Scattering for similarity search of playing technique

Mathieu Lagrange

March 5, 2018

Context

- Technical need: explore a database of recordings of musical instrument playing techniques using computational similarity and automatic ranking techniques
- Representing of musical sound signals mostly focus on the frequency distribution of energy
- For playing techniques, this is probably not sufficient due to intensive use of modulations
- Aim: have a representation that is expressive enough to flexicably adapt to different aspects of musical instrument perception

Data

- audio recordings of individual tones of:
- instruments, with appendums
- playing techniques
- nuances
- pitch

Numbers

- number of items is large: ¿ 10 000
- number of classes is medium: ¡ 200
- number of dimensions too: ¡ 1000

Processing steps

- features: mel, mfccs, time / frequency wavelet scattering
- projection: linear discriminant analysis (lda), large margin nearest neighbors (lmnn)
- metric: precision @ 5 (p@5)

Features

- mel: spectral features on log frequency scale
- mfccs: mel projected on a DCT basis
- time / frequency wavelet scattering

Projection

- lda: projection in a (C-1) dimensional space that best separate the classes
- lmnn: malahanobis distance metric with projection matrix optimized to achieve best p@k on the training dataset

`http://jmlr.csail.mit.edu/papers/volume10/weinberger09a/weinberger09a.pdf`

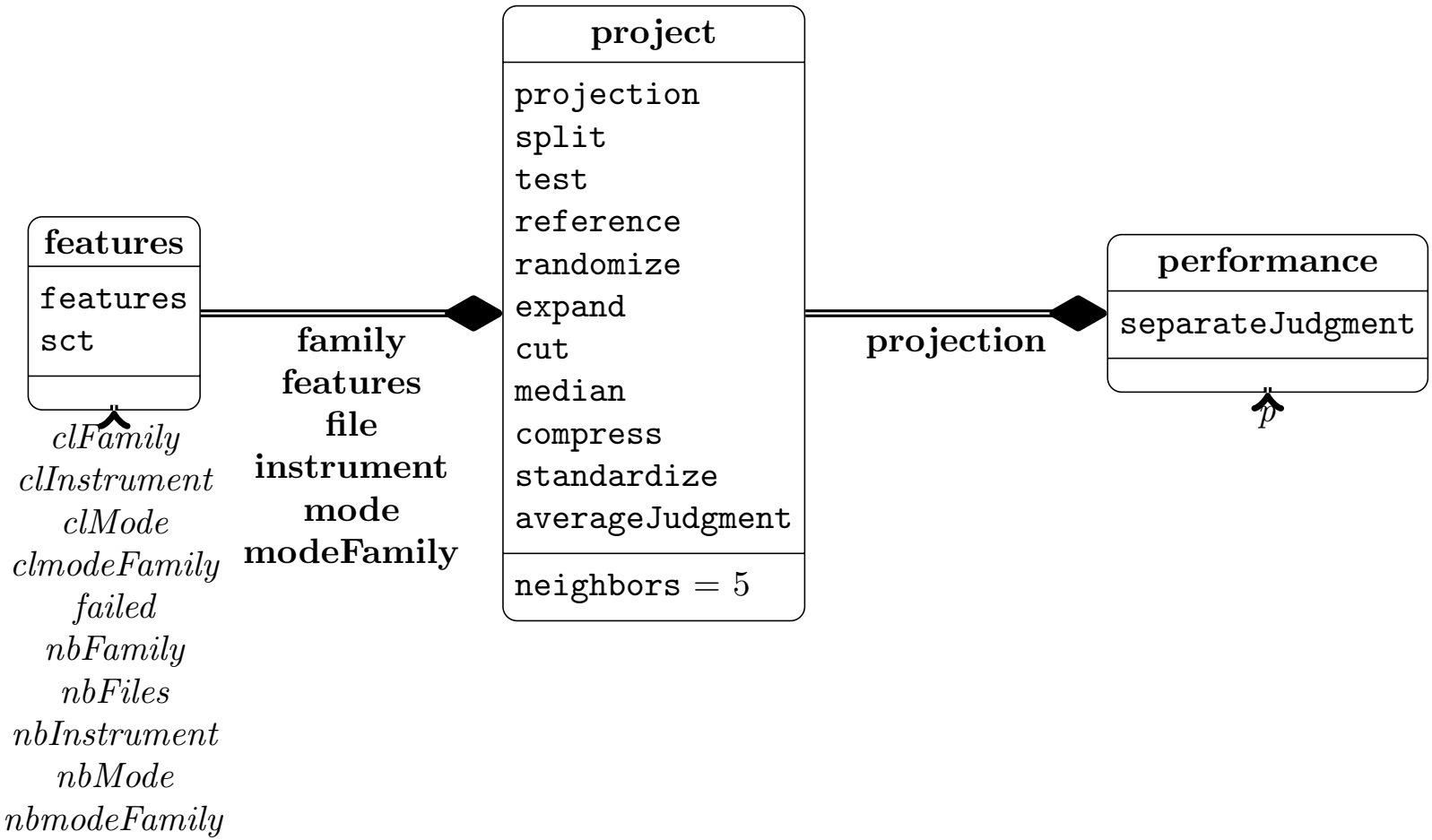
Metric

- the precision @ k is a ranking metric
- it counts the number of items closest to a query item of the same class than the query item

