On the Optimization of SVM Kernel Parameters for Improving Audio Classification Accuracy

Lăcrimioara GRAMA, Liana TUNS, **Corneliu RUSU**

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

14th International Conference on Engineering of Modern Electric Systems 2017

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI–UEFISCDI, project number PNIII-P2-2.1-BG-2016-0378, 54BG/2016, within PNCDI III.

Outline

- Why this research?
- Intruder Detection System
- Wildlife Database
- Mel Frequency Cepstral Coefficients
- Support Vector Machines
- Classification by Optimizing SVM Kernel Parameters
- Results
- Conclusion

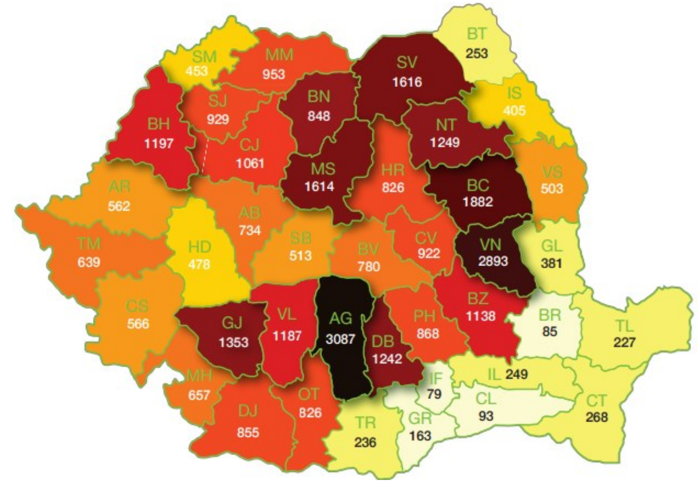
Why this research?

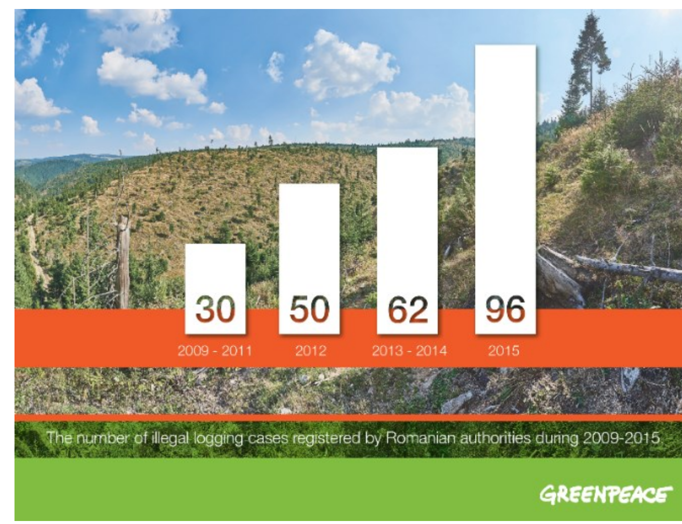
- The number of events that imply illegal logging, hunting, or trespassing of natural reservations, parks, or forests increased so much in the past decade, that on a high demand became the design of wildlife surveillance systems

- These systems are meant to detect in time such type of unwanted activities within the protected areas and help the authorities to take an action

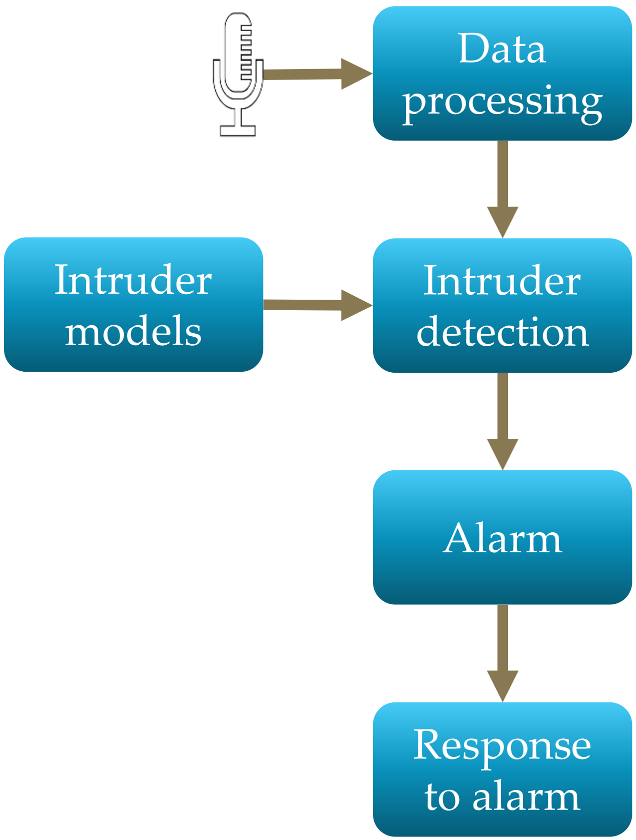
Why this research?

