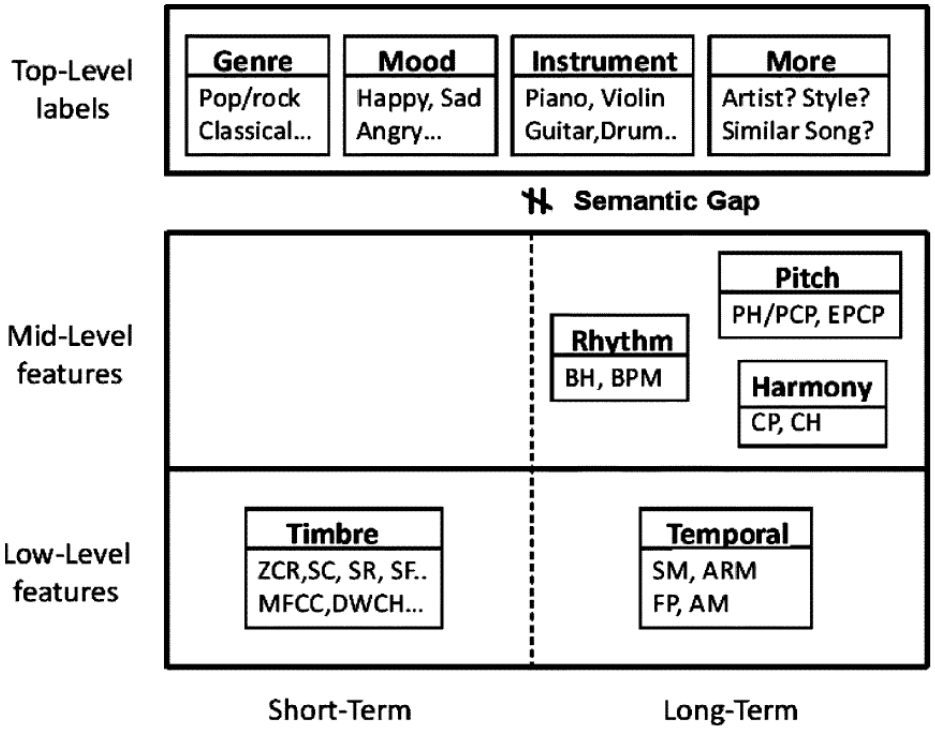
Music Classification and Annotation

Janik-Vasily Benzin, 16.01.18

Motivation

- Music Information Retrieval (MIR)
- Growing size of digital music libraries on- and offline



