

CLASSIFICAÇÃO EM GÊNEROS MUSICAIS

Gabriel BRITTO
Luiz CHAMON

- Introdução
 - Contexto e justificativa
 - O problema e a solução
- Features
- RNA
 - Estrutura
 - Aprendizado
- Resultados e discussões
- Demonstração
- Conclusão e perspectivas

CONTEXTO E JUSTIFICATIVA

- Aumentaram enormemente as bases de dados musicais
(iTunes: 14 milhões de músicas)
- Gêneros são a principal fonte de classificação musical ...
... mas é difícil defini-los



O que é rock?
O que é jazz?

(30 musicólogos levaram 1 ano
para categorizar 100.000 músicas
do *MSN music search*)



Discordância entre
taxonomias

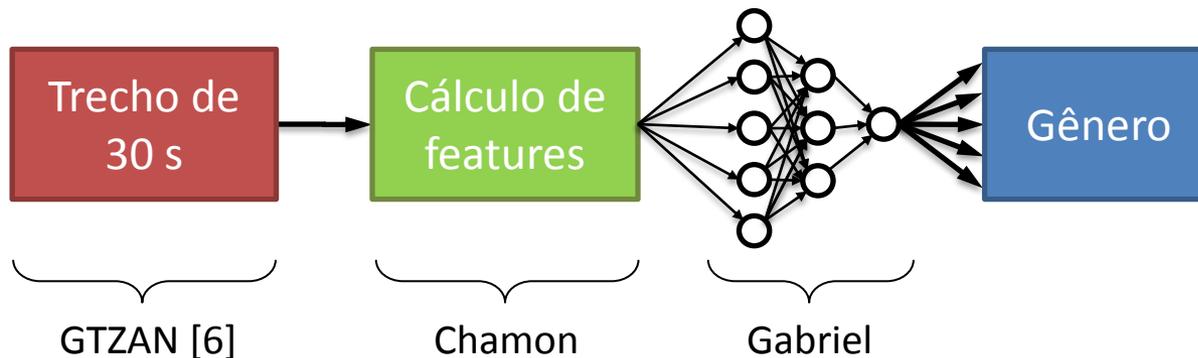
(Amazon: 719 gêneros
iTunes: 244 gêneros)

CLASSIFICAÇÃO MUSICAL É UM PROBLEMA NÃO-TRIVIAL

DESCRIÇÃO DO PROBLEMA

Categorizar músicas em gêneros automaticamente baseado em trechos de suas gravações e exemplos rotulados *a priori*

A SOLUÇÃO



FEATURES

20 MFCCs [2]

Muito usado em reconhecimento de voz. É uma medida de variações no tom.

$$MFCC = | \mathcal{F} \{ \log[mel(|S(f)|^2)] \} |^2$$

RMS [2]

Usado como medida da energia média da música.

$$RMS = \sqrt{\frac{1}{T} \int_{-T/2}^{T/2} s^2(t) dt}$$

Spectral Centroid [2]

Uma espécie de centro de massa do espectro, determina uma relação entre altas e baixas frequências.

$$C = \frac{\sum_{f=0}^M f \cdot |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}$$

FEATURES

Spectral bandwidth [2]

Mede a variação do espectro em torno da centróide espectral.

$$B^2 = \frac{\sum_{f=0}^M (f - c)^2 |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}.$$

Zero-crossing [2]

Representa uma medida do “ruído” de fundo em uma música. Trata-se simplesmente do número de cruzamentos por zero no domínio do tempo.

Band energy ratio [2]

Razão da energia no primeiro quarto do espectro com relação ao espectro todo.

$$BER = \frac{\sum_{f=0}^{M/4} |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}$$

FEATURES

Octave spectral contrast [3]

Medida da diferença entre espectros em bandas de oitava.

$$Valley_k = \log\left(\frac{1}{\alpha N} \sum_{f=0}^{\alpha N} |S(f)|\right), \quad Peak_k = \log\left(\frac{1}{\alpha N} \sum_{f=0}^{\alpha N} |S(N - f + 1)|\right),$$

$$SC_k = Peak_k - Valley_k.$$

Loudness [4]

Modelo psicoacústico de sensação sonora (“intensidade” percebida).

Sharpness [4]

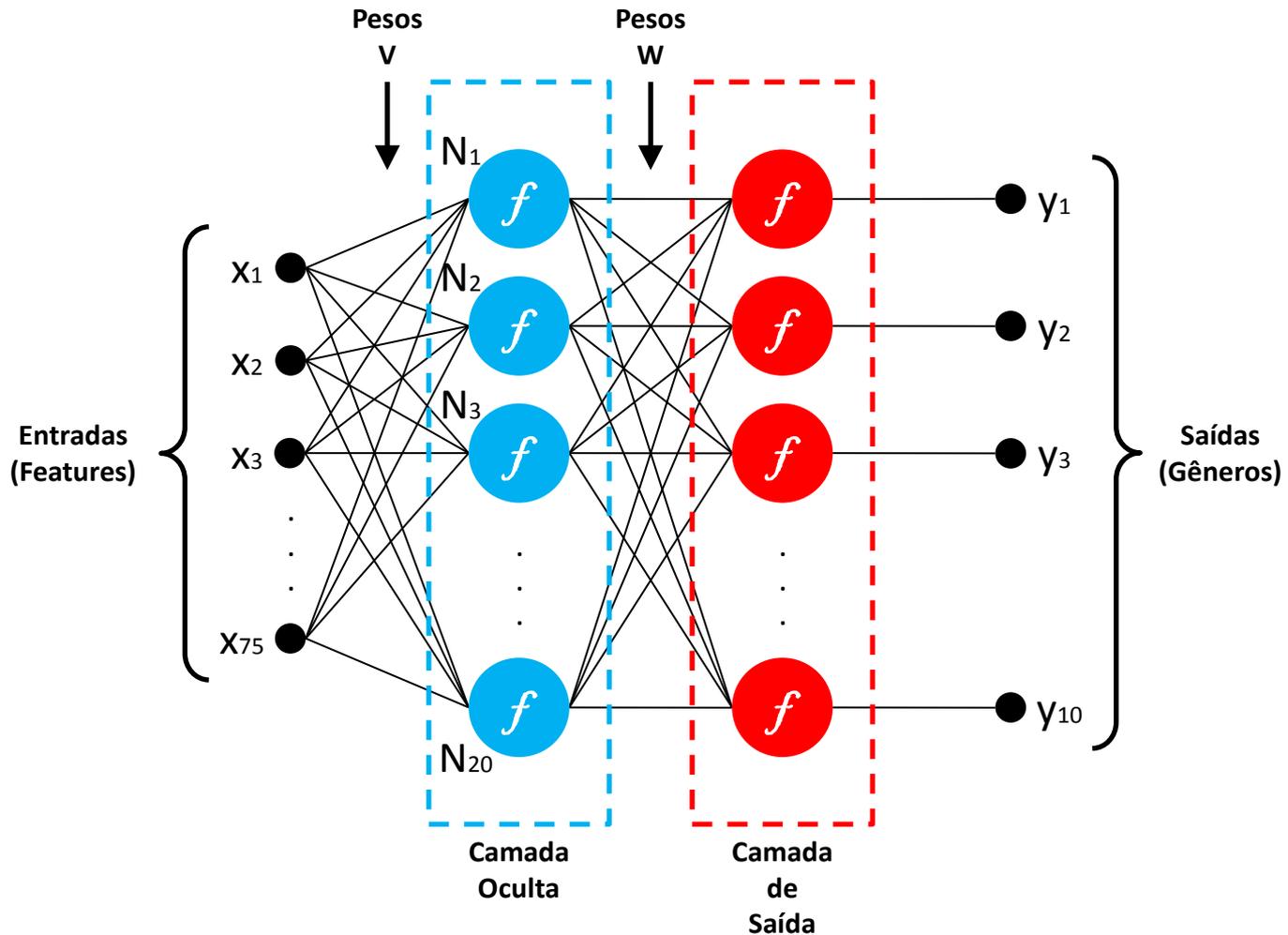
Medida de conteúdo espectral de altas frequências. Avaliação psicoacústica do quão “afiado”, “agudo” um som é.