Agenda

- experimental design
- dataset
- features
- projection technique
- control potential overfit of learnt projection with randomization and dimensionality expansion
- control potential overfit of learnt projection with dataset splitting
- study the impact of the observation window size

Factors flow graph



Sol db

nbFiles	25444
nbFamily	16
nbInstrument	33
nbMode	469
nbmodeFamily	143

Sol db

clFamily	1590 ±936
clInstrument	771 ±814
clMode	54 ±59
clmodeFamily	178 ±429

Mel / mfcc: sct: 25, projection: none, split: none, reference: family, randomize: 0, expand: 0

features	cut	standardize	p (%)
mfcc	0	0	85
mfcc	0	1	84
mfcc	1	0	88
mfcc	1	1	**89**
mel		0	53
mel		1	50

Mel / mfcc: sct: 25, projection: none, split: none, reference: modeFamily, randomize: 0, expand: 0

features	cut	standardize	p (%)
mfcc	0	0	35
mfcc	0	1	32
mfcc	1	0	**46**
mfcc	1	1	45
mel		0	19
mel		1	19

Scattering: features: scat, sct: 25, projection: none, split: none, reference: family, randomize: 0

median	compress	standardize	p (%)
0	0	0	64
0	0	1	76
0	1	0	84
0	1	1	83
1	0	0	77
1	0	1	76
1	1	0	**89**
1	1	1	89

Scattering: features: scat, sct: 25, projection: none, split: none, reference: modeFamily, randomize: 0

median	compress	standardize	p (%)
0	0	0	28
0	0	1	38
0	1	0	43
0	1	1	43
1	0	0	40
1	0	1	38
1	1	0	50
1	1	1	**50**

Projection: sct: 25, split: none, reference: family, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1

features	mfcc	scat
none	89	89
lmnn	90	98
lda	87	96

Projection: sct: 25, split: none, reference: modeFamily, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1

features	mfcc	scat
none	45	50
lmnn	48	53
lda	50	52

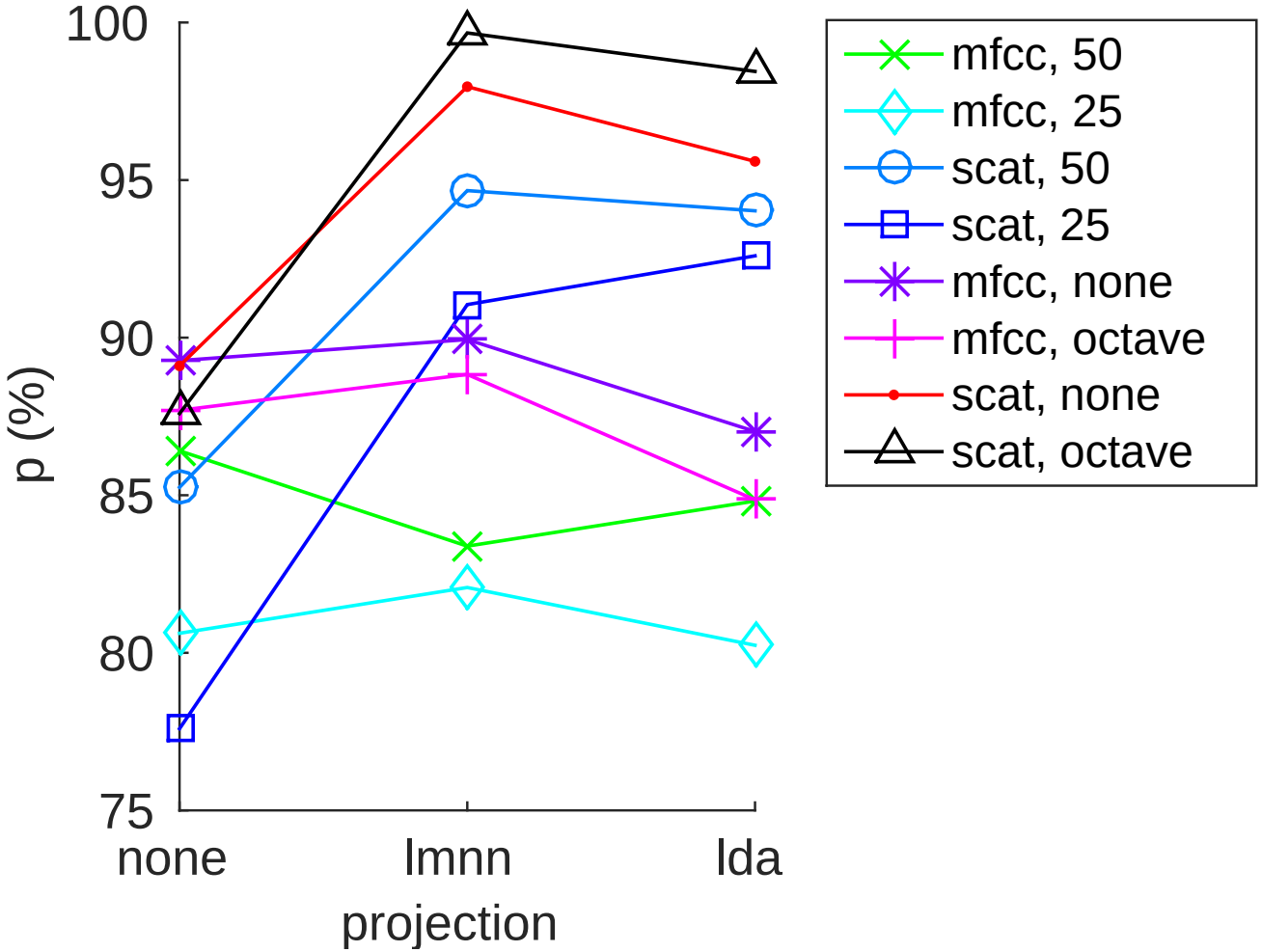
Control learning: sct: 25, split: none, reference: family, cut: 1, median: 1, compress: 1, standardize: 1