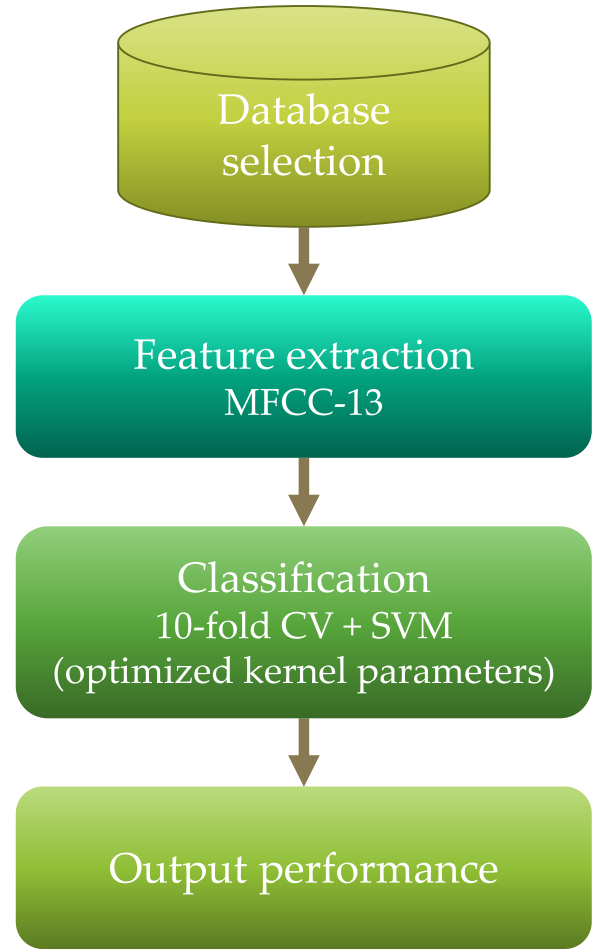
- Over 25 environmental agencies and organizations world wide, are being proactive in tracking illegal logging and hunting
- About 25 million birds are killed illegally in the Mediterranean every year, according to a first-of-its-kind scientific review carried out by *BirdLife International*
- Romania: in 2015 the authorities registered 34.870 cases of illegal logging, which means 96 cases/day (*Greenpeace 2015*)





Audio Intruder Detection System





Wildlife Database

Birds dataset – 654 audio files originated from 70 different species of birds (Internet)

Chainsaws dataset – 356 audio files originated from 18 different types of chainsaws (SPG)