ESTRUTURA

➤ Gêneros Musicais Considerados

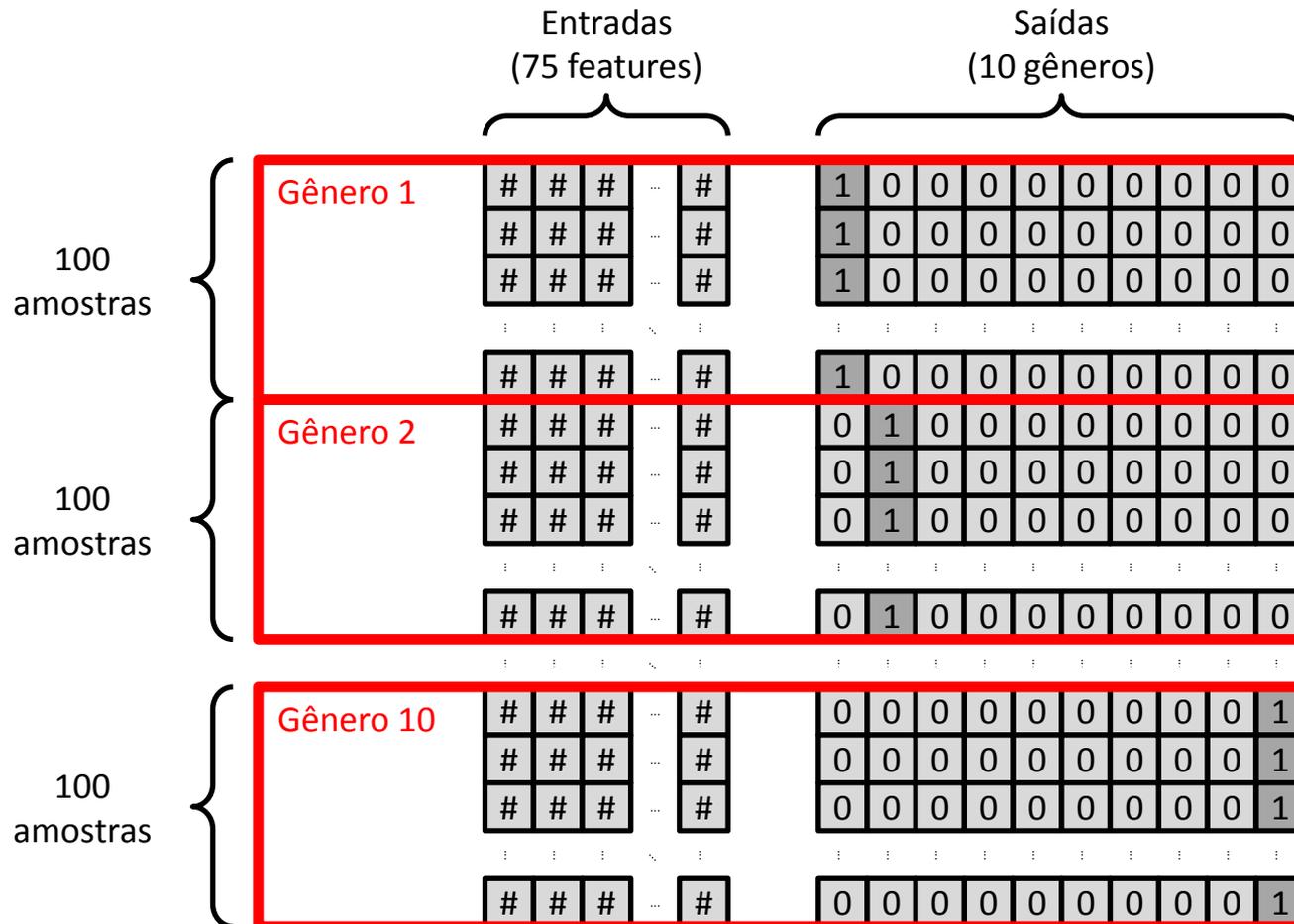
- 1) Blues
- 2) Classical
- 3) Country
- 4) Disco
- 5) Hip-Hop
- 6) Jazz
- 7) Metal
- 8) Pop
- 9) Reggae
- 10) Rock

