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

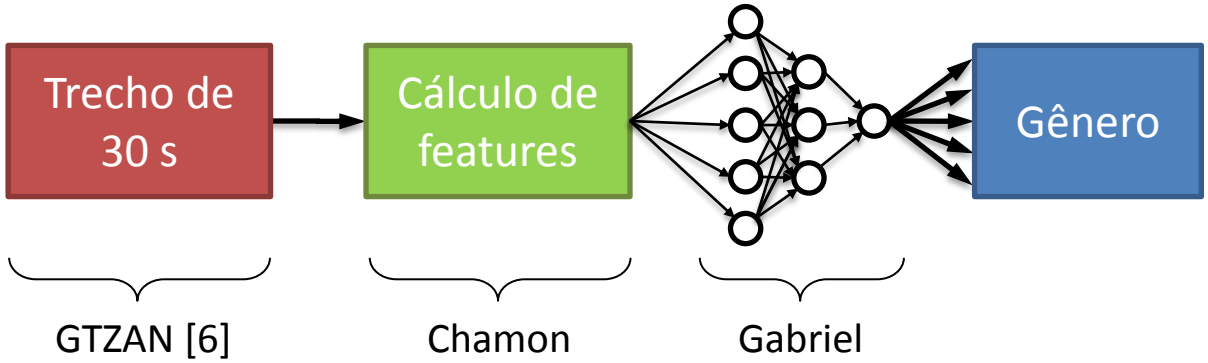
(Amazon: 719 gêneros
iTunes: 244 gêneros)

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.

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

Spectral Centroid [2]

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

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

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]

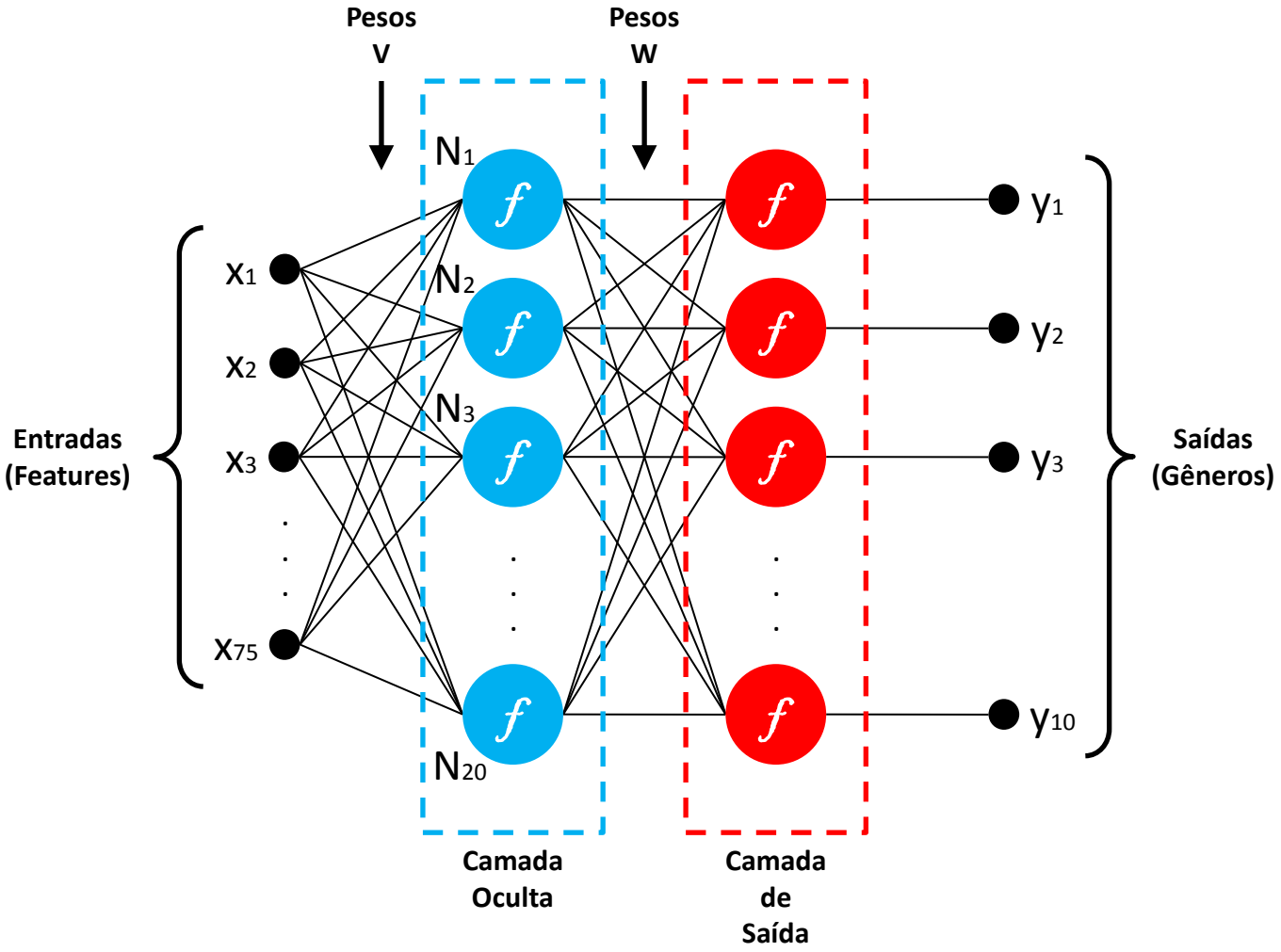
Medida de conteúdo espectral de altas freqüências. Avaliação psicoacústica do quão “afiado”, “agudo” um som é.

ESTRUTURA

Gêneros Musicais Considerados

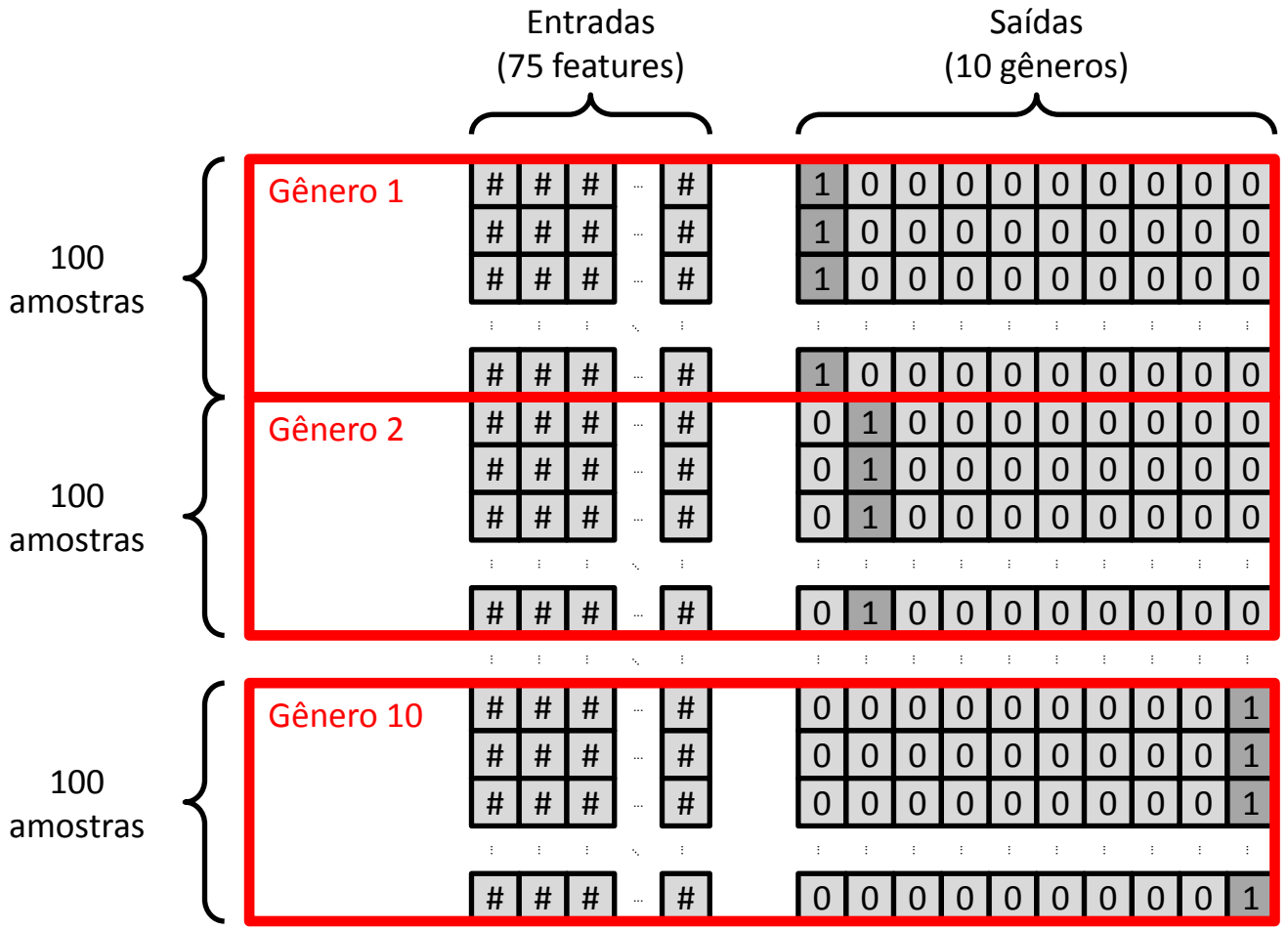
1) Blues
2) Classical
3) Country
4) Disco
5) Hip-Hop
6) Jazz
7) Metal
8) Pop
9) Reggae
10) Rock