features	randomize	expand	none	lmnn	lda
mfcc	0	0	89	90	87
mfcc	0	494	88	91	89
mfcc	1	0	8	8	8
mfcc	1	494	8	9	9
scat	0		89	98	96
scat	1		8	9	9

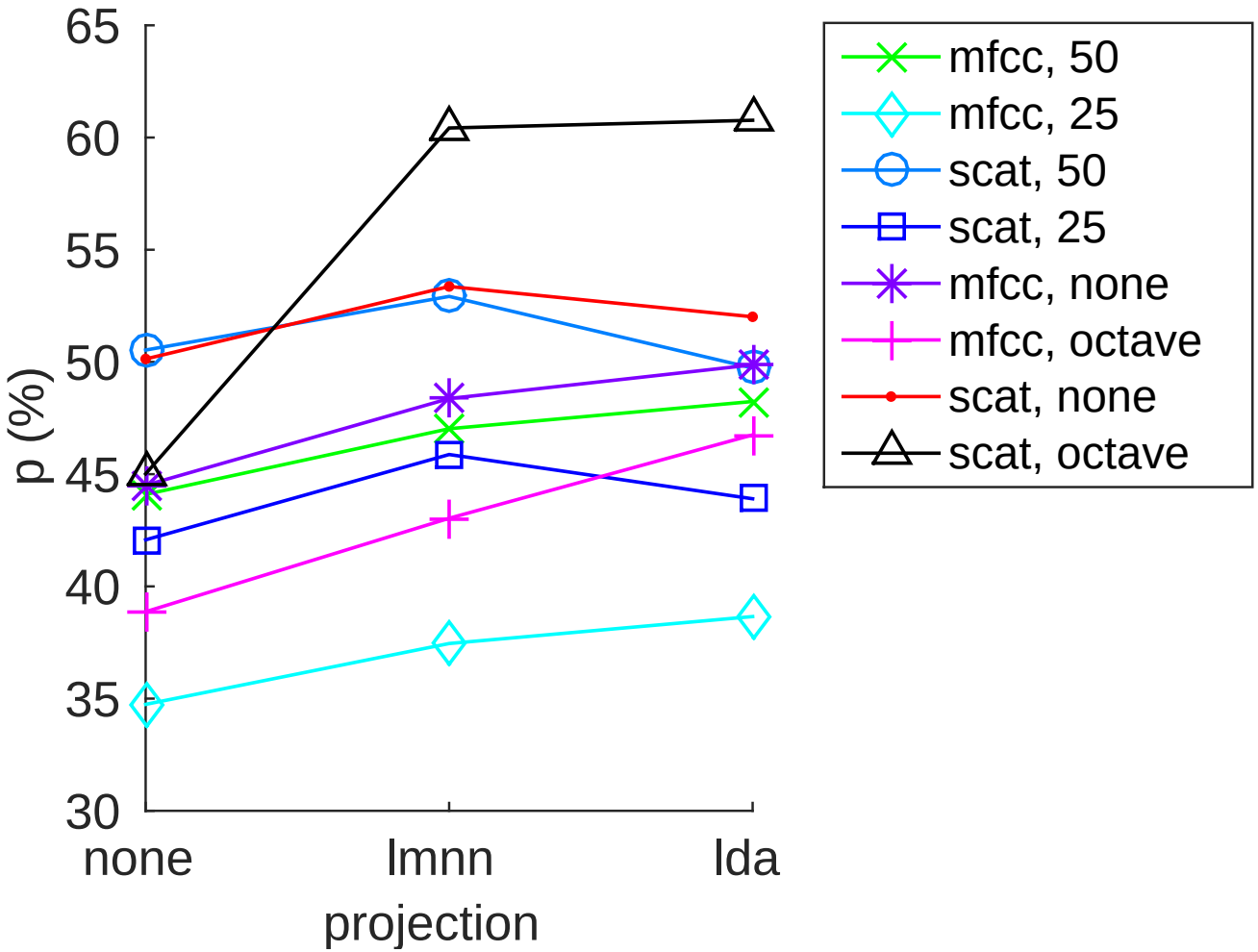
Control learning: sct: 25, split: none, reference: modeFamily, cut: 1, median: 1, compress: 1, standardize: 1

