



DECOMPOSIÇÃO DE PROBLEMAS MULTICLASSE



CAPÍTULO 17


Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	?	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

O QUE É?

- CLASSE - DESTINO DE UM RÓTULO
- DECOMPOSIÇÃO - SEPARAR EM PROBLEMAS MENORES (BINÁRIOS)
- DUAS FASES - DECOMPOSIÇÃO E RECONSTRUÇÃO



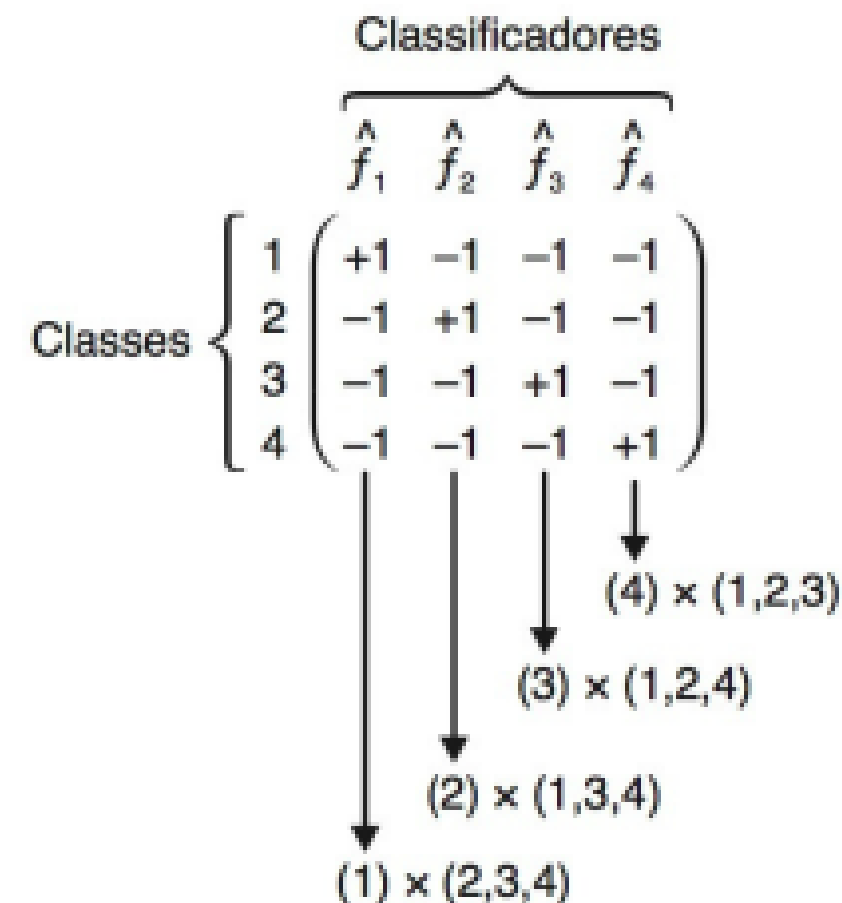
POR QUE DECOMPOR?

- MUITAS TÉCNICAS SÃO ORIGINALMENTE BINÁRIAS
 - ALGORITMOS POUCO ADEQUADOS À MULTICLASSE
 - ALGORITMOS COM PROCEDIMENTOS INTERNOS RESTRITOS A PROBLEMAS BINÁRIOS
 - REDUZIR COMPLEXIDADE COMPUTACIONAL
(SUBDIVISÃO EM TAREFAS)
- 



FASE DE DECOMPOSIÇÃO

- OCORRE ANTES DO APRENDIZADO
- OBTENÇÃO DE SUBPROBLEMAS BINÁRIOS
- $m_{ij} = \{-1, 0, 1\}$
 - -1 = RÓTULO NEGATIVO
 - 1 = RÓTULO POSITIVO
 - 0 = NÃO PARTICIPA
- UM-CONTRA-TODOS (OAA)
- TODOS-CONTRA-TODOS (OAO)
- DECOMPOSIÇÃO HIERÁRQUICA
- CÓDIGOS DE CORREÇÃO DE ERRO (ECOC)



UM-CONTRA-TODOS

- SEMPRE APRESENTA FORMA $K \times K$
- CADA CLASSIFICADOR COMPARA UMA CLASSE COM TODAS AS OUTRAS
- DIAGONAL +1, DEMAIS VALORES -1