ESTRUTURA



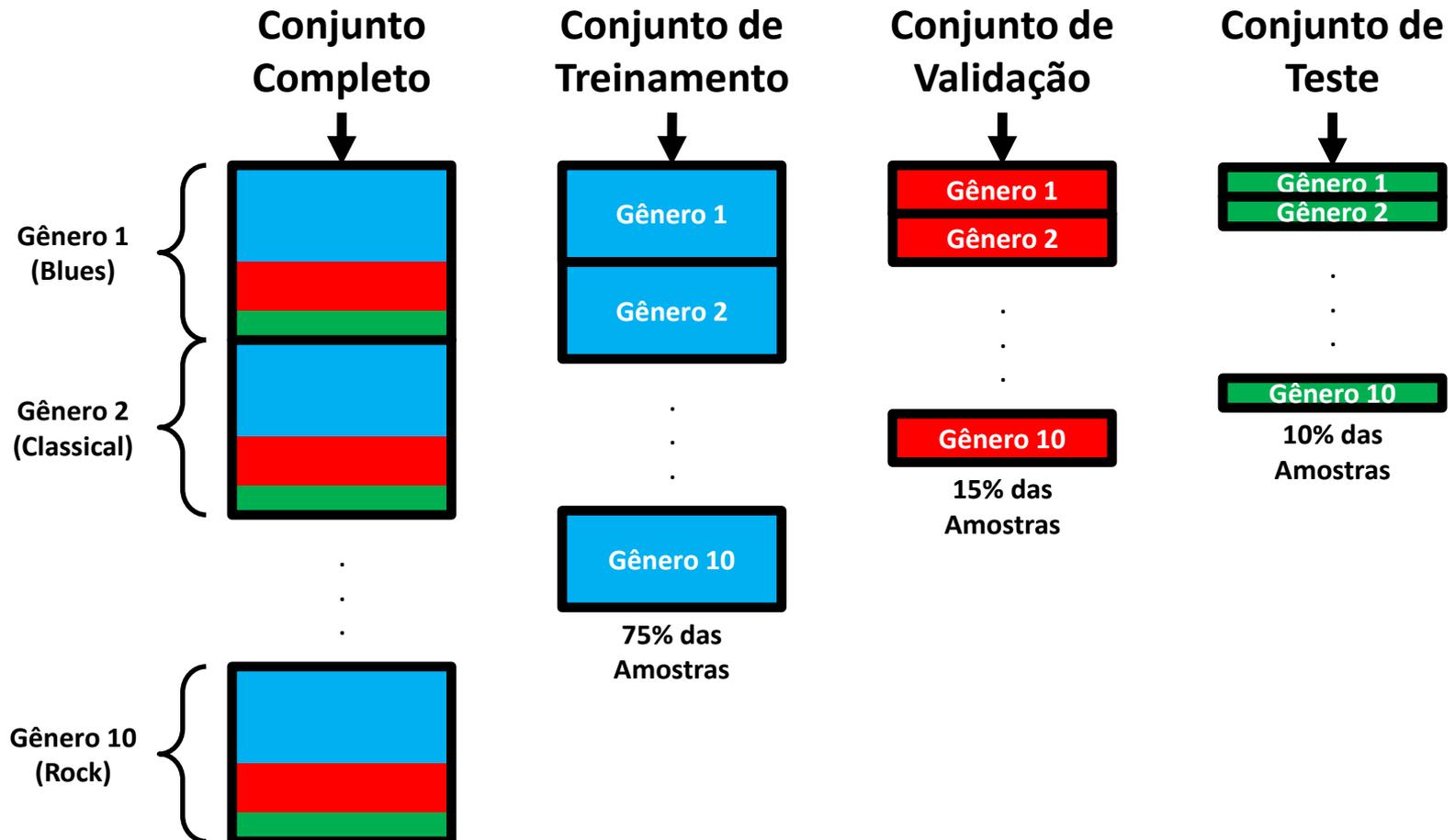
ESTRUTURA

➤ Formato das Entradas e Saídas



APRENDIZADO

➤ Conjuntos de Treinamento, Validação e Teste



APRENDIZADO

2 abordagens

- **Gradiente Descendente com “Cross-Validation”**
- **Gradiente Normalizado com “Early Stopping”**

APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”

- Algoritmo:

- “Backpropagation” e “Cross-Validation”

- Parâmetros:

- Número de nós:

- 20 na camada oculta

- 10 na camada de saída (1 para cada gênero)

- Taxa de aprendizagem:

- $\eta = 0,35$

- Erro máximo no aprendizado:

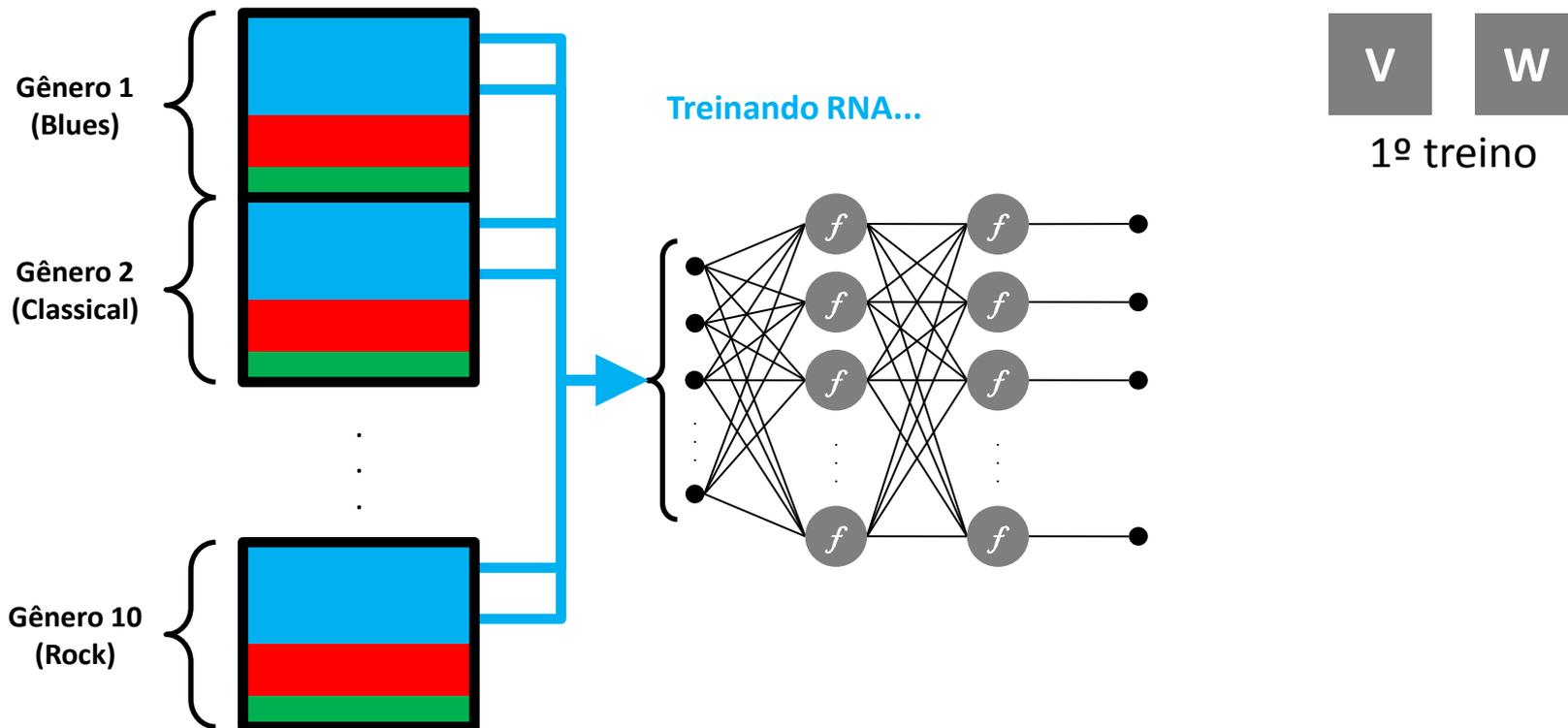
- erro = 0.1

- Número máximo de iterações:

- numMaxIt = 50000

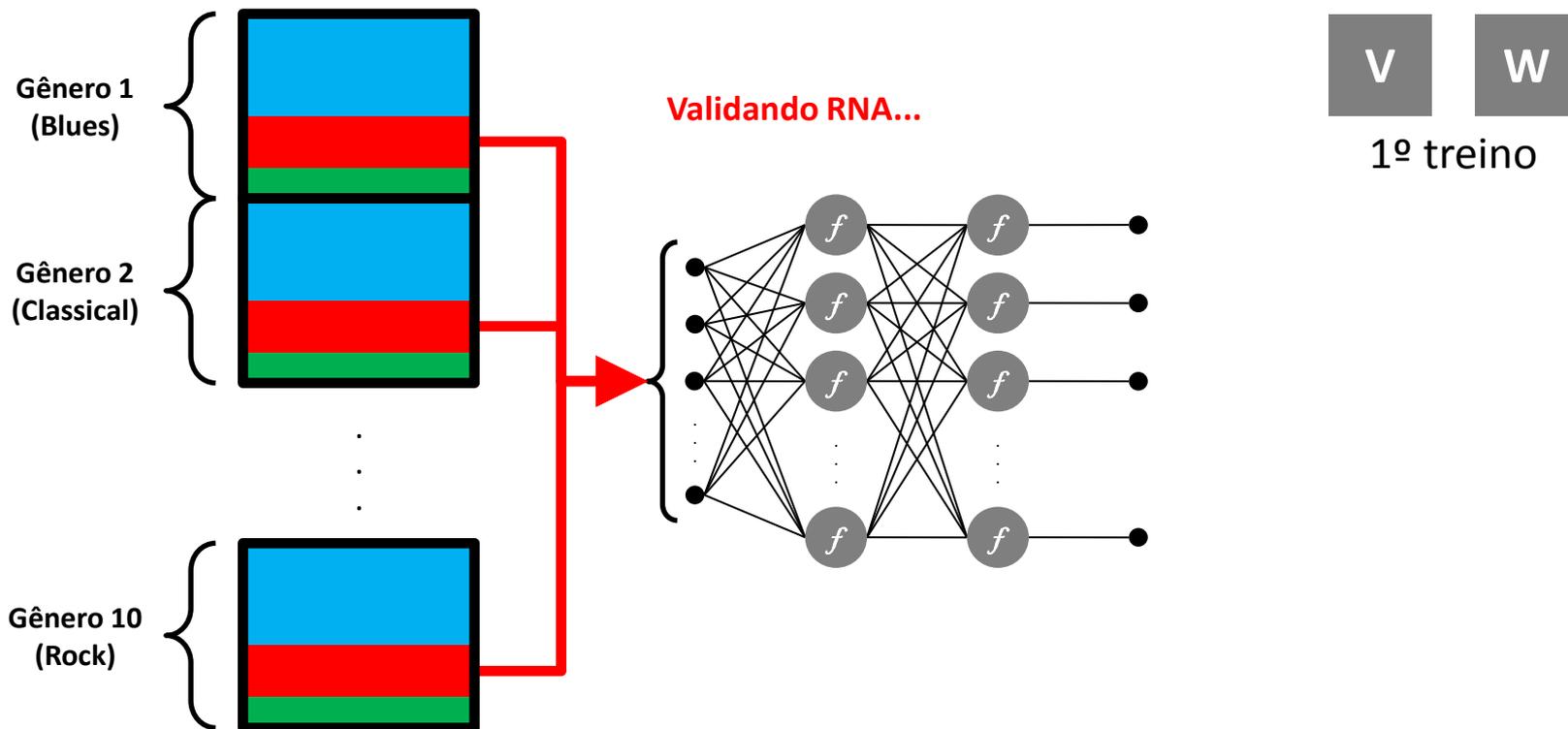
APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”



