CLASSIFICAÇÃO DE GÊNEROS MUSICAIS utilizando FEATURES GLOBAIS DE ÁUDIO e CLASSIFICADORES TRADICIONAIS

Roberto Piassi Passos Bodo

Instituto de Matemática e Estatística
Universidade de São Paulo

28 de Março de 2017

GÊNEROS MUSICAIS

- termos que identificam recorrências e semelhanças;

- descrevem características intrínsecas e extrínsecas;

- difíceis de serem descritos sistematicamente;

- atribuídos por membros de uma comunidade;

- não existe um consenso de opinião;

GÊNEROS MUSICAIS

- se tornam cada vez mais específicos;

- vão sendo mesclados (*fusions*);

- são de extrema utilidade para tarefas de MIR;

- devem ser atribuídos a bancos de dados enormes;

- serão determinados a partir de classificação automática.

GTZAN: CONJUNTO DE DADOS

- criado por George Tzanetakis e Perry Cook;

- 1000 arquivos de áudio;

- duração: 30 segundos;

- taxa de amostragem: 22050Hz;

- resolução: 16 bits;

- 10 gêneros musicais:

blues	jazz
classical	metal
country	pop
disco	reggae
hiphop	rock

FEATURES DE ÁUDIO

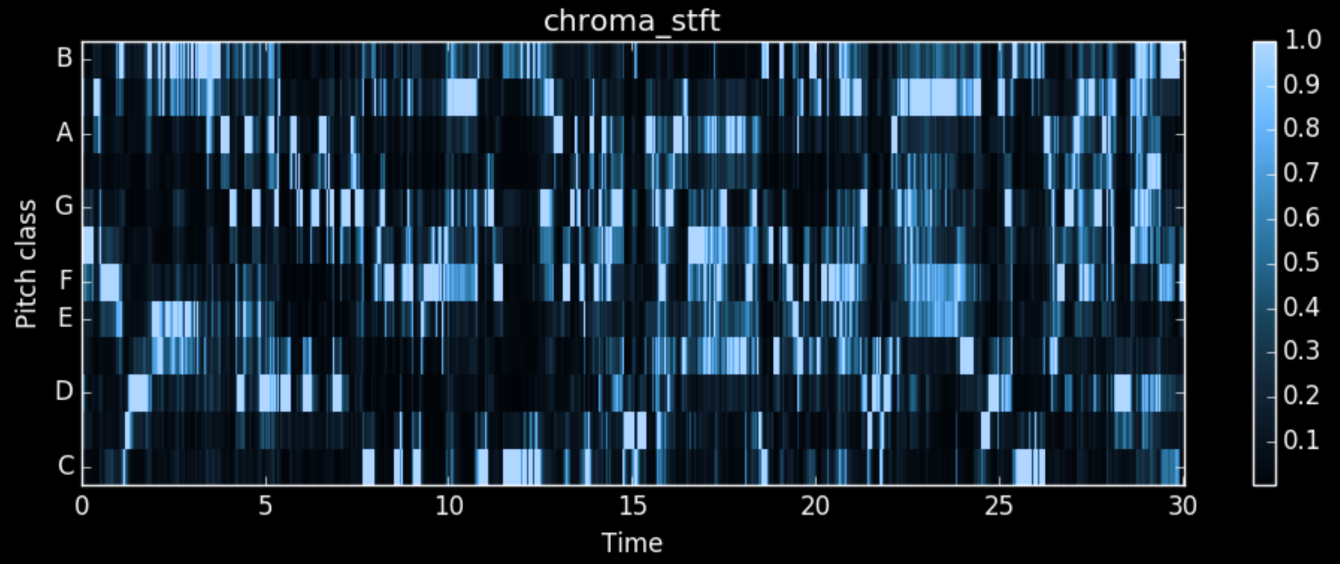
- representações alternativas para os áudios;

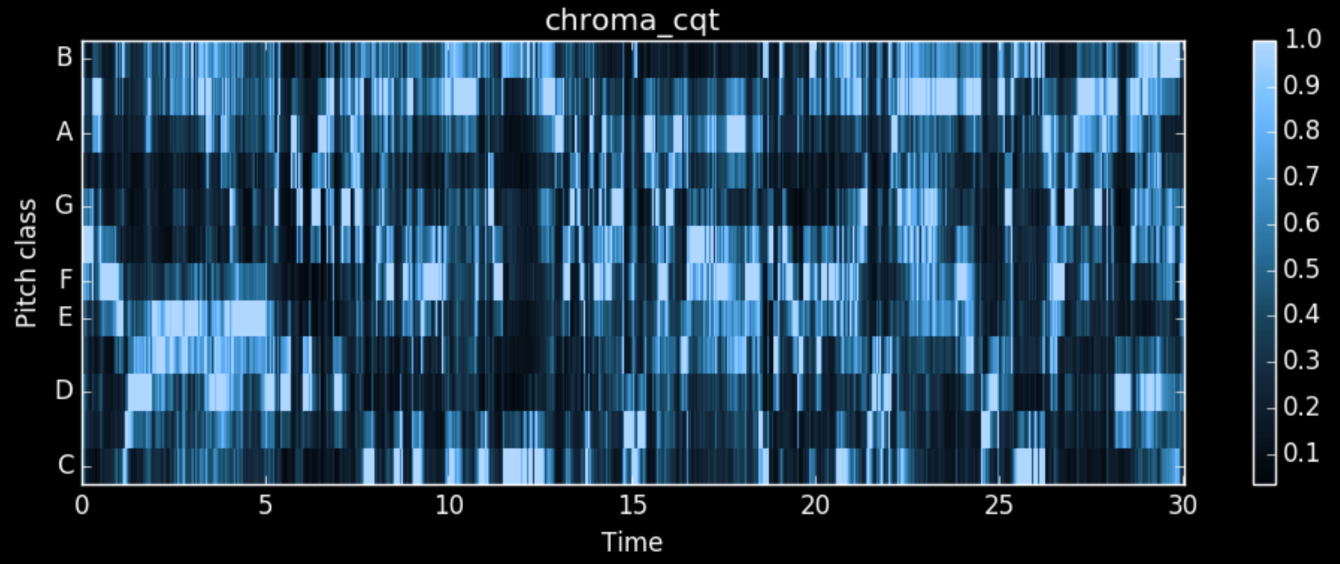
- extraídas com LibROSA;

- diversas taxonomias;

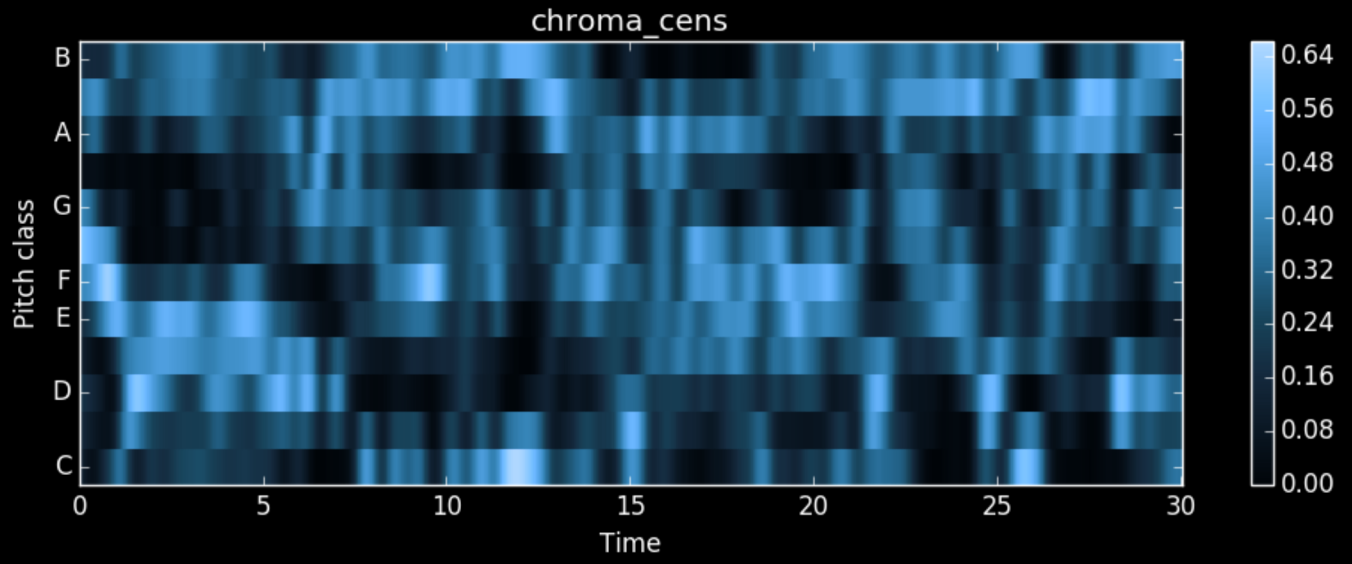
- pensaremos nos seguintes aspectos:
    1. domínio do tempo ou frequência;
    2. propriedades musicais:
        - melódico-harmônicas
        - timbrísticas
        - de andamento
        - de dinâmica