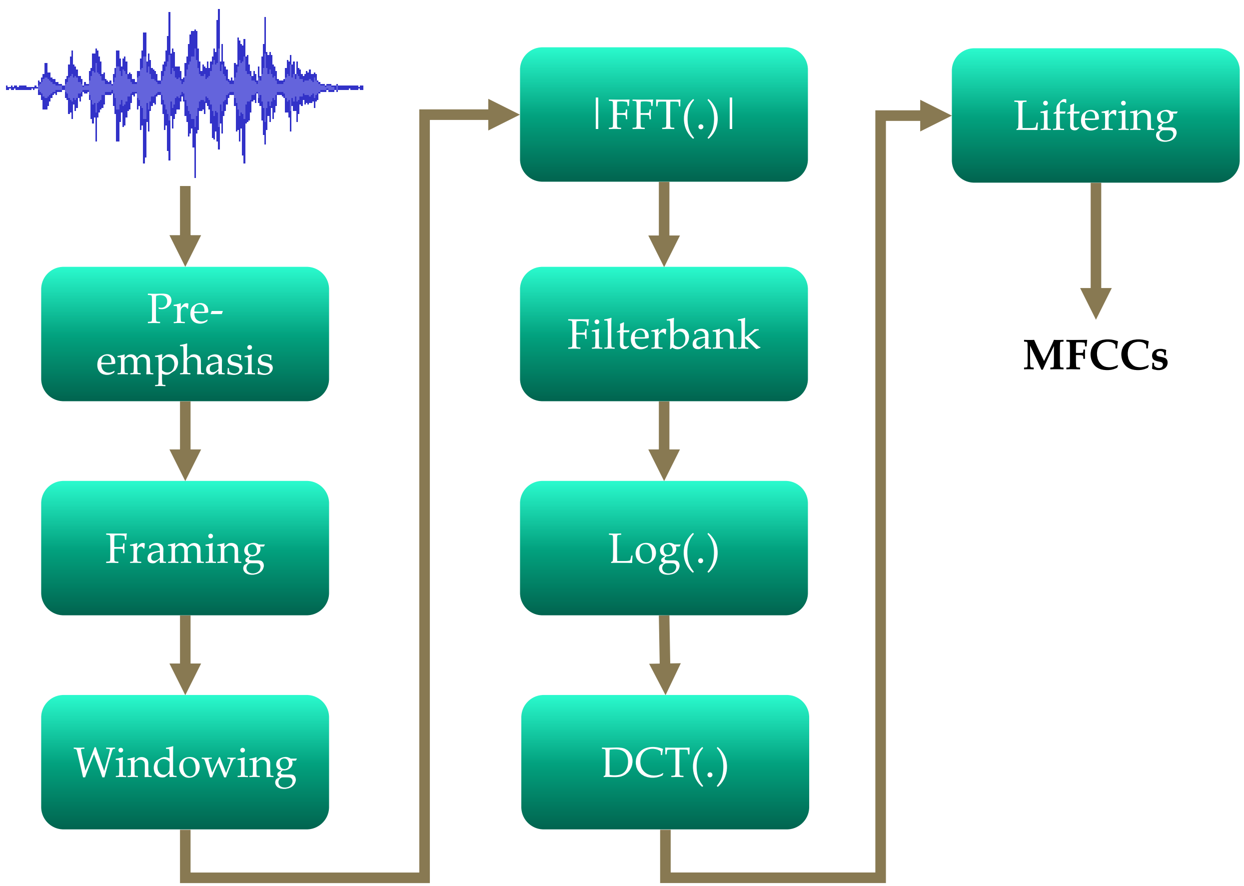
Gunshots dataset – 120 audio files originated from 40 different types of guns (Internet)

Human voice dataset – 207 speech sounds originated from 50 different former students from the TUCN

Tractors dataset – 260 audio files originated from 17 different types of tractors (SPG)

- 16 kHz, 16-bit
- None of the audio signals are studio recordings $\Rightarrow$ they are subject to some additive noise from surroundings

Mel Frequency Cepstral Coefficients



$$H(z) = 1 - 0.97z^{-1}$$

- 25 ms frames (60% overlap)
- Hamming window
- 512-FFT
- 20 triangular filters (0÷8 kHz)

$$c_n = \sqrt{\frac{2}{M}} \sum_{i=1}^{M} \ln S_i \cos\left[\frac{\pi n}{M}(i - 0.5)\right]$$

$$c'_n = \left(1 + \frac{L}{2}\sin\frac{\pi n}{L}\right)c_n, \quad n = \overline{0,12}$$

Support Vector Machines

- SVMs are supervised learning methods used for classification, regression or outliers' detection
- SVMs optimization problem involves the minimization of the error function $\frac{1}{2}\mathbf{w}^T\mathbf{w}+C\sum_{i=1}^{N}\xi_i$, subject to $y_i\left[\mathbf{w}^T\phi\left(\mathbf{x}_i\right)+b\right]\geq 1-\xi_i,\ \ \xi_i\geq 0,\ \ i=\overline{1,N}$
  - $\mathbf{x}_i\in\Re^p$, i=1, 2, …, n are training vectors,
  - $\mathbf{y}\in\{-1,1\}^n$ represent the class labels,
  - $\mathbf{w}$ is an n-dimensional weight vector,
  - C>0 is the penalty parameter of the error term (cost/capacity constant),
  - b is a scalar (bias value),
  - ϕ is the kernel which maps the training vectors (inputs) into a higher dimensional space (the feature space)
  - ξ_i represents parameters for handling nonseparable inputs

Support Vector Machines

- We denote the kernel function by $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$

  - Linear kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$

  - Radial basis kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2)$, $\gamma > 0$
    - Most popular because of its localized and finite responses across the entire real x-axis range

