70, 71.

[2] Agrawal, R.; Chatterjee, J. M.; Jena, A. K. (eds.). Machine Learning for Healthcare: Handling and Managing Data. CRC Press, 2020.

Arquivo local: pdfs/1-Machine Learning/Machine Learning for Healthcare _ Handling and Managing Data -- edited by Rashmi Agrawal, Jyotir Moy Chatterjee, Abhishek -- 1, PT, 2020 -- CRC Press, -- isbn13 9780367352332 -- 6284b1970cc3abcc85ea2ecea9ac7116 --.pdf · páginas PDF utilizadas: 26, 27, 28, 30.

[3] Bellazzi, R.; Dagliati, A.; Nicora, G. Predicting Medical Outcomes. In: Intelligent Systems in Medicine and Health. Springer, 2022. [Link externo](#)

Arquivo local: pdfs/3-Intelligent Systems/XI-Predicting Medical Outcomes.pdf · páginas PDF utilizadas: 9, 17, 24, 25.