ESTRUTURA


Pesos
v
Pesos
w
N1
N2
N3
N20
x1
x2
x3
x75
Entradas
(Features)
y1
y2
y3
y10
Saídas
(Gêneros)
Camada
Oculta
Camada
de
Saída

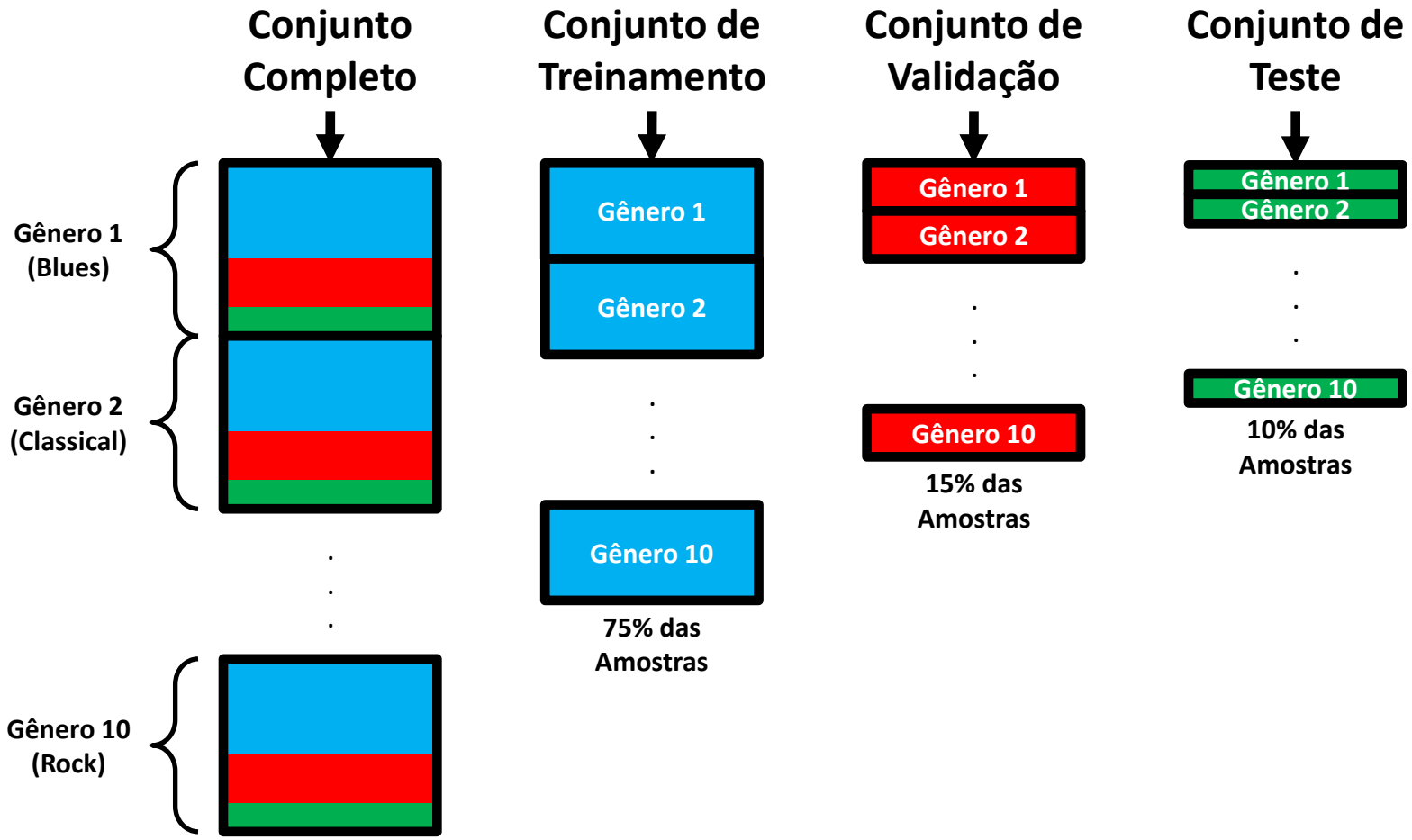
ESTRUTURA

➢ 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”

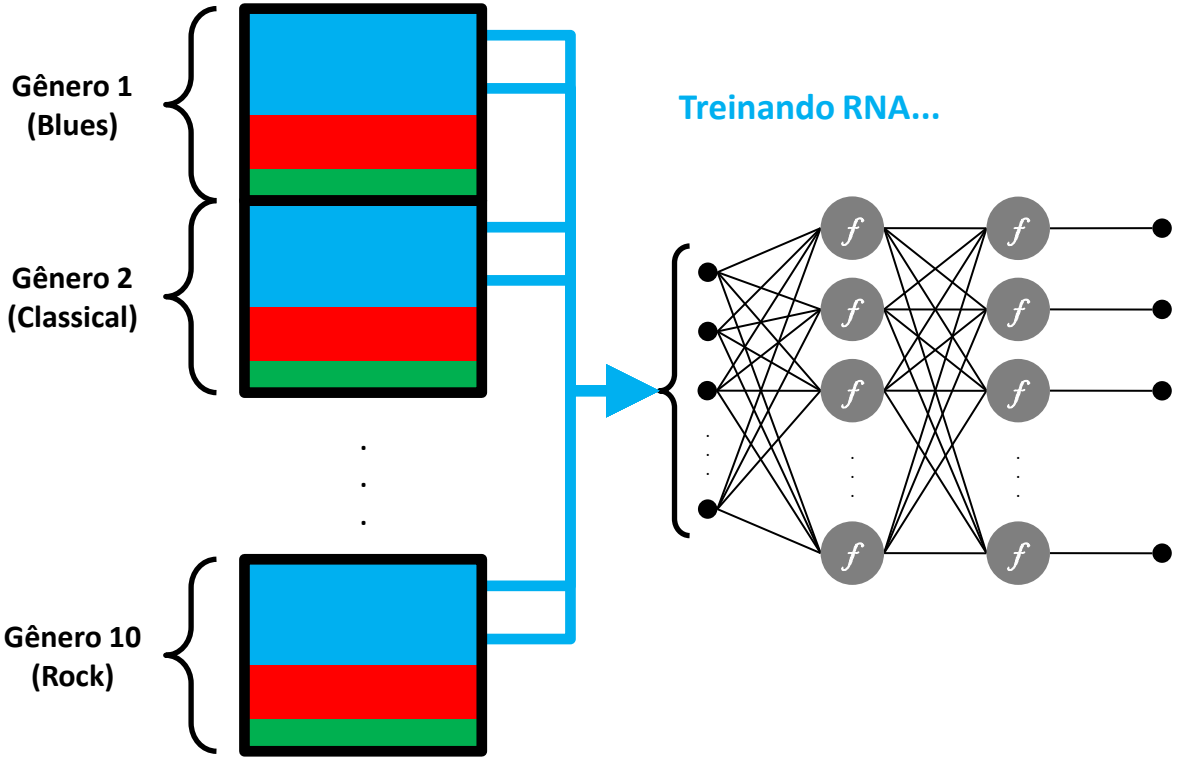
APRENDIZADO

➢ 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”

