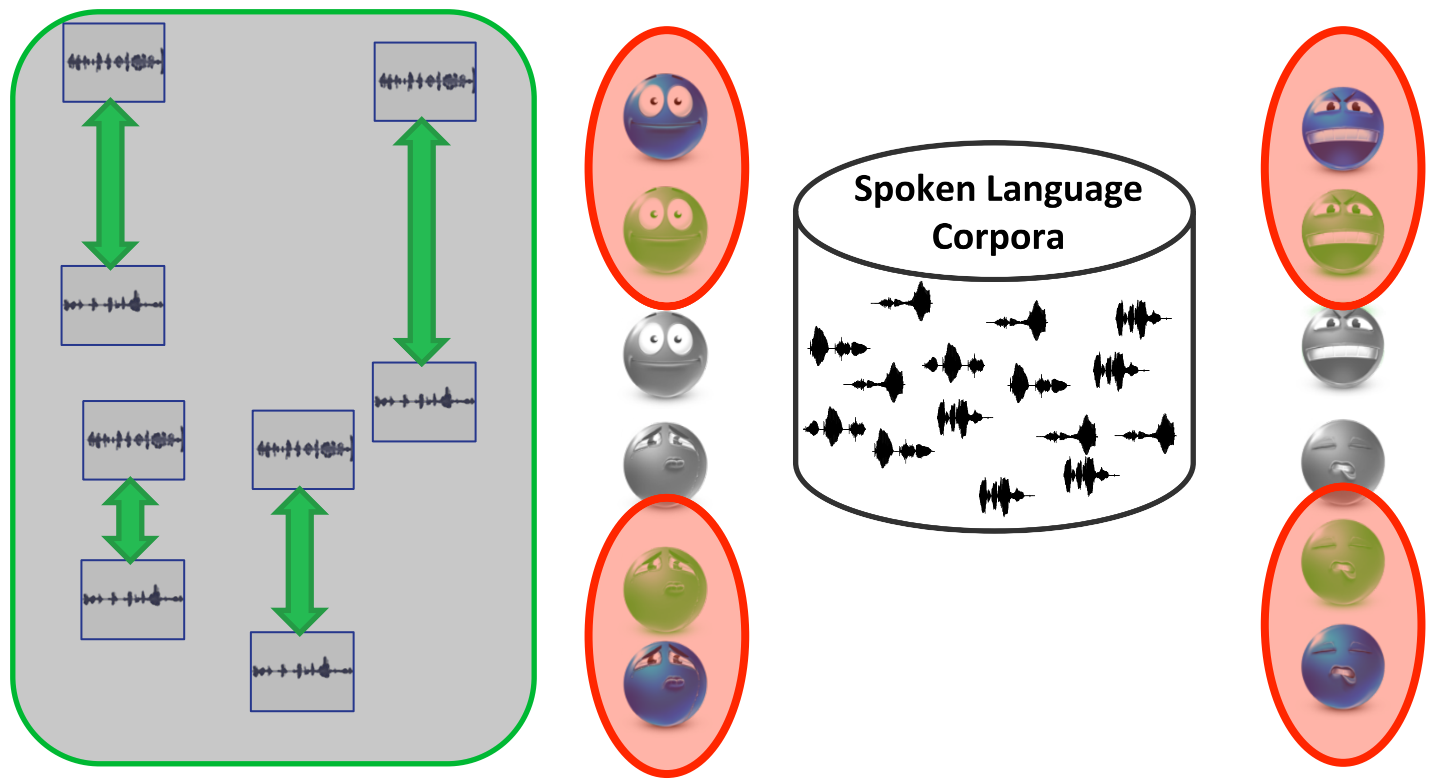
Practical Considerations on the Use of Preference Learning for Ranking Emotional Speech

REZA LOTFIAN AND CARLOS BUSSO

Multimodal Signal Processing (MSP) lab
The University of Texas at Dallas
Erik Jonsson School of Engineering and Computer Science



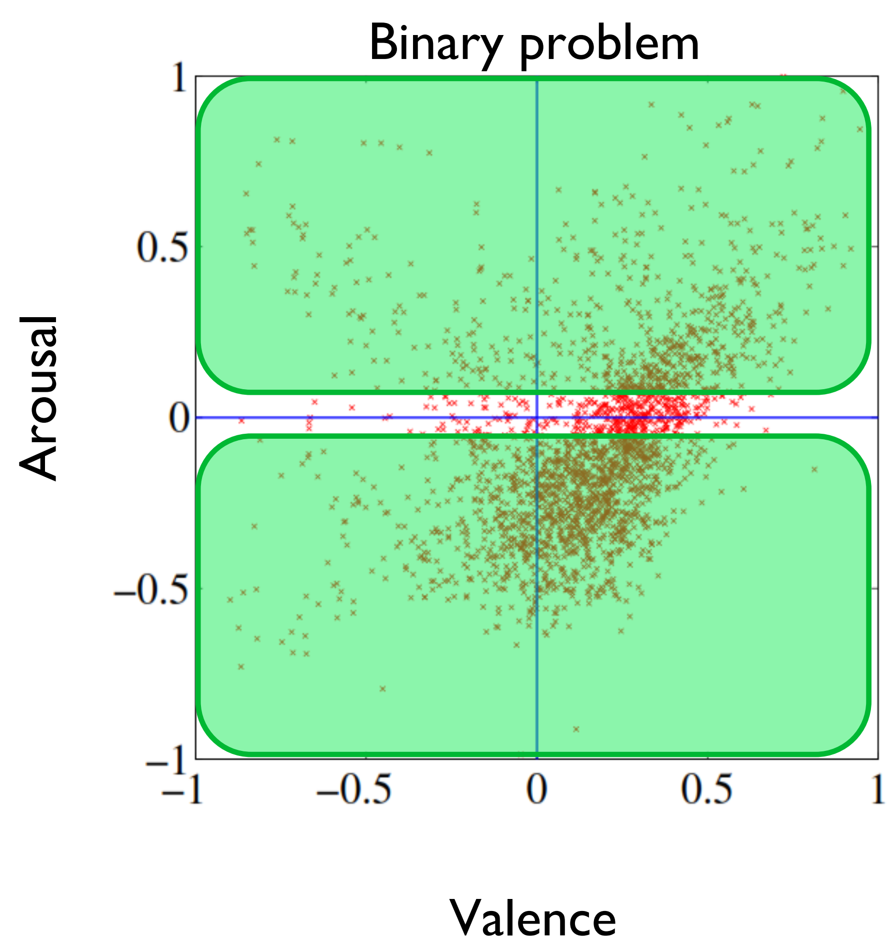
March. 25th, 2016

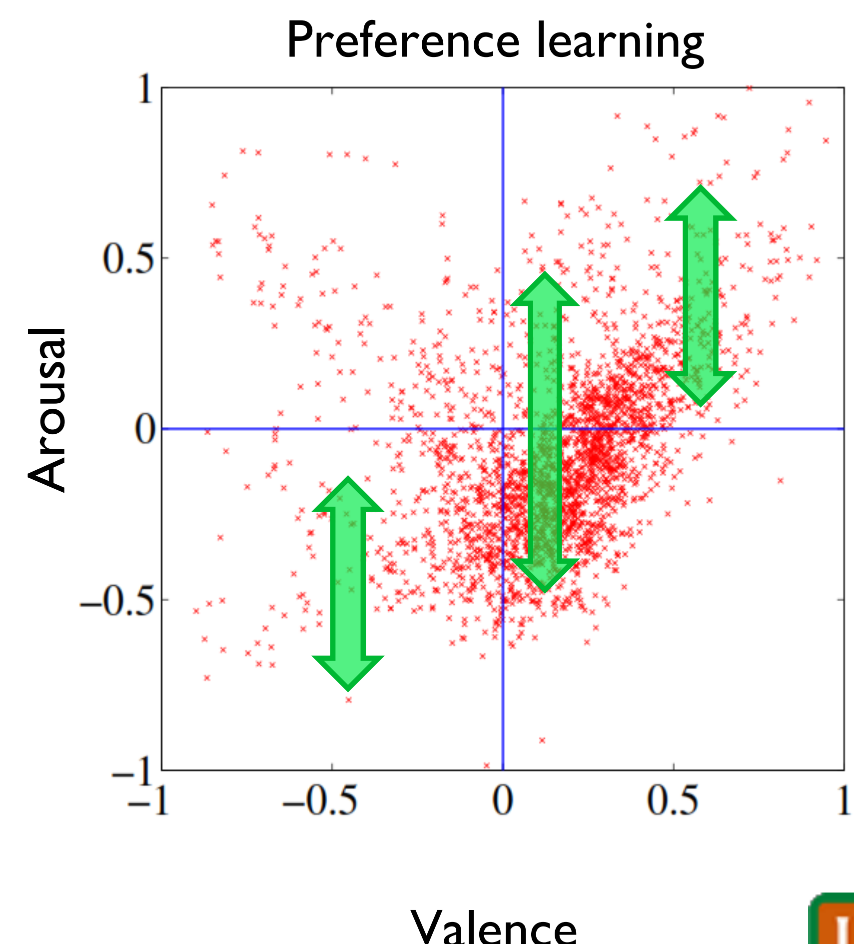
Motivation

- Creating emotions aware human computer interaction
  - Binary or multi-class speech emotion classification
- Preference learning offers an appealing alternative
  - Widely explored in images, music, video, text
  - Few studies on preference learning for emotion recognition
- Emotion retrieval from speech
  - Call centers
  - Healthcare applications

Definition of the problem

- Binary/multiclass classification versus preference learning
  - Binary class: $O(n)$training samples
    - low or high arousal?
  - Preference learning: $O(n^2)$ training samples
    - Is the arousal level of *sample1* higher than arousal level of *sample2*?