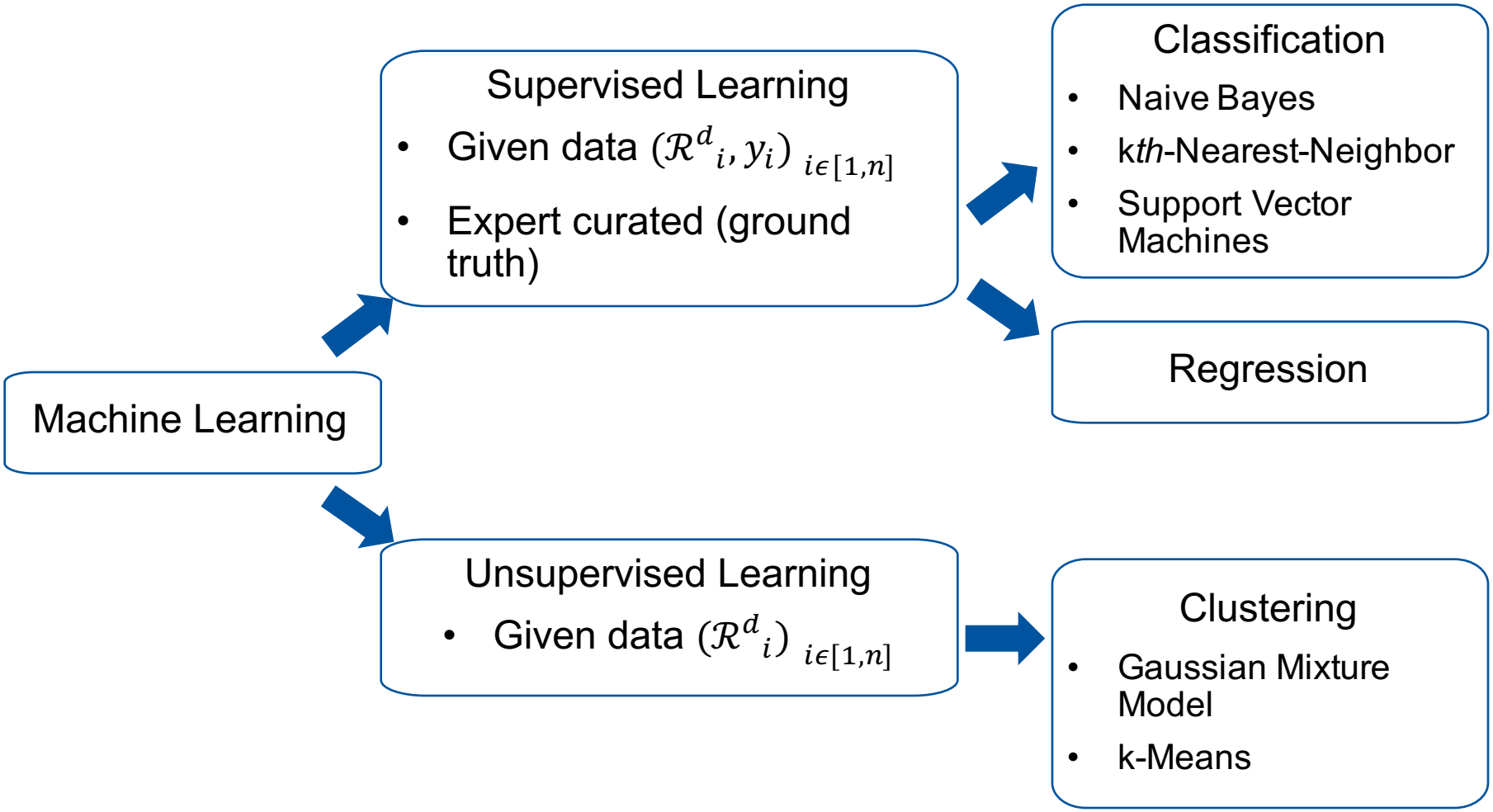
http://hpac.rwth-aachen.de/teaching/sem-mus-15/

Outline

Machine Learning

- Supervised vs. unsupervised learning
- Classification setting
- General & demo steps
  - Feature Extraction & Matlab feature demo
  - “Train“ k*th-Nearest-Neighbor classifier*
- Demo results
- Discussion