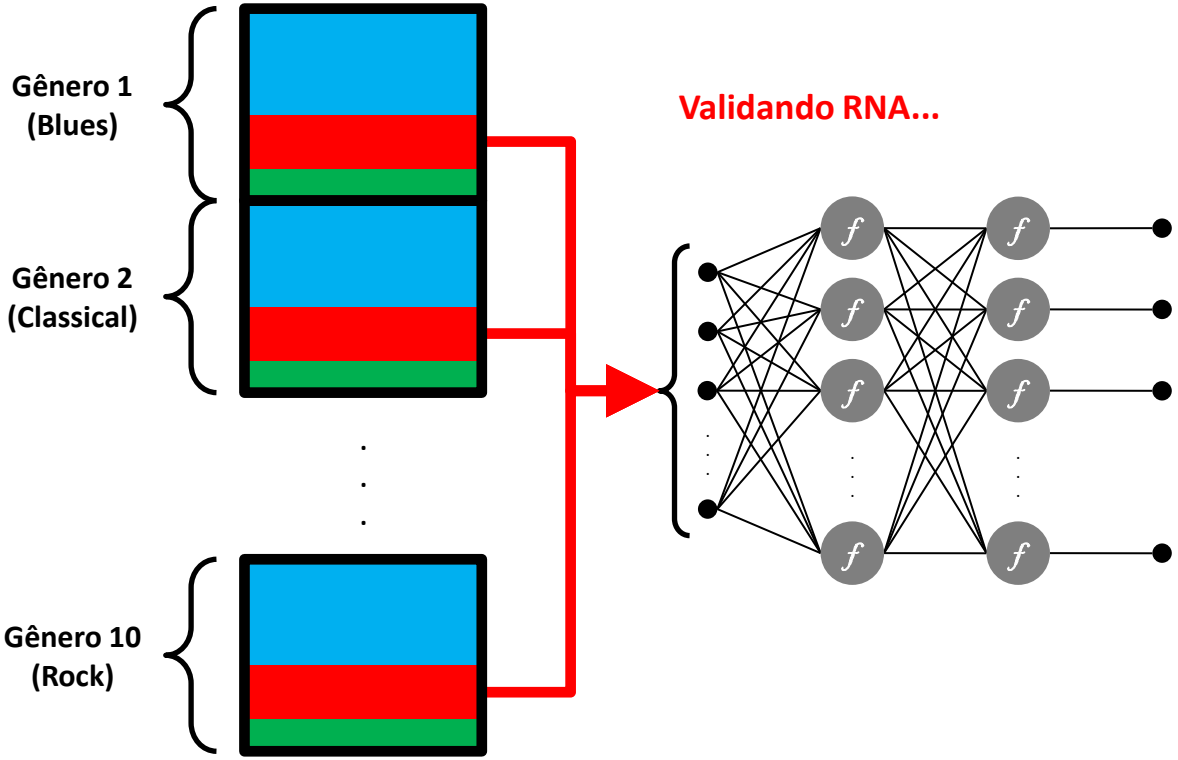


V W

1º treino

APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”

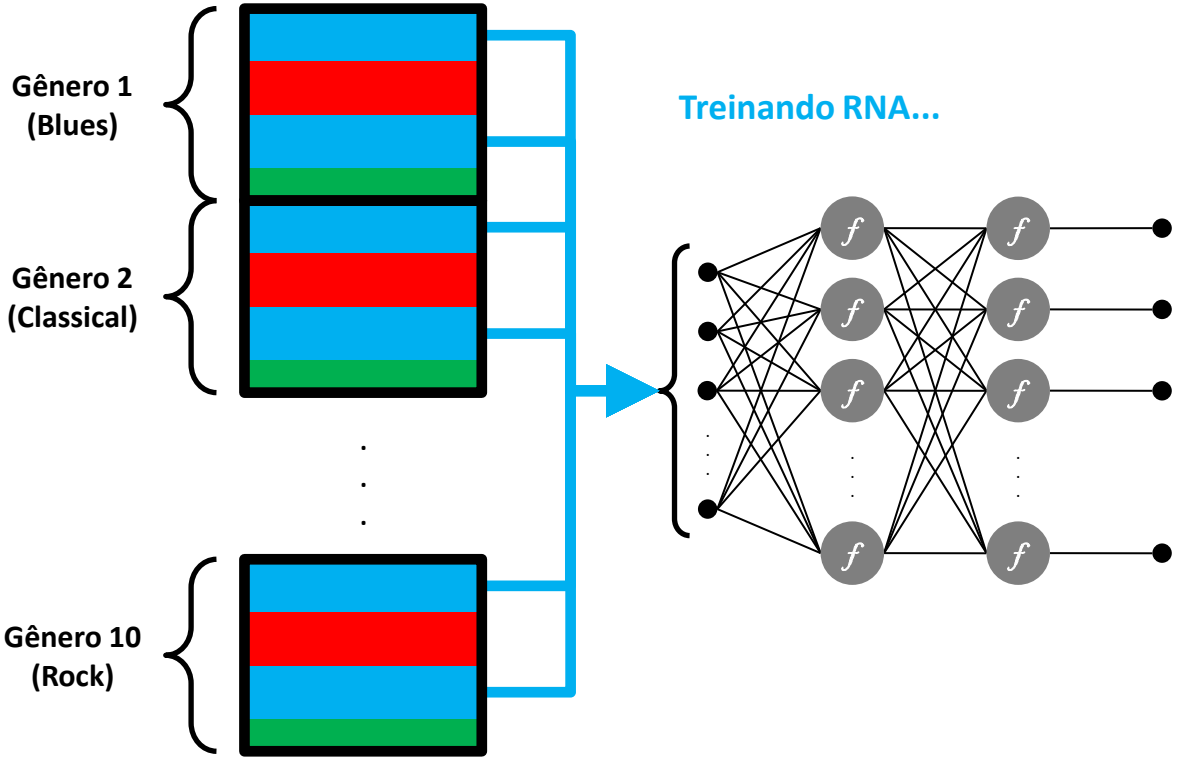


V W

1º treino

APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”

