

GCT634/AI613: Musical Applications of Machine Learning (Fall 2021)

Homework #1 Review



Graduate School of
Culture Technology

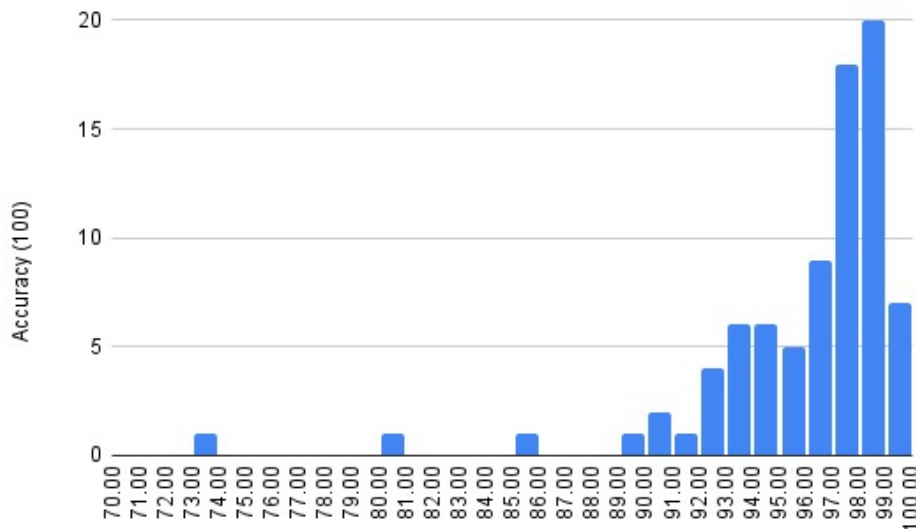
Professor Juhan Nam

TA Minsuk Choi

Statistics of Accuracy

- 2021: Best = 99.7% / Mean = 95.77% / Median = 97%
2020: Best = 97% / Mean = 81.13% / Median = 81.5%
2019: Best = 88% / Mean = 74.66% / Median = 75.85%

Performance



Grading Criteria

- Performance (5)
 - Accuracy
- Breadth (5)
 - Extensiveness in Experiment
- Depth (5)
 - Detail and Rationality in Discussion
- Novelty and Thoroughness (extra scores)

Audio Features

- MFCC with various dimensions
 - Delta, double-delta
- Spectral Centroid / Contrast / Bandwidth / Roll-off
- RMS energy (amplitude envelope)
- Mel-spec with various dimensions
- Chroma with various time-frequency representations

Feature Summarization

- Summary vectors
 - Mean, variance, ..
 - Context size: the entire frames, initial N frames ($N = 50, 100, \dots$)
- Bag-frames by K-means
 - Frequency-wise histogram
 - Time-wise histogram

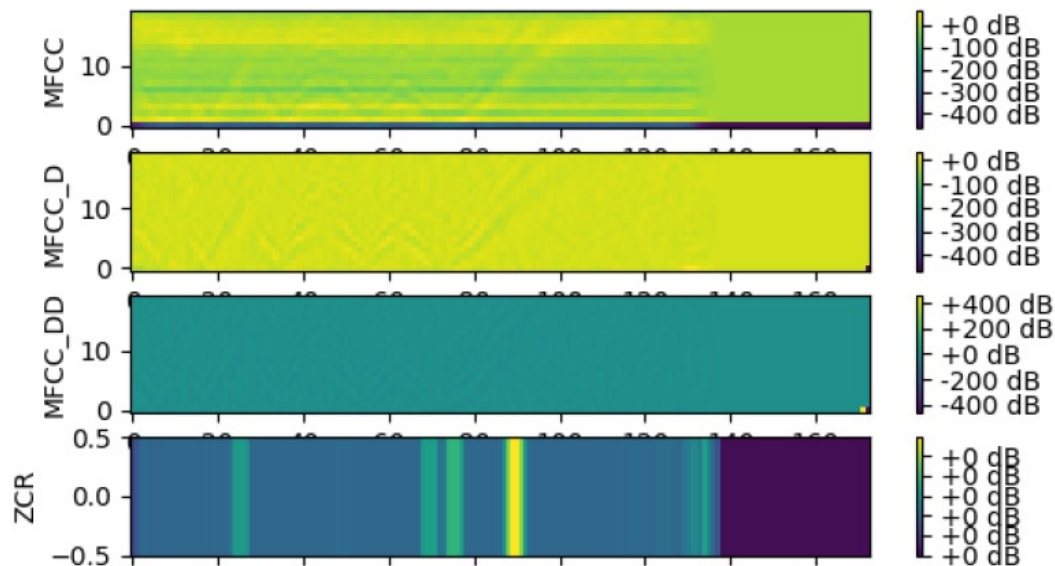
Classifiers

- K-NN
- SVM (linear, RBF)
- MLP
- Logistic Regression
- GMM
- Random forest
- CNNs