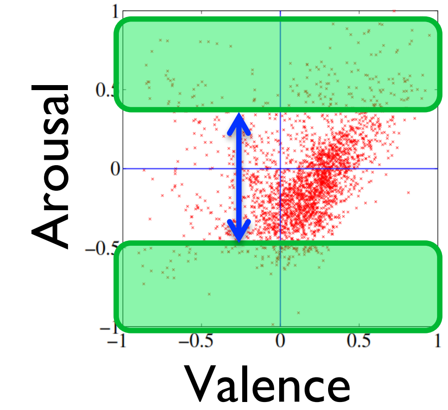


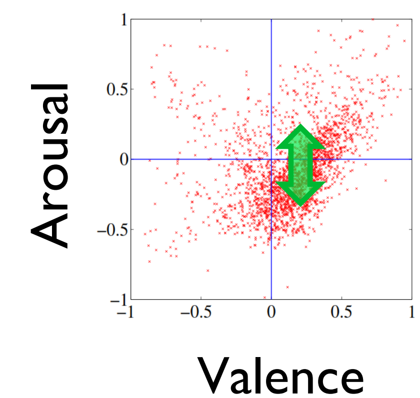
Definition of the problem

- Absolute ratings of the emotions are noisy
- Binary problem
  - Remove samples close to boundary of different classes
- Preference learning

$$e^{s_1}_{arousal} - e^{s_2}_{arousal} > margin \longrightarrow \quad s_1 \succ_{arousal} s_2$$

- Questions
  - How many samples are available for training?
  - How reliable are the labels?
  - What are the optimum parameters? (margin + size of training set)
  - How does it compare to alternative methods?



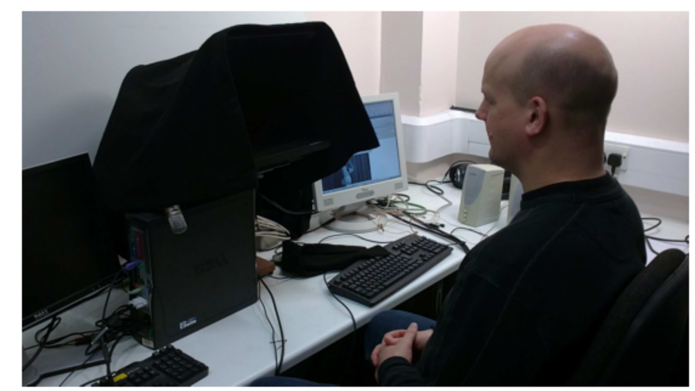


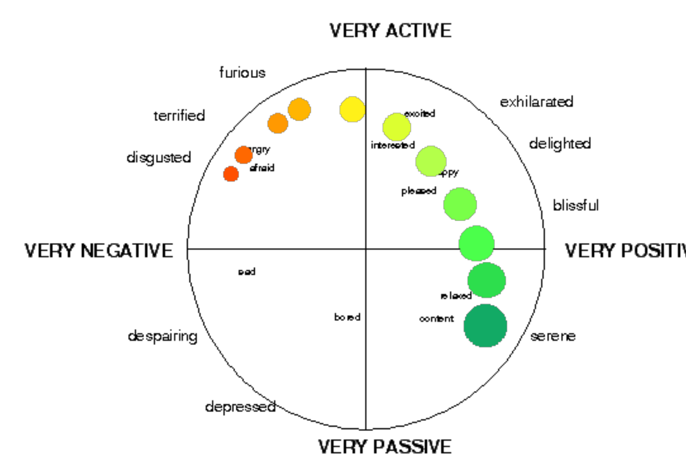
SEMAINE database

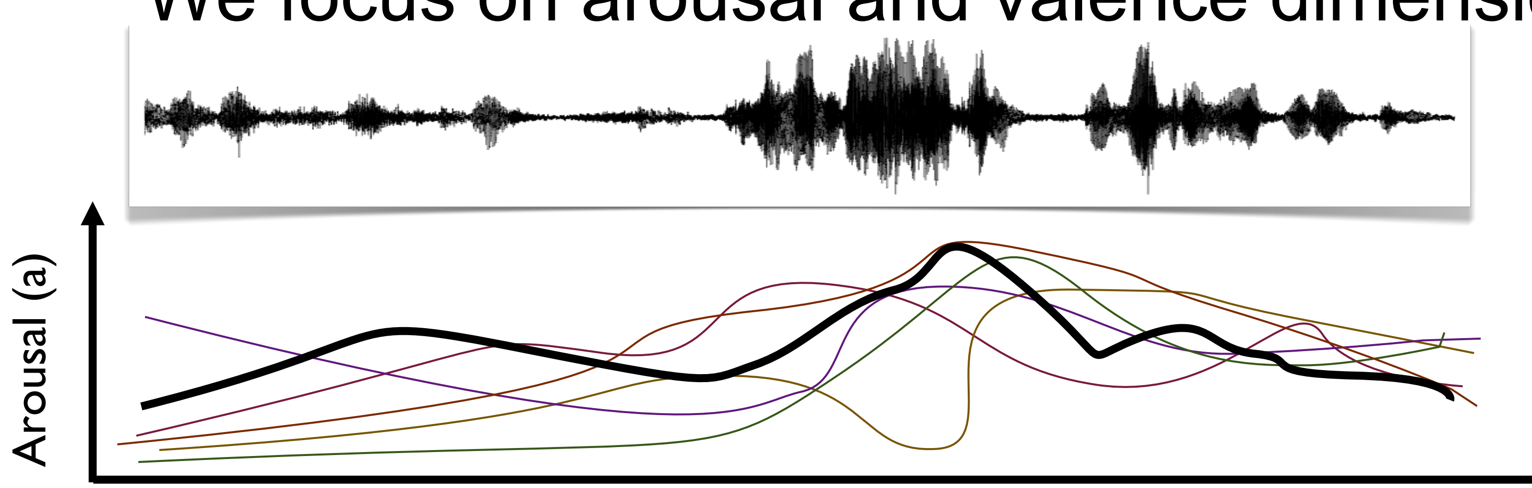
- Emotionally colored machine-human interaction
  - Sensitive artificial listener framework
  - Only solid SAL used (operator was played with another human)
  - 91 sessions, 18 subjects (user)
- Time-continuous dimensional labels
  - Annotated by FEELTRACE
  - We focus on arousal and valence dimensions

User

Operator







Acoustic features

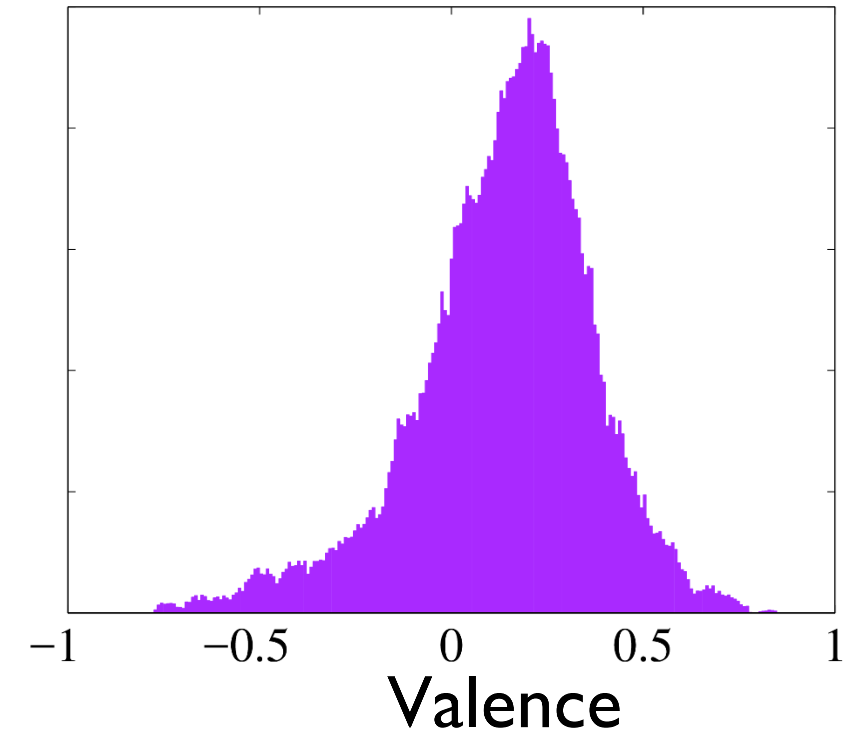
- Speaker state challenge feature set at INTERSPEECH 2013
  - 6308 high level descriptors
  - OpenSMILE toolkit
  - Feature selection (separate for arousal and valence)
  - Step 1: 6308→500
    - Information gain separating binary labels (e.g., low vs high arousal)
  - Step 2: 500→50
    - Floating forward feature selection
    - Maximizing the precision of retrieving 10% top and 10% bottom

How many samples are available for training?