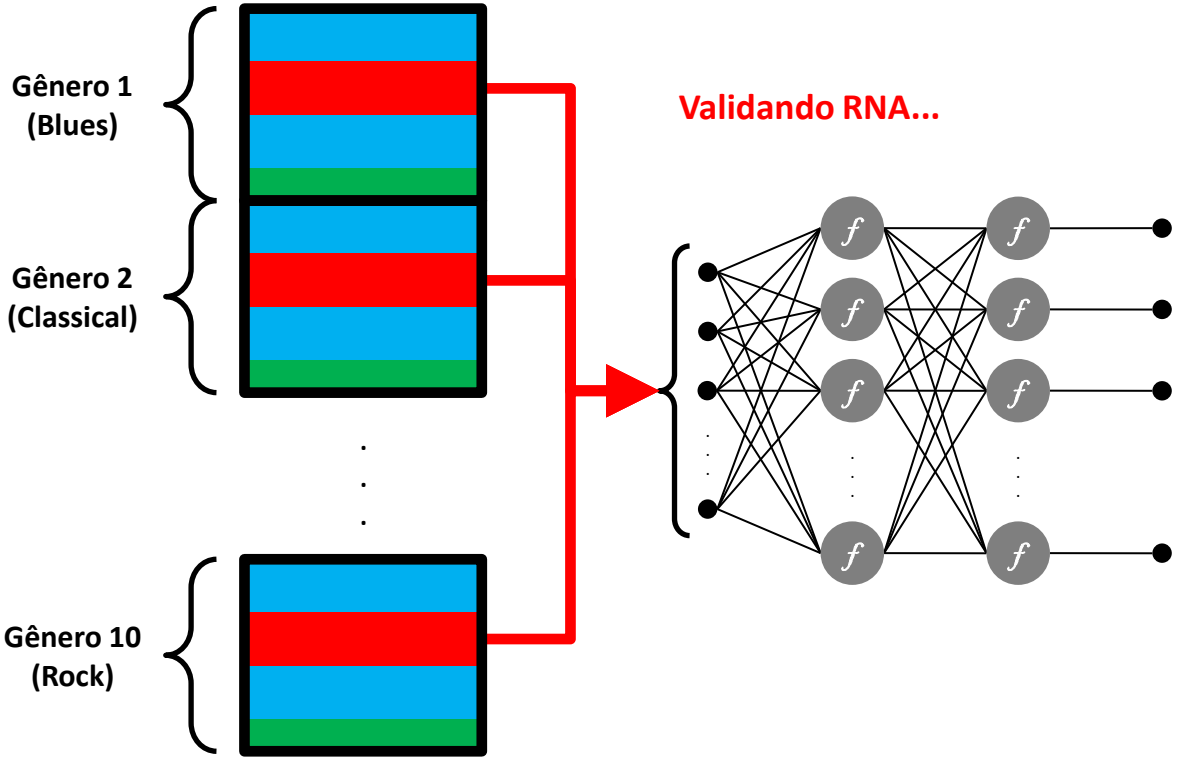


V W

2º treino

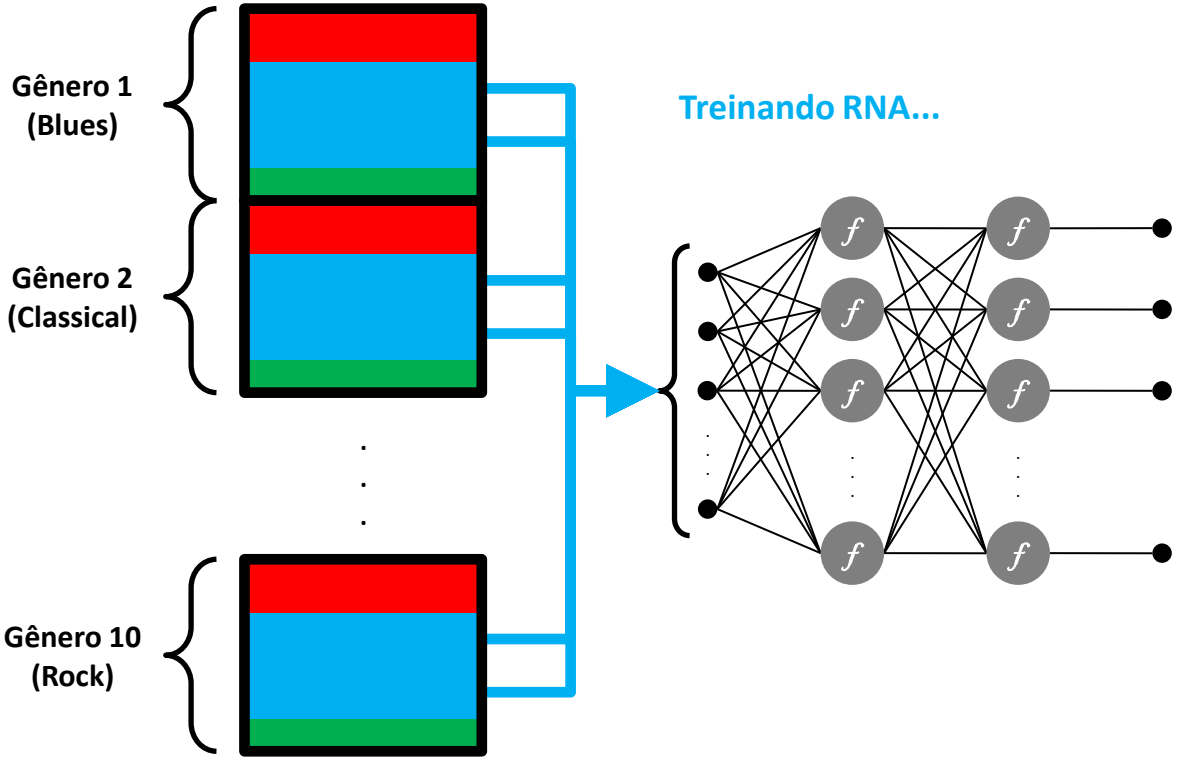
APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”



APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”

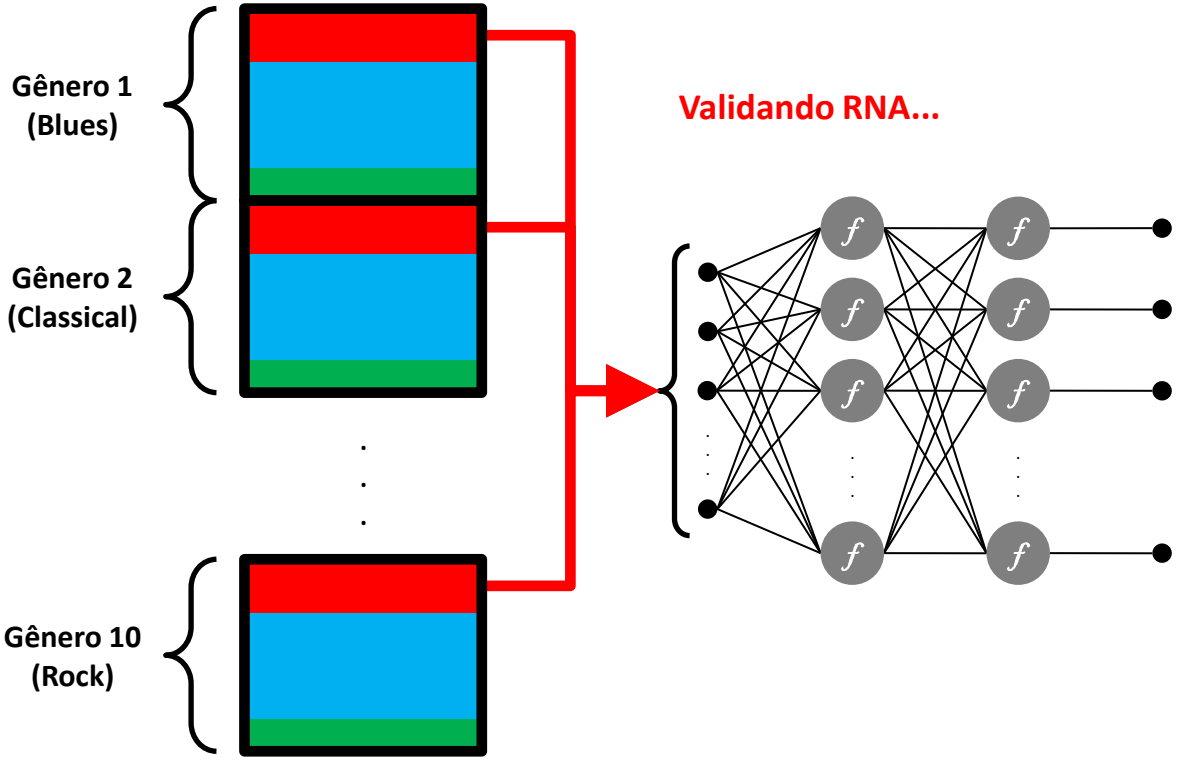


V W

3º treino

APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”