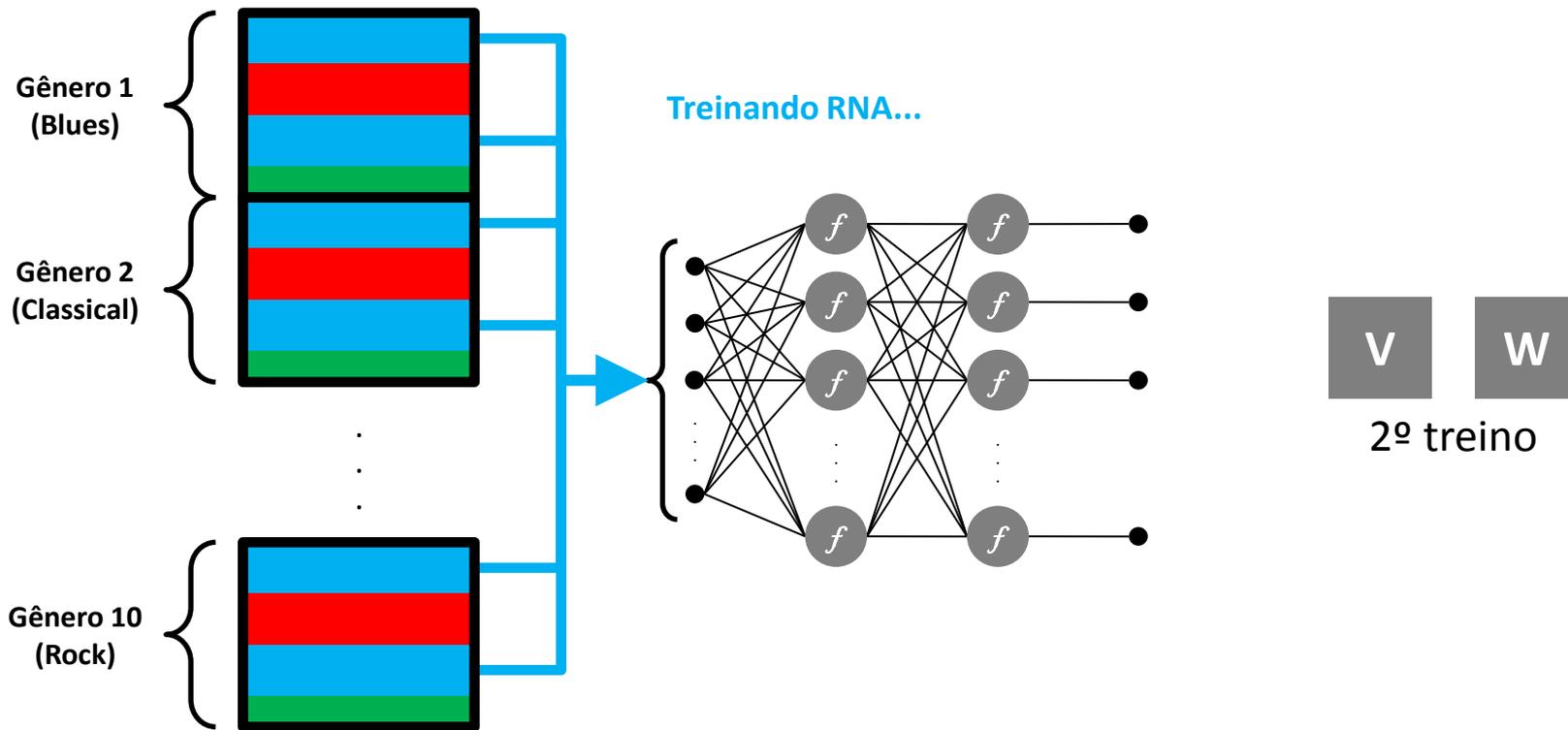
APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”



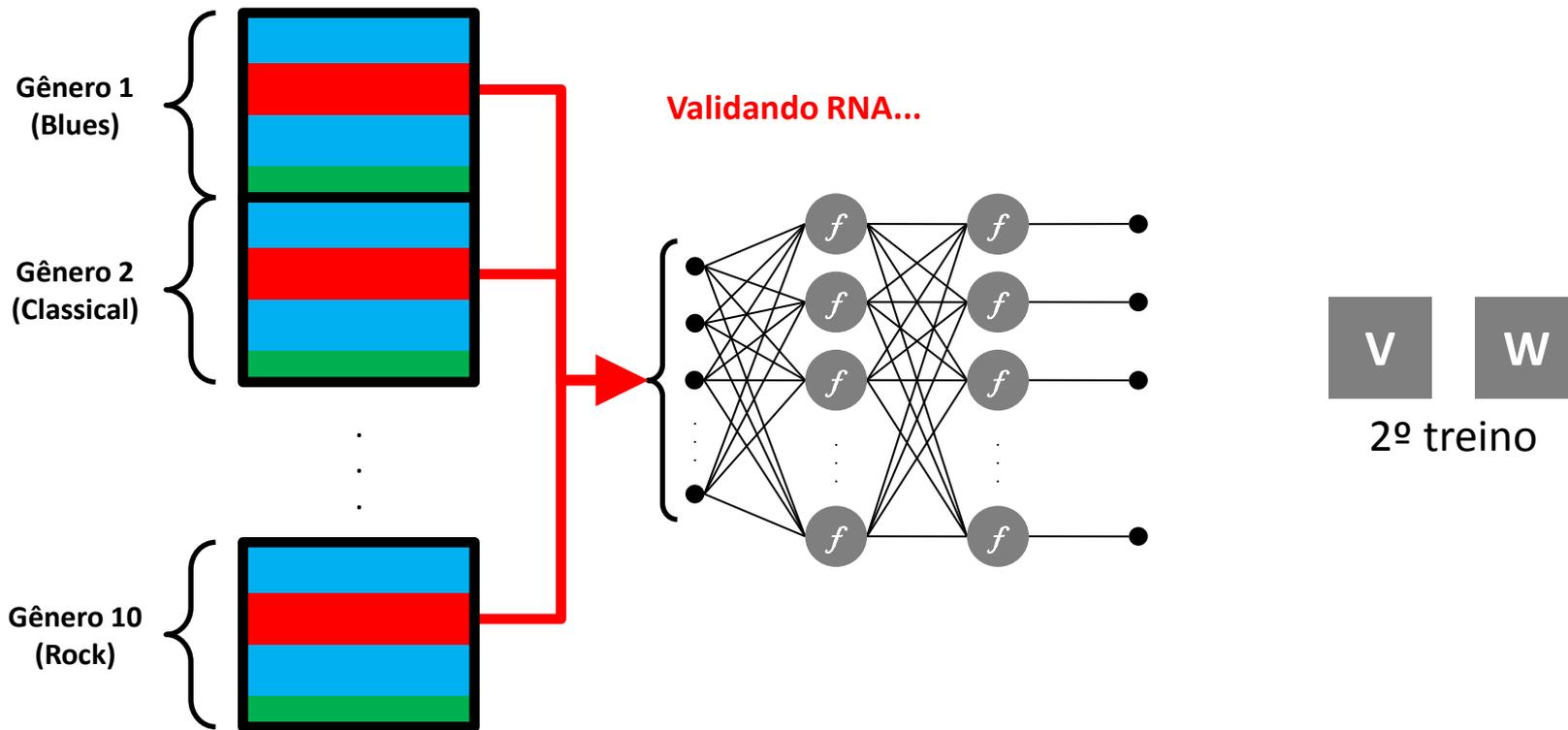
APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”