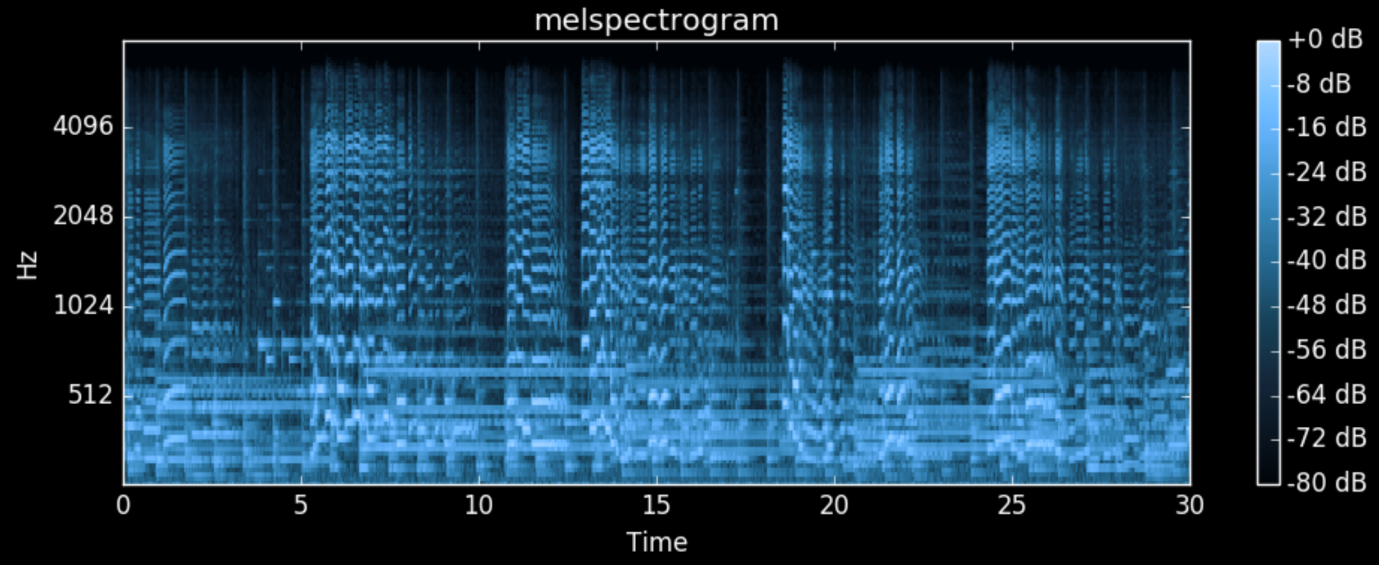
FEATURES DE ÁUDIO



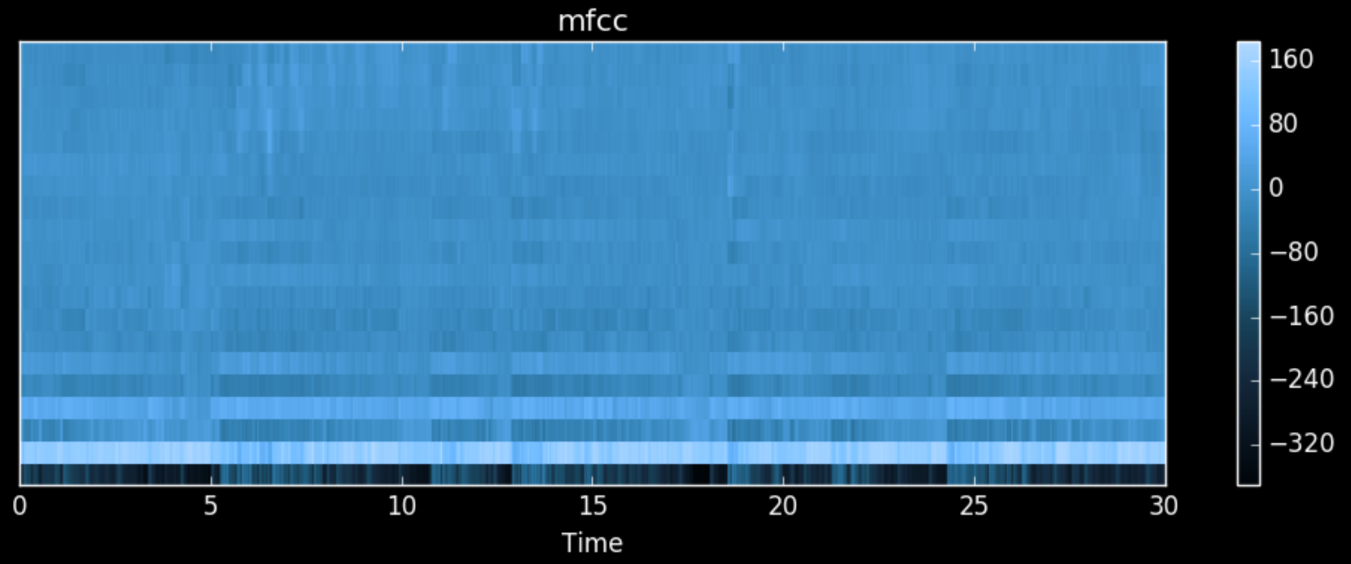


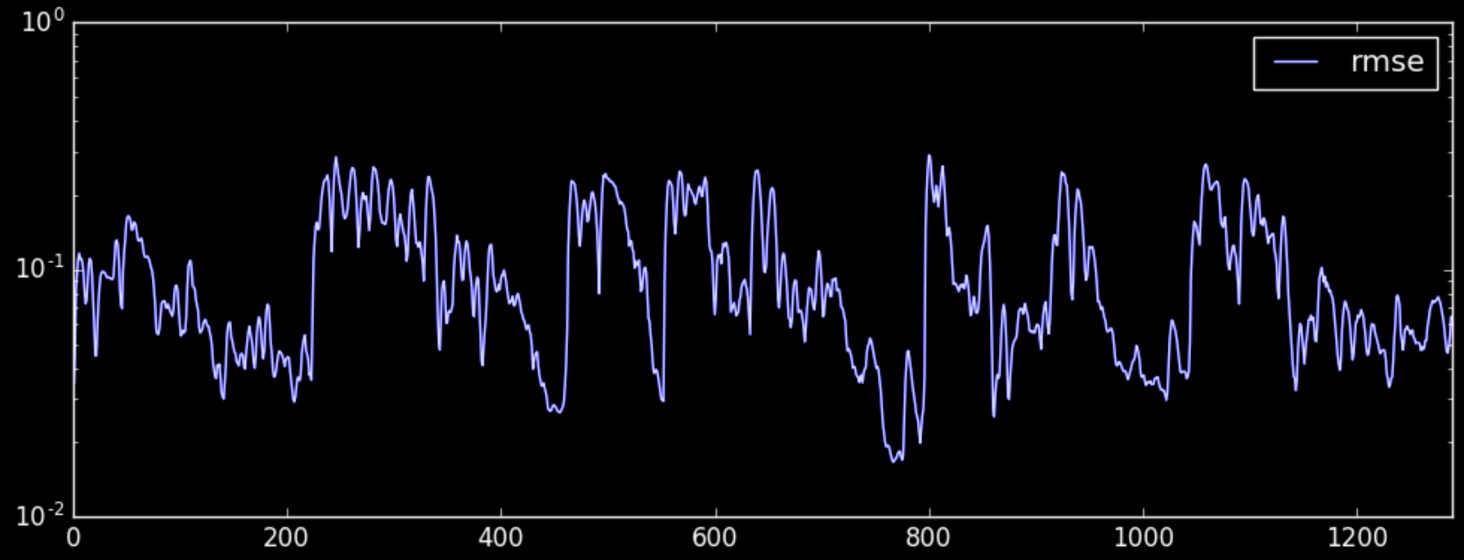
FEATURES DE ÁUDIO



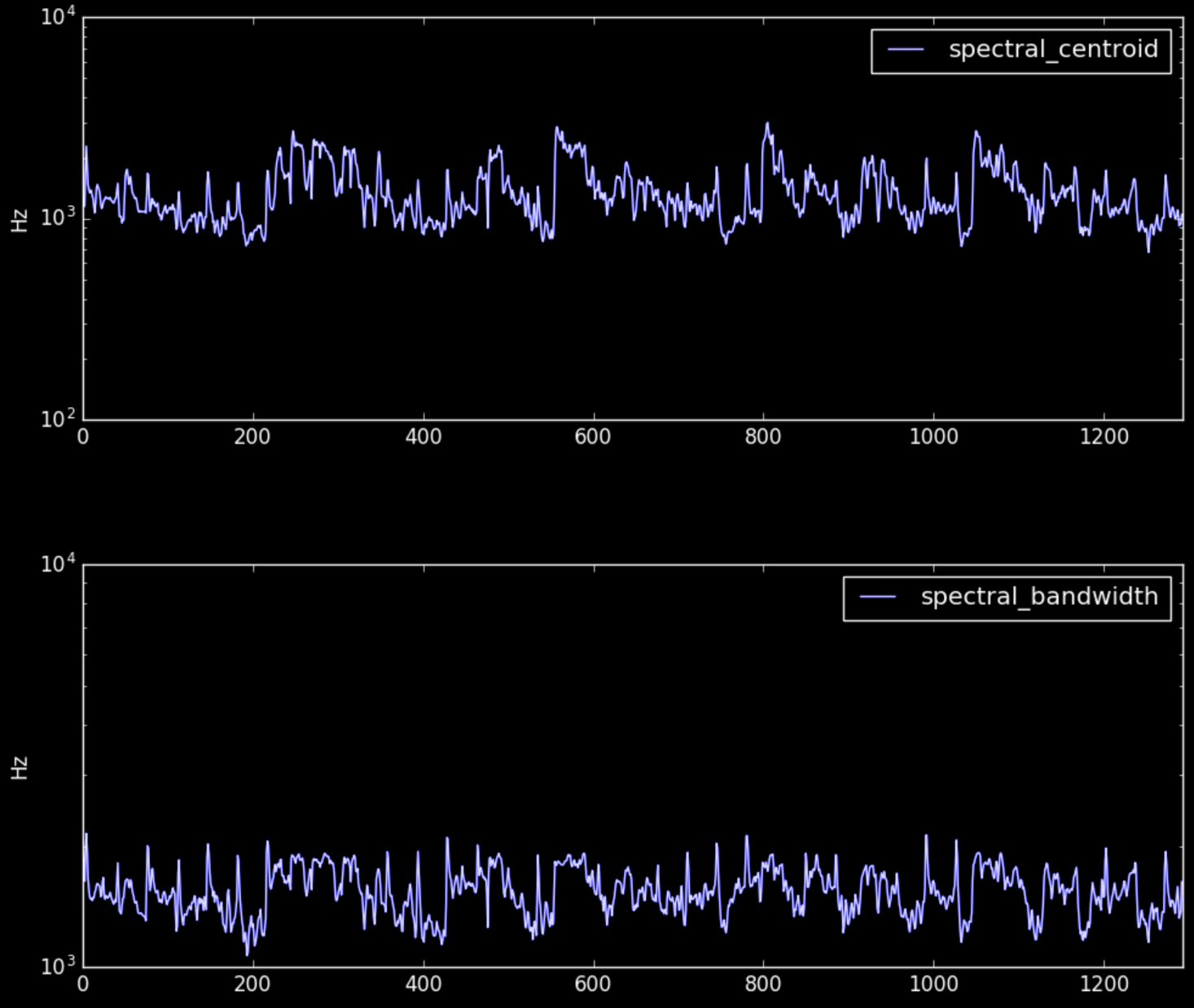


FEATURES DE ÁUDIO

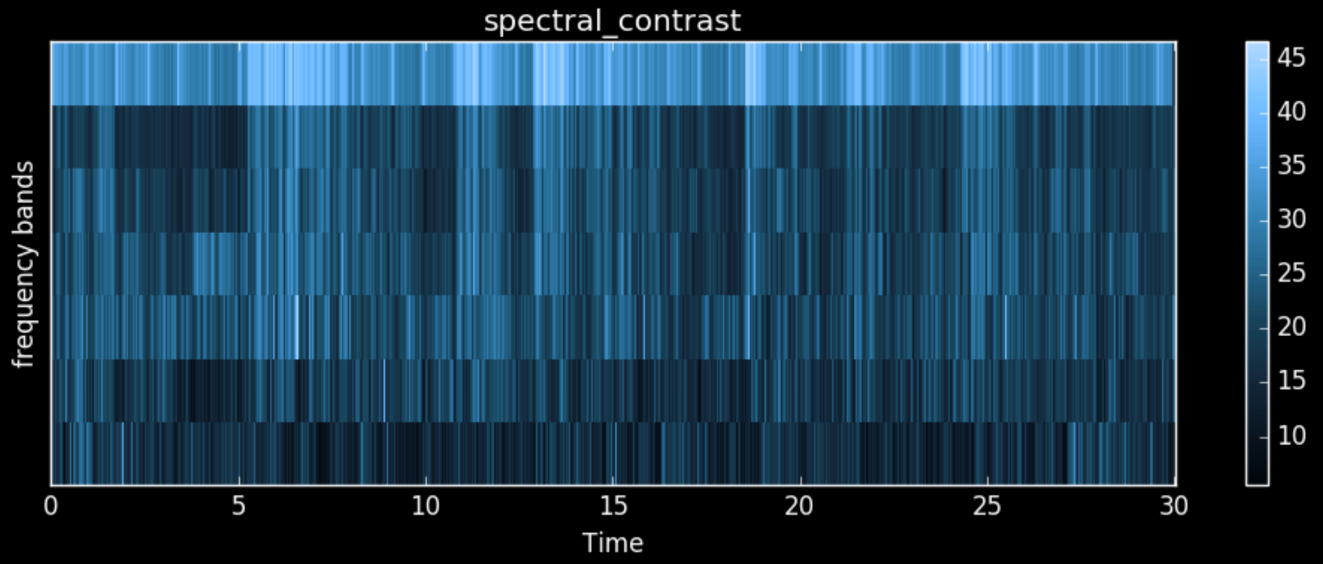


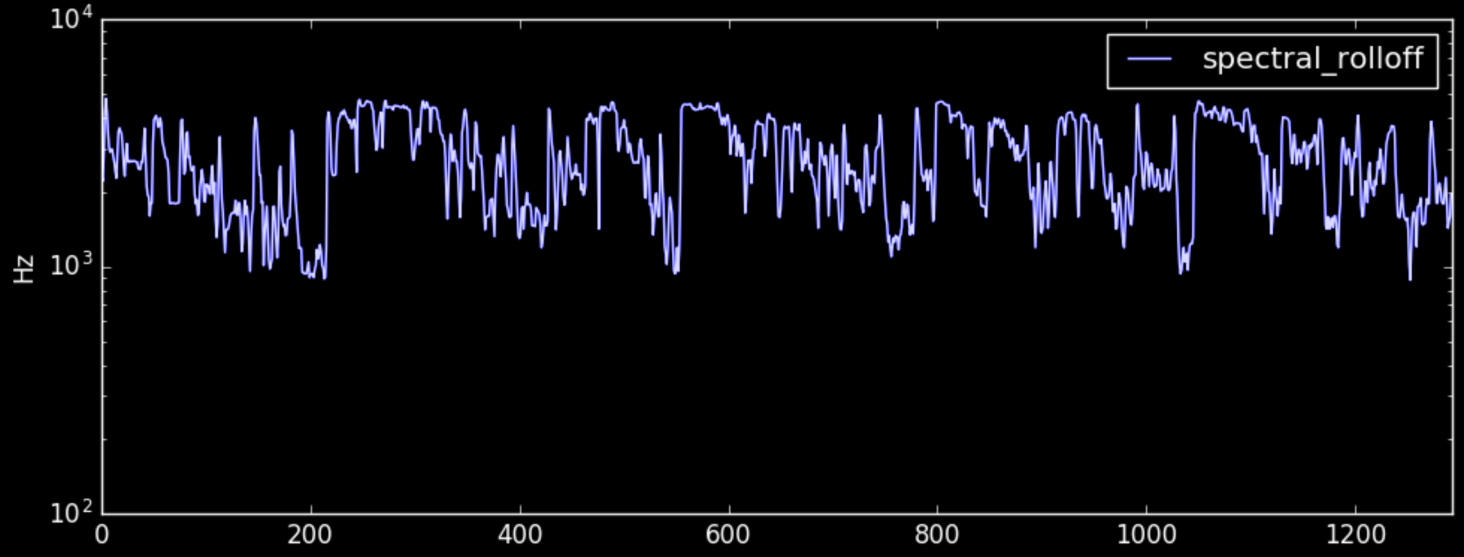


FEATURES DE ÁUDIO

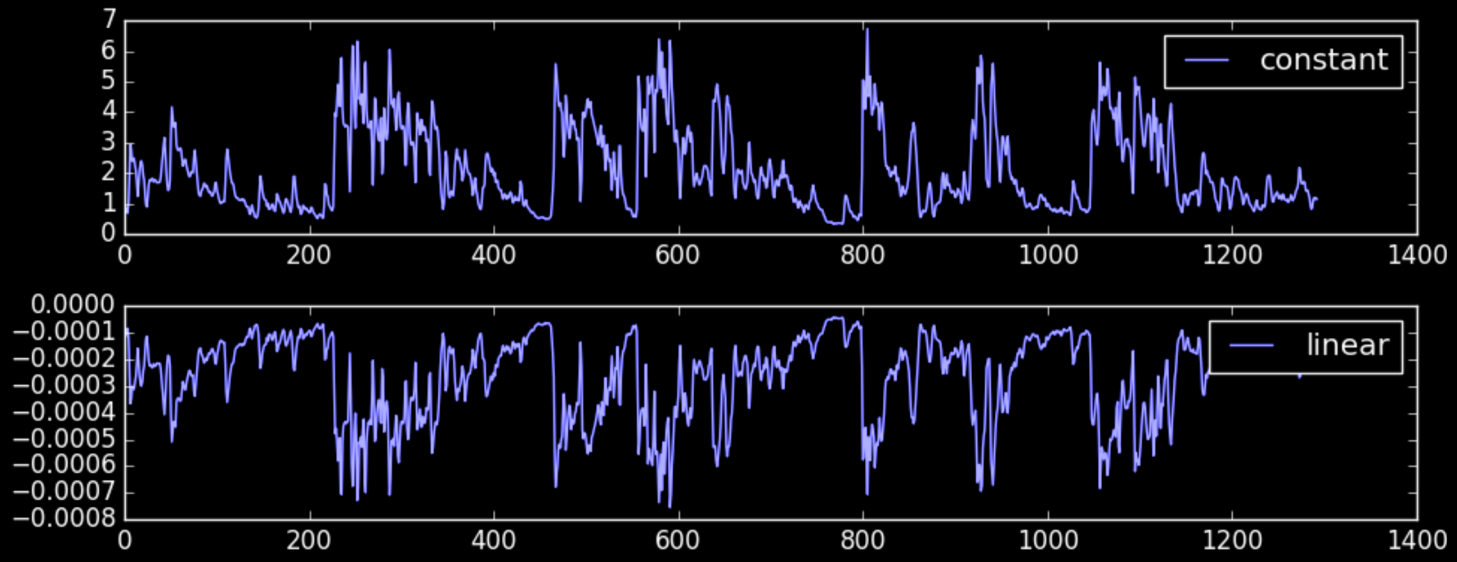


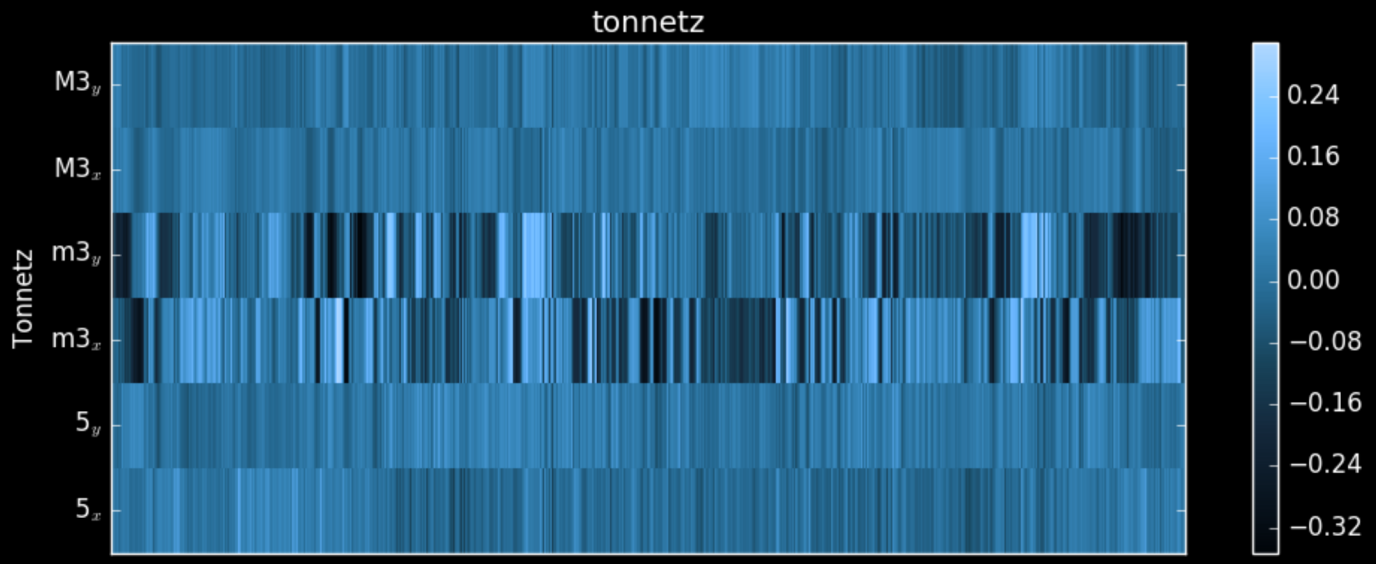
FEATURES DE ÁUDIO



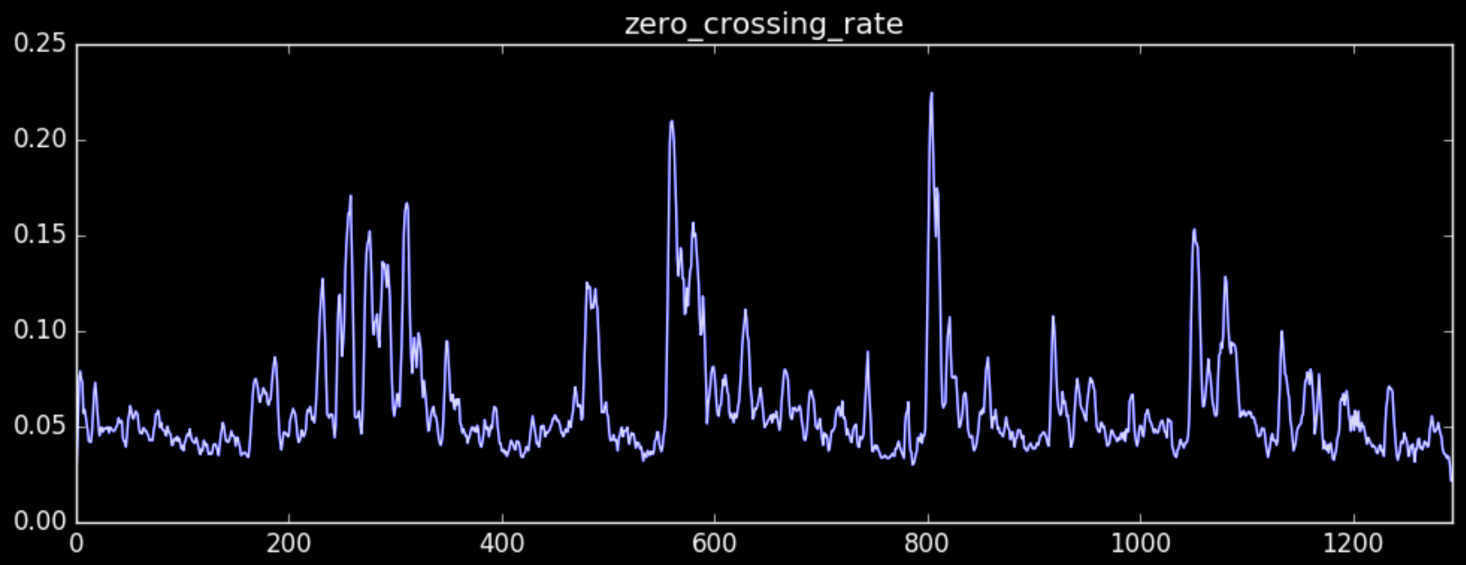


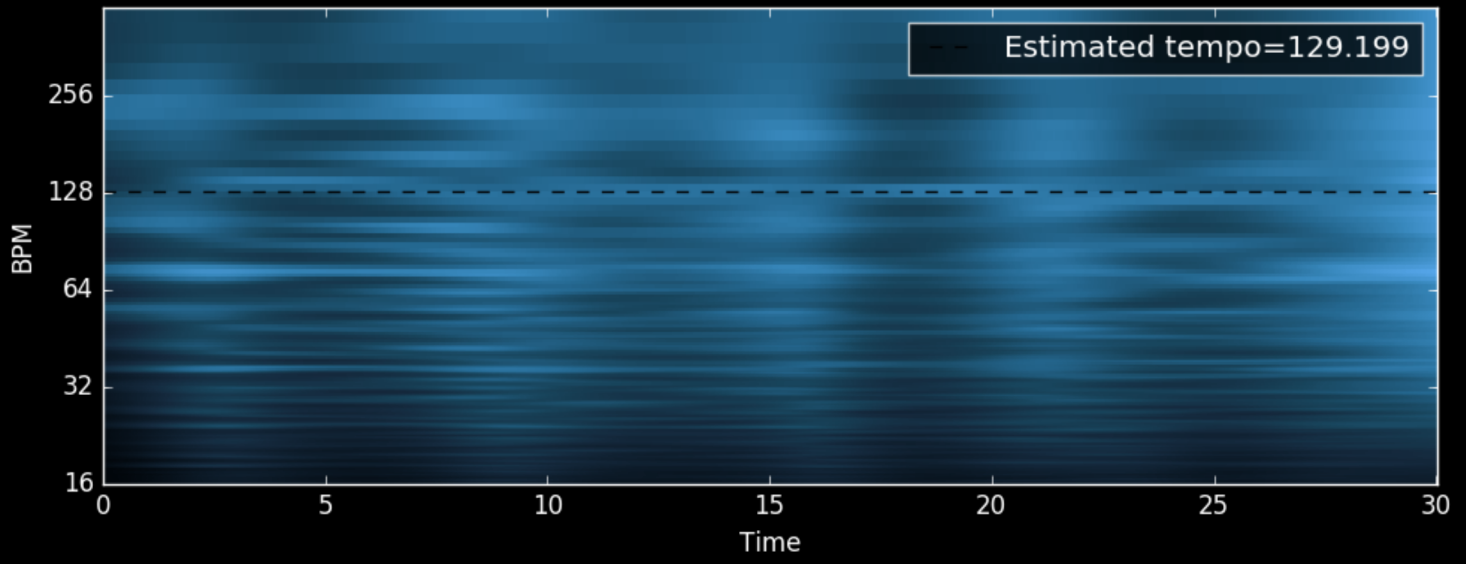
FEATURES DE ÁUDIO





FEATURES DE ÁUDIO





APRENDIZAGEM DE MÁQUINA

- aprendizagem supervisionada;

- problema de classificação;

- classificadores implementados por scikit-learn:
    1. GaussianNB (GNB);