Machine Learning – Supervised vs. Unsupervised



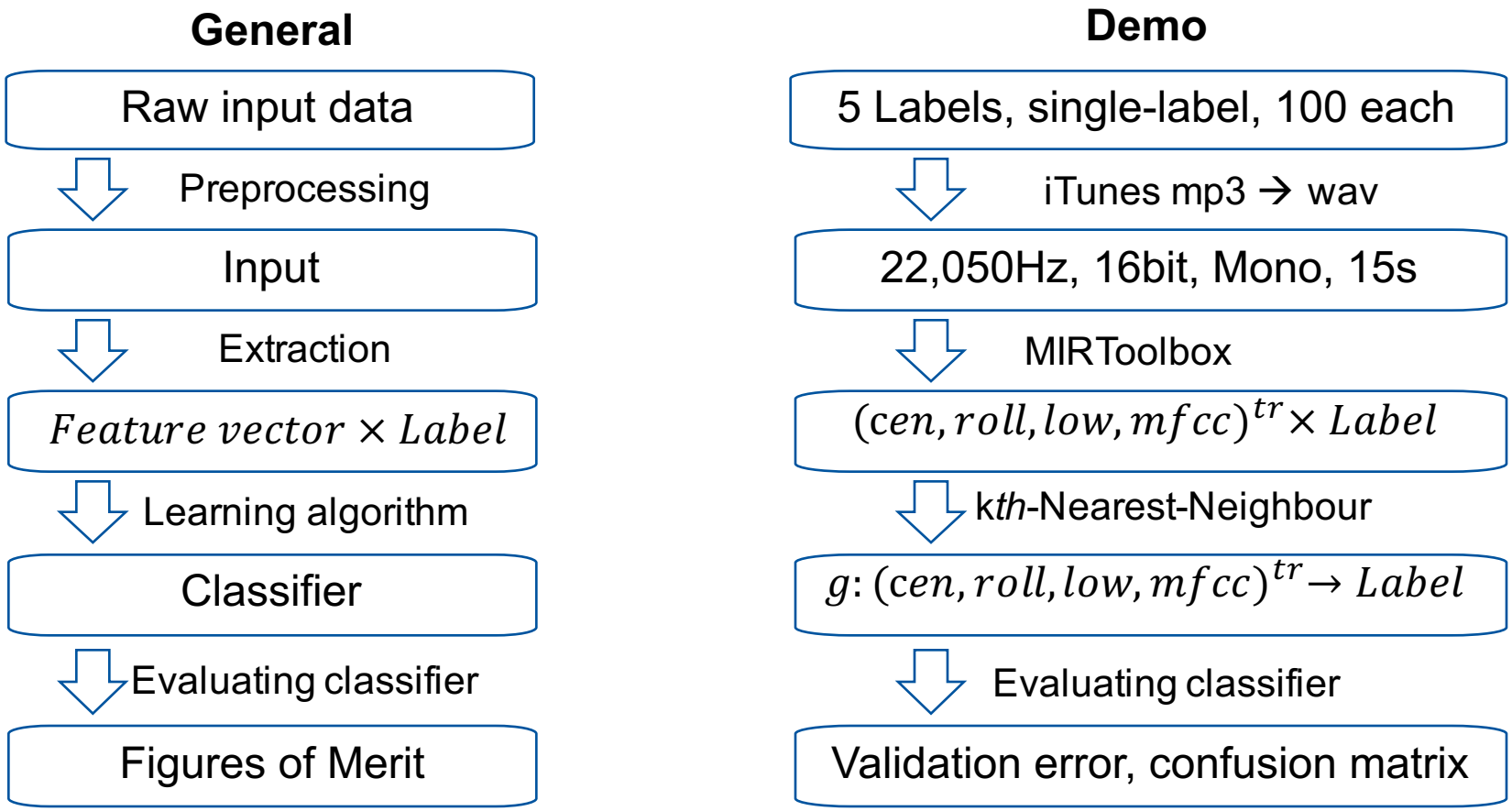
Machine Learning – Classification Setting

Classification Setting

- Given $(\mathcal{R}^d{}_i, y_i \epsilon \{1, \dots, M\})_{i\epsilon[1,n]}$
- Train classifier $g: \mathcal{R}^d \rightarrow \{1, \dots, M\}$
- Evaluate
- Use it to classify unseen data
- Will g be deterministic?

  Not in most cases → Probabilistic setting

Machine Learning – General steps – Classification



Machine Learning – Feature Extraction

22,025Hz, 16bit, Mono, 15s

MIRToolbox
50ms frames, half-overlapping

- mircentroid: $cen \in \mathbb{R}$
- mirrolloff: $roll \in \mathbb{R}$
- mirlowenergy: $low \in \mathbb{R}$
- mirmfcc: $mfcc \in f: [1,13] \rightarrow \mathbb{R}$

Frame → Spectrum (FFT) → Mel-scaled Spectrum (Triangular Windowing) → Log Mel-scaled (Logarithm) Spectrum (Discrete Cosine Transform) → mfcc

$$(cen, roll, low, mfcc)^{tr} \times Label$$

Machine Learning – Matlab feature demo

Matlab demo

Machine Learning – “Train“ kNN Classifier

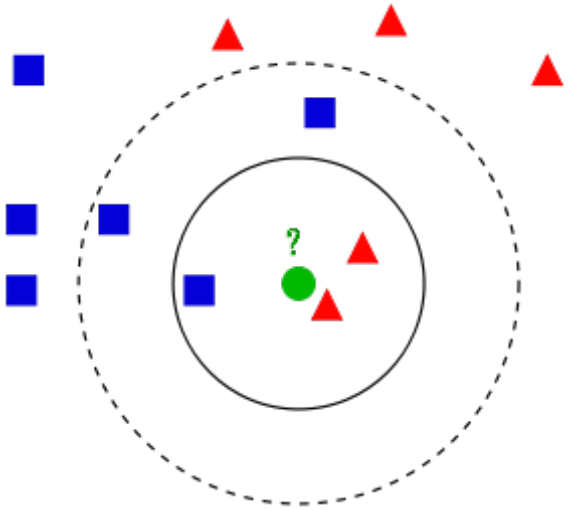
$$g_n: (cen, roll, low, mfcc)^{tr} \rightarrow Label$$