- Applying thresholds increases the reliability of training labels
  - Removing ambiguous labels
- Larger margin:

  \+ more reliable labels

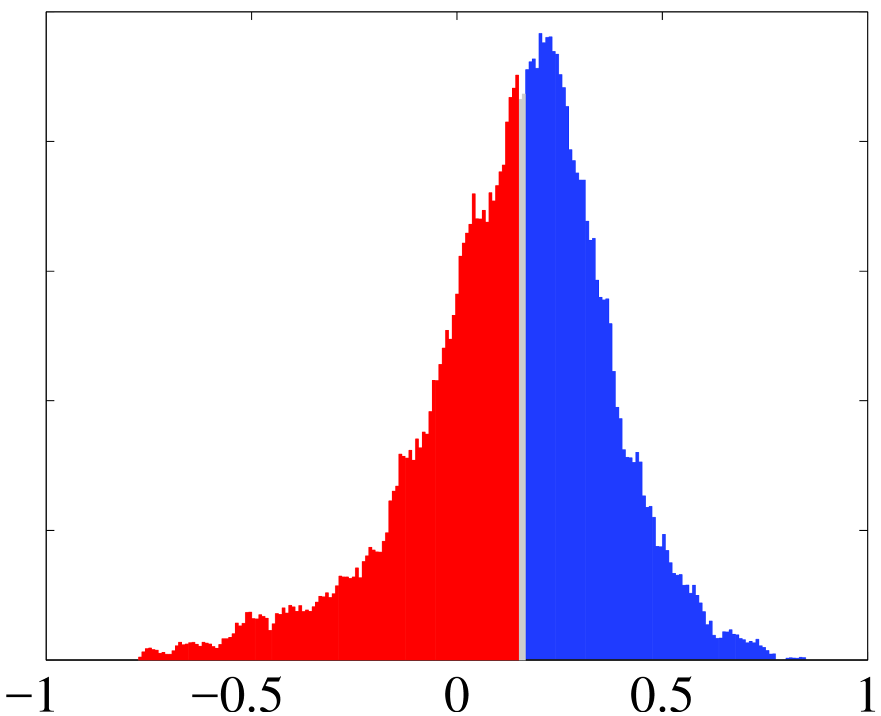
  − less samples for training

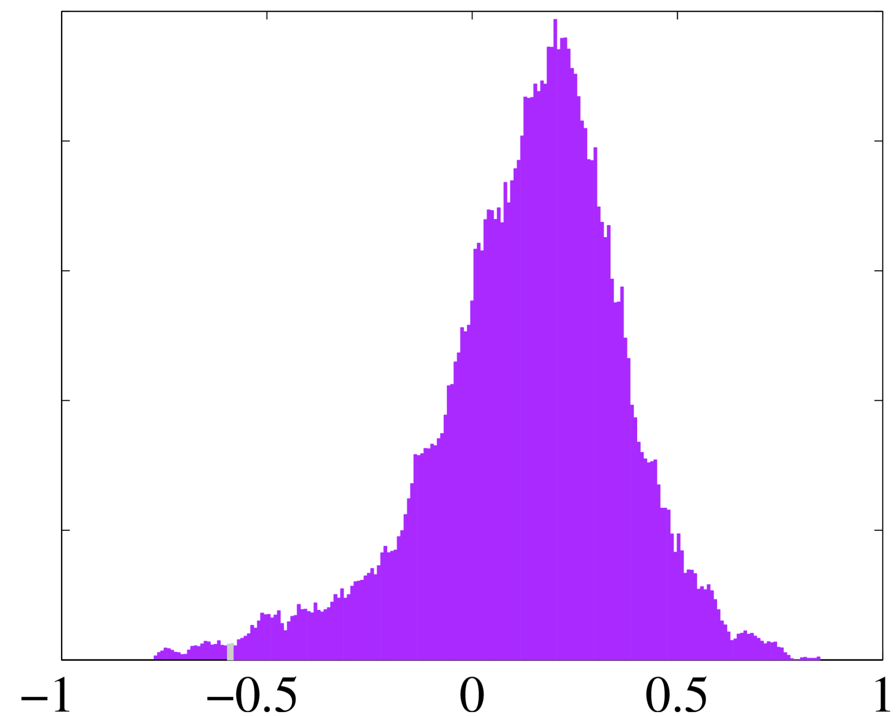


- How does different margins affect available training samples in binary and pairwise problems?

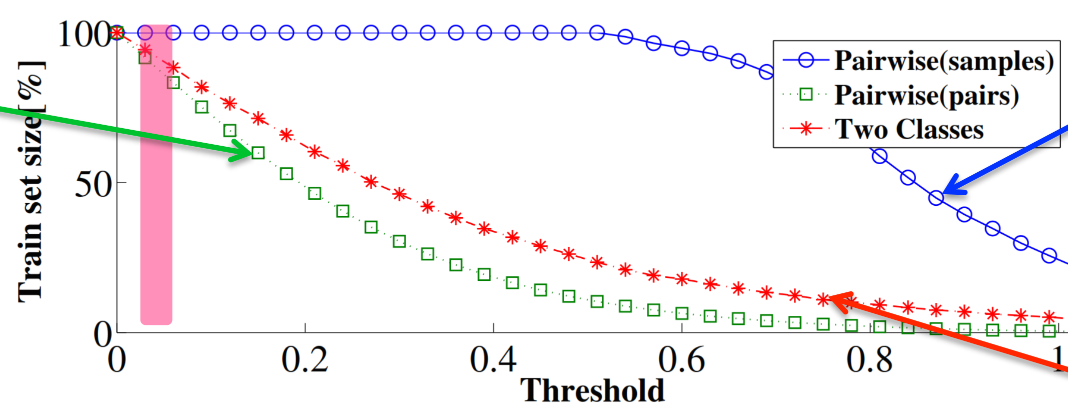
How many samples are available for training?