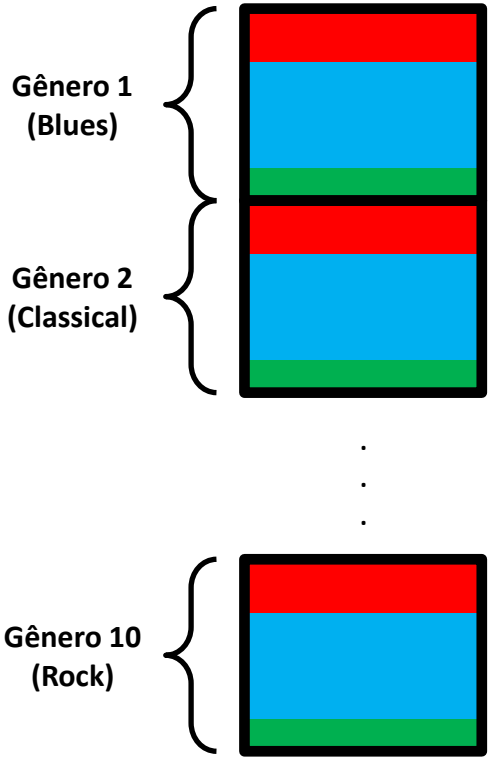


V W

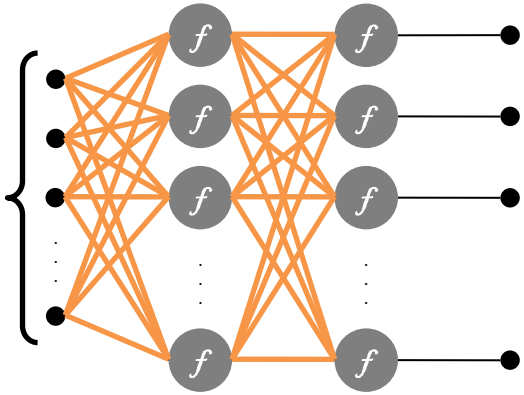
3º treino

APRENDIZADO

➢Gradiente Descendente com “Cross-Validation”

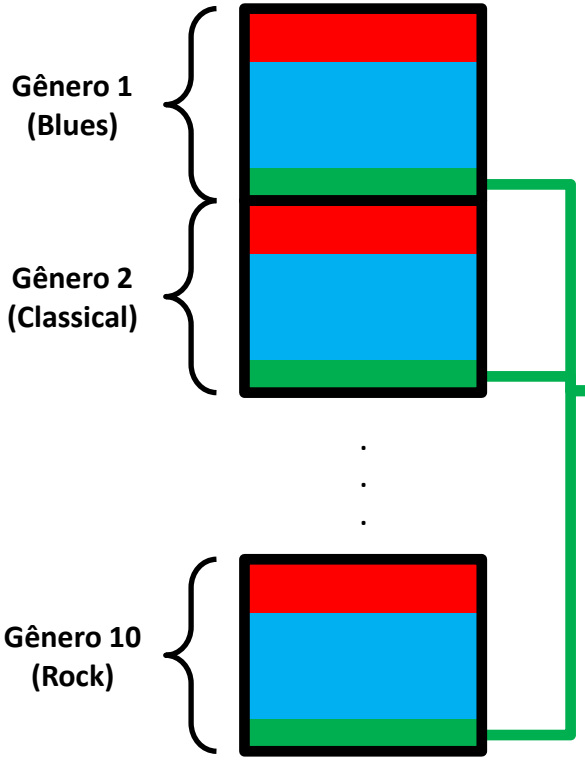


Escolhendo o melhor conjunto de pesos...

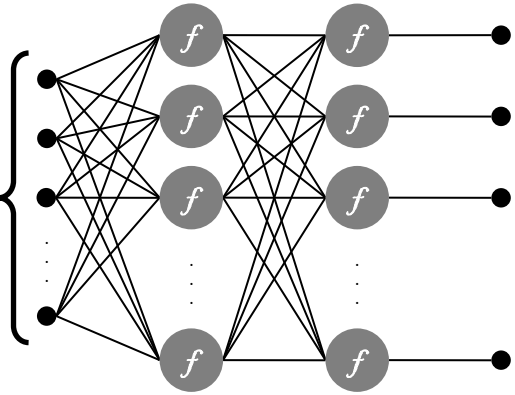


APRENDIZADO

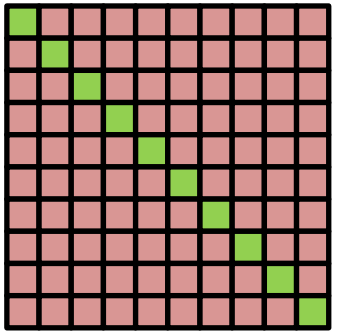
➢Gradiente Descendente com “Cross-Validation”



Testando RNA...



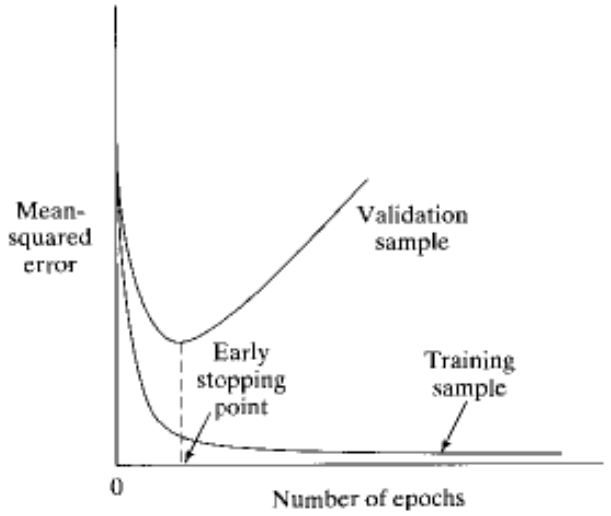
Matriz de Confusão



APRENDIZADO

➢Gradiente Normalizado com “Early Stopping”