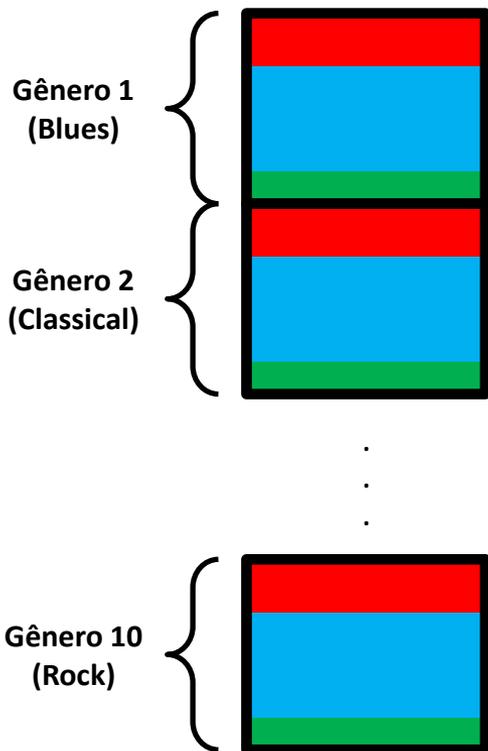
APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”

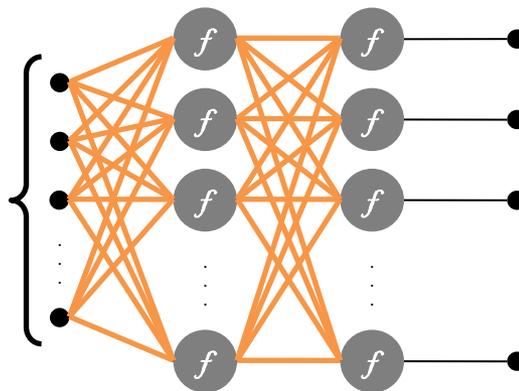


APRENDIZADO

➤ Gradiente Descendente com “Cross-Validation”



Escolhendo o melhor conjunto de pesos...



1º treino



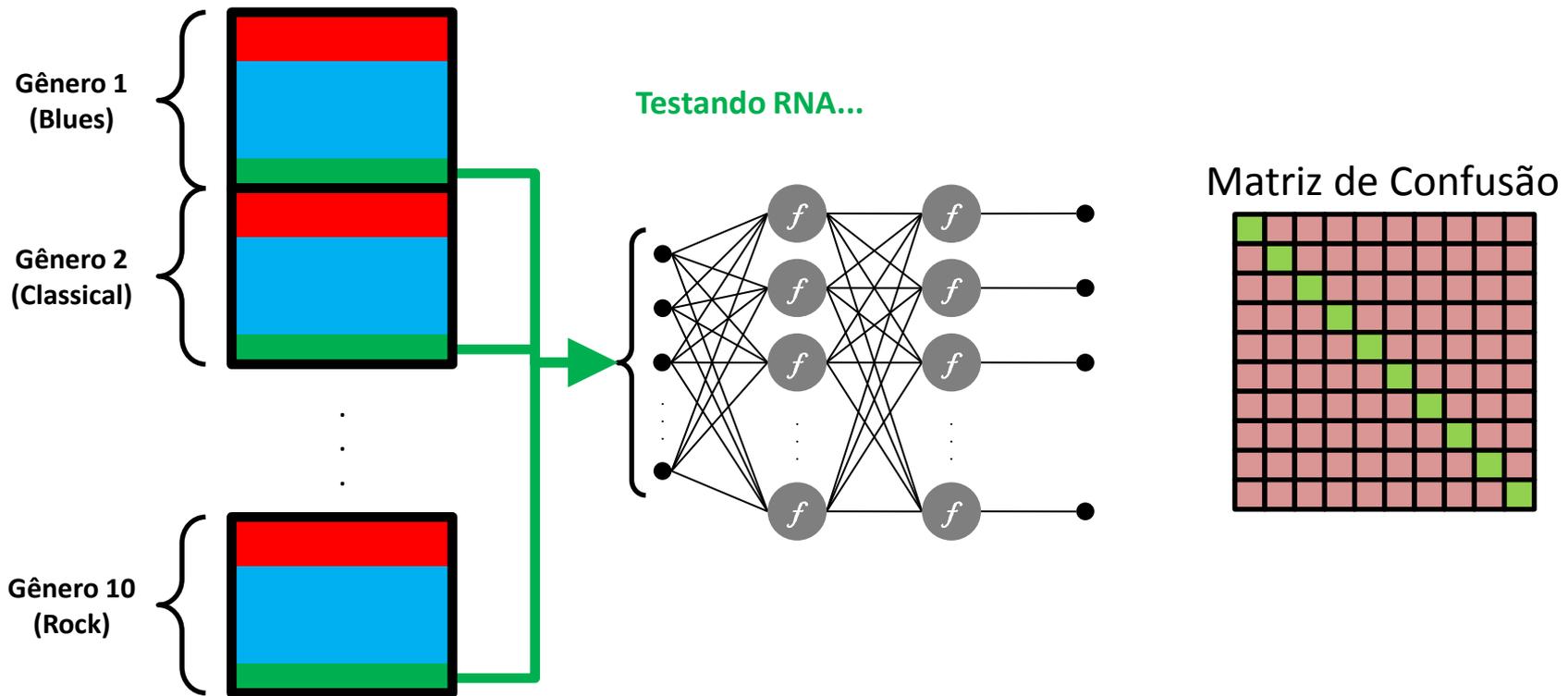
2º treino



3º treino

APRENDIZADO

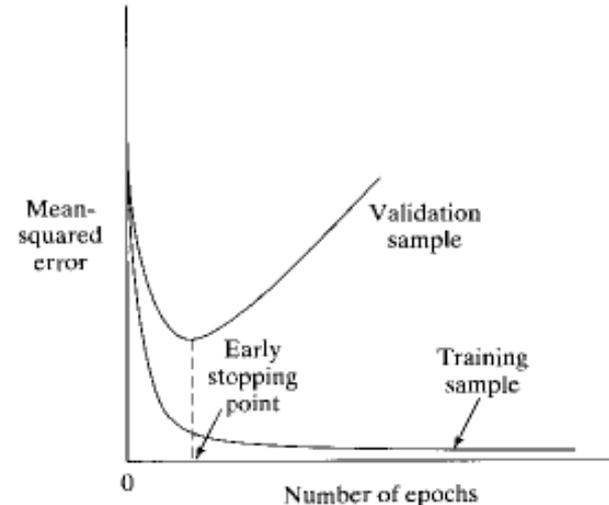
➤ Gradiente Descendente com “Cross-Validation”