- Binary labels
- Pairwise labels





Proportion of potential pairwise comparisons included in training testing sets



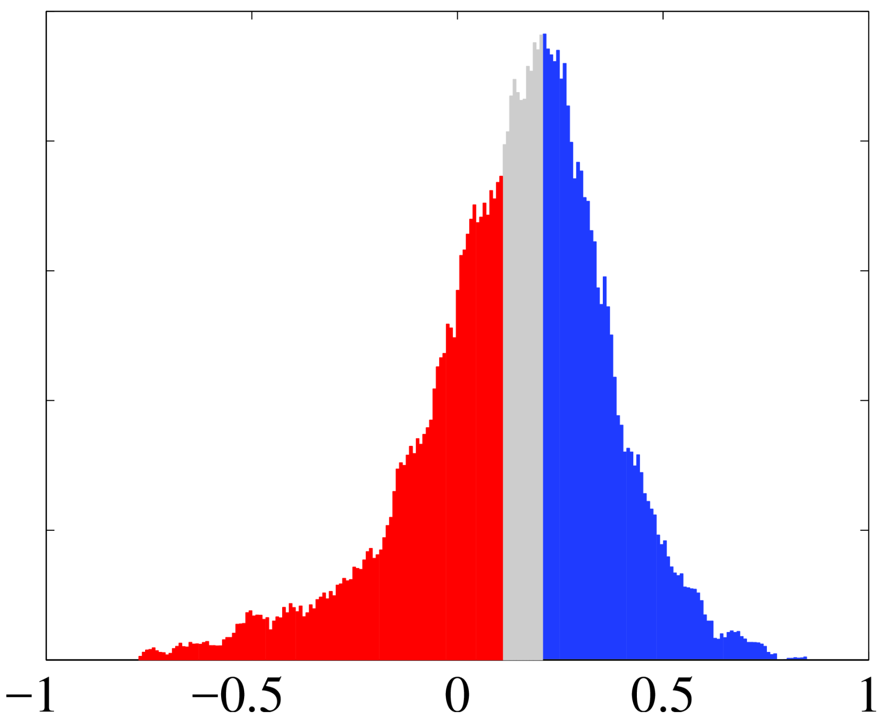
Valence

Samples included in training/testing sets

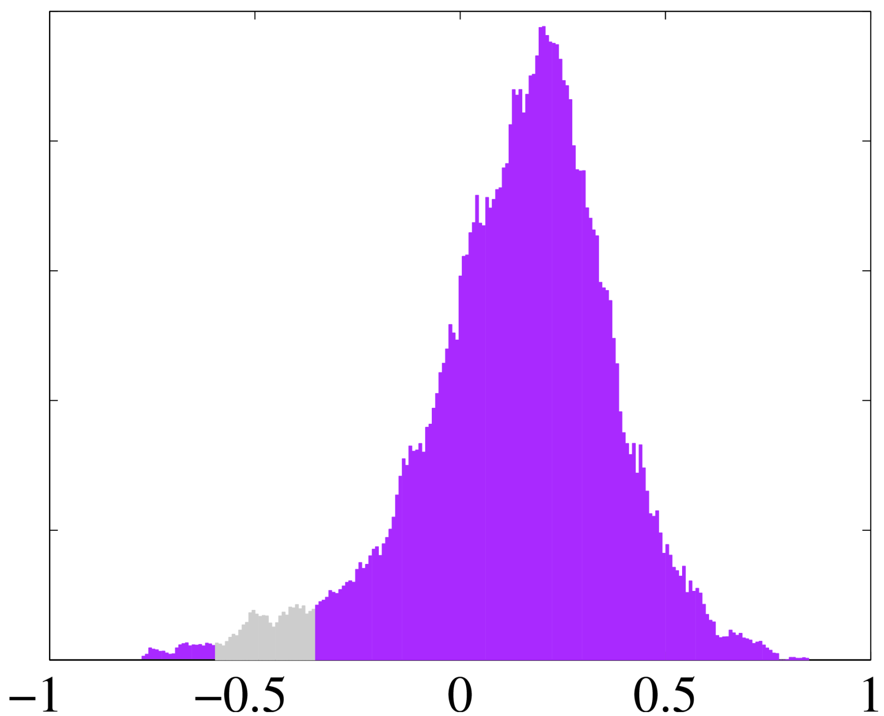
Samples included in binary classification

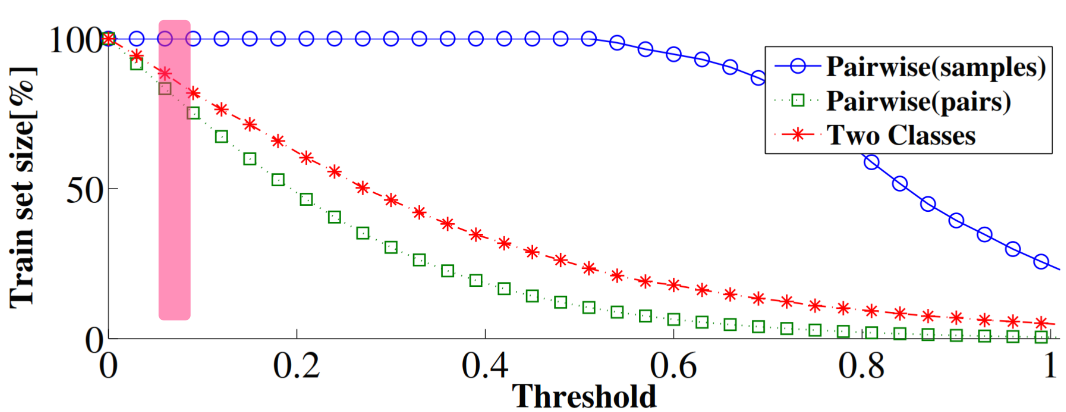
How many samples are available for training?

- Binary labels



- Pairwise labels

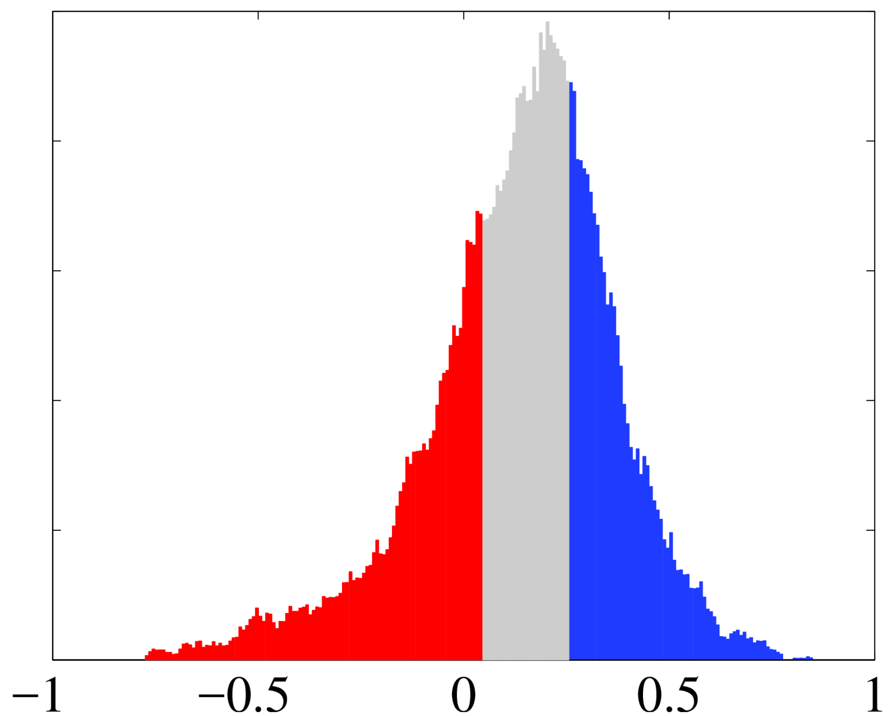


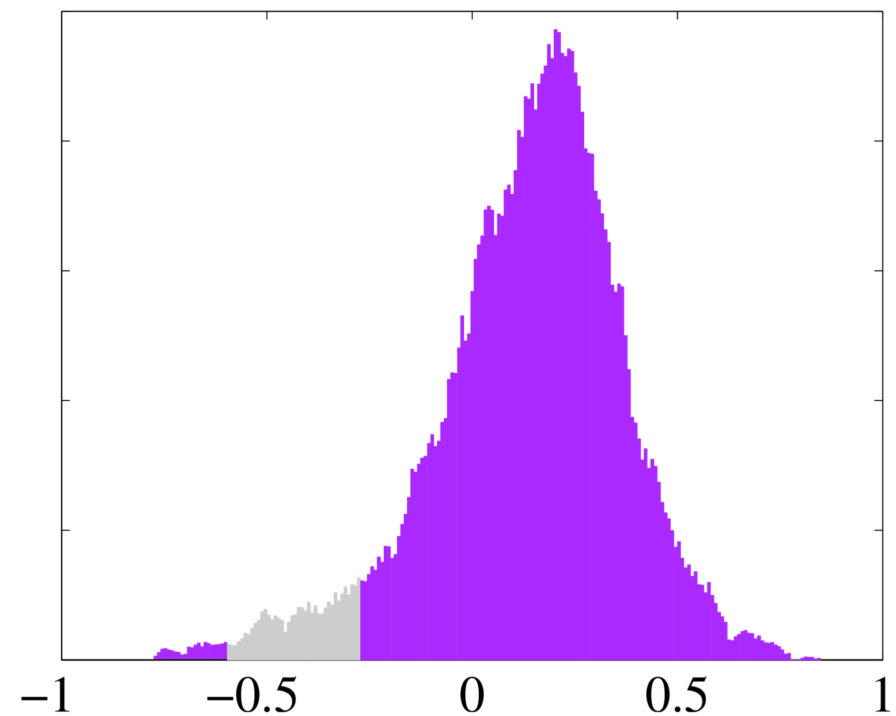


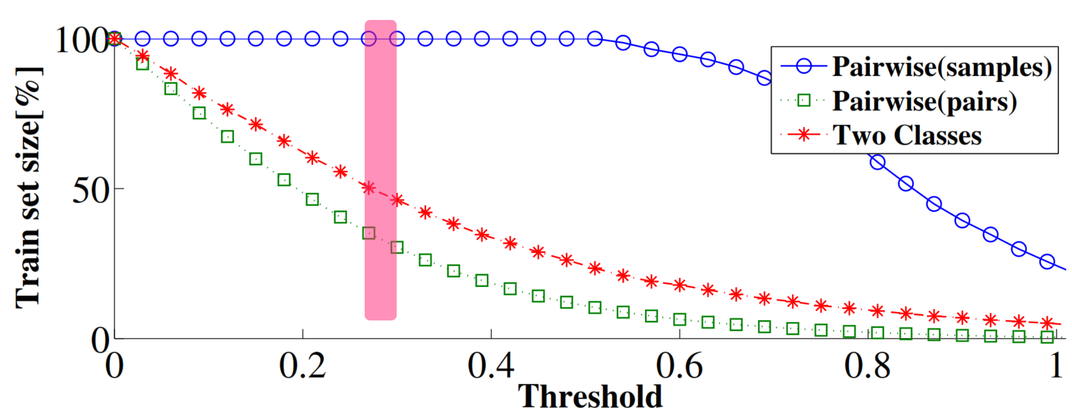
Valence

How many samples are available for training?