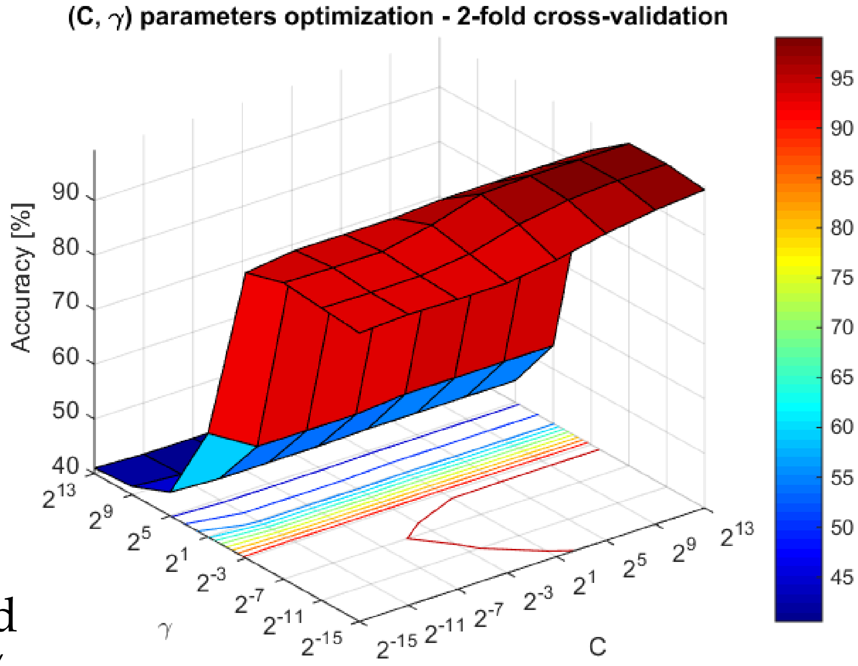
  - Sigmoid kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$, $\gamma > 0$

Optimizing SVMs Kernel Parameters

- To find the best kernels' parameters, based on the overall accuracy, we use a grid search algorithm, in log2-space
- First a 2-fold CV is employed for parameters selection, then, the best point in the space is taken as center and a 10-fold CV is performed with the adjacent parameters
  - If better parameters are found, they will act as a new center and a 10-fold CV is apply again
  - The process should be repeated again and again, until no better parameters are found, or until the parameters are at the border of the grid

Optimizing SVMs Kernel Parameters

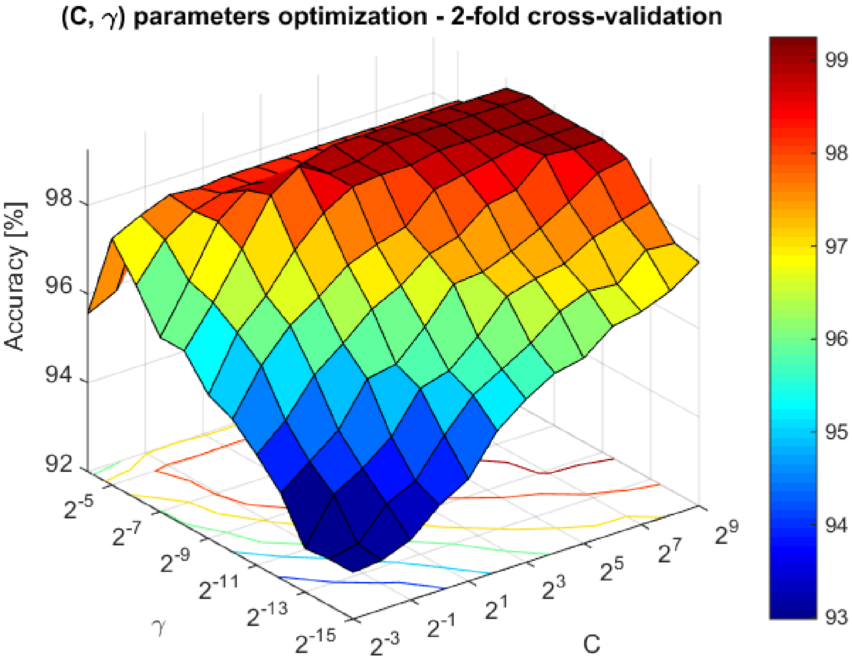
- Radial basis kernel
  - Two parameters should be optimized
    - The cost parameter C jointly with γ
  - The initial considered values
    - $C = \gamma = 2^i$, i = -15, -11, …, 13
    - C and γ values are uniformly distributed in log2-space
    - A 2-fold CV is used
    - We keep only those values for C and γ, for which the accuracies are >90%
    - There is no need to retain $\gamma > 2^{-4}$ and $C < 2^{-3}$



