

Árvores de Decisão e Redes Neurais Artificiais

Prof. Rodrigo Pedrosa

2 de agosto de 2025

1 Árvores de decisão

1. O que é uma árvore de decisão e como ela é usada na aprendizagem de máquina?
2. Qual é a diferença entre os algoritmos de construção de árvore de decisão para classificação e para regressão?
3. O que podemos fazer para evitar o *overfitting* em árvores de decisão?
4. Considere um conjunto de dados com 100 exemplos, dos quais 60 pertencem à classe A e 40 pertencem à classe B. Calcule o índice de Gini desse conjunto de dados.
5. Em um conjunto de dados com 80 exemplos, dos quais 45 pertencem à classe X e 35 pertencem à classe Y. Calcule o índice de Gini desse conjunto de dados.
6. Suponha que um conjunto de dados seja dividido em dois subconjuntos, onde o subconjunto A contém 30 exemplos, dos quais 20 pertencem à classe P e 10 pertencem à classe Q, e o subconjunto B contém 70 exemplos, dos quais 40 pertencem à classe P e 30 pertencem à classe Q. Calcule o ganho de Gini para essa divisão com base no índice de Gini inicial do conjunto de dados.
7. Considere um conjunto de dados com duas características, "Altura"(com valores "Alto" e "Baixo") e "Idade"(com valores "Jovem" e "Adulto"), e uma classe "Classe"(com valores "A" e "B").

Considere a seguinte tabela de dados:

Altura	Idade	Classe
Alto	Jovem	A
Alto	Adulto	A
Baixo	Jovem	B
Baixo	Adulto	B
Alto	Jovem	B
Baixo	Adulto	A

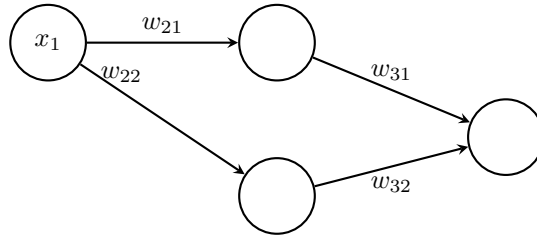
Construa uma árvore de decisão para classificar os exemplos com base nessas características, usando o critério de Gini.

8. Dado um conjunto de dados com 100 exemplos, onde 70 pertencem à classe X e 30 pertencem à classe Y, avalie duas divisões possíveis com base no índice de Gini, e determine qual delas é mais preferível em termos de impureza.
9. Considere a seguinte base de dados:
 - (a) Apresente uma árvore de decisão para a classificação das *User-actions* e calcule o grau de impureza (I_G) médio do nó raiz da sua árvore. (Obs: $I_G(p) = 1 - \sum_{i=1}^J p_i^2$)
 - (b) De acordo com a árvore apresentada, qual a classificação dos exemplos e_{19} e e_{20} ?

<i>Example</i>	<i>Author</i>	<i>Thread</i>	<i>Length</i>	<i>Where_read</i>	<i>User_action</i>
e_1	known	new	long	home	skips
e_2	unknown	new	short	work	reads
e_3	unknown	followup	long	work	skips
e_4	known	followup	long	home	skips
e_5	known	new	short	home	reads
e_6	known	followup	long	work	skips
e_7	unknown	followup	short	work	skips
e_8	unknown	new	short	work	reads
e_9	known	followup	long	home	skips
e_{10}	known	new	long	work	skips
e_{11}	unknown	followup	short	home	skips
e_{12}	known	new	long	work	skips
e_{13}	known	followup	short	home	reads
e_{14}	known	new	short	work	reads
e_{15}	known	new	short	home	reads
e_{16}	known	followup	short	work	reads
e_{17}	known	new	short	home	reads
e_{18}	unknown	new	short	work	reads
e_{19}	unknown	new	long	work	?
e_{20}	unknown	followup	short	home	?

2 Redes Neurais Artificiais

1. Considere a rede neural abaixo:



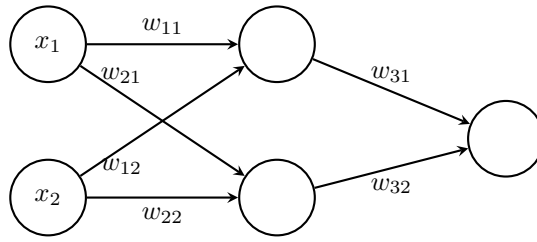
Esta rede não possui termos de viés (bias) e tem como funções de ativação a função ReLU (Rectified Linear Unit) que pode ser definida como:

$$\text{ReLU}(x) = \max(0, x) \quad (1)$$

A derivada da ReLU é definida como:

$$\frac{d}{dx}(\text{ReLU}(x)) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (2)$$

Obtenha o gradiente do erro quadrado em relação aos pesos da rede. Todos os passos da derivação da gradiente devem ser apresentados.