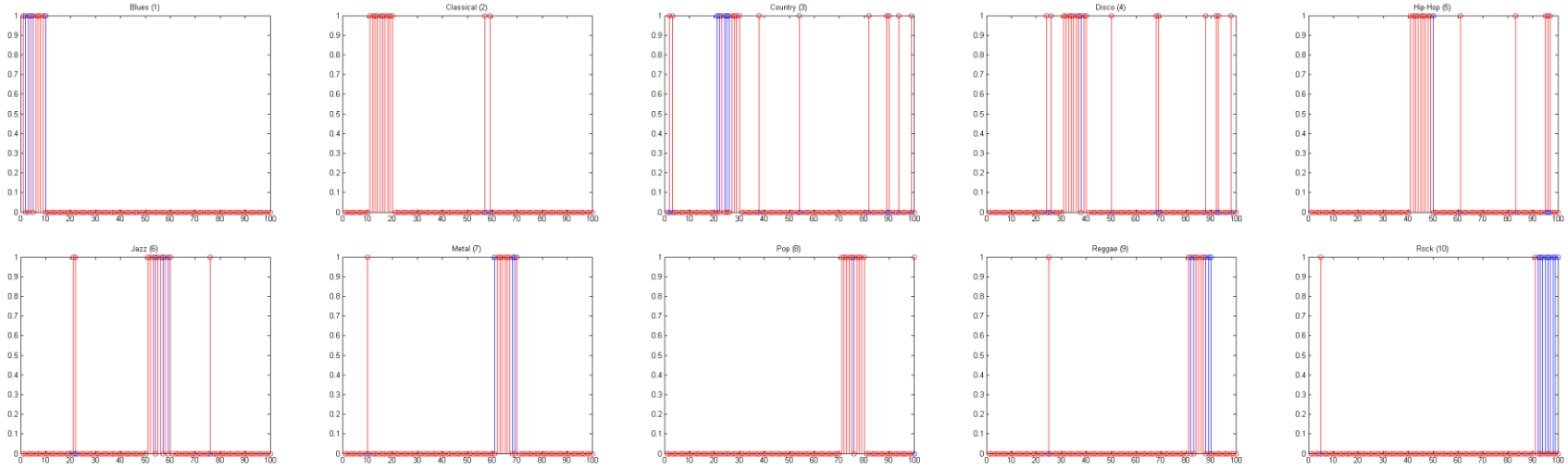
- Algoritmo:
  - “Scaled conjugate gradient” e “early stopping”
- Parâmetros:
  - Número de nós:
    - 20 na camada oculta
    - 10 na camada de saída
  - Regularização:
    - λ = 5e-7
  - Erros na validação:
    - max_fail = 10
  - Número máximo de iterações:
    - epochs = 50000



Fonte: Haykin [7]

AMOSTRAS DE TESTE x RNA

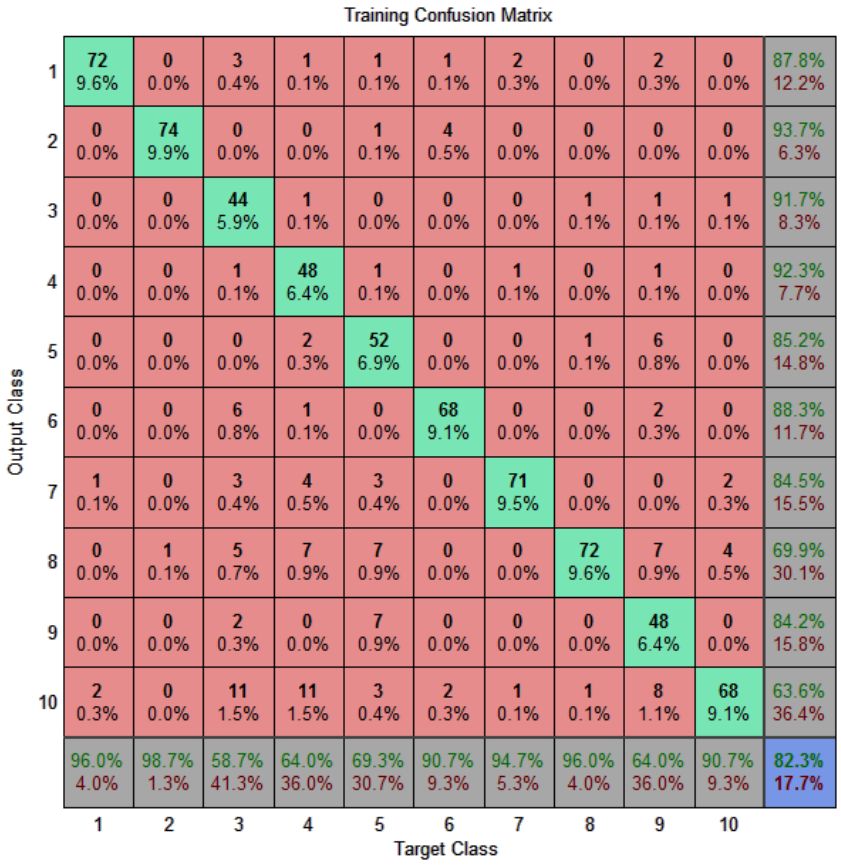
➢Gradiente Descendente com “Cross-Validation”

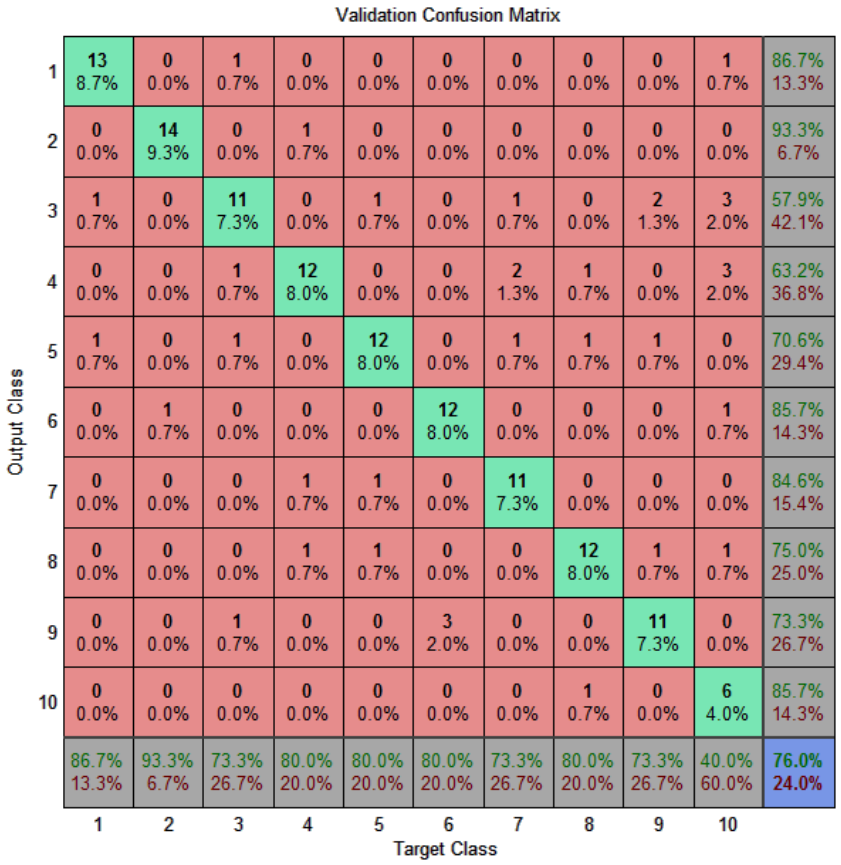


Na figura acima, amostras de teste de cada gênero (azul) comparadas aos resultados obtidos com a RNA treinada (vermelho).

MATRIZES DE CONFUSÃO (TREINO E VALIDAÇÃO)

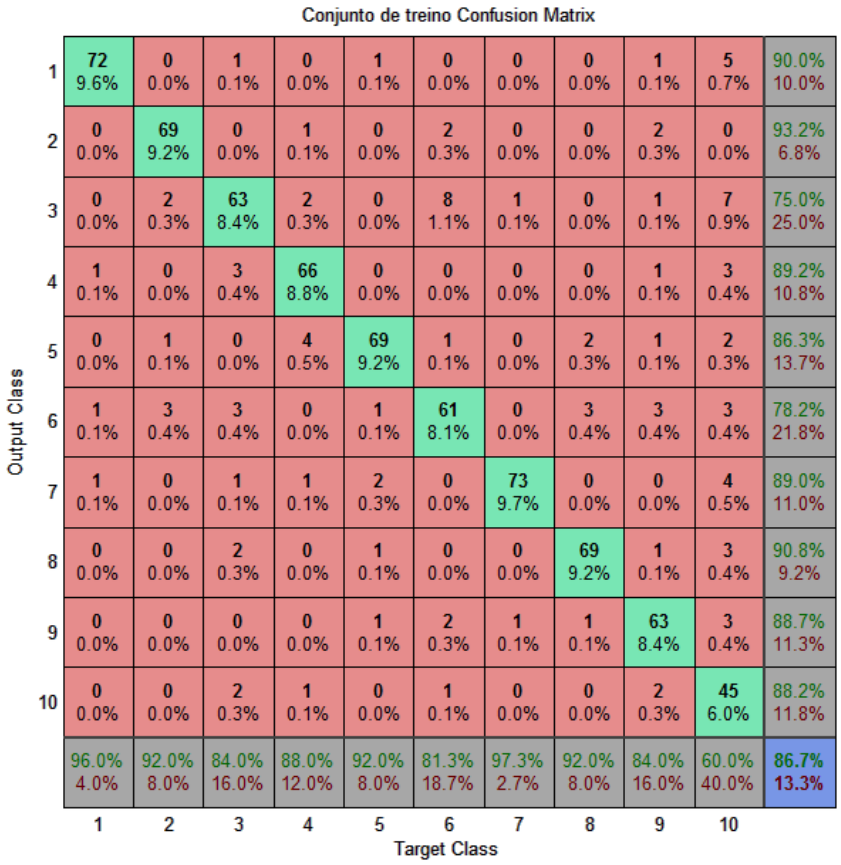
➢Gradiente Descendente com “Cross-Validation”