Radial basis kernel – (C, γ) – 1st step: 2-fold CV

Optimizing SVMs Kernel Parameters

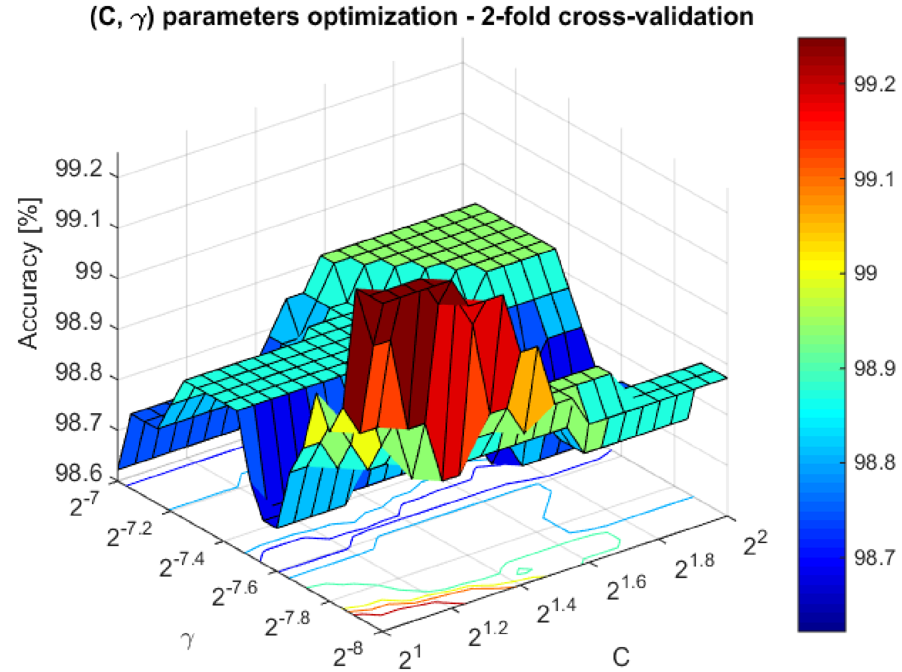
- Radial basis kernel
  - We perform another 2-fold CV, using a smaller step-size
    - $C = 2^i$, i = -3, -2, …, 9 and $\gamma = 2^i$, i = -15, -14, …, -4
  - If we want to keep only those pairs for an accuracy >98.5%, we should retain
    - for C the values from 2^1 to 2^2 together with γ between 2^{-8} and 2^{-7},
    - or C between 2^4 and 2^9 together with γ between 2^{-11} and 2^{-7}



Radial basis kernel – (C, γ) – 2nd step: 2-fold CV

Optimizing SVMs Kernel Parameters

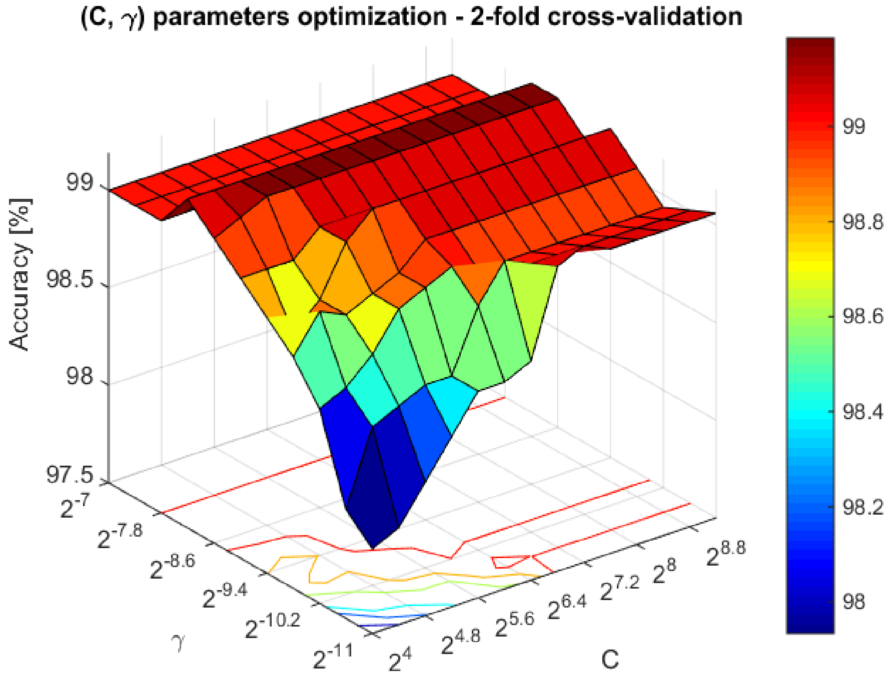
- Radial basis kernel
  - First case
    - $C = 2^i$, $i = 0.9, 1, \ldots, 2.1$ and $\gamma = 2^i$, $i = -8.1, -8, \ldots, -6.9$
    - Now the grid is enough smaller, thus this was the last step
    - There are ten "best pairs" with a 99.25% accuracy
      - $C = 2$, $\gamma \in \{2^{-8}, 2^{-7.95}, 2^{-7.9}\}$
      - $C = 2^{1.05}$, $\gamma \in \{2^{-8}, 2^{-7.95}\}$
      - $C = 2^{1.1}$, $\gamma \in \{2^{-8}, 2^{-7.95}\}$
      - $C = 2^{1.15}$, $\gamma \in \{2^{-8}, 2^{-7.95}\}$
      - $C = 2^{1.2}$, $\gamma = 2^{-8}$



