

## *Estudo Dirigido III*

**Aluno:** Nathann Zini dos Reis

**Matrícula:** 19.2.4007

---

Questão 01 -

- a) **Overfitting:** ocorre quando um modelo tem um desempenho excelente com os dados de treino e acaba decorando as regras do mesmo, que, quando aplicado ao cenário de teste, o desempenho não é bom devido a ele tentar usar as regras criadas no cenário de teste que não se aplica a tipos diferentes de dados. Algoritmo muito complexo para os dados: podemos simplificar nosso modelo escolhendo um algoritmo mais simples, com menos parâmetros, caso seja possível. Isso reduz as chances do modelo sofrer overfitting.
- b) **Underfitting:** ocorre quando o modelo tem um desempenho péssimo ainda com base de dados de treino. Então o mesmo pode ser descartado já nessa fase, pois não encontra uma boa relação entre os dados. Algoritmo inadequado, pouco poderoso para os dados: aqui podemos amplificar o poder do nosso algoritmo escolhendo outro com mais parâmetros para solucionar o underfitting.
- c) **Viés de busca:** O viés de busca de um algoritmo é a forma como o algoritmo busca a hipótese que melhor se ajusta aos dados de treinamento. Ele define como as hipóteses são pesquisadas no espaço de hipóteses. Por exemplo, o algoritmo ID3, que é utilizado para indução de árvores de decisão, tem como viés de busca a preferência por árvores de decisão com poucos nós.
- d) **Viés de representação:** Cada algoritmo utiliza uma forma ou representação para descrever a hipótese induzida. Por exemplo, árvores de decisão utilizam uma estrutura de árvore em que cada nó interno é representado por uma pergunta referente ao valor de um atributo e cada nó externo está associado a uma classe. A representação utilizada define o viés de representação do algoritmo e pode restringir o conjunto de hipóteses que podem ser induzidas pelo algoritmo.

Questão 2 -

Os modelos de aprendizado de composição funcionam de modo que pegam diferentes modelos de predição mais simples e os combinam em um modelo de predição mais complexos e robustos, diminuindo a suscetibilidade a bias e a variance. Portanto, os métodos de Ensemble devem levar em conta a maneira com a qual eles agrupam os

modelos, associando os algoritmos de forma a minimizar suas desvantagens individuais no modelo final.

Questão 3 -

Total 6

Fórmula:  $-\sum(p_i * \log_2(p_i))$

a) Classe 1 : 0 ; Classe 2 : 6

$p_i[1]: 0/6 = 0;$

$p_i[2]: 6/6 = 1;$

Entropia:  $-(1 \log_2(1)) = 0;$

Erro:  $1 - 1 = 0;$

b) Classe 1 : 1 ; Classe 2 : 5

$p_i[1]: 1/6 = 0,1666$

$p_i[2]: 5/6 = 0,8333$

Entropia:  $-(0,166 * \log_2(0,166) + 0,833 * \log_2(0,833)) = 0,65$

Erro:  $1 - 0,8333 = 0,1666$

c) Classe 1: 2 ; Classe 2: 4

$p_i[1]: 2/6 = 0,333$

$p_i[2]: 4/6 = 0,666$

Entropia:  $-(0,333 * \log_2(0,333) + 0,666 * \log_2(0,666)) = 0,918$

Erro:  $1 - 0,666 = 0,333$

d) Classe 1: 3 ; Classe 2: 3

$p_i[1]: 3/6 = 0,5$

$p_i[2]: 3/6 = 0,5$

Entropia:  $-2*(0,5 * \log_2(0,5))$

Erro:  $1 - 0,5 = 0,5$

Questão 4 -

a)

