One-Class Classification with Gaussian Processes

Michael Kemmler, Erik Rodner, and Joachim Denzler

Chair for Computer Vision
Friedrich Schiller University of Jena

{michael.kemmler,erik.rodner,joachim.denzler}@uni-jena.de
www.inf-cv.uni-jena.de

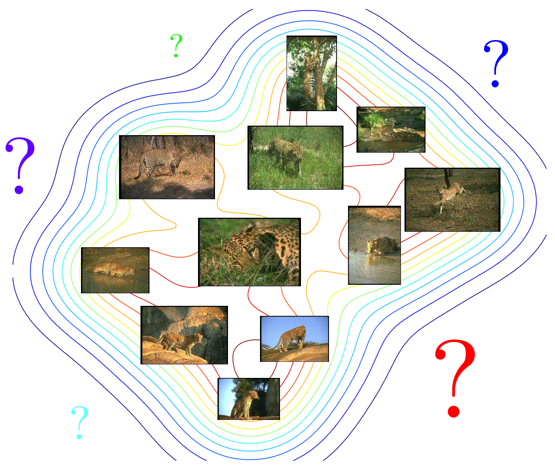
The work was funded by the German Excellence Initiative "MikroPlex"

seit 1558

JSMC

MIKROPlex Microbial Identification

One-Class Classification



In many tasks, only positive training examples are available!

This problem is known as

- One-Class Classification
- Outlier and Failure Detection
- Novelty Detection

Gaussian Processes Regression

- non-parametric regression: $y = f(\mathbf{x}) + \varepsilon$, given data $(\mathbf{X}, \mathbf{y})$
- assumptions: $\mathbf{f} = f(\mathbf{X}) \sim \mathcal{N}(\mathbf{0}, \kappa(\mathbf{X}, \mathbf{X}))$, $\mathbf{y} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I})$
- Inference [1] for unknown points $\mathbf{x}_*$: $p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) \sim \mathcal{N}(\mu_*, \sigma_*^2)$, where

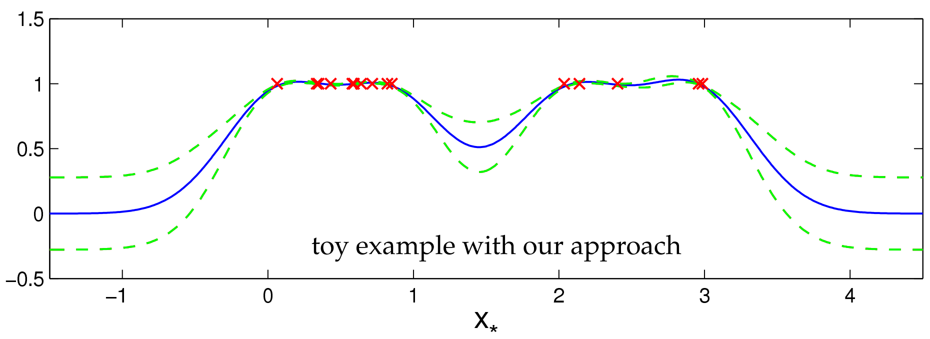
$$\begin{aligned} \mu_* &= \mathbf{k}_*^T \left(\mathbf{K} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{y} \\ \sigma_*^2 &= k_{**} - \mathbf{k}_*^T \left(\mathbf{K} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{k}_* \end{aligned}$$

- with shorthands $\mathbf{K} = \kappa(\mathbf{X}, \mathbf{X})$, $\mathbf{k}_* = \kappa(\mathbf{X}, \mathbf{x}_*)$ and $k_{**} = \kappa(\mathbf{x}_*, \mathbf{x}_*)$
- using some kernel κ, e.g. squared exponential kernel:

$$\kappa_{SE}(\mathbf{x}, \mathbf{x}') = \nu_0^2 \exp\left(-\frac{1}{2\nu_1^2} ||\mathbf{x} - \mathbf{x}'||^2\right)$$

One-Class Classification with Gaussian Processes

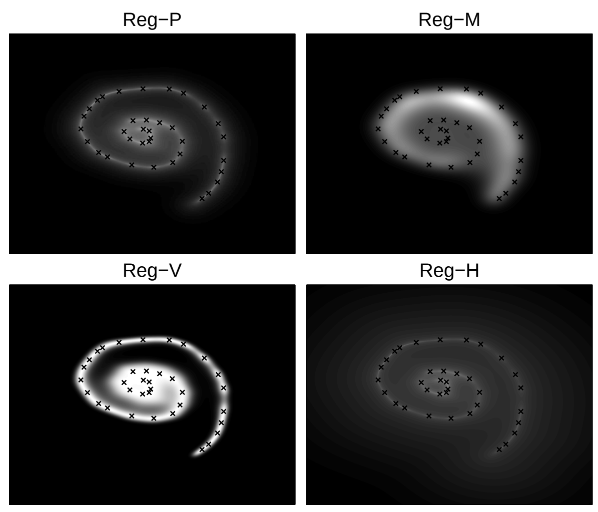
- **Our contribution:** new OCC methods using GP priors
- Fitting the data to labels $y = 1$ and using a **zero-mean** GP prior favours functions suitable for OCC



- various OCC measures can be derived from the predictive distribution

1.	Probability (P)	$p(y_* = 1 \mid \mathbf{X}, \mathbf{y}, \mathbf{x}_*)$
2.	Mean (M)	$\mu_* = \mathcal{E}(y_* \mid \mathbf{X}, \mathbf{y}, \mathbf{x}_*)$
3.	neg. Variance (V)	$-\sigma_*^2 = -\mathcal{V}(y_* \mid \mathbf{X}, \mathbf{y}, \mathbf{x}_*)$
4.	Heuristic (H)	$\mu_* \cdot \sigma_*^{-1}$