Radial basis kernel – (C, γ) – 3rd step: 2-fold CV (case 1)

Optimizing SVMs Kernel Parameters

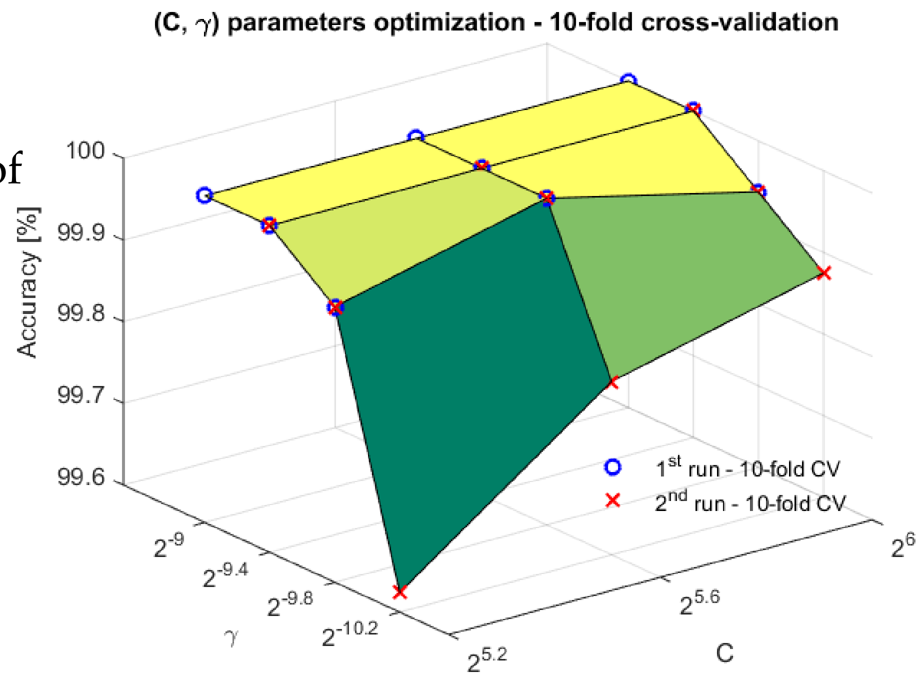
- Radial basis kernel
  - Second case
    - $C = 2^i$, $i = 4, 4.4, \ldots, 9.2$ and $\gamma = 2^i$, $i = -11, -10.6, \ldots, -7$
    - 14x11=154 combinations are tested within the 2-fold CV
    - The highest accuracy attained is 99.19% (for 36 pairs)
    - We consider only the pair $(C, \gamma) = (2^{5.6}, 2^{-9.4})$



Radial basis kernel – (C, γ) – 3rd step: 2-fold CV (case 2)