	Família		Esporte		Luxo
Classe 1:	1		2		1
Classe 2:	4		1		1

$gini(f):$

$$1 - ((\frac{1}{6})^2 - (\frac{5}{6})^2) = 0,32$$

$gini(e):$

$$1 - ((\frac{2}{3})^2 - (\frac{1}{3})^2) = 0,44$$

gini(l):

$$1 - ((\frac{1}{2})^2 - (\frac{1}{2})^2) = 0,5$$

$$\text{gini(d)}: (5/10) * 0,32 + (3/10) * 0,44 + (2/10) * 0,5 = 0,393$$

b)                      Baixa | Alta

Classe 1:        3 |    1

Classe 2:        2 |    4

gini(b):

$$1 - ((\frac{3}{5})^2 - (\frac{2}{5})^2) = 0,48$$

gin(a)

$$1 - ((\frac{1}{5})^2 - (\frac{4}{5})^2) = 0,32$$

$$\text{gin(d)}: (5/10) * 0,48 + (5/10) * 0,32 = 0,40$$

Questão 5 -

Estudo Dirigido III  
Nathann Zini dos Reis 19.2.4005

Questão 5.

$$Gini = 1 - \sum_{i=0}^{c-1} C(p_i)^2$$

Total: 6

Doente: 3

Saudável: 3

Dores	Sim	Não
$D_1$	1	3
$D_2$	2	0

$$Gini_{D_1} = 1 - (1/4)^2 - (3/4)^2 = 0,375$$

$$Gini_{D_2} = 1 - (2/2)^2 - (0/2)^2 = 0$$

$$Gini(w) = (4/6) * 0,375 + (2/6) * 0 = 0,25$$

$$Gini_{Dores} = 0,5 - 0,25 = 0,25$$

Manchas	Pequenas	Grandes
$M_1$	1	2
$M_2$	2	1

$$Gini_{M_1} = 1 - (1/3)^2 - (2/3)^2 = 0,445$$

$$M_2 = 1 - (2/3)^2 - (1/3)^2 = 0,444$$

$$Gini(w) = (3/6) * 0,445 + (3/6) * 0,444 = 0,444$$

$$Gini_{Manchas} = 0,5 - 0,444 = 0,055$$

Nathann Zini: das Reis 19.2.4007

• Emjeio:	Sim	Não
E1	3	3
E2	2	2

$$Gini_{E1} = 1 - (3/2)^2 - (3/2)^2 = 1$$

$$Gini_{E2} = 1 - (2/4)^2 - (2/4)^2 = 1$$

$$Gini_{(E)} = (2/6)^2 \cdot 3 + (4/6)^2 \cdot 3 = 1$$

$$Gini_{Emjeio} = 0,5 - 1 = -0,5$$

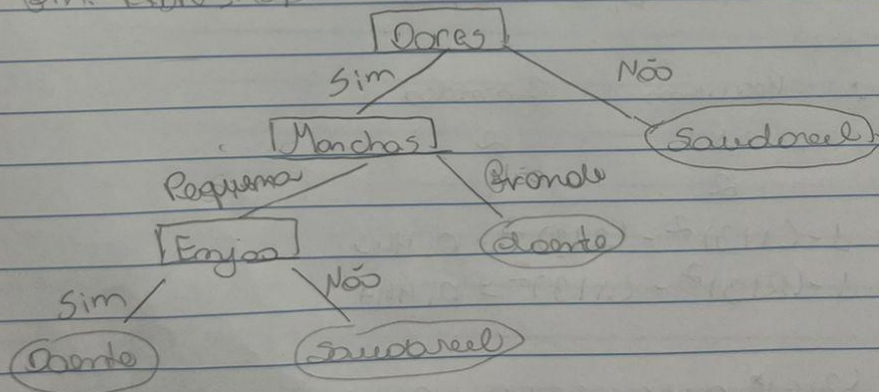
• Febre:	Sim	Não
F1	2	2
F2	3	3

$$Gini_{F1} = 1$$

$$Gini_{F2} = 1$$

$$Gini_{(F)} = 1$$

$$Gini_{Febre} = 0,5$$





Nathann Zini dos Reis 192.4004

teste

• Chuis, não, não, pequenos, sim)

Luis:

Dores  $\xleftrightarrow{\text{sim}}$  Monchas  $\xleftrightarrow{\text{Pequenos}}$  Saloborel

(Saloborel)

• Chuis, sim, sim, grande, sim)

Dores  $\xleftrightarrow{\text{sim}}$  Monchas  $\xleftrightarrow{\text{Grande}}$  Engas  $\xleftrightarrow{\text{sim}}$  Dente