APRENDIZADO

➤ Gradiente Normalizado com “Early Stopping”

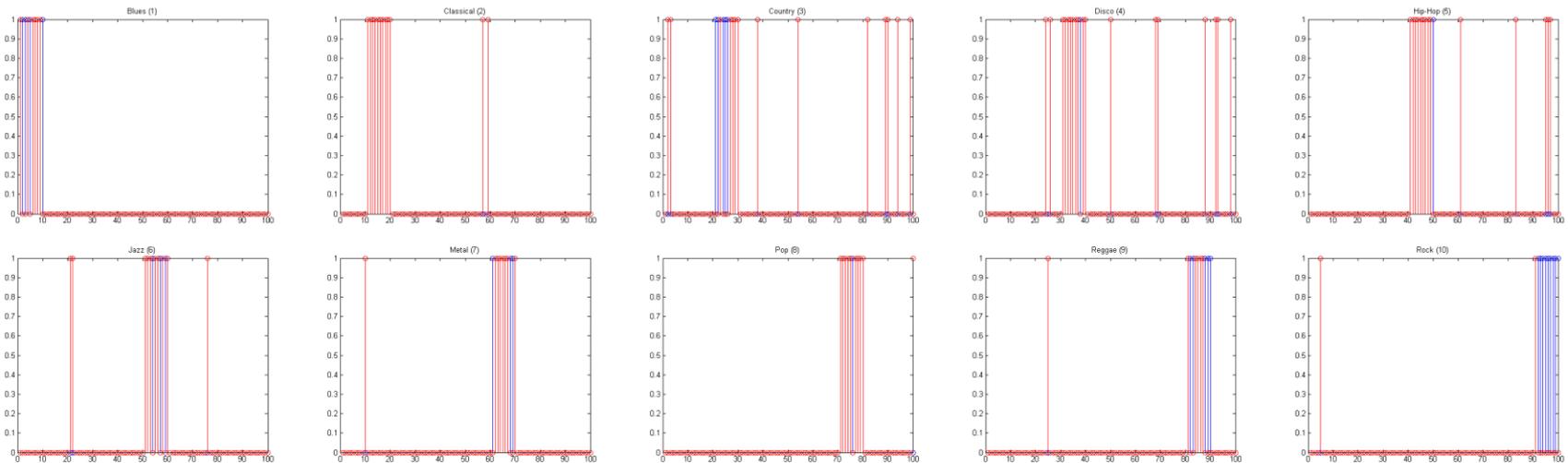
- Algoritmo:
 - “Scaled conjugate gradient” e “early stopping”
- Parâmetros:
 - Número de nós:
 - 20 na camada oculta
 - 10 na camada de saída
 - Regularização:
 - $\lambda = 5e-7$
 - Erros na validação:
 - max_fail = 10
 - Número máximo de iterações:
 - epochs = 50000



Fonte: Haykin [7]

AMOSTRAS DE TESTE x RNA

➤ Gradiente Descendente com “Cross-Validation”



Na figura acima, amostras de teste de cada gênero (azul) comparadas aos resultados obtidos com a RNA treinada (vermelho).

MATRIZES DE CONFUSÃO (TREINO E VALIDAÇÃO)

➤ Gradiente Descendente com “Cross-Validation”

Training Confusion Matrix

1	72 9.6%	0 0.0%	3 0.4%	1 0.1%	1 0.1%	1 0.1%	2 0.3%	0 0.0%	2 0.3%	0 0.0%	87.8% 12.2%
2	0 0.0%	74 9.9%	0 0.0%	0 0.0%	1 0.1%	4 0.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	93.7% 6.3%
3	0 0.0%	0 0.0%	44 5.9%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	1 0.1%	1 0.1%	91.7% 8.3%
4	0 0.0%	0 0.0%	1 0.1%	48 6.4%	1 0.1%	0 0.0%	1 0.1%	0 0.0%	1 0.1%	0 0.0%	92.3% 7.7%
5	0 0.0%	0 0.0%	0 0.0%	2 0.3%	52 6.9%	0 0.0%	0 0.0%	1 0.1%	6 0.8%	0 0.0%	85.2% 14.8%
6	0 0.0%	0 0.0%	6 0.8%	1 0.1%	0 0.0%	68 9.1%	0 0.0%	0 0.0%	2 0.3%	0 0.0%	88.3% 11.7%
7	1 0.1%	0 0.0%	3 0.4%	4 0.5%	3 0.4%	0 0.0%	71 9.5%	0 0.0%	0 0.0%	2 0.3%	84.5% 15.5%
8	0 0.0%	1 0.1%	5 0.7%	7 0.9%	7 0.9%	0 0.0%	0 0.0%	72 9.6%	7 0.9%	4 0.5%	69.9% 30.1%
9	0 0.0%	0 0.0%	2 0.3%	0 0.0%	7 0.9%	0 0.0%	0 0.0%	0 0.0%	48 6.4%	0 0.0%	84.2% 15.8%
10	2 0.3%	0 0.0%	11 1.5%	11 1.5%	3 0.4%	2 0.3%	1 0.1%	1 0.1%	8 1.1%	68 9.1%	63.6% 36.4%
	96.0% 4.0%	98.7% 1.3%	58.7% 41.3%	64.0% 36.0%	69.3% 30.7%	90.7% 9.3%	94.7% 5.3%	96.0% 4.0%	64.0% 36.0%	90.7% 9.3%	82.3% 17.7%
	1	2	3	4	5	6	7	8	9	10	

Validation Confusion Matrix

1	13 8.7%	0 0.0%	1 0.7%	0 0.0%	1 0.7%	86.7% 13.3%						
2	0 0.0%	14 9.3%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	93.3% 6.7%
3	1 0.7%	0 0.0%	11 7.3%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	1 0.7%	0 0.0%	1 0.7%	2 1.3%	57.9% 42.1%
4	0 0.0%	0 0.0%	1 0.7%	12 8.0%	0 0.0%	0 0.0%	2 1.3%	1 0.7%	0 0.0%	0 0.0%	3 2.0%	63.2% 36.8%
5	1 0.7%	0 0.0%	1 0.7%	0 0.0%	12 8.0%	0 0.0%	1 0.7%	1 0.7%	1 0.7%	1 0.7%	0 0.0%	70.6% 29.4%
6	0 0.0%	1 0.7%	0 0.0%	0 0.0%	0 0.0%	12 8.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	85.7% 14.3%
7	0 0.0%	0 0.0%	0 0.0%	1 0.7%	1 0.7%	0 0.0%	11 7.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	84.6% 15.4%
8	0 0.0%	0 0.0%	0 0.0%	1 0.7%	1 0.7%	0 0.0%	0 0.0%	12 8.0%	1 0.7%	1 0.7%	1 0.7%	75.0% 25.0%
9	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	3 2.0%	0 0.0%	0 0.0%	11 7.3%	0 0.0%	0 0.0%	73.3% 26.7%
10	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	6 4.0%	85.7% 14.3%
	86.7% 13.3%	93.3% 6.7%	73.3% 26.7%	80.0% 20.0%	80.0% 20.0%	80.0% 20.0%	73.3% 26.7%	80.0% 20.0%	73.3% 26.7%	40.0% 60.0%	76.0% 24.0%	
	1	2	3	4	5	6	7	8	9	10		