Optimizing SVMs Kernel Parameters

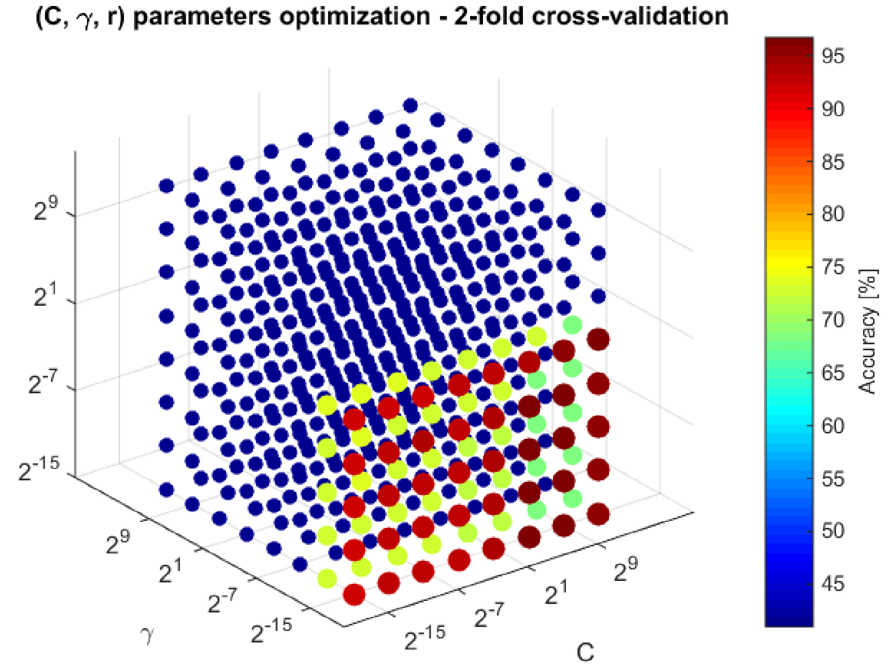
- Radial basis kernel
  - Second case
    - For validation, $(C, \gamma) = (2^{5.6}, 2^{-9.4})$ will be the center for the first run of 10-fold CV
    - Its adjacent pairs are $(2^{5.2}, 2^{-9.8})$, $(2^{5.2}, 2^{-9.4})$, $(2^{5.2}, 2^{-9})$, $(2^{5.6}, 2^{-9.8})$, $(2^{5.6}, 2^{-9})$, $(2^{6.2}, 2^{-9.8})$, $(2^{6.2}, 2^{-9.4})$, and $(2^{6.2}, 2^{-9})$
    - After this run, the best accuracy was 100% for $(2^{5.6}, 2^{-9.8})$
    - This will be the center for the second run
    - With an accuracy of 100% the declared winner is $(C, \gamma) = (2^{5.6}, 2^{-9.8})$



Radial basis kernel – (C, γ) – validation: 10-fold CV

Optimizing SVMs Kernel Parameters

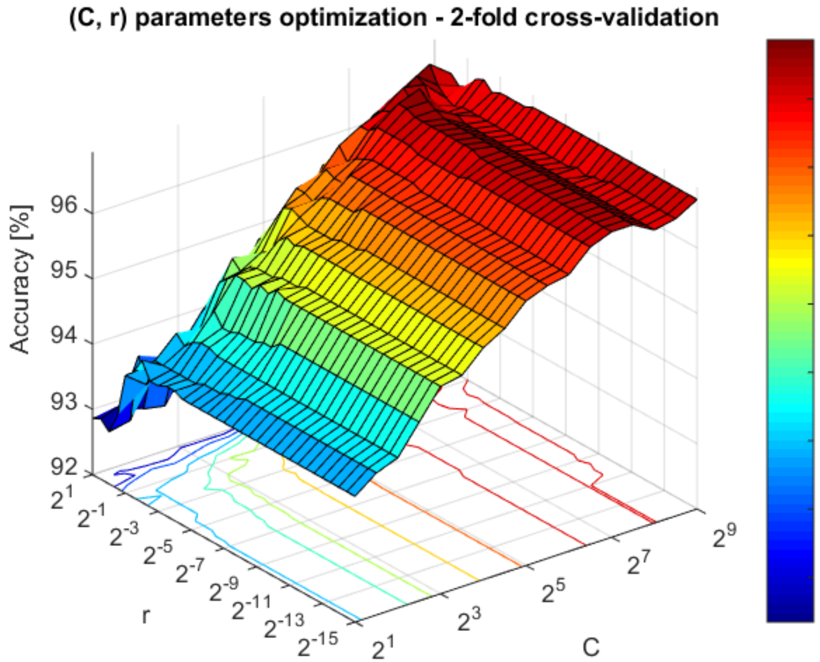
- Sigmoid kernel
  - Three parameters should be optimized
    - The cost parameter C jointly with γ, and r
  - The initial considered values
    - $C = \gamma = r = 2^i$, i = -15, -11, …, 13
    - A 2-fold CV is used
    - For accuracies >95% the only suitable values are $\gamma = 2^{-15}$, $r < 2$, and $C > 2$