- Binary labels
- Pairwise labels



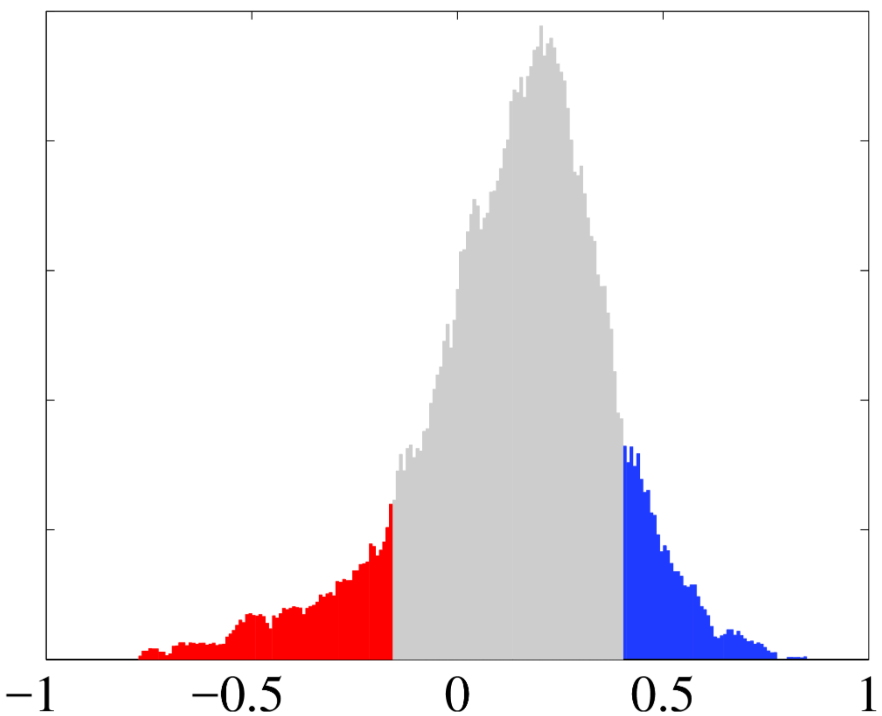




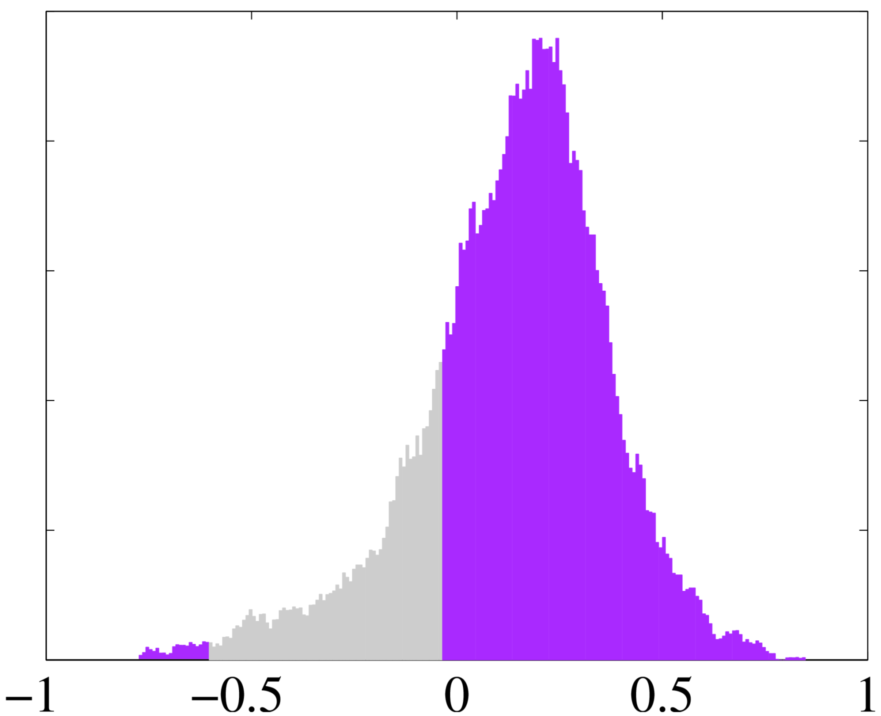
Valence

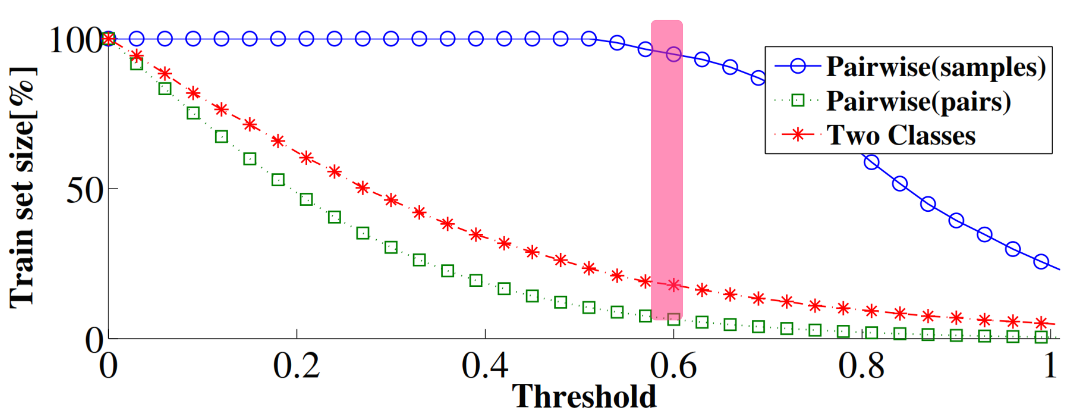
How many samples are available for training?