    2. DecisionTreeClassifier (DTC);
    3. KNeighborsClassifier (KNC);
    4. SVC;

NAIVEBAYES

- probabilidade conjunta:

$$P(C, X_1, ..., X_N) = P(C) * \prod_{i=1 \rightarrow N} P(X_i \mid C)$$

- fase de treinamento:

cálculo das probabilidades condicionais $P(Xi \mid C = c)$

- classe atribuída:

$$c^* = \text{argmax} [P(C = c) * \prod_{i=1 \rightarrow N} P(X_i \mid C = c)]$$

ou

$$c^* = \text{argmax} [\log(P(C = c)) + \sum_{i=1 \rightarrow N} \log(P(X_i \mid C = c))]$$

ÁRVORES DE DECISÃO

- cada nó contém um critério de separação dos dados;

- #ramificações de um nó = #valores possíveis do atributo;

- cada folha possui uma classe a ser atribuída;

- fase de treinamento:
    - criação da árvore de classificação;
    - atributos que possuem os maiores poderes de discriminação são escolhidos primeiro;
    - métricas usualmente baseadas em entropia;

K-VIZINHOS MAIS PRÓXIMOS

- normalmente utilizado para clusterização;

- é adaptável para classificação;

- amostras inseridas no espaço de características;

- voto de maioria das k-classes vizinhas;

- k muito pequeno: pode gerar *overfitting;*

- k muito grande: pode gerar *underfitting*.

SUPPORT VECTOR MACHINE

- algoritmo de classificação linear e binária;

- busca hiperplano que maximiza a margem ;

- classificação multi-classes:
  abordagens “um-contra-todos”, “um-contra-um”, etc

- classificação não-linear:
  transformações nos dados (*kernels*)