Sigmoid kernel – (C, γ, r) – 1st step: 2-fold CV

Optimizing SVMs Kernel Parameters

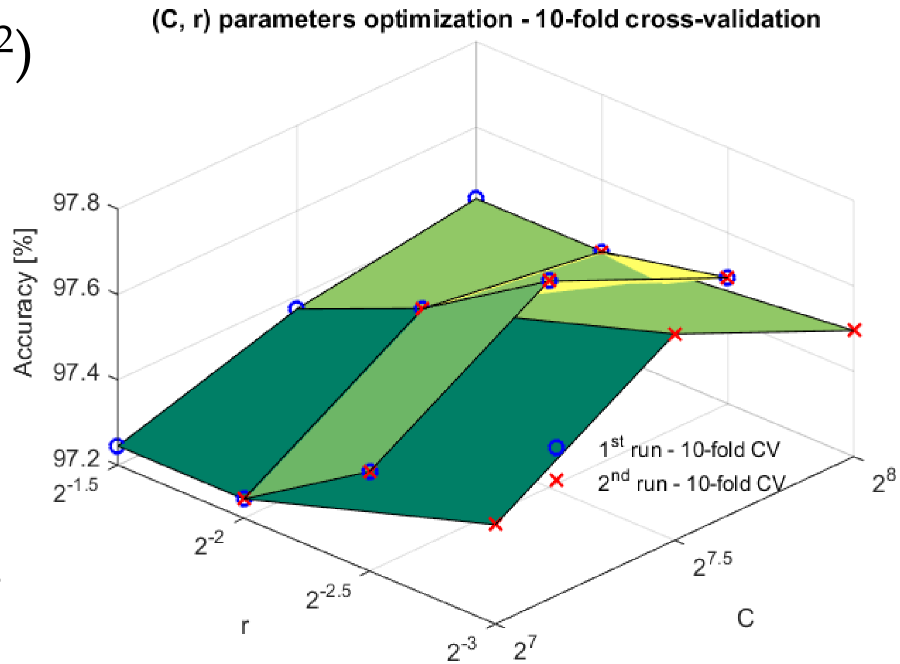
- Sigmoid kernel
  - We perform another 2-fold CV, using a smaller step-size
    - $C = 2^i$, $i = 1, 1.5, \ldots, 9$ and $\gamma = 2^i$, $i = -15, -14.5, \ldots, 1$
    - After the second 2-fold CV the highest accuracy is 96.93% for $(C, r) = (2^{7.5}, 2^{-2})$



Sigmoid kernel – (C, γ, r) – 2nd step: 2-fold CV

Optimizing SVMs Kernel Parameters

- Sigmoid kernel
  - For validation, $(C, r) = (2^{7.5}, 2^{-2})$ will be the center of the first run of 10-fold CV
    - After the first run, the highest accuracy is 97.68% for $(C, r) = (2^{7.5}, 2^{-2.5})$
    - This will be the center of the second run
    - There was no improvement in the accuracy, thus the declared trio winner is $(C, \gamma, r) = (2^{7.5}, 2^{-15}, 2^{-2.5})$



Sigmoid kernel – (C, γ, r) – validation: 10-fold CV

Optimizing SVMs Kernel Parameters

- To validate the proposed method for SVMs kernels' parameters optimization we evaluate the average accuracy and the standard deviation (Std.Dev.) over 10 runs of stratified 10-fold CV

Kernel		**Accuracy [%] (Std.Dev.)**
Linear	Default parameters	97.64 (1.14)
Radial basis	Default parameters	98.98 (0.81)
	$(C, \gamma) = (2^{1.1}, 2^{-8})$	99.46 (0.65)
	$(C, \gamma) = (2^{5.6}, 2^{-9.8})$	99.72 (0.51)
Sigmoid	Default parameters	40.95 (0.32)
	$(C, \gamma, r) = (2^{7.5}, 2^{-15}, 2^{-2.5})$	97.32 (1.20)

- The highest improvement is obtained for the sigmoid kernel (~56%)

Conclusion

- We have presented a step-by-step grid search approach in log2-space to optimize the kernels' parameters for SVMs
- We have shown that the parameters optimization improves the recognition performance for audio classification, especially when using the sigmoid kernel
- We have compared the accuracies obtained with and without kernel's parameters optimization
- As features we have used MFCCs
- For five classes, using 10-fold cross validation, we have obtained average accuracies of 99.74% for radial basis kernel, and 97.18% for sigmoid kernel

On the Optimization of SVM Kernel Parameters for Improving Audio Classification Accuracy

Lăcrimioara GRAMA, Liana TUNS, **Corneliu RUSU**

Technical University of Cluj-Napoca
Faculty of Electronics, Telecommunications and Information Technology
Basis of Electronics Department
Signal Processing Group

14th International Conference on Engineering of Modern Electric Systems 2017

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/ CCCDI–UEFISCDI, project number PNIII-P2-2.1-BG-2016-0378, 54BG/2016, within PNCDI III.