Good Directions

- Feature Ensembles

- Concatenate multiple features: e.g. MFCC delta + double-delta (capture local timbre dynamic patterns)



Good Directions

- Temporal Envelope Features
 - ADSR-based feature summary: down-sampled RMS energy

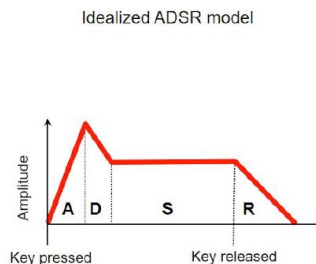
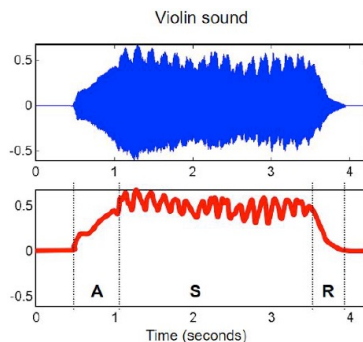
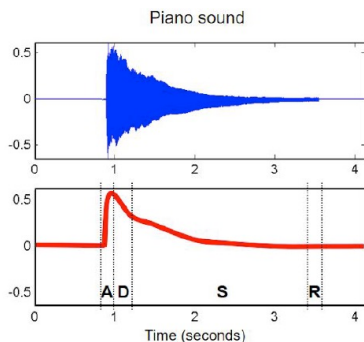


Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]

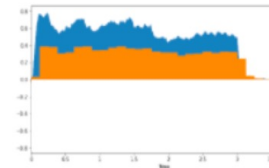
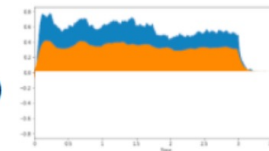
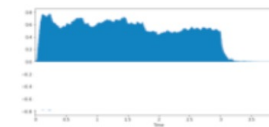
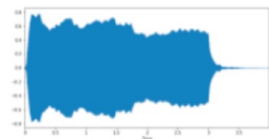


$$y = f(x)$$

$$y_2 = |y|$$

$$M(t) = \text{mean}(y_2(t \sim t + \Delta t))$$

$$M_2(t) = M[t/a]$$

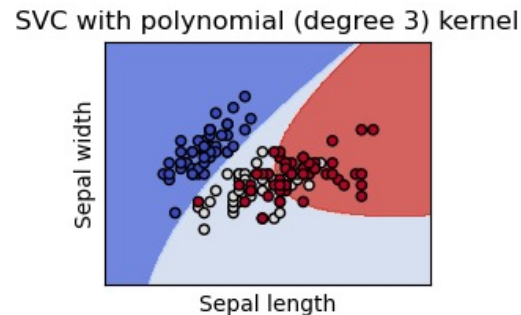
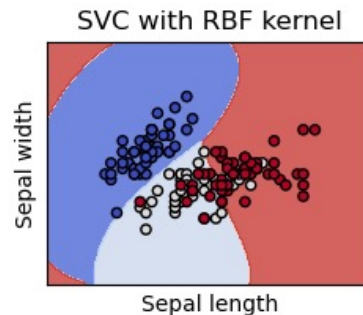
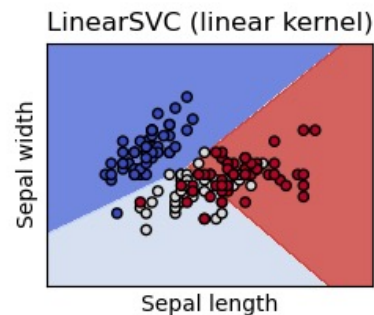
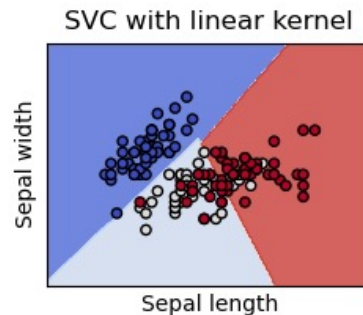