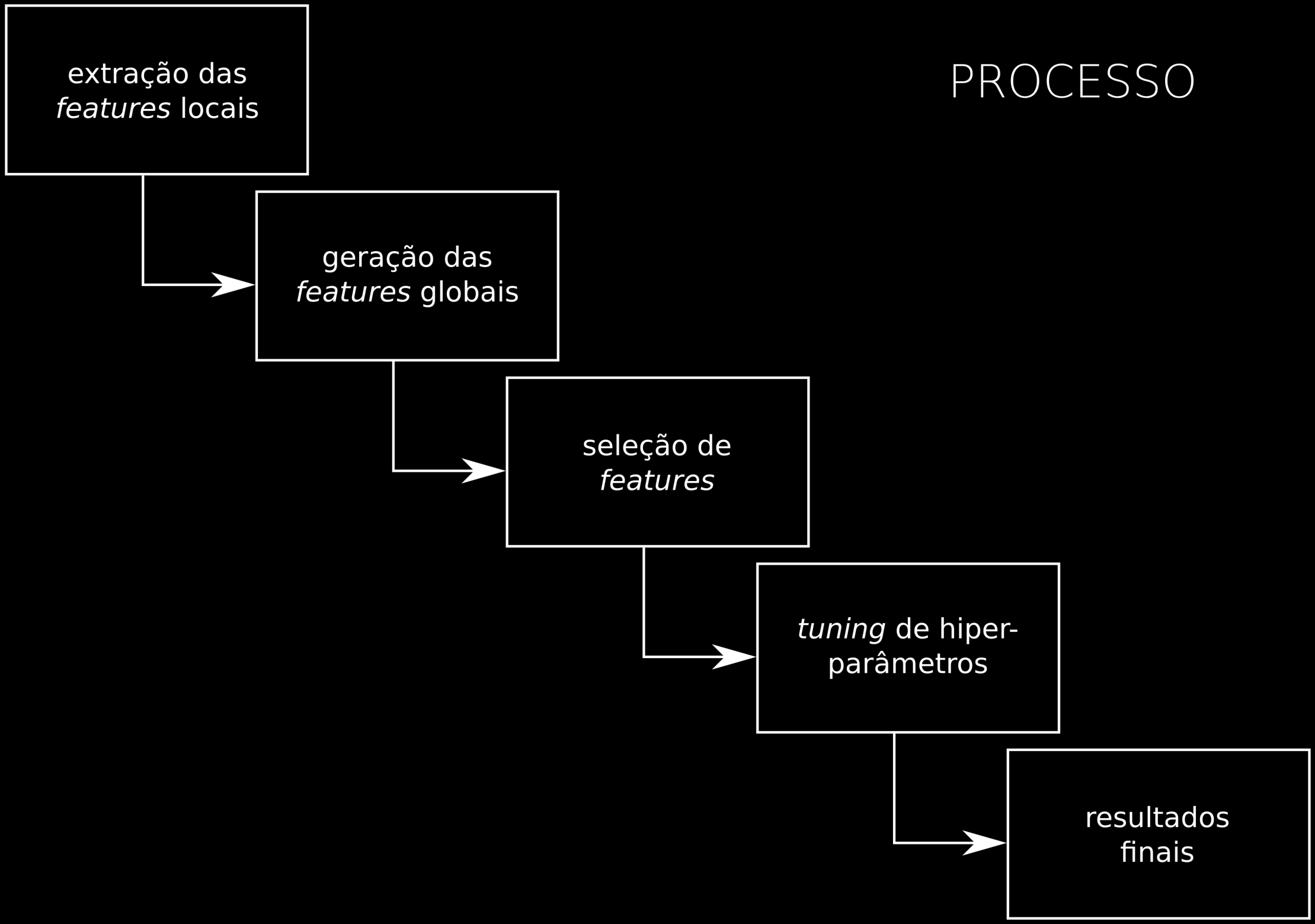
HIPÓTESES

- Alguns gêneros musicais possuem conjuntos padrões de instrumentos bem específicos, será que as *features* timbrísticas os classificarão bem?

- Alguns gêneros musicais possuem harmonias típicas, será que as *features* harmônicas os classificarão bem?

- Alguns gêneros podem conter músicas mais rápidas do que outros, será que as *features* de andamento os classificarão bem?

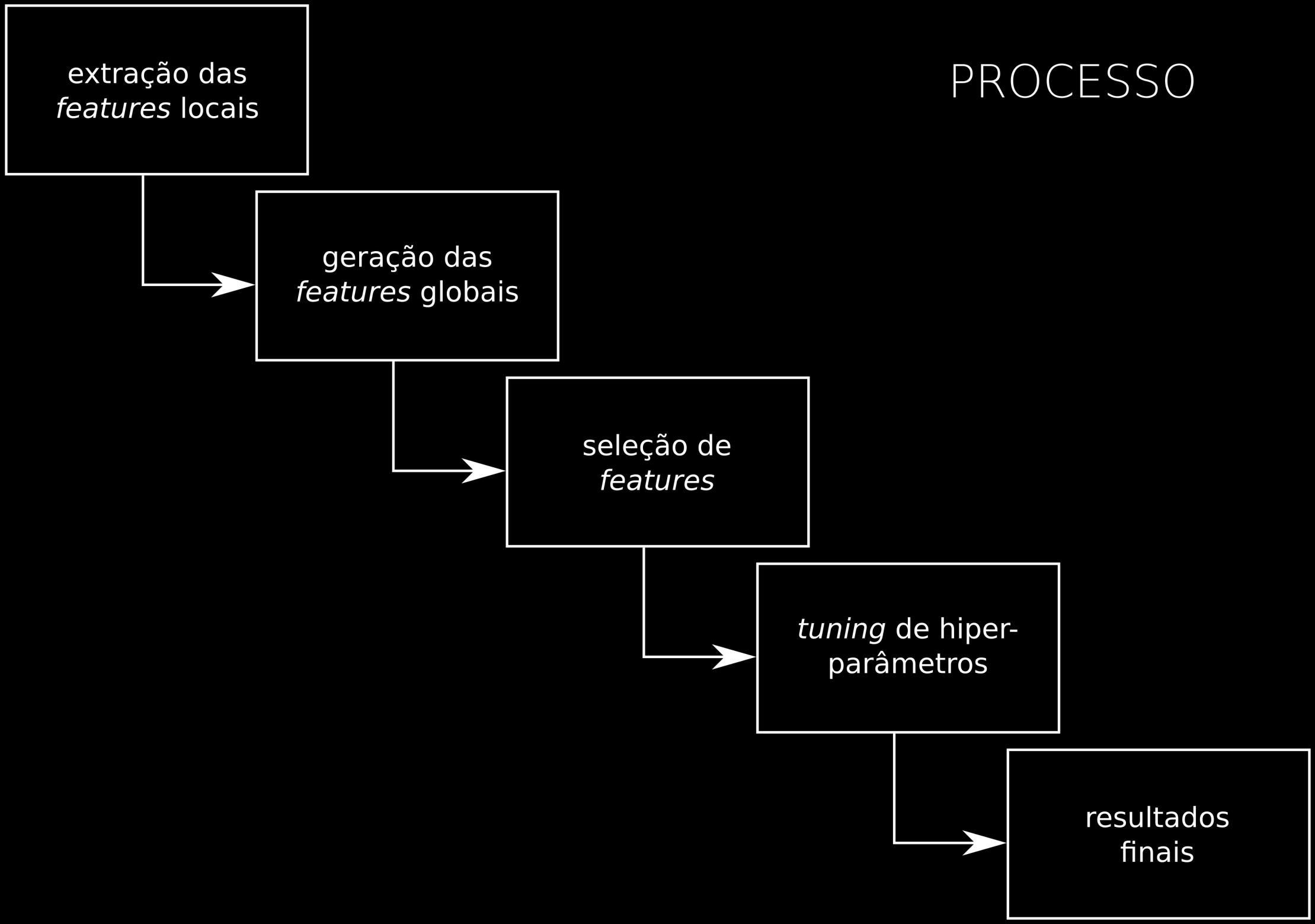
- Ainda, hipóteses sobre o processo em si:
    - Será que as *features* sincronizadas ao beat serão melhores?
    - Será que conjuntos de treinamento maiores terão melhores resultados?
    - Será que algum método de seleção de *features* irá aprimorar os resultados?


PROCESSO
extração das *features* locais
geração das *features* globais
seleção de *features*
tuning de hiper-parâmetros
resultados finais

FEATURES LOCAIS

- 30 * 22050 / 512 ~ 1292

- shape das features (jazz.00042):
    - chroma_cens (12, 1293)
    - chroma_cqt (12, 1293)
    - chroma_stft (12, 1293)
    - melspectrogram (128, 1293)
    - mfcc (20, 1293)
    - poly_features (2, 1293)
    - rmse (1289,)
    - spectral_bandwidth (1293,)
    - spectral_centroid (1293,)
    - spectral_contrast (7, 1293)
    - spectral_rolloff (1293,)
    - tempogram (384, 1293)
    - tonnetz (6, 1293)
    - zero_crossing_rate (1293,)