MATRIZES DE CONFUSÃO (TREINO E VALIDAÇÃO)

➤ Gradiente Normalizado com “Early Stopping”

Conjunto de treino Confusion Matrix

1	72 9.6%	0 0.0%	1 0.1%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	5 0.7%	90.0% 10.0%
2	0 0.0%	69 9.2%	0 0.0%	1 0.1%	0 0.0%	2 0.3%	0 0.0%	0 0.0%	2 0.3%	0 0.0%	93.2% 6.8%
3	0 0.0%	2 0.3%	63 8.4%	2 0.3%	0 0.0%	8 1.1%	1 0.1%	0 0.0%	1 0.1%	7 0.9%	75.0% 25.0%
4	1 0.1%	0 0.0%	3 0.4%	66 8.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	3 0.4%	89.2% 10.8%
5	0 0.0%	1 0.1%	0 0.0%	4 0.5%	69 9.2%	1 0.1%	0 0.0%	2 0.3%	1 0.1%	2 0.3%	86.3% 13.7%
6	1 0.1%	3 0.4%	3 0.4%	0 0.0%	1 0.1%	61 8.1%	0 0.0%	3 0.4%	3 0.4%	3 0.4%	78.2% 21.8%
7	1 0.1%	0 0.0%	1 0.1%	1 0.1%	2 0.3%	0 0.0%	73 9.7%	0 0.0%	0 0.0%	4 0.5%	89.0% 11.0%
8	0 0.0%	0 0.0%	2 0.3%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	69 9.2%	1 0.1%	3 0.4%	90.8% 9.2%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.1%	2 0.3%	1 0.1%	1 0.1%	63 8.4%	3 0.4%	88.7% 11.3%
10	0 0.0%	0 0.0%	2 0.3%	1 0.1%	0 0.0%	1 0.1%	0 0.0%	0 0.0%	2 0.3%	45 6.0%	88.2% 11.8%
	96.0% 4.0%	92.0% 8.0%	84.0% 16.0%	88.0% 12.0%	92.0% 8.0%	81.3% 18.7%	97.3% 2.7%	92.0% 8.0%	84.0% 16.0%	60.0% 40.0%	86.7% 13.3%
	1	2	3	4	5	6	7	8	9	10	

Conjunto de validação Confusion Matrix

1	14 9.3%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	87.5% 12.5%
2	0 0.0%	12 8.0%	0 0.0%	0 0.0%	0 0.0%	2 1.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	85.7% 14.3%
3	0 0.0%	0 0.0%	11 7.3%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	1 0.7%	1 0.7%	1 0.7%	73.3% 26.7%
4	0 0.0%	0 0.0%	2 1.3%	13 8.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	86.7% 13.3%
5	0 0.0%	0 0.0%	0 0.0%	1 0.7%	10 6.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	71.4% 28.6%
6	1 0.7%	3 2.0%	0 0.0%	0 0.0%	0 0.0%	11 7.3%	0 0.0%	0 0.0%	0 0.0%	1 0.7%	68.8% 31.3%
7	0 0.0%	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	15 10.0%	0 0.0%	0 0.0%	2 1.3%	83.3% 16.7%
8	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	14 9.3%	1 0.7%	0 0.0%	87.5% 12.5%
9	0 0.0%	0 0.0%	0 0.0%	0 0.0%	4 2.7%	1 0.7%	0 0.0%	0 0.0%	10 6.7%	3 2.0%	55.6% 44.4%
10	0 0.0%	0 0.0%	1 0.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 1.3%	5 3.3%	62.5% 37.5%
	93.3% 6.7%	80.0% 20.0%	73.3% 26.7%	86.7% 13.3%	66.7% 33.3%	73.3% 26.7%	100% 0.0%	93.3% 6.7%	66.7% 33.3%	33.3% 66.7%	76.7% 23.3%
	1	2	3	4	5	6	7	8	9	10	

MATRIZES DE CONFUSÃO (TESTE)

➤ Gradiente Normalizado com “Early Stopping”

Conjunto de teste Confusion Matrix

1	9 9.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	2 2.0%	75.0% 25.0%
2	0 0.0%	10 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	90.9% 9.1%
3	1 1.0%	0 0.0%	6 6.0%	0 0.0%	0 0.0%	1 1.0%	0 0.0%	1 1.0%	0 0.0%	4 4.0%	46.2% 53.8%
4	0 0.0%	0 0.0%	0 0.0%	8 8.0%	0 0.0%	0 0.0%	1 1.0%	0 0.0%	1 1.0%	0 0.0%	80.0% 20.0%
5	0 0.0%	0 0.0%	0 0.0%	2 2.0%	9 9.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	81.8% 18.2%
6	0 0.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	8 8.0%	0 0.0%	0 0.0%	1 1.0%	0 0.0%	80.0% 20.0%
7	0 0.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	9 9.0%	0 0.0%	1 1.0%	0 0.0%	81.8% 18.2%
8	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	8 8.0%	0 0.0%	0 0.0%	100% 0.0%
9	0 0.0%	0 0.0%	1 1.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	7 7.0%	1 1.0%	70.0% 30.0%
10	0 0.0%	0 0.0%	1 1.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	3 3.0%	75.0% 25.0%
	90.0% 10.0%	100% 0.0%	60.0% 40.0%	80.0% 20.0%	90.0% 10.0%	80.0% 20.0%	90.0% 10.0%	80.0% 20.0%	70.0% 30.0%	30.0% 70.0%	77.0% 23.0%
	1	2	3	4	5	6	7	8	9	10	

Output Class

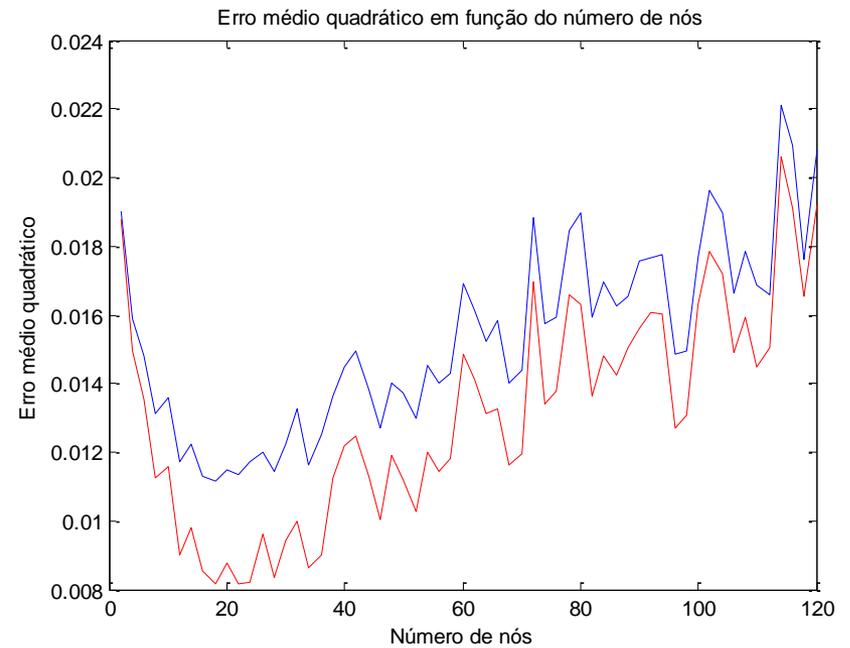
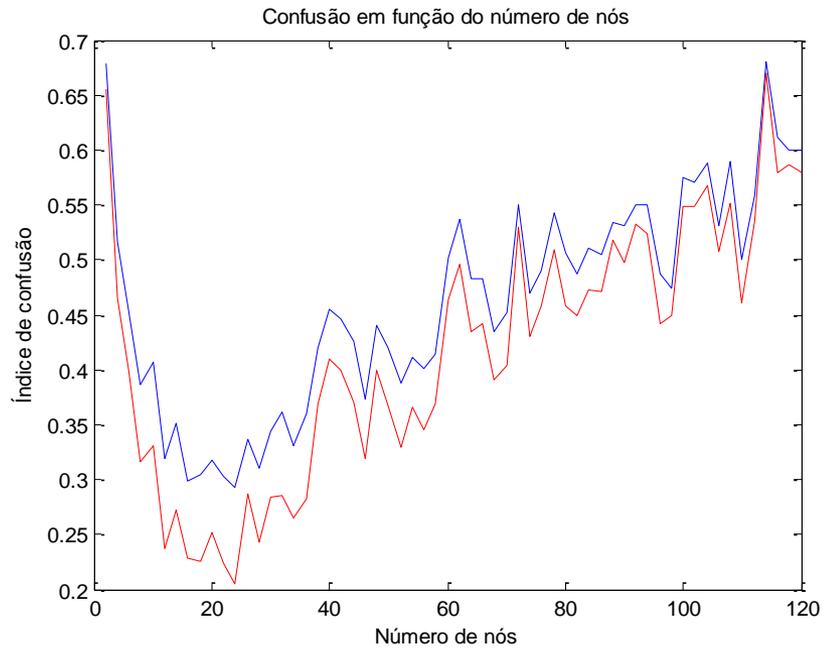
Target Class