Good Directions

- Feature Summarization
 - Standard deviation pooling, variation pooling
 - Capture temporal dynamics of features
 - VQ over time: statistics of temporal patterns
- Feature Normalization
 - Zero-mean and unit-variance
 - PCA whitening (reduce dimensionality)

Good Directions

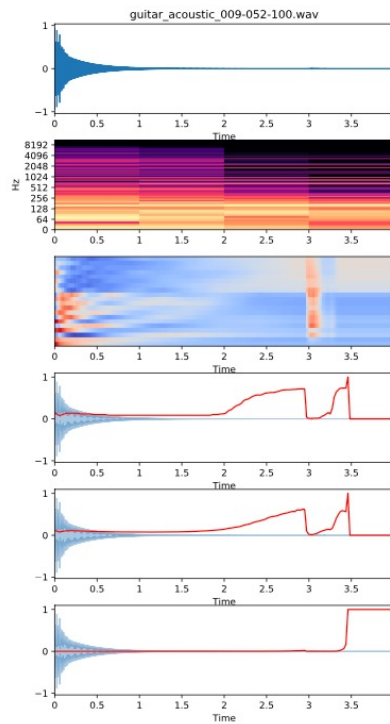
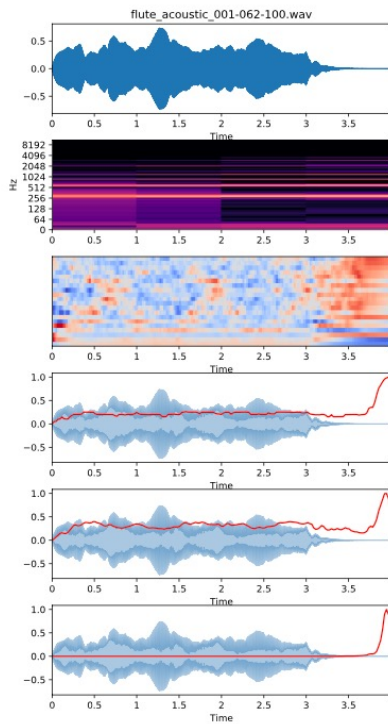
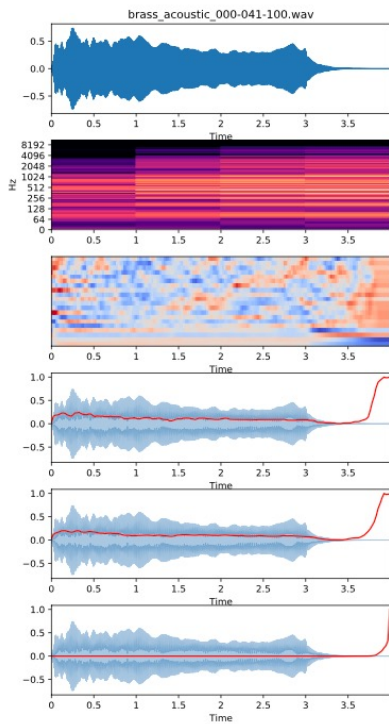
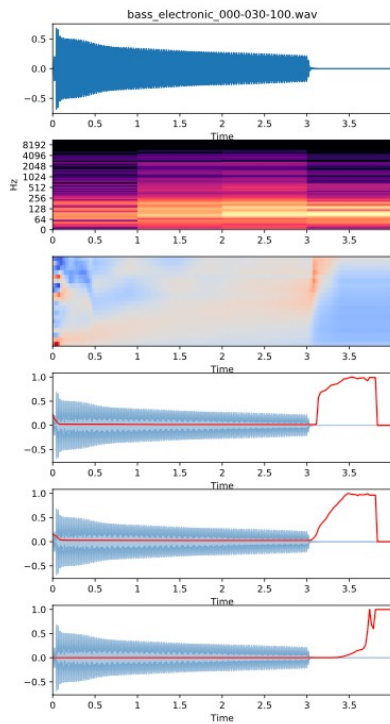
- Non-linear Classifiers
 - SVM with RBF
 - MLP
 - K-Nearest Neighbor