features	randomize	expand	none	lmnn	lda
mfcc	0	0	45	48	50
mfcc	0	494	43	49	49
mfcc	1	0	5	5	5
mfcc	1	494	5	4	5
scat	0		50	53	52
scat	1		5	4	5

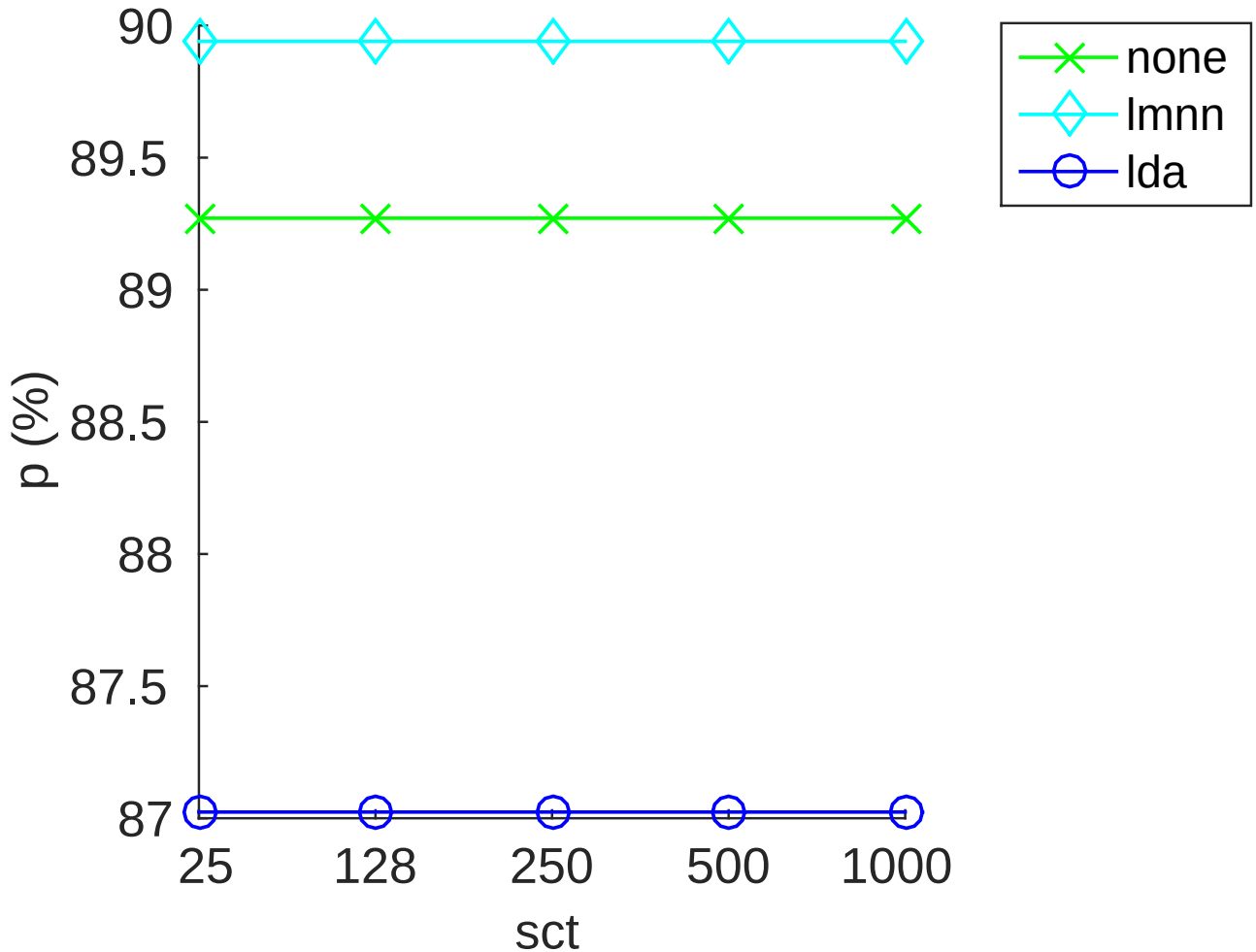
db splitting: sct: 25, test: 1, reference: family, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1