2. Considere a rede neural abaixo:

$$w_{11} = w_{21} = w_{12} = w_{22} = w_{31} = w_{32} = 1$$

Esta rede tem como funções de ativação a função ReLU (Rectified Linear Unit) que pode ser definida como:

$$\text{ReLU}(x) = \max(0, x) \quad (3)$$

A derivada da ReLU é definida como:

$$\frac{d}{dx}(\text{ReLU}(x)) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (4)$$

- Calcule o gradiente do erro quadrado em relação à w_{32} quando $\mathbf{x} = [2, 1]$ e $y = 20$.
 - Calcule o gradiente do erro quadrado em relação à w_{22} quando $\mathbf{x} = [2, 1]$ e $y = 20$.
 - Como w_{32} e w_{22} devem ser alterados de forma a diminuir o erro?
- Quais são as funções de ativação mais comuns usadas nas redes neurais artificiais e como elas afetam o processo de treinamento?
 - Como o *overfitting* pode ser evitado em redes neurais artificiais?

3 Overfitting, Underfitting e Avaliação de Modelos

3.1 Conceitos fundamentais

- Defina *overfitting* e *underfitting*. Dê um exemplo de situação que tipicamente leva a cada um deles.
- Explique o *trade-off* viés-variância. Como esse *trade-off* se manifesta quando aumentamos a complexidade do modelo?
- Diferencie erro de aproximação (viés) e erro de estimação (variância). Por que ambos importam?
- Em que sentido as regularizações L1 e L2 ajudam a mitigar overfitting? Compare seus efeitos esperados sobre os coeficientes.
- O que é *capacidade do modelo*? Relacione capacidade, número de parâmetros e risco de overfitting.
- Explique *early stopping*. Por que ele pode ser visto como uma forma de regularização?

3.2 Diagnóstico e detecção

7. Descreva como curvas de aprendizado (*learning curves*) podem distinguir overfitting de underfitting. Que padrões você esperaria ver?
8. Você observa alta acurácia em treino e baixa em teste. Quais hipóteses explicam esse comportamento e como investigá-las?
9. Em um problema desbalanceado, por que acurácia pode mascarar overfitting? Sugira métricas alternativas.
10. O que é *data leakage*? Dê dois exemplos comuns de vazamento e como evitá-los.
11. Qual o papel de um conjunto de validação separado na detecção de overfitting durante ajuste de hiperparâmetros?

3.3 Particionamento e validação

12. Diferencie conjunto de treino, validação e teste em termos de propósito. O que ocorre quando usamos o teste para escolher hiperparâmetros?
13. Explique *validação cruzada k-fold*. Quais são suas vantagens em relação a um único *holdout*?
14. Quando a validação cruzada *estratificada* é necessária? O que ela preserva e por quê?
15. Por que a *nested cross-validation* (CV aninhada) é recomendada para comparação justa de modelos? Descreva a estrutura externa e interna.
16. Em séries temporais, por que k-fold aleatório é inadequado? Descreva um esquema de validação temporal apropriado.
17. Em quais situações o *leave-one-out* CV (LOOCV) é vantajoso e quando pode ser problemático?
18. Como você escolheria k no k-fold? Quais impactos de k muito pequeno ou muito grande sobre viés/variancia da avaliação?

3.4 Seleção de modelos e métricas

19. Diferencie *seleção de modelo* de *avaliação final*. Por que não devemos reportar no artigo o melhor resultado observado durante a busca de hiperparâmetros sem correção?
20. Como escolher métricas quando os custos de erro são assimétricos? Dê exemplos de métricas apropriadas para classificação com custos diferenciados e para regressão com *outliers*.
21. Você treinou um modelo com ajuste de hiperparâmetros via CV e agora quer relatar o desempenho final. Qual procedimento correto para estimar o desempenho fora da amostra?

4 Questões Práticas [opcional]

1. Reproduza o tutorial disponível em <https://www.analyticsvidhya.com/blog/2021/06/end-to-end-machine-learning-with-logistic-regression/> com os dados <https://www.kaggle.com/datasets/mriganekjswal/student-performance-in-mathematics-and-science>
2. Substitua o modelo do tutorial por uma Árvore de Decisão e faça uma análise comparativa quantitativa e qualitativa entre o modelo de Regressão Linear e a Árvore de decisão.