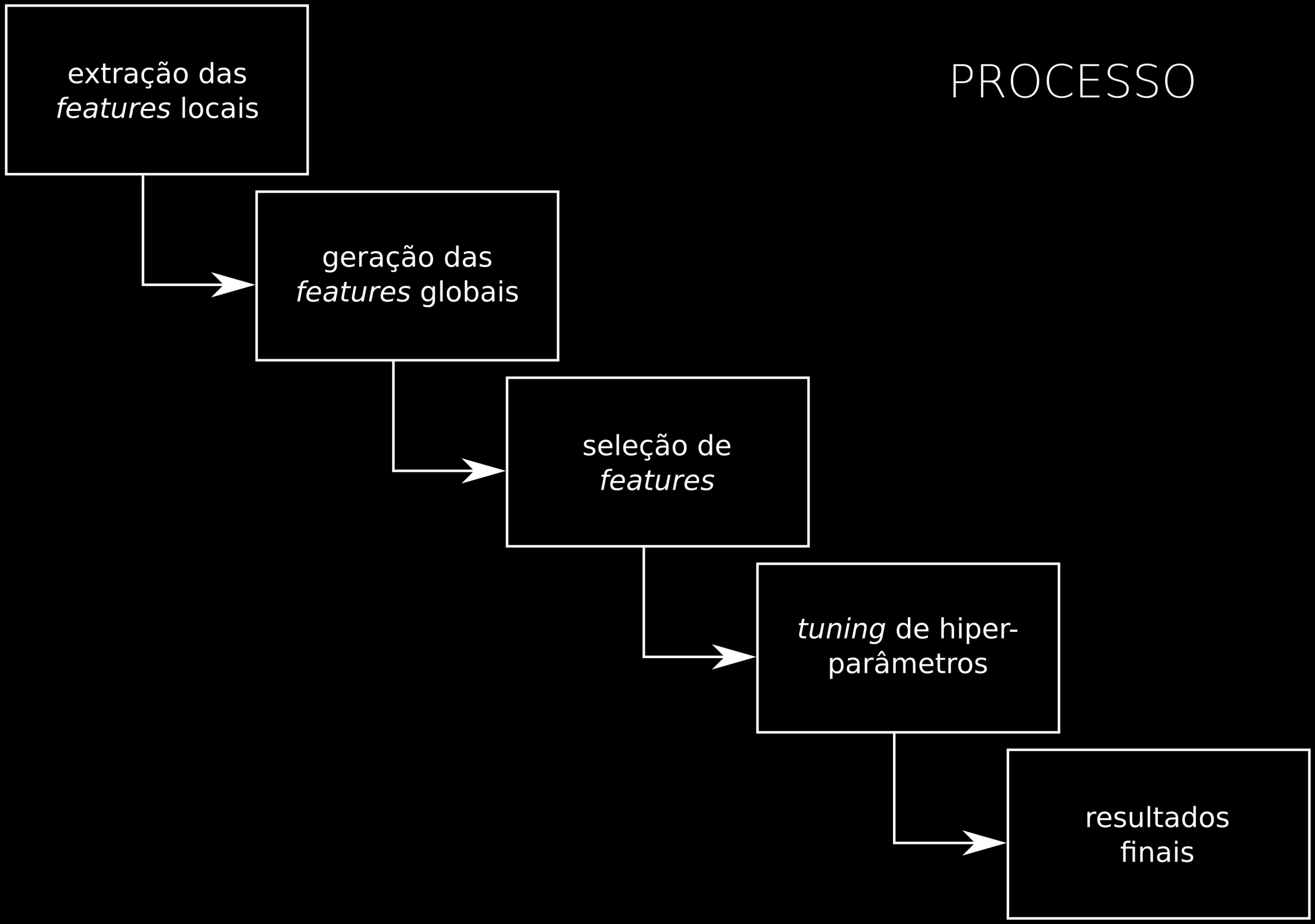

PROCESSO
extração das *features* locais
geração das *features* globais
seleção de *features*
tuning de hiper-parâmetros
resultados finais

FEATURES GLOBAIS

- dados estatísticos das features nos frames;

- medidas escolhidas:
    - média;
    - desvio padrão;
    - mínimo;
    - máximo.

- exemplo com MFCC:
    - antes: 20 * 1293 = 25860
    - depois: 4 * 20 = 80


PROCESSO
extração das *features* locais
geração das *features* globais
seleção de *features*
tuning de hiper-parâmetros
resultados finais

PRINCIPAL COMPONENT ANALYSIS

- cria novas componentes a partir de transformações nos dados;

    - primeiramente, cria um vetor com a maior variância possível dos dados originais;

    - depois, cria um segundo vetor com a próxima maior variância e, ainda, ortogonal ao primeiro;

    - assim por diante...

- componentes principais formam uma base ortogonal de dimensão menor;

- explica grande parte da variância.

LINEAR DISCRIMINANT ANALYSIS

- pode ser visto como um método de aprendizagem em si;

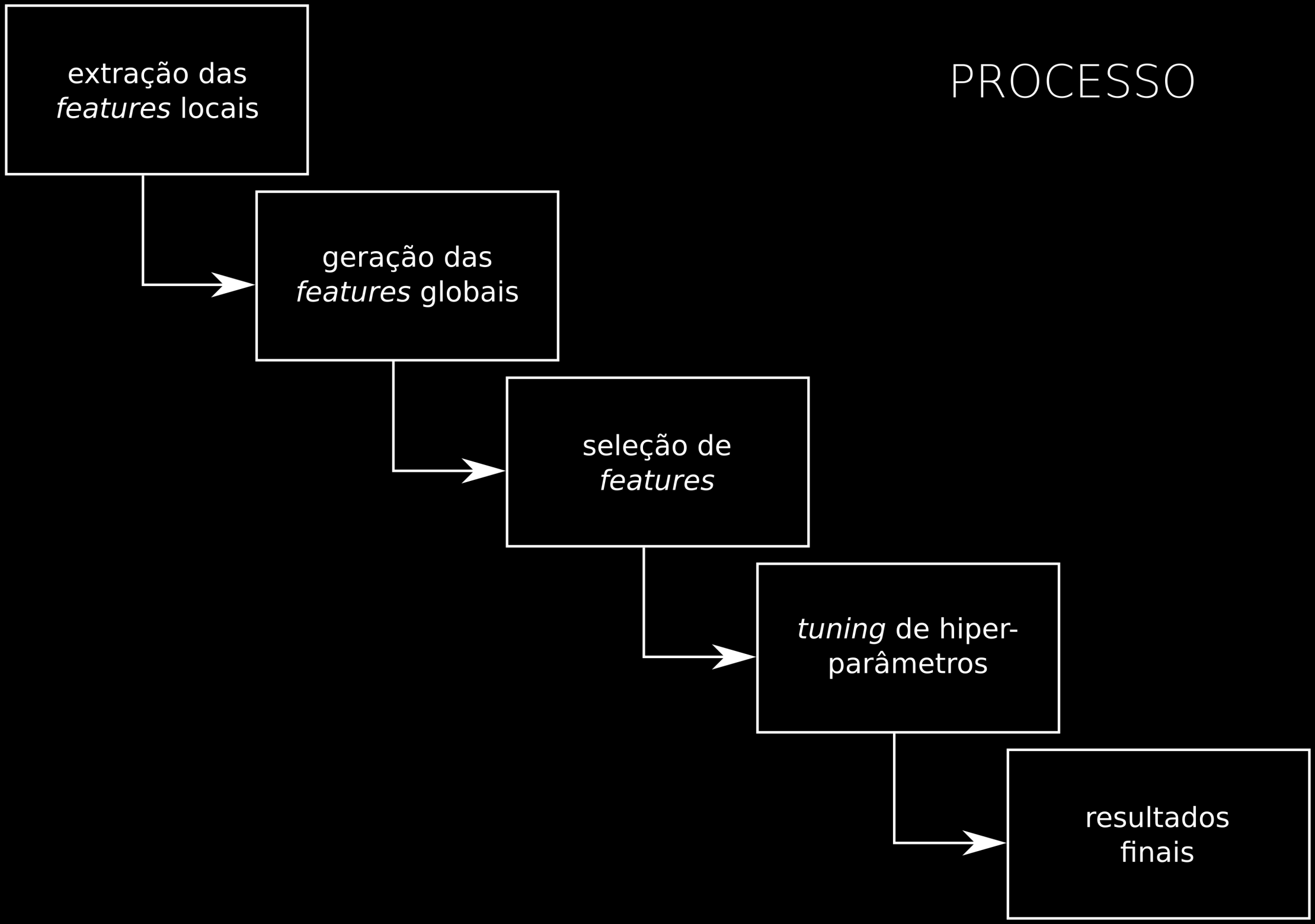
- recebe as amostras rotuladas e faz separação dos dados;

- define um subespaço onde:
    - itens de mesma classe fiquem mais próximos uns dos outros;
    - itens de diferentes classes fiquem mais distantes uns dos outros.

- projetamos os dados nos eixos mais discriminantes desse subespaço;

- obtemos uma nova representação dos dados com menor dimensão;

- preserva grande parte da variância original dos dados.


PROCESSO
extração das *features* locais
geração das *features* globais
seleção de *features*
tuning de hiper-parâmetros
resultados finais

TUNING DE PARÂMETROS

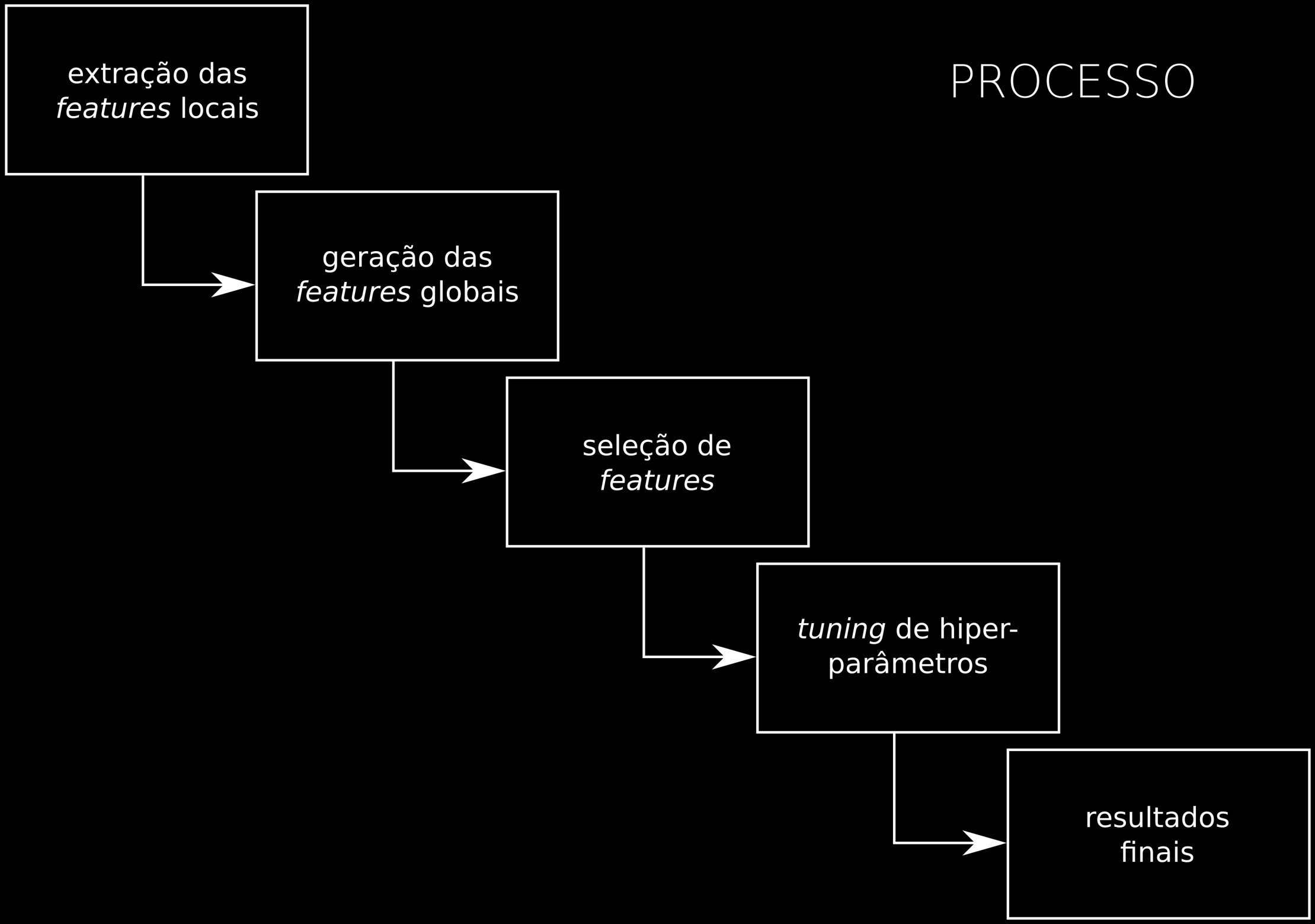
- variação dos hiper-parâmetros dos classificadores altera consideravelmente os resultados;