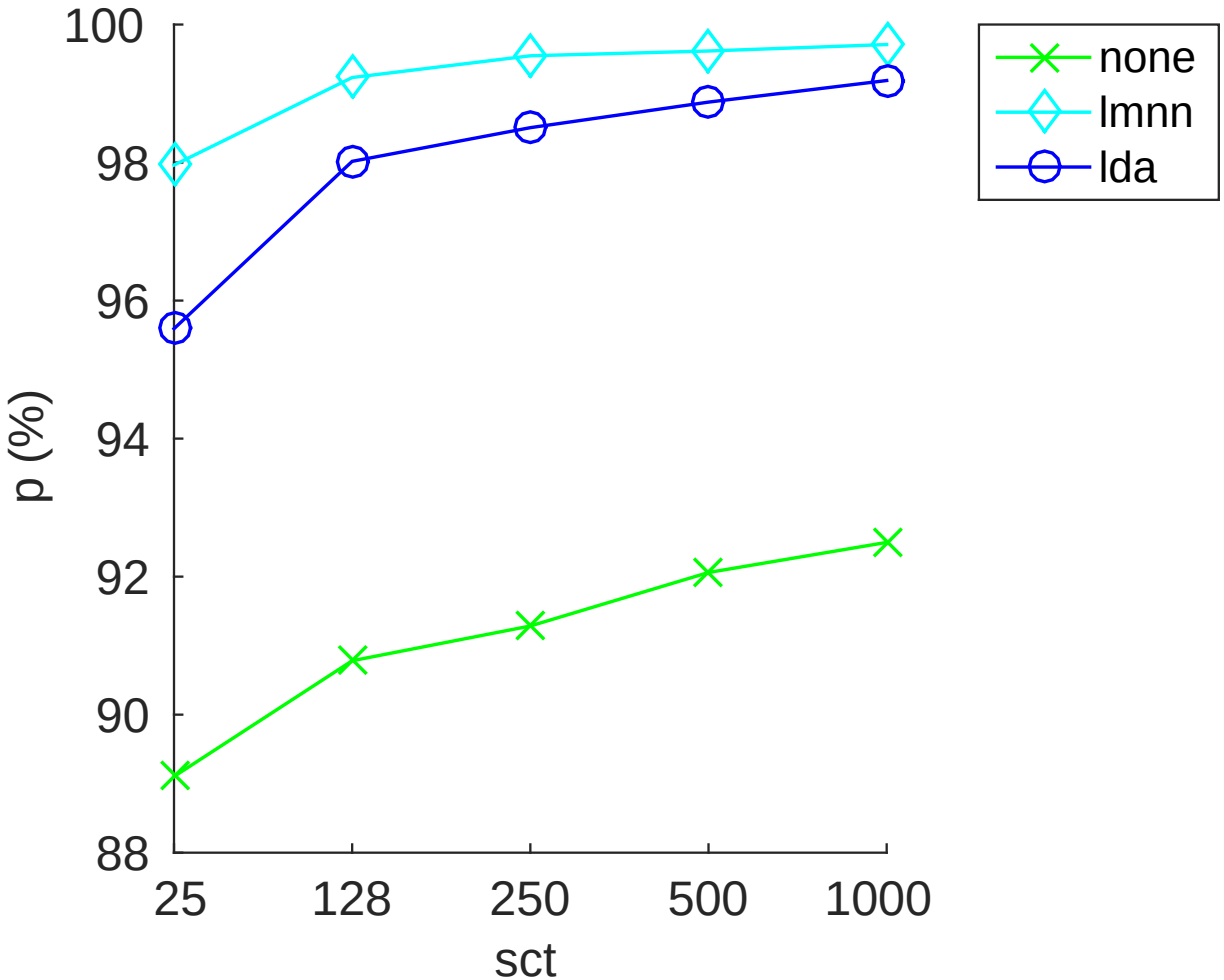
db splitting: sct: 25, test: 1, reference: modeFamily, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1