- use OCC measures as class membership scores



Visualization of some methods on a spiral example

Visual Object Recognition

Task 1: separate image categories from background

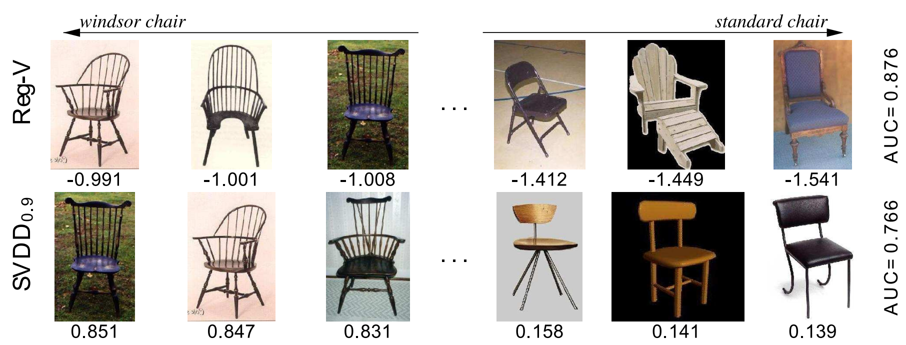
- utilize image kernels: PHoG [2] and Color [3] spatial pyramids
- compare against SVDD [4] with outlier fraction $\nu \in \{0.1, 0.2, \ldots, 0.9\}$
- avg. AUCs over all class-background problems (15 training samples) :

Caltech 101	Reg-P	Reg-M	Reg-V	Reg-H	$\text{SVDD}_{0.5}$	$\text{SVDD}_{0.9}$
PHoG	*0.696*	0.693	0.692	*0.696*	0.690	0.685
Color	*0.761*	0.736	**0.766**	*0.755*	0.739	0.746

- additional experiments show: GP regression outperforms GP classification

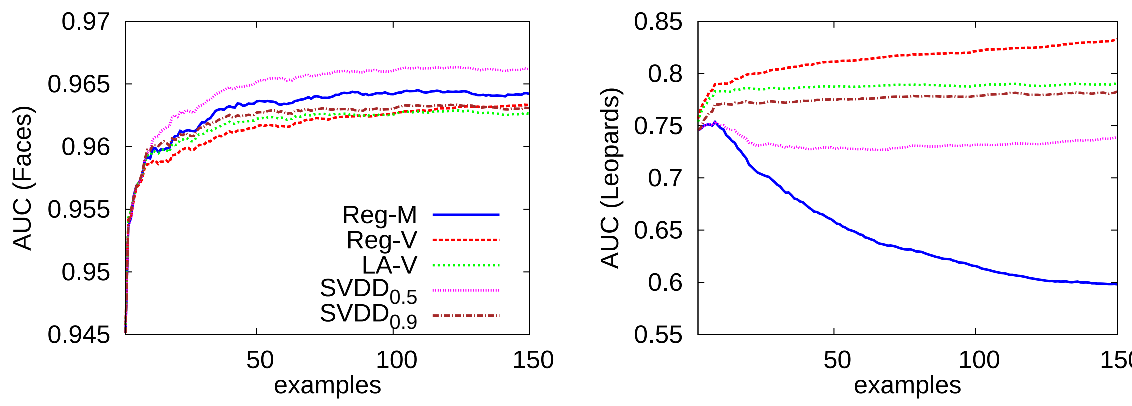
Task 2: separate between different sub-categories

- e.g. class *windsor chair* versus *chair* (30 training samples)



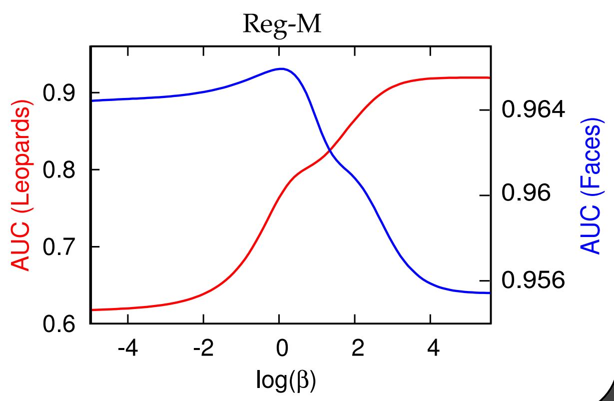
Impact of training size (Task 1)

- performance can decrease with more training examples (Reg-M, $\text{SVDD}_{0.5}$)
- potential problem: fixed (implicit) scale parameter of the kernel



Introduction of an Explicit Scale Parameter

- we use a generalized rbf kernel [5]: $\kappa_\beta(\mathbf{x}, \mathbf{x}') = \exp\left(-\beta d_\kappa^2(\mathbf{x}, \mathbf{x}')\right)$, where $d_\kappa^2(\mathbf{x}, \mathbf{x}') = \kappa(\mathbf{x}, \mathbf{x}) - 2\kappa(\mathbf{x}, \mathbf{x}') + \kappa(\mathbf{x}', \mathbf{x}')$
- performance of Reg-M highly depends on β
- optimal β is task specific
- tuning the scale parameter leads to significant performance benefits



Conclusion

- investigates non-parametric OCC scores via Gaussian processes
- general method is easy to implement
- significantly better results compared to SVDD
- suitable for object recognition using image kernels
- depends on scale parameter (hard to obtain)

References

[1] Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press (2005)

[2] Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval. (2007) 401–408

[3] van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. (2010)

[4] Tax, D.M.J., Duin, R.P.W.: Support vector data description. Mach. Learn. **54** (2004) 45–66

[5] Vedaldi, A., Soatto, S.: Relaxed matching kernels for object recognition. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition. (2008)