- Binary labels



- Pairwise labels

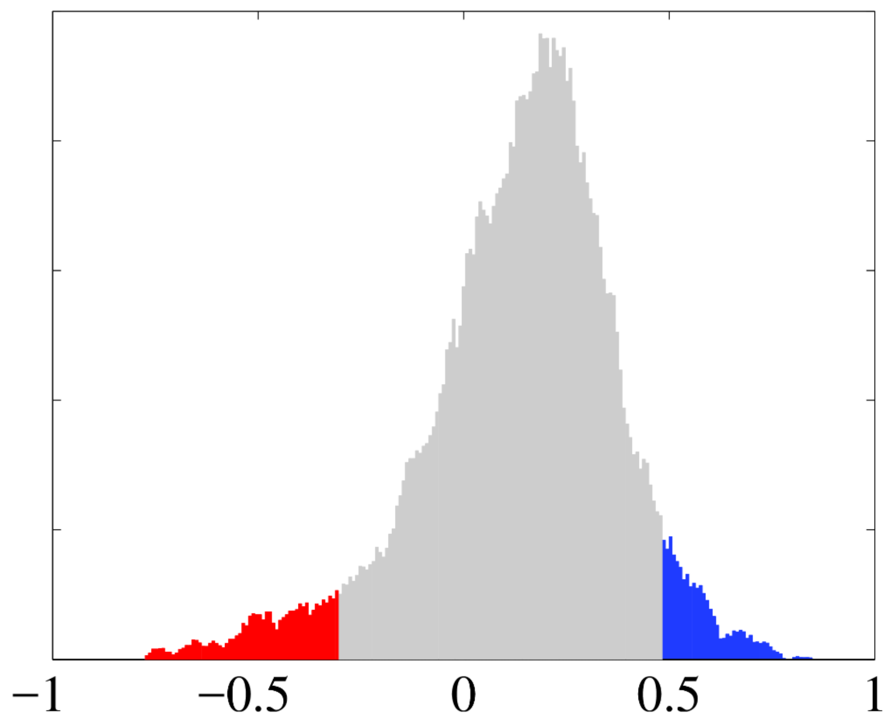


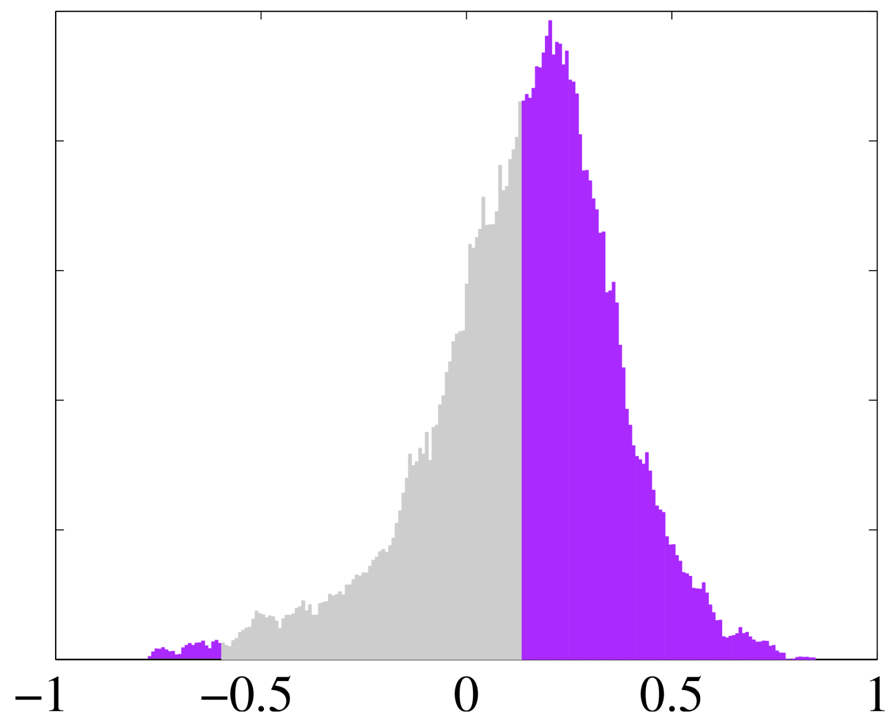


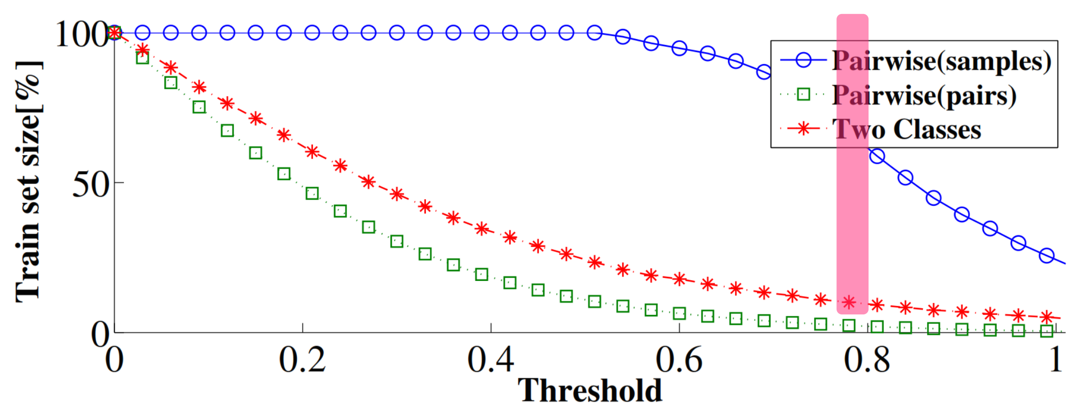
Valence

How many samples are available for training?

- Binary labels
- Pairwise labels



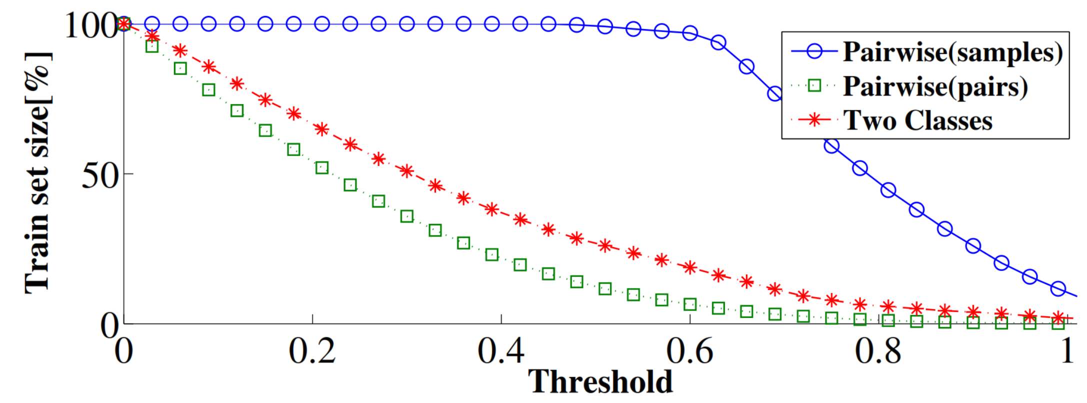




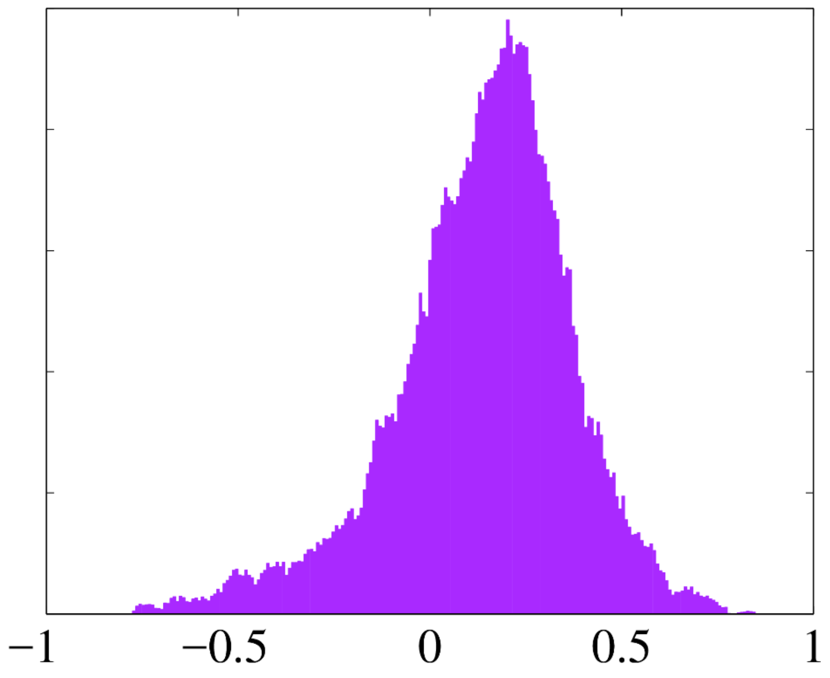
Valence

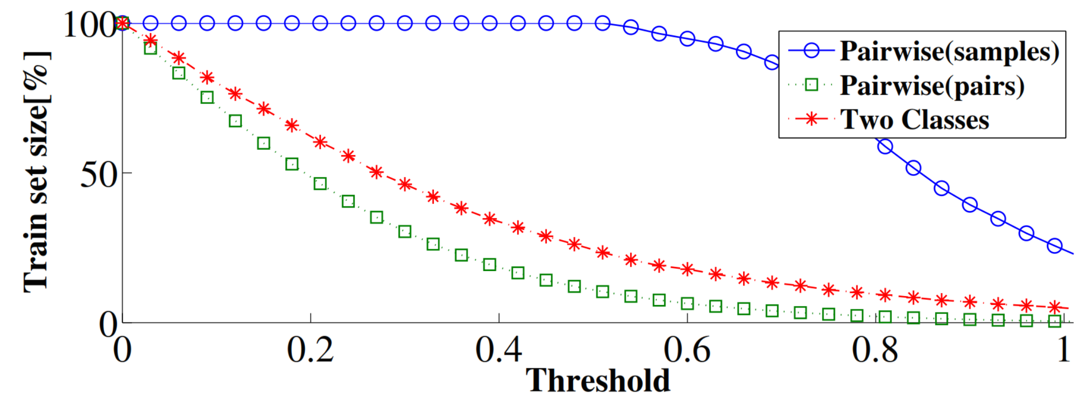
How many samples are available for training?