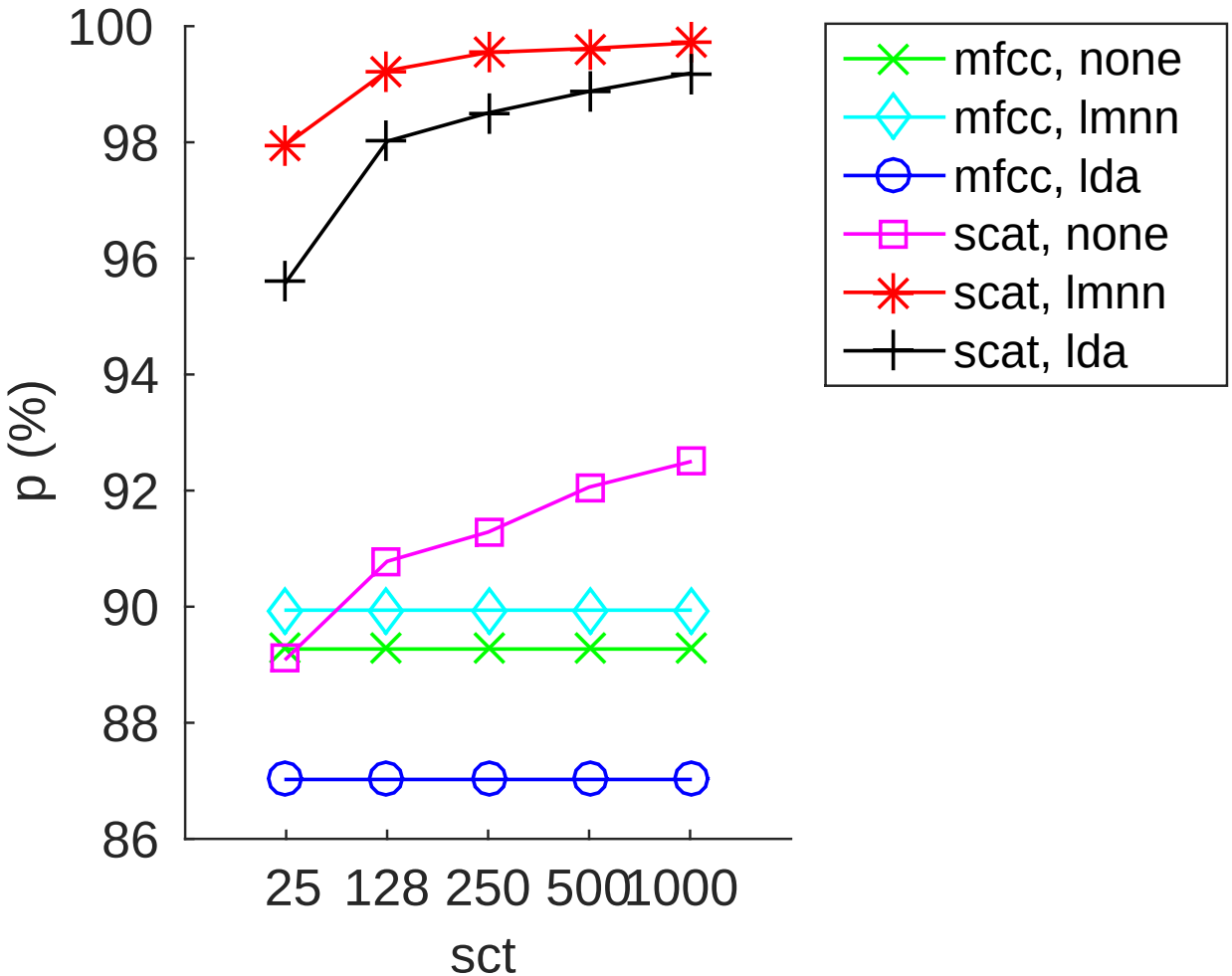
T: features: mfcc, reference: family, split: none, randomize: 0, expand: 0, cut: 1, standardize: 1



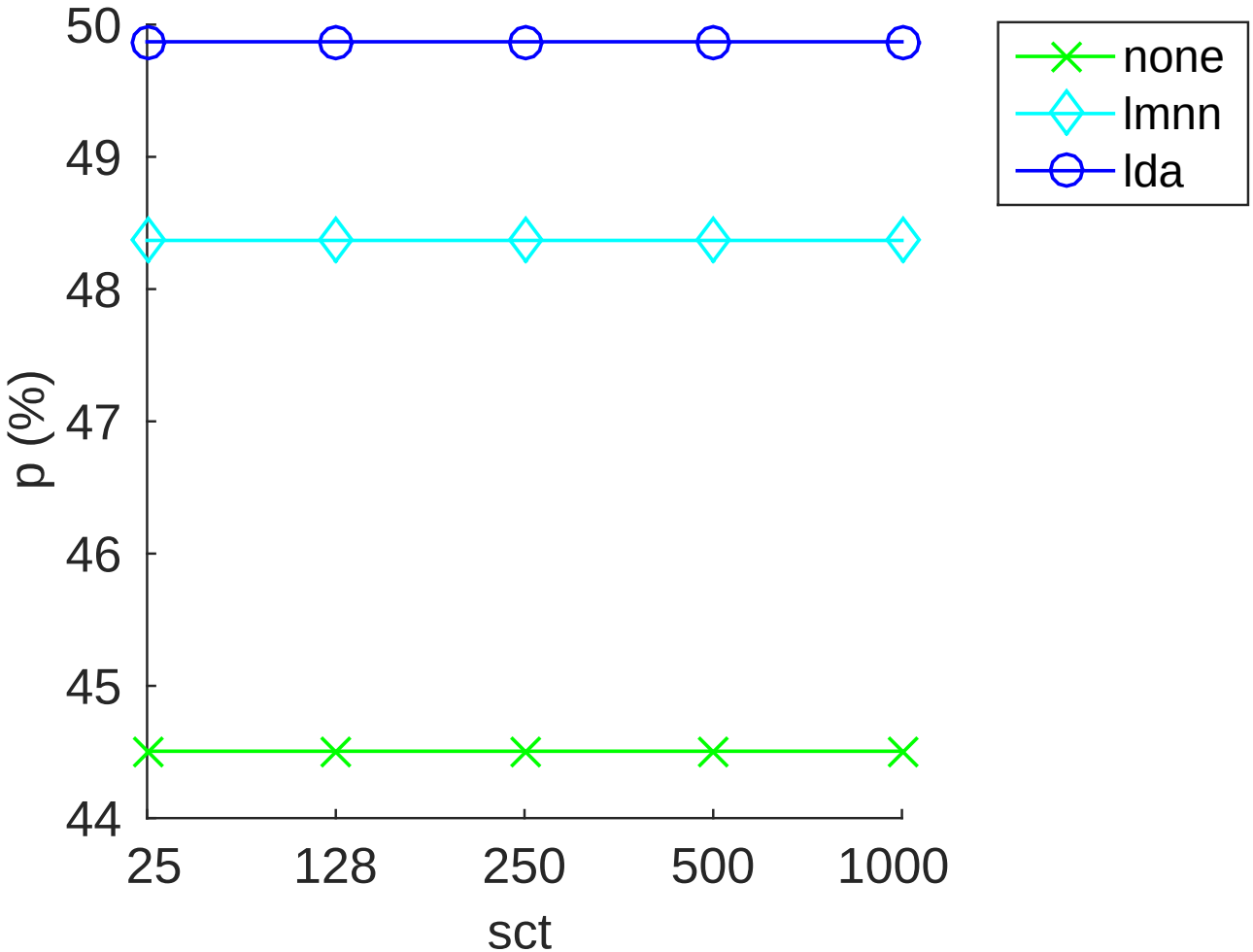
T: features: scat, reference: family, split: none, randomize: 0, median: 1, compress: 1, standardize: 1