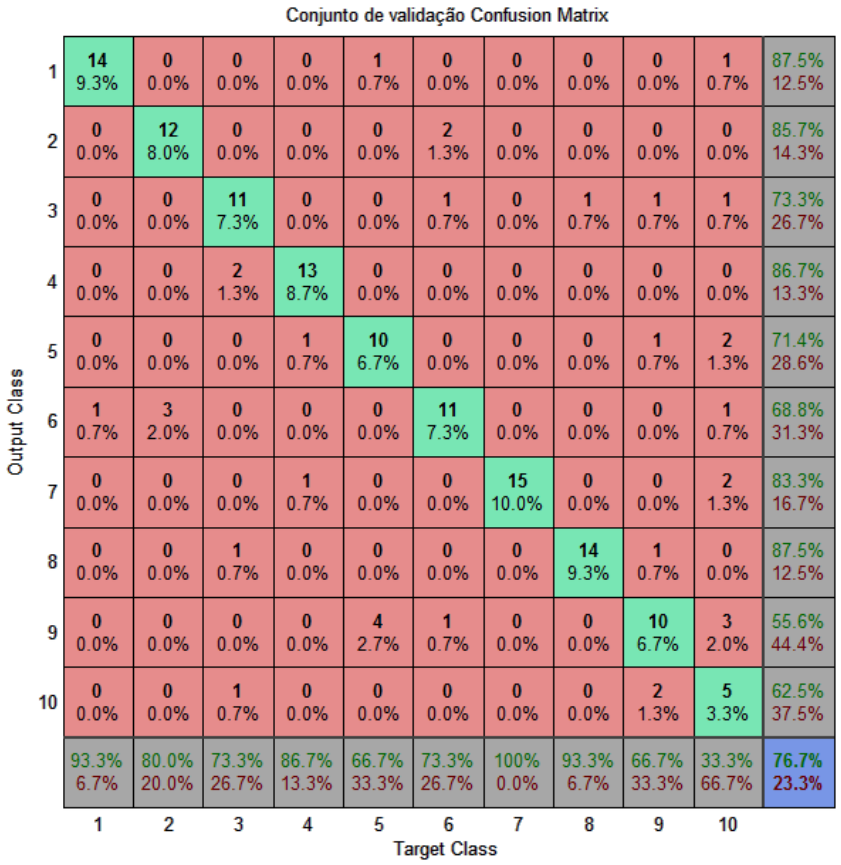




MATRIZES DE CONFUSÃO (TREINO E VALIDAÇÃO)

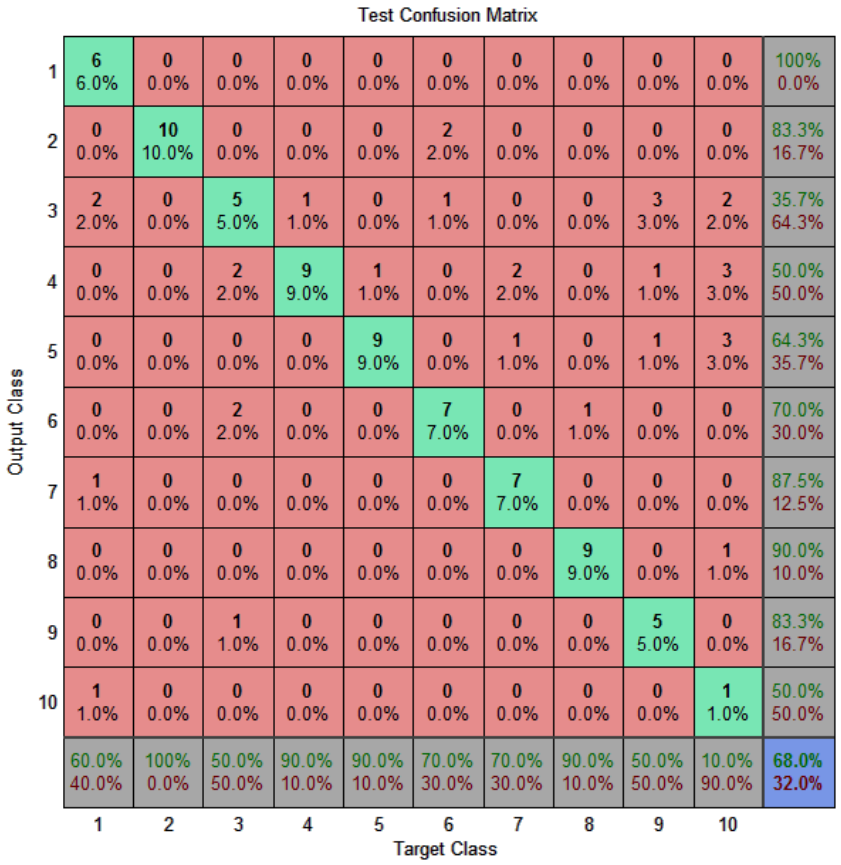
➢Gradiente Normalizado com “Early Stopping”





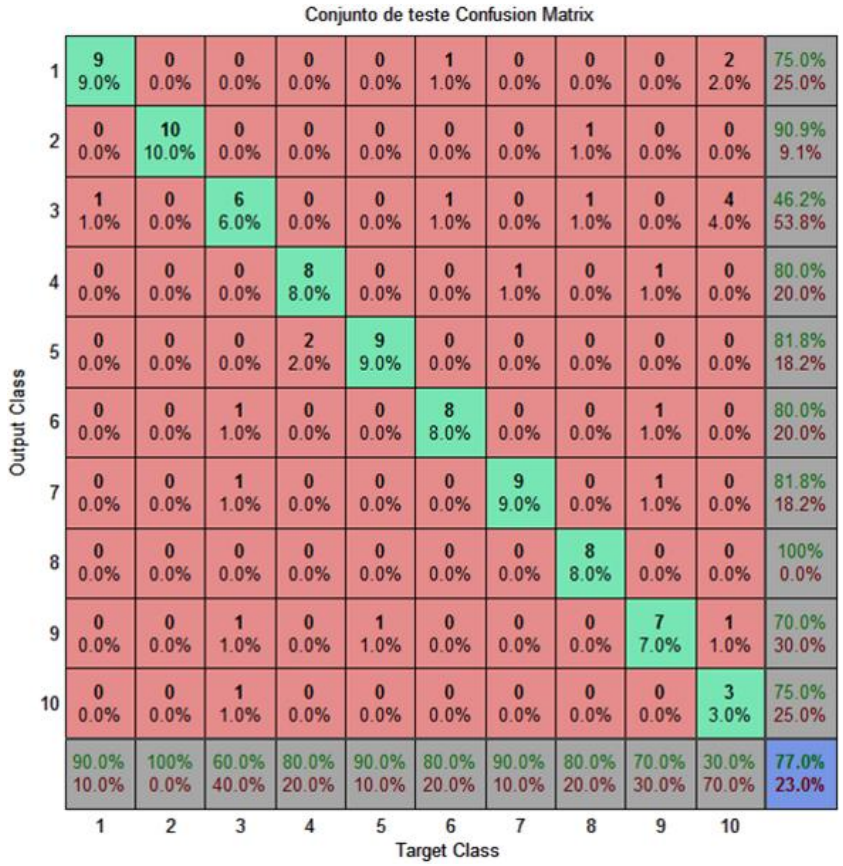
MATRIZES DE CONFUSÃO (TESTE)