- More samples remain in training set in pairwise classification



Arousal



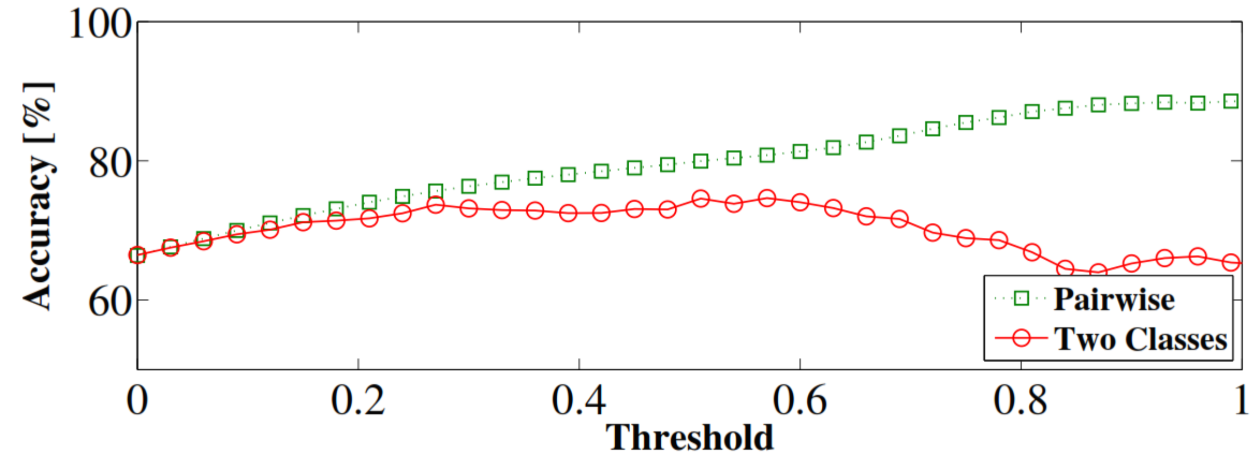


Valence

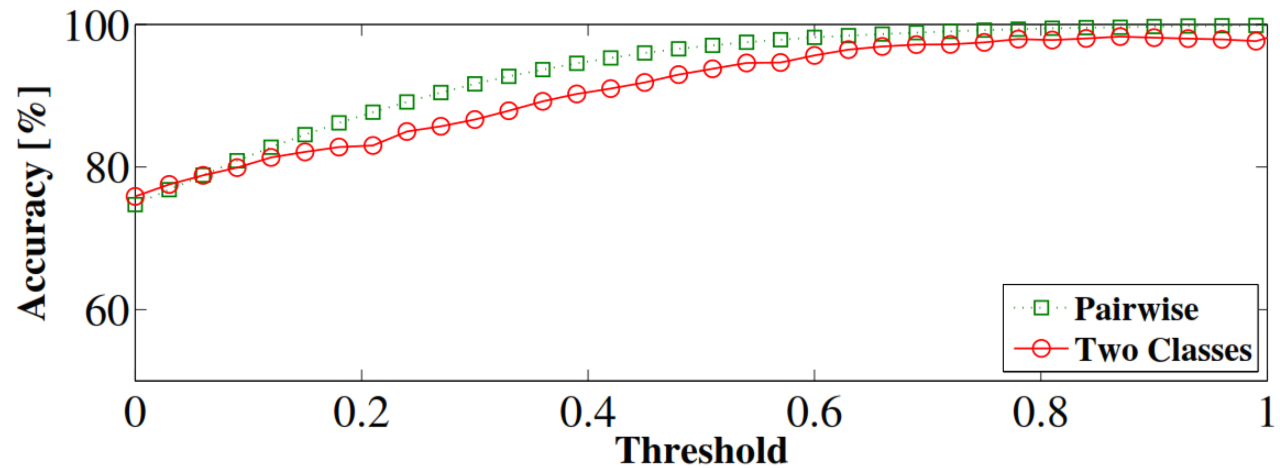
How reliable are the labels?

- Precision of subjective evaluations
  - Find the average of ratings for all evaluators except one
  - Compare his/her labels to aggregated score
- Pairwise labels: higher agreement between subjective evaluations for different thresholds evaluations
  - Few sample for margin >0.7 lead to noisy binary labels

Arousal



Valence



What are the optimum parameters?

- Rank SVM problem
  - $x_i^{(1)}$ and $x_i^{(2)}$ are feature vectors of pair i where $s{\downarrow}1$ is preferred over $s{\downarrow}2$

$$\min_{w,\zeta} \qquad \frac{1}{2}\|w\|^2 + C\sum\nolimits_i \zeta_i$$

$$\text{subject to } \langle w, \boxed{(x_i^{(1)} - x_i^{(2)})}\rangle \geq 1 - \zeta_i, \zeta_i \geq 0 \text{ for } i \in [l]$$

ζ_i: nonzero slack variable

C: soft margin variable

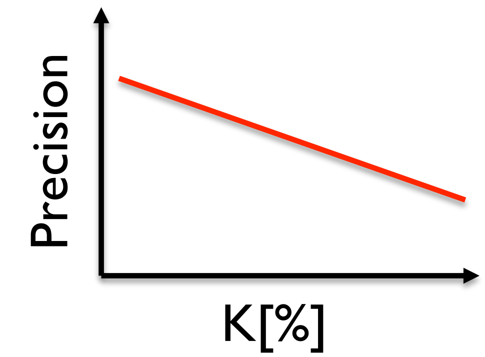
- Testing: s_1 is preferred over s_2 if $\langle w, (x_i^{(1)} - x_i^{(2)})\rangle \geq 0$

Preference learning

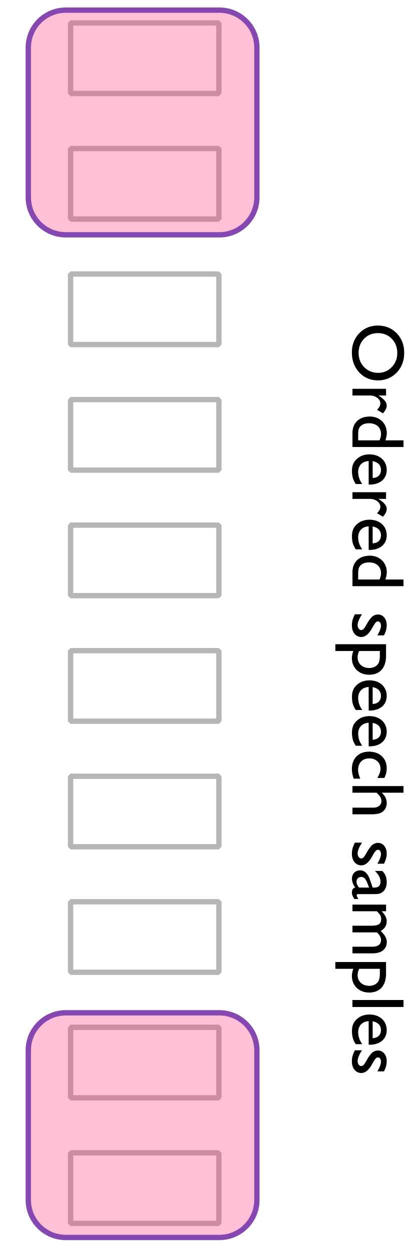
- Training samples
  - Speaker independent partitioning for
    - Development (feature selection): 8 randomly selected speakers
    - Cross validation: 5 speakers for training, 5 speakers for testing
  - Set of pairwise preferences (rankings of length 2)
    - Samples that satisfy the margin's threshold are selected
    - Different sample size is evaluated

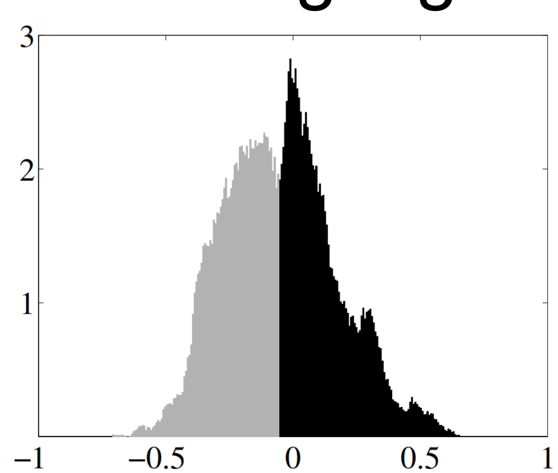
Measure of retrieval performance

Precision at K (P@K)

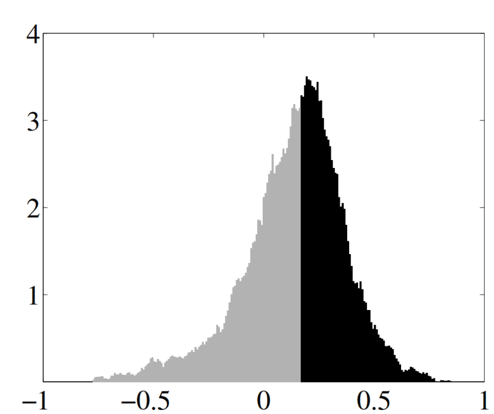


- Speech samples ordered by Rank-SVM
- Select K/2 samples from top, K/2 samples from bottom
  - Example: P@100 → binary classification
- Success if the sample is in the right side
  - Black → high; Gray → bottom
  - We can compare this approach to other machine learning algorithm





