Aim
The classification of musical instruments using supervised learning is studied. A combined feature set including psychoacoustically relevant spectral features is used. A fairly small dataset consisting of 496 sound examples from 4 instruments - violin, clarinet, saxophone and bassoon is used. Perfect classification accuracy is achieved on the test set with multi-class logistic regression and SVM with RBF kernel.

Features
- Each file is broken into frames of size 1024 samples, with overlap of 512 samples.
- 15 MFCCs\(^2\) computed per frame.
- 15 Warped LPCs\(^3\) computed per frame.
- Median values across all frames selected as features.

Dataset
- **Bach**\(^1\) dataset – audio recordings of the ensemble of ten pieces of four-part J.S. Bach chorales, and their MIDI scores.
- Soprano, Alto, Tenor and Bass of each piece are performed by violin, clarinet, saxophone and bassoon.
- 2s excerpts of each instrument – total 496 sound files, 164 labelled sounds for each instrument.

Methods
- **Multi-class Logistic Regression**
  - Softmax probability distribution
  \[
  P(y = k|x; \theta) = \frac{\exp(\theta_k^T x)}{1 + \sum_{j=1}^{K-1} \exp(\theta_j^T x)}
  \]
  - Maximum log likelihood with L1 regularization.
- **SVM with RBF kernel**
  - Non-linear SVM with infinite dimensional feature map, given by the RBF kernel
  \[
  K(x^i, x^j) = \exp\left(-\gamma \|x^i - x^j\|^2\right)
  \]
  - Kernel determines similarity between two feature vectors.

Results
- Dataset randomly split into training (75%) and test set (25%)
- 41 sound files in the test set and 123 sound files in the training set for each instrument.
- 3-Fold Cross-validation done with a log-spaced grid search to select regularization parameters for both logistic regression and SVM, and \(\gamma\), the kernel parameter for SVM.
- 100% classification accuracy is achieved on the test set with both logistic regression and SVM.

Discussion
No misclassifications between same family of instruments (eg:single-reed, woodwind).
Instruments purely harmonic, Not percussive.
Method should be extended to include many more instruments.

References