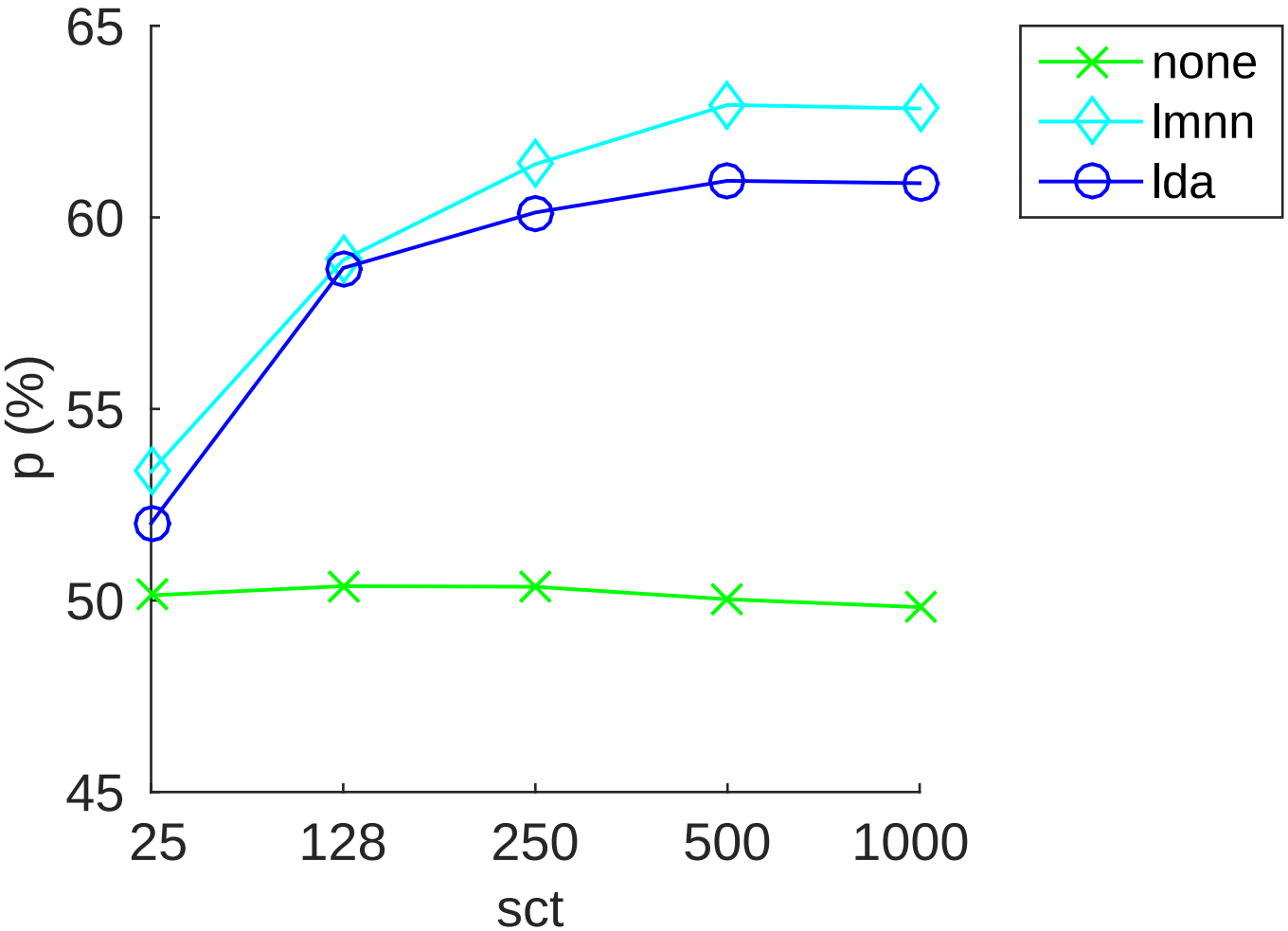
T: reference: family, split: none, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1