- GridSearchCV (scikit-learn): recebe uma lista de parâmetros, intervalos de valores e faz uma busca exaustiva;

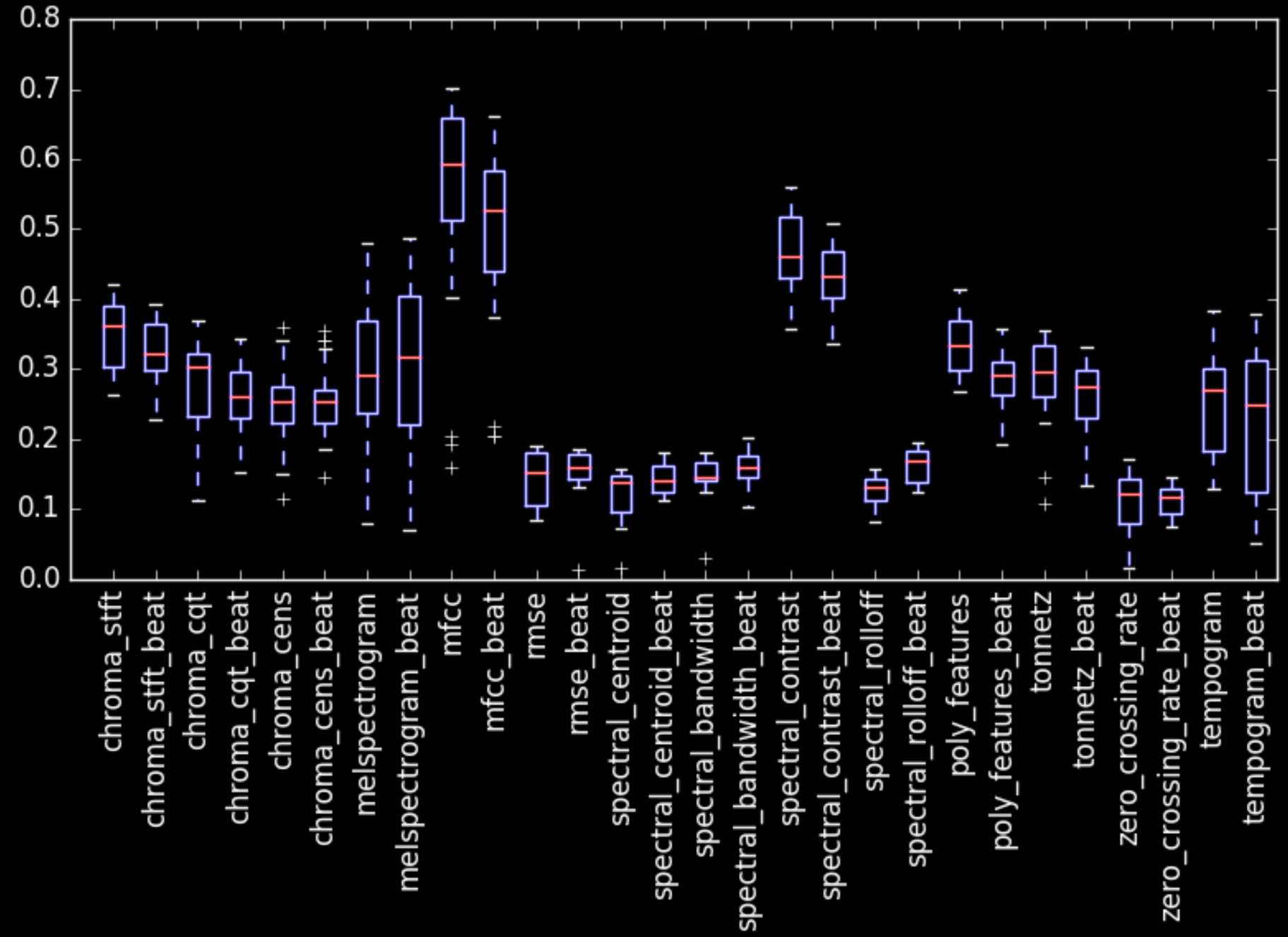
- todos os algoritmos, menos o GNB, passaram pela fase de afinação.

DIVISÃO TREINAMENTO/TESTE

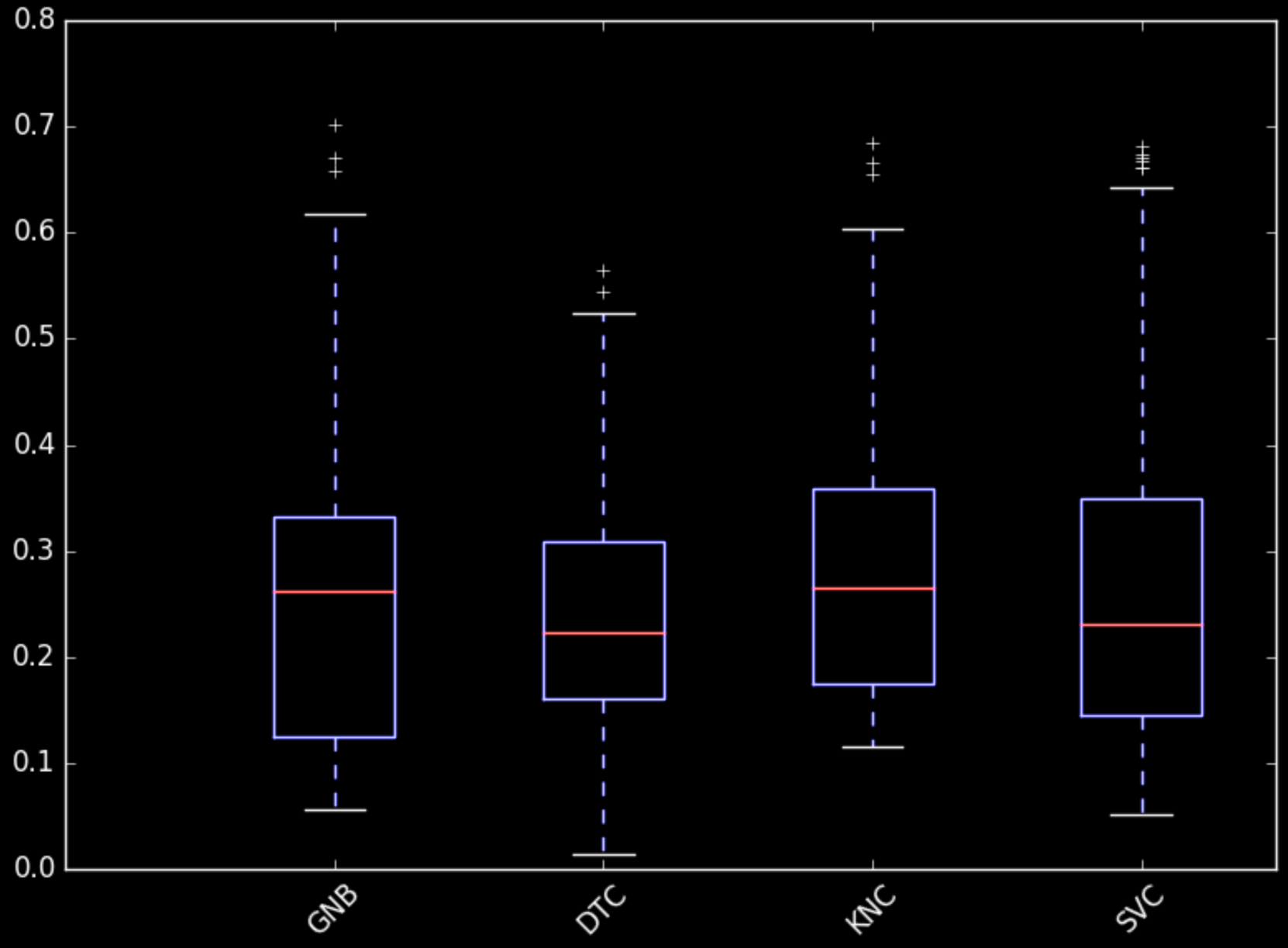
- foram escolhidas 3 divisões:
    - 50% / 50%
    - 60% / 40%
    - 70% / 30%


PROCESSO
extração das *features* locais
geração das *features* globais
seleção de *features*
tuning de hiper-parâmetros
resultados finais