$$g_n(x) = c_j \, , if \; c_j = \underset{c_{j \in [1,5]}}{\operatorname{argmax}} \sum_{i=1}^{n} w_{ni} \, \mathrm{I}_{\{Y_i = c_j\}}$$

$$where \; w_{ni} = {}^{1}/_{k} \; if \; X_i \; is \; among \; the \; k \; neirest \; of \; x,$$

$$otherwise \; 0$$

- Metric feature space
- Euclidian distance, however asymptotic properties are independent of distance measure
- No assumption about distribution properties needed (universally consistent)
- Very simple



Machine Learning – Results

Leave-one-out-Cross-Validation (LOO-CV)

- p = 1
- 500 times fitting model

Validation accuracy 0.3740 (1NN), 0.3700 (5NN), 0.4120 (GMM1)

Confusion matrix

kNN1	atmo	dark	dreamy	happy	trippy
atmo	0.51	0.21	0.12	0.1	0.03
dark	0.29	0.31	0.18	0.13	0.06
dreamy	0.26	0.18	0.28	0.18	0.07
happy	0.29	0.12	0.24	0.3	0.02
trippy	0.2	0.1	0.11	0.09	0.47

Machine Learning – Discussion

- LOO-CV Advantages/Disadvantages
- Low accuracy
- Dataset problem - benchmarking
- Definitional problem with Labels
- Reproducibility of results & published results
- Multitude of features & learning algorithm due to ongoing research

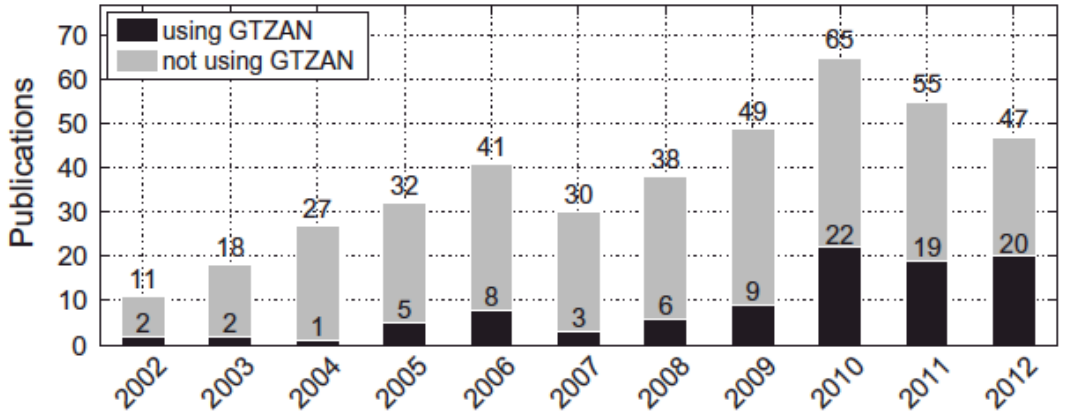


Fig. 3. Annual numbers of published works in MGR with experimental components, divided into ones that use and do no use *GTZAN*.

Machine Learning – Sources

Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013.

Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang. A survey of audio-based music classification and annotation. *IEEE transactions on multimedia*, 13(2):303–319, 2011.

Olivier Lartillot, Petri Toiviainen, and Tuomas Eerola. *A Matlab Toolbox for Music Information Retrieval*, pages 261–268. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

William H Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

Bob L Sturm. The state of the art ten years after a state of the art: Future research in music information retrieval. *Journal of New Music Research*, 43(2):147–172, 2014.

Thank you for your attention

RWTH AACHEN UNIVERSITY

Questions?

RWTH AACHEN UNIVERSITY