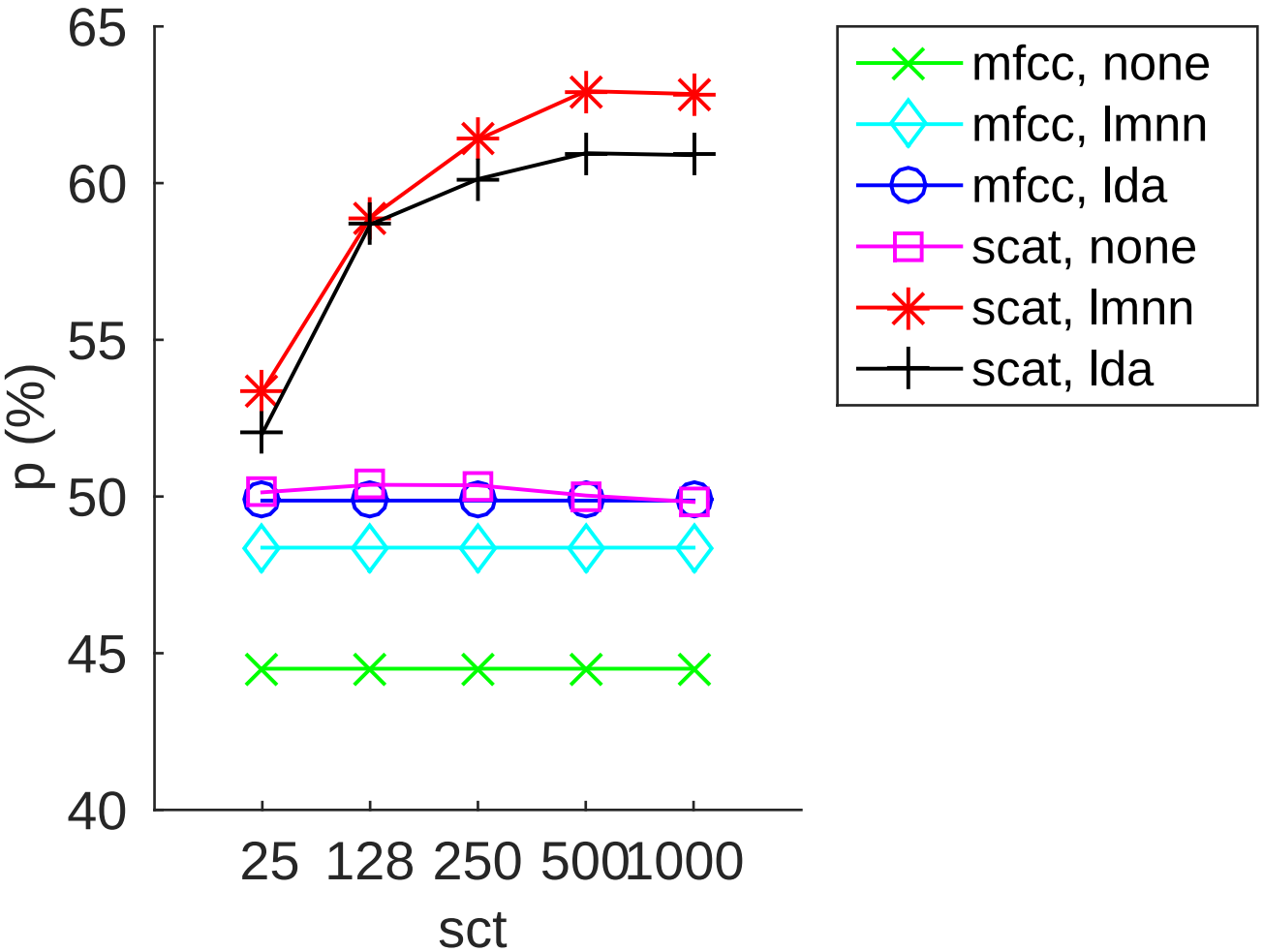
T: features: mfcc, reference: modeFamily, split: none, randomize: 0, expand: 0, cut: 1, standardize: 1



T: features: scat, reference: modeFamily, split: none, randomize: 0, median: 1, compress: 1, standardize: 1



T: reference: modeFamily, split: none, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1



conclusion: split: none, randomize: 0, expand: 0, cut: 1, median: 1, compress: 1, standardize: 1