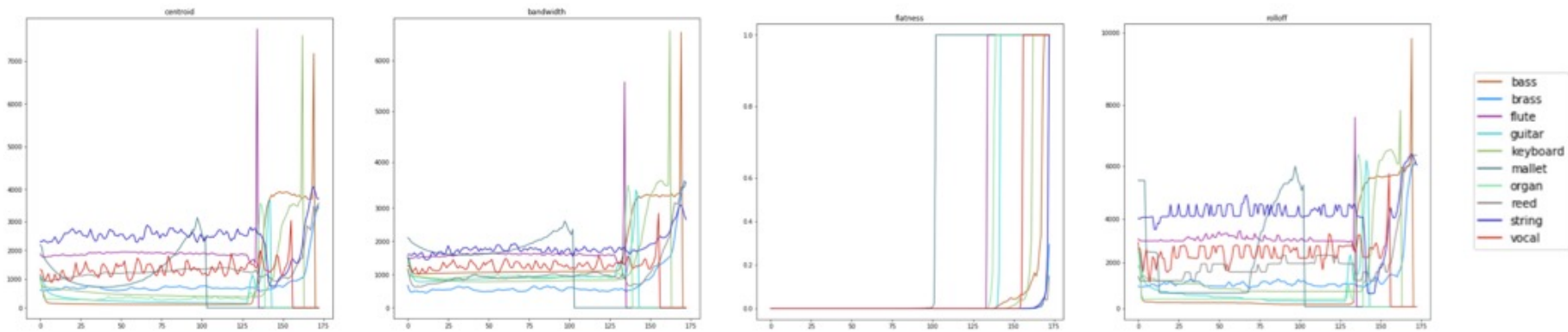
Good Directions

- Visualizing and Observing features
 - Spectrogram
 - RMS; Spectral Features
 - MFCC

Good Directions

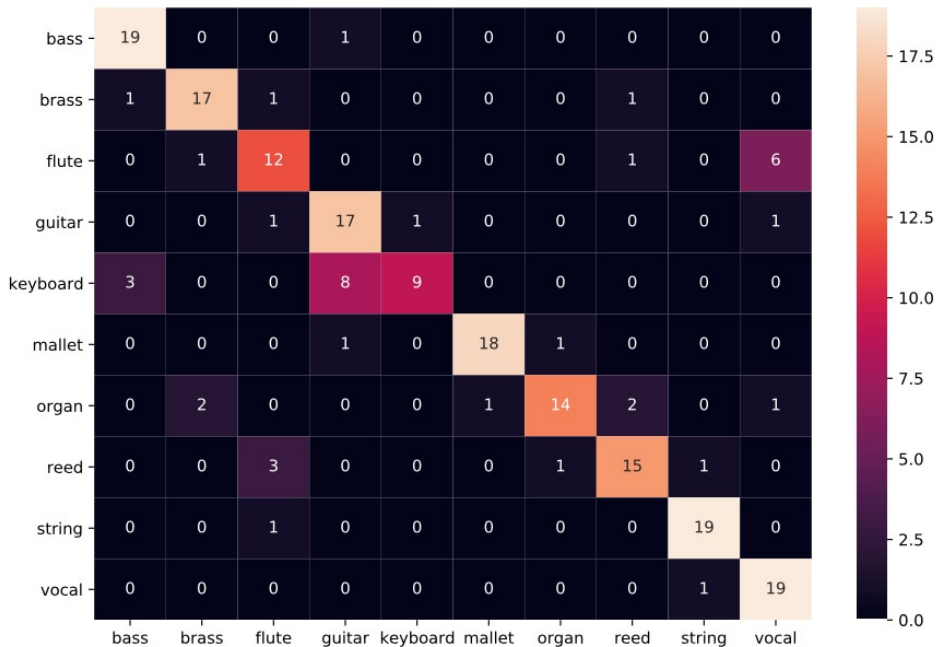


Good Directions



Good Directions

- Confusion matrix
 - Figure out which pair of instruments are confused more often than others



Best Report

- Sojin Shin
- Haemin Kim
- Hyunsong Kwon
- Joon Kyu Park
- Pilsun Eu
- Gihoon Kim
- Houn Su Kim
- Hyeonho Na
- Franck Meyer
- Hyeongjin Byeon

The best reports will be shared in KLMS under their permission !