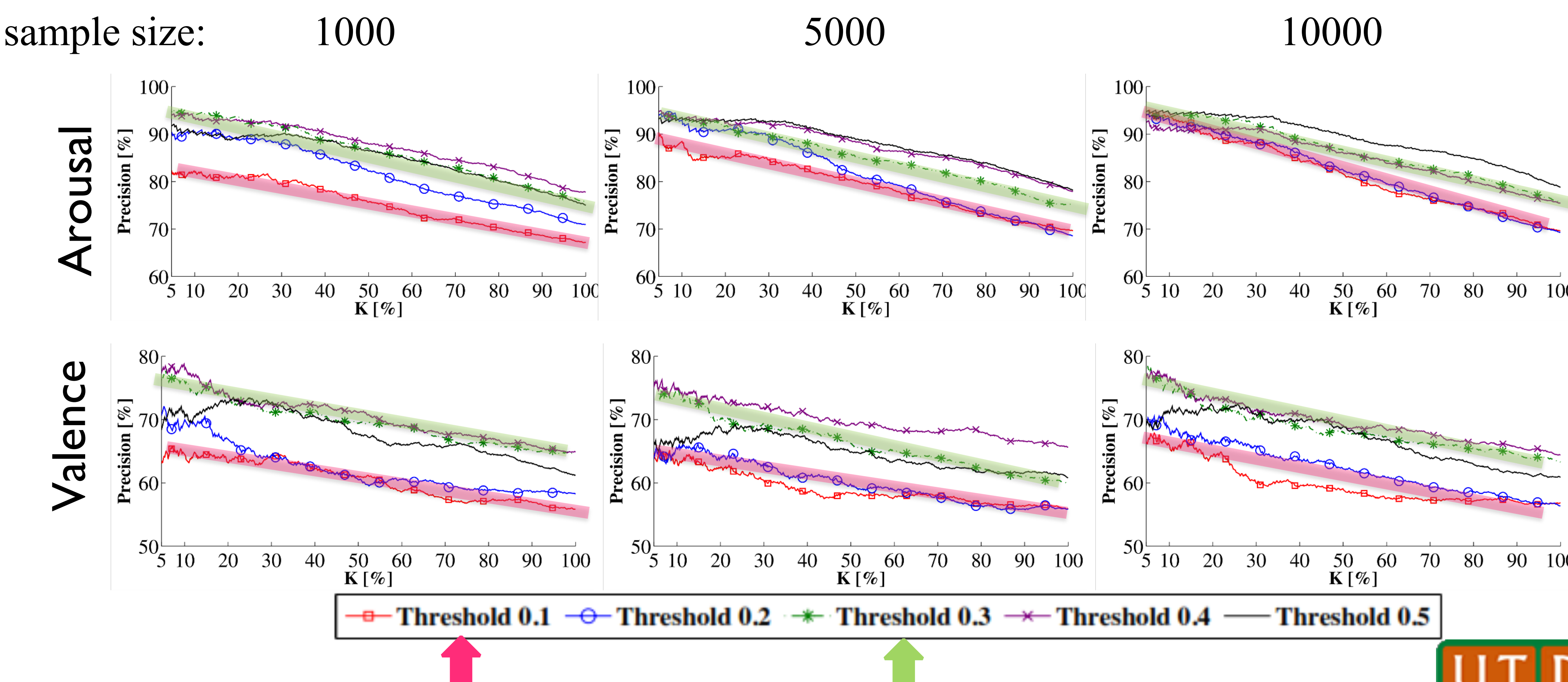
Arousal



Valence

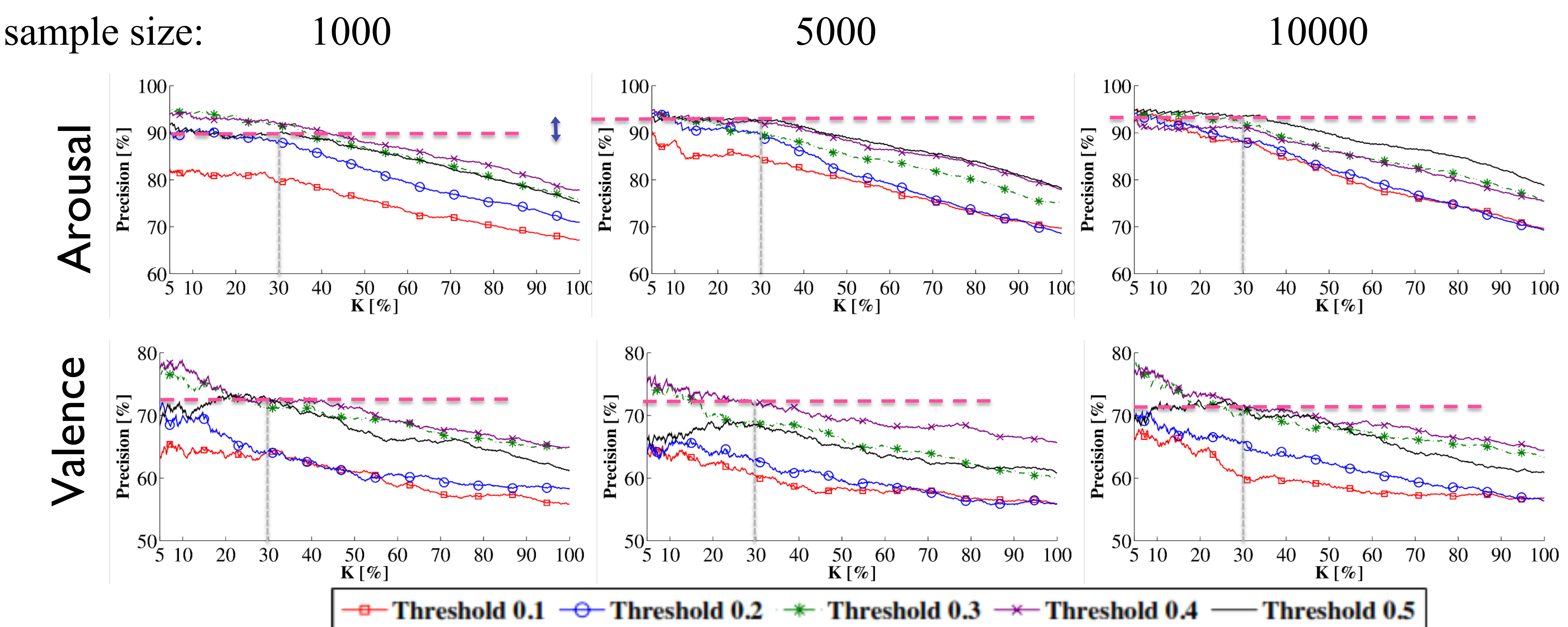
What are the optimum parameters?

- Optimum margin threshold
  - Arousal → 0.5
  - Valence → 0.4



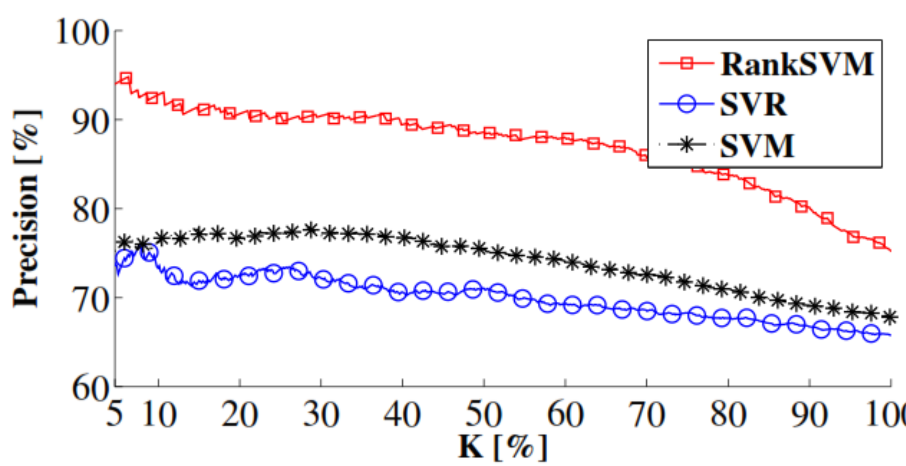
What are the optimum parameters?

- Optimum sample size
  - ~5000

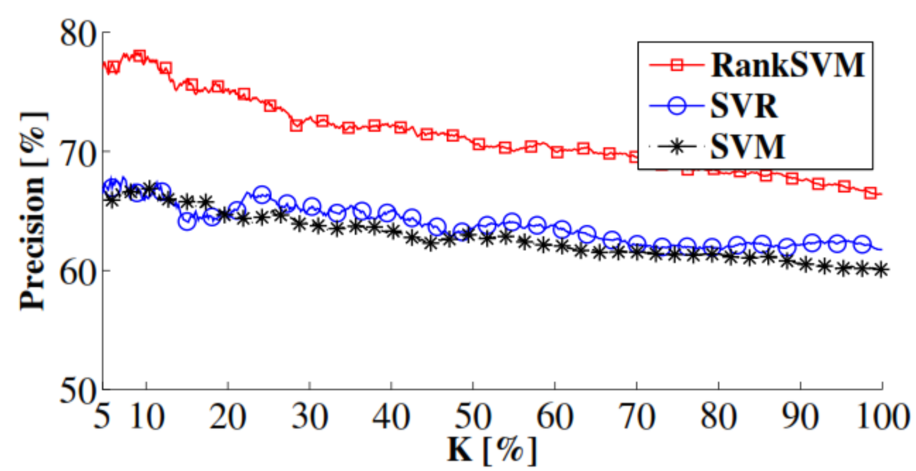


How does it compare to alternative methods?

- Support vector machine (SVM) → Binary classifiers
- Support vector regression (SVR) → Regression



Arousal



Valence

(P@100)

Dimension	Rank-SVM [%]	SVR [%]	SVM [%]
Arousal	**77.1**	**65.5**	**68.1**
Valence	**66.8**	**62.1**	**61.7**

Conclusion

- Considerations in preference training for emotion retrieval
- Trade-offs
  - Label reliability vs training size
  - Optimize the margin between emotion labels in training samples
  - Preference learning provides more reliable labels and larger training set
- Preference learning has higher precision in retrieval
- Higher performance in binary classification
  - 7% arousal
  - 5.1% valence

Thanks for your attention!

Reza Lotfian
Ph.D. Student
Affective computing

http://msp.utdallas.edu/