COMPARAÇÃO COM A LITERATURA

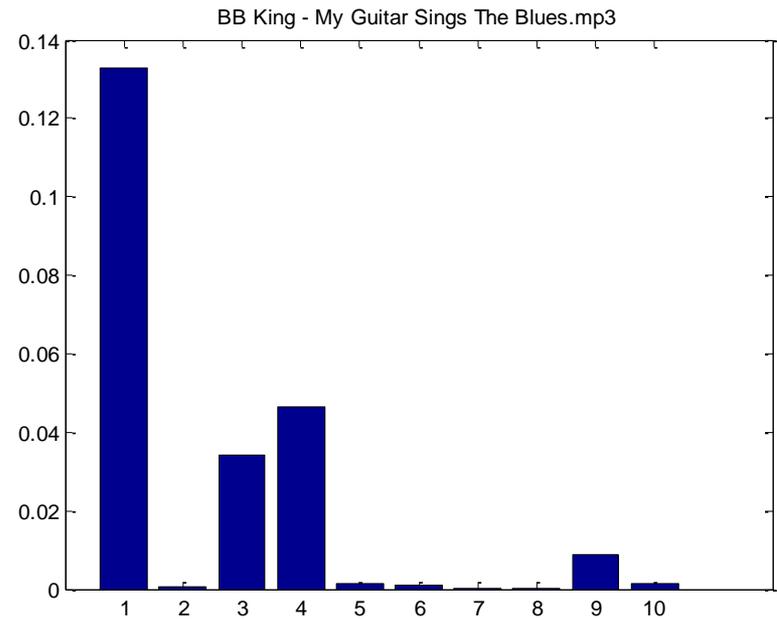
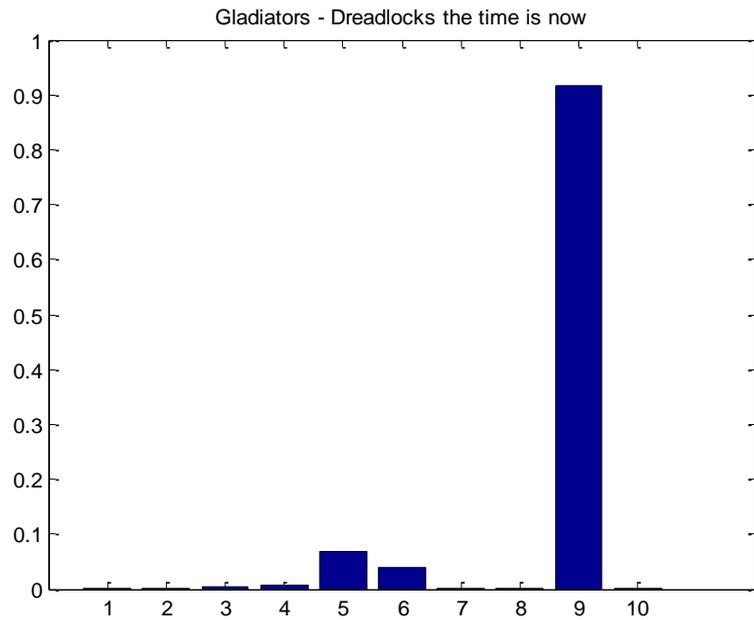
Referência	Base de Dados	Precisão
Bergstra	GTZAN	82.50%
Li	GTZAN	78.50%
Panagakis	GTZAN	78.20%
Britto, Chamon ¹	GTZAN	77.00%
Lidy	GTZAN	76.80%
Benetos	GTZAN	75.00%
Holzapfel	GTZAN	74.00%
Britto, Chamon ²	GTZAN	68.00%
Tzanetakis	GTZAN	61.00%

Fonte: Panagakis *et al.* [5]

CONFUSÃO x NÚMERO DE NÓS



DEMONSTRAÇÃO



CONCLUSÃO E PERSPECTIVAS

O problema de classificação em gêneros musicais é não-trivial, devido à indefinição e subjetividade destes.

É fundamental para um bom desempenho da solução que os conjuntos de treinamento, validação e teste sejam adequados, assim como a técnica de aprendizagem adotada.

Os resultados talvez pudessem ser melhorados com a inclusão de novos “features” (e.g. ritmo) e uma base de dados mais extensa.



REFERÊNCIAS

- [1] SCARINGELLA, N.; ZOIA, G.; MLYNEK, D. Automatic genre classification of music content: a survey. **Signal Processing Magazine**. Piscataway, v. 23[2], p. 133-141, 2006.
- [2] LI, D.; SETHI, I. K.; DIMITROVA, N.; MCGEE, T. Classification of general audio data for content-based retrieval. **Pattern Recognition Letters**. Amsterdam, v. 22, p. 533-544, 2001.
- [3] JIANG, D.-N.; LU, L.; ZHANG, H.-J.; TAO, J.-H.; CAI, L.-H. Music type classification by spectral contrast feature. In: **IEEE ICME, 2002**, Lausanne. *Anais...* Lausanne: Swiss Federal Institute of Technology, 2002.
- [4] MCKINNEY, M. F.; BREEBAART, J. Features for audio and music classification. In: **ISMIR, 2003**, Baltimore. *Anais...*
- [5] PANAGAKIS, Y.; KOTROPOULOS, C.; ARCE, G. R. Music genre classification via sparse representations of auditory temporal modulations. In: **EUSIPCO, 2009**, Glasgow. *Anais...*
- [6] TZANETAKIS, G.; COOK, P. **GTZAN Genre Collection**. Disponível em: http://marsyas.info/download/data_sets. Acesso em: 7 de jun. 2011.
- [7] HAYKIN, S. **Neural Networks: A comprehensive foundation**. Upper Saddle River: Prentice-Hall, 1999. 842 p.
- [8] REFAEILZADEH, P.; TANG, L.; LIU, H. Cross-Validation. In: ÖSZU, M. T.; LIU, L. **Encyclopedia of Database Systems**. New York: Springer, 2009. Disponível em: <http://www.public.asu.edu/~ltang9/papers/ency-cross-validation.pdf>. Acesso em: 7 de jun. 2011.