➢Gradiente Descendente com “Cross-Validation”



MATRIZES DE CONFUSÃO (TESTE)

➢Gradiente Normalizado com “Early Stopping”

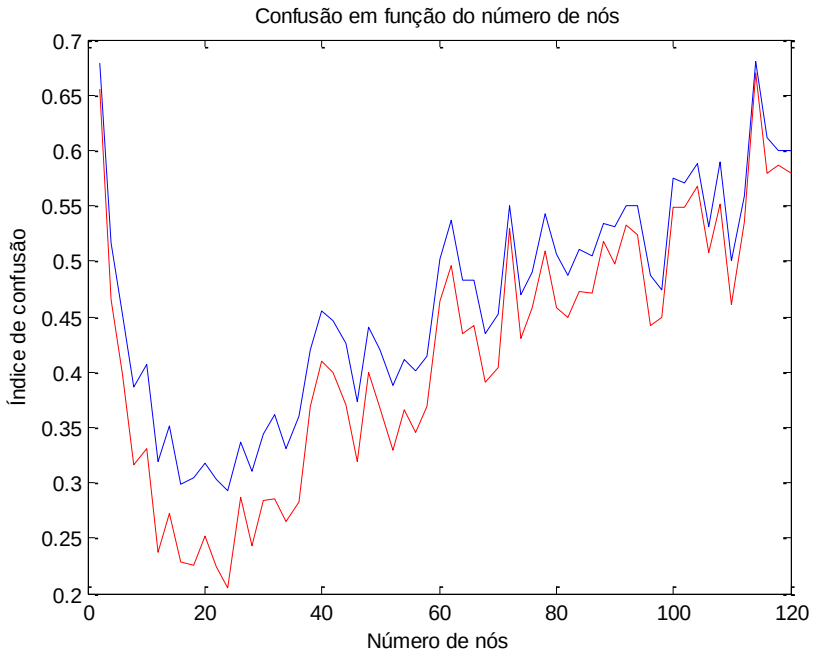


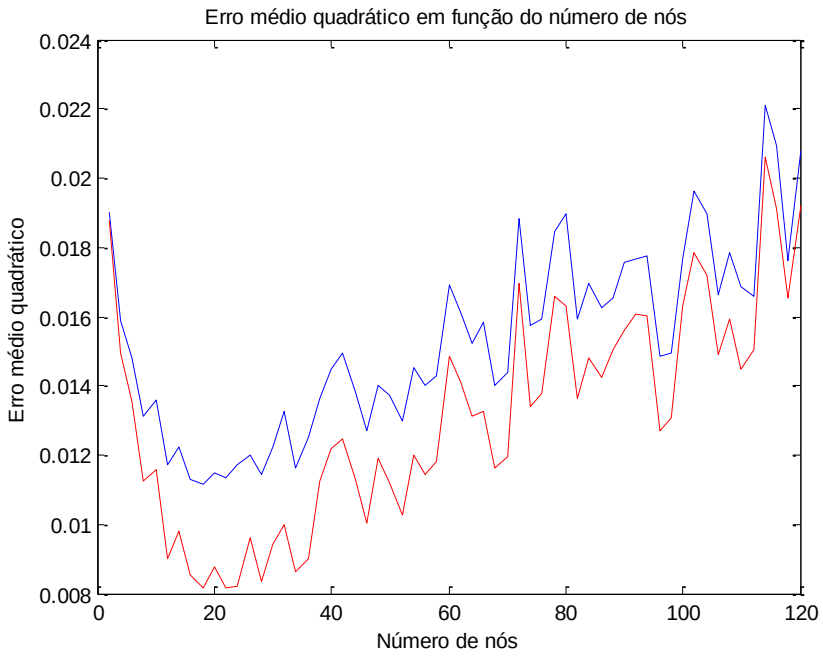
COMPARAÇÃO COM A LITERATURA

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%

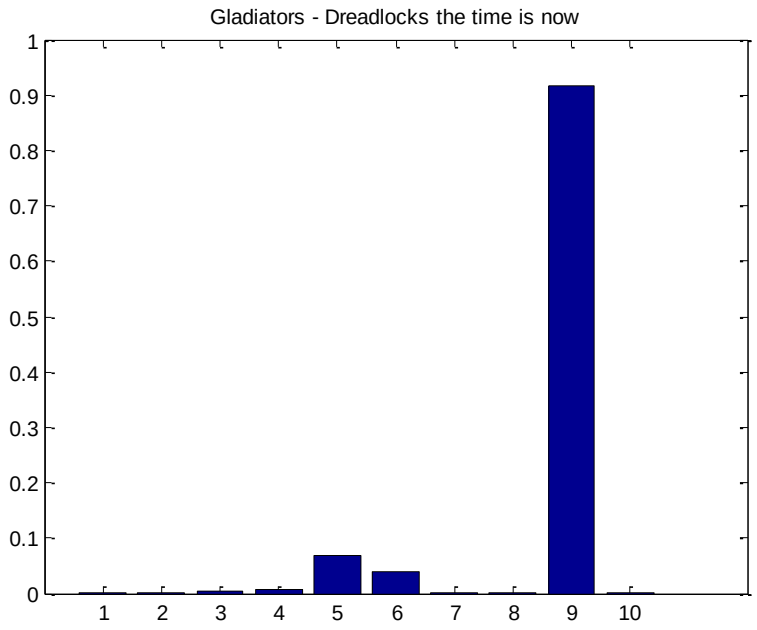
Fonte: Panagakis *et al.* [5]

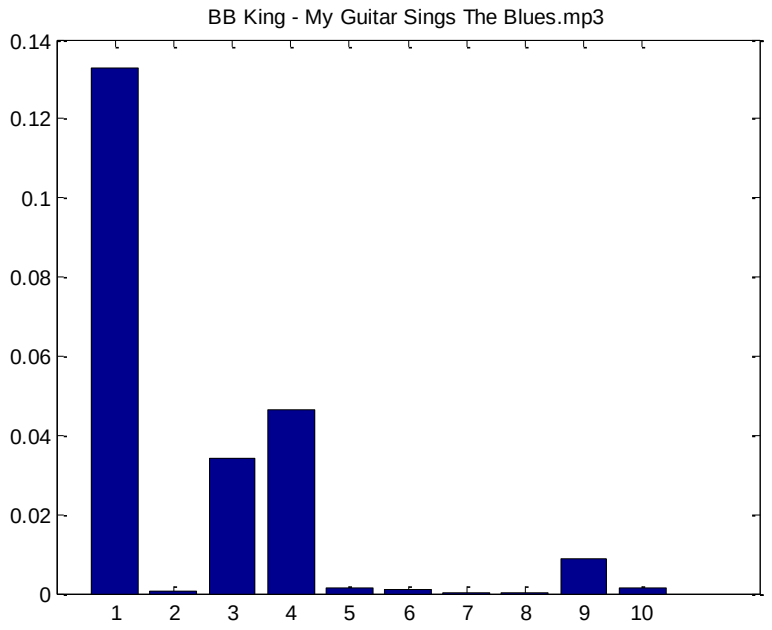
CONFUSÃO x NÚMERO DE NÓS





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 Techonology, 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.