$$\begin{pmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix}$$

- PROBLEMA: DESEMPENHO PREJUDICADO EM CASO DE DESBALANCEAMENTO DOS DADOS

TODOS-CONTRA-TODOS

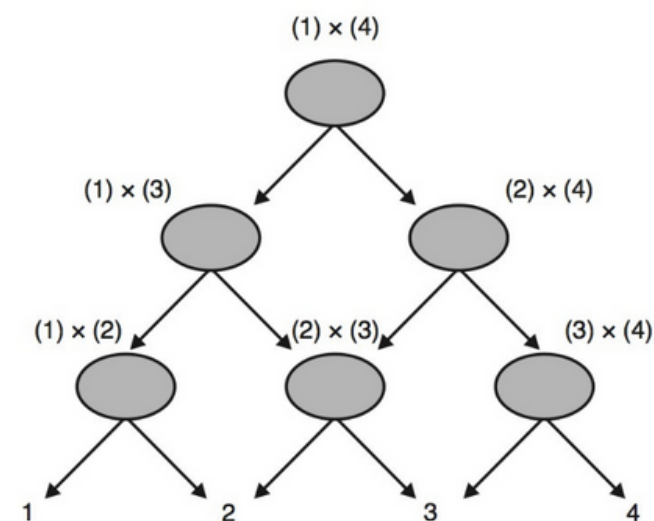
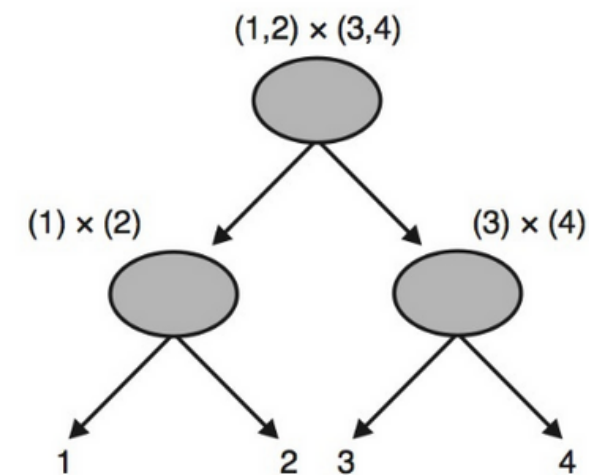
- K CLASSES $\rightarrow K(K-1)/2$ CLASSIFICADORES
- CADA CLASSIFICADOR COMPARA DUAS CLASSES
- MUITOS CLASSIFICADORES (ORDEM DE K^2), MAS TREINAMENTO DOS CLASSIFICADORES É RÁPIDO

$$\begin{pmatrix} +1 & +1 & +1 & 0 & 0 & 0 \\ -1 & 0 & 0 & +1 & +1 & 0 \\ 0 & -1 & 0 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix}$$

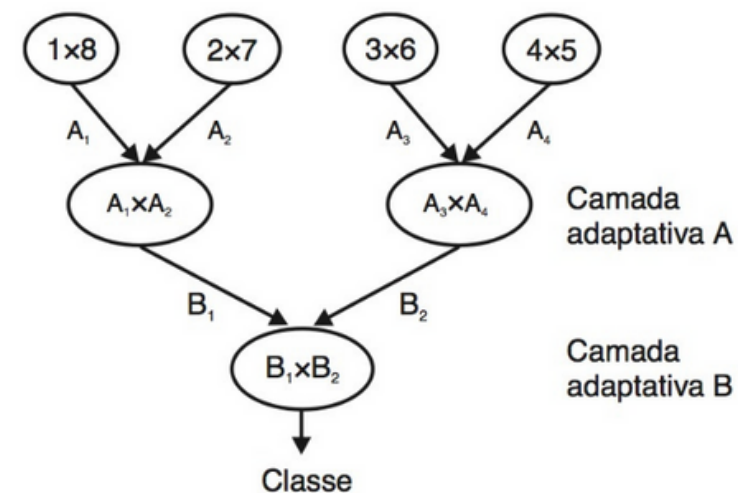
- PROBLEMA: DADO $K > 4$, MAIORIA DOS CLASSIFICADORES NÃO OFERECE INFORMAÇÃO NENHUMA PARA UMA CLASSE ALEATÓRIA ESCOLHIDA

DECOMPOSIÇÃO HIERÁRQUICA

- REDUZ COMPLEXIDADE DA SOLUÇÃO
- GRAFOS DIRECIONADOS ACÍCLICOS:
 - K-1 CLASSIFICADORES
 - DDAG (OAO EM FORMA DE GRAFO)



- ADAG (DDAG REVERSO)



CÓDIGOS DE CORREÇÃO DE ERROS

- UTILIZAÇÃO DOS CÓDIGOS COMO BASE PARA A DECOMPOSIÇÃO

- MATRIZ HAMMING: $H = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$

- VALORES ENTRE $\{-1, +1\}$
- TÉCNICAS DIFERENTES DE ACORDO COM NUM K DE CLASSES:
 - $K \leq 7 \rightarrow$ EXAUSTIVO (COMBINAR CLASSIFICADORES)
 - $8 \leq K \leq 11 \rightarrow$ SELECIONA COLUNAS
 - $K > 11 \rightarrow$ HILL-CLIMBING OU BCH

$$\begin{pmatrix} +1 & +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 \\ -1 & -1 & +1 & +1 & -1 & -1 \\ -1 & +1 & -1 & +1 & -1 & +1 \end{pmatrix}$$

- SUBPROBLEMAS DE DIFÍCIL APRENDIZADO

FASE DE RECONSTRUÇÃO

- OCORRE APÓS PREDIÇÃO
- COMBINAÇÃO DOS CLASSIFICADORES BINÁRIOS EM CLASSES FINAIS
 - 1 - L CLASSIFICADORES GERAM VETOR
 - 2 - VETOR É COMPARADO COM AS LINHAS DA MATRIZ (CÁLCULO DA DISTÂNCIA VARIA)
 - 3 - LINHA DA CLASSE MAIS PRÓXIMA É A CLASSE FINAL
- ESTRATÉGIA HIERÁRQUICA TENDE A SER MAIS RÁPIDA (CONSULTA PARTE DOS PREDITORES)

FERRAMENTAS E BIBLIOTECAS

PANDAS & NUMPY

*manipulação da base
de dados*

SEABORN & MATPLOT

*visualização gráfica
dos resultados*

SKLEARN

*algoritmos de
machine learning*

EXPERIMENTOS

- *UM-CONTRA-TODOS*
- *TODOS-CONTRA-TODOS*
- *DECOMPOSIÇÃO HIERÁRQUICA*
 - *ADAG*