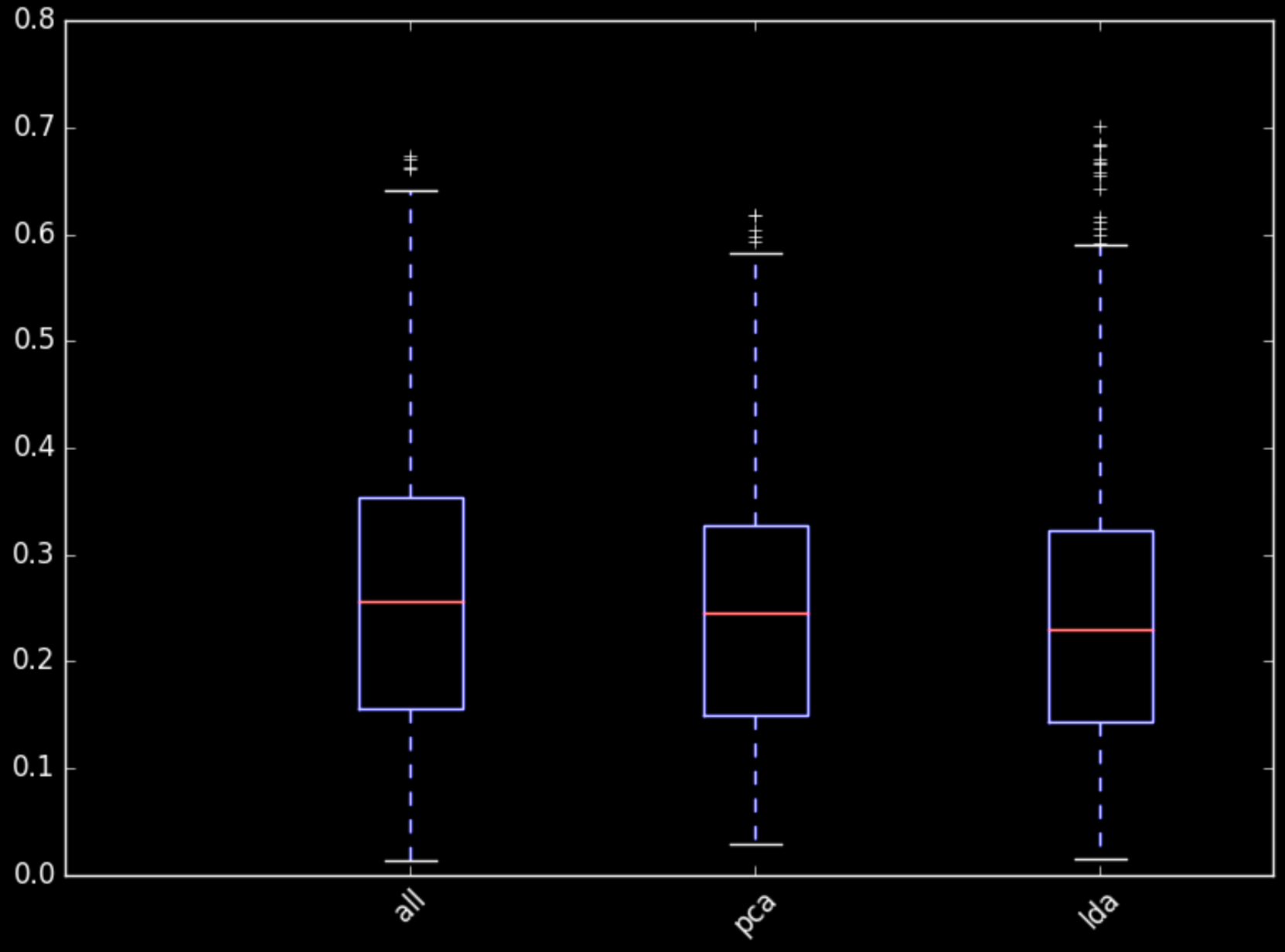
RESULTADOS FINAIS



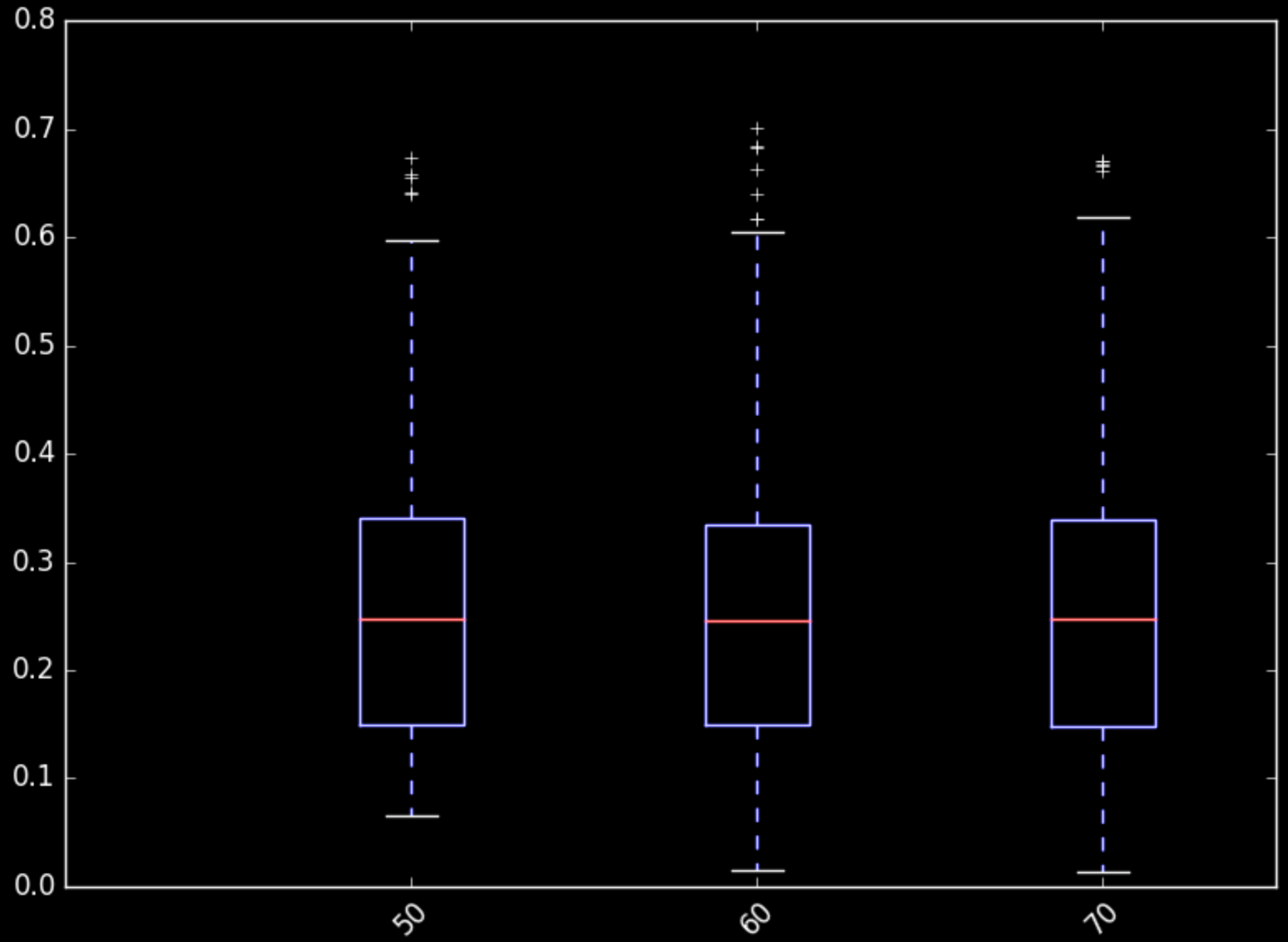
RESULTADOS FINAIS



RESULTADOS FINAIS



RESULTADOS FINAIS



MELHORES RESULTADOS GERAIS

feature	divisão	classificador	dados	f-score
mfcc_beat	70	SVC	all	0.661805
mfcc	60	SVC	all	0.662028
mfcc	70	KNC	lda	0.665517
mfcc	70	SVC	lda	0.667908
mfcc	70	GNB	lda	0.670536
mfcc	70	SVC	all	0.670784
mfcc	50	SVC	all	0.673348
mfcc	60	SVC	lda	0.682115
mfcc	60	KNC	lda	0.684708
mfcc	60	GNB	lda	0.701322

MELHORES RESULTADOS POR FEATURE

feature	divisão	classificador	dados	f-score
chroma_stft	60	SVC	all	0.421104
chroma_stft_beat	70	KNC	pca	0.392945
chroma_cqt	50	SVC	all	0.369784
chroma_cqt_beat	70	SVC	all	0.343861
chroma_cens	60	SVC	all	0.359509
chroma_cens_beat	70	SVC	all	0.355407
melspectrogram	60	DTC	all	0.479174
melspectrogram_beat	70	DTC	all	0.487913
mfcc	60	GNB	lda	0.701322
mfcc_beat	70	SVC	all	0.661805
rmse	50	KNC	lda	0.190610
rmse_beat	70	DTC	lda	0.184516
spectral_centroid	70	DTC	all	0.156766
spectral_centroid_beat	70	SVC	pca	0.181168

MELHORES RESULTADOS POR FEATURE

feature	divisão	classificador	dados	f-score
spectral_bandwidth	60	DTC	all	0.179901
spectral_bandwidth_beat	60	KNC	lda	0.202414
spectral_contrast	60	SVC	lda	0.560253
spectral_contrast_beat	50	SVC	pca	0.507445
spectral_rolloff	60	KNC	pca	0.155932
spectral_rolloff_beat	50	KNC	pca	0.194903
poly_features	60	KNC	lda	0.414029
poly_features_beat	70	KNC	lda	0.357469
tonnetz	70	KNC	lda	0.354125
tonnetz_beat	50	SVC	all	0.330817
zero_crossing_rate	70	KNC	lda	0.172082
zero_crossing_rate_beat	50	DTC	lda	0.145372
tempogram	70	SVC	all	0.382497
tempogram_beat	60	SVC	all	0.378859

COMPARAÇÃO DE RESULTADOS

Classifier	Features	Acc. (%)
CSC	Many features [6]	92.7
SRC	Auditory cortical feat. [25]	92
RBF-SVM	Learned using DBN [12]	84.3
Linear SVM	Learned using PSD on octaves	$\mathbf{83.4 \pm 3.1}$
AdaBoost	Many features [2]	83
Linear SVM	Learned using PSD on frames	$\mathbf{79.4 \pm 2.8}$
SVM	Daubechies Wavelets [19]	78.5
Log. Reg.	Spectral Covariance [3]	77
LDA	MFCC + other [18]	71
Linear SVM	Auditory cortical feat. [25]	70
GMM	MFCC + other [29]	61

OBRIGADO! =)

https://github.com/rppbodo/musical-genre-classification