Authentication

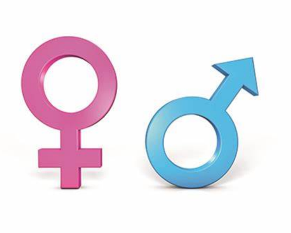
Prof. Dr. Helene Dörksen

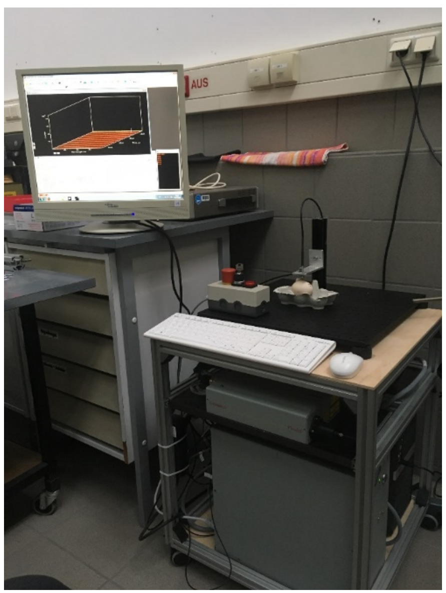
helene.doerksen@th-owl.de

Current Information

Due to very important business trip

next week 28.06.2021
no lecture/exercises/labs





Learned before

Combinatorial Refinement of Feature Weighting for Linear Classification

$$h(x) = \boxed{a_1 x_1 + a_2 x_2 + \cdots + a_i\, x_i} + \boxed{a_{i+1} x_{i+1} + \cdots + a_d x_d} + c_h \qquad (I)$$

$$h(x) = \underbrace{\boxed{\sum_{j \in I_1} a_j\, x_j}}_{\alpha} + \underbrace{\boxed{\sum_{j \in I_2} a_j x_j}}_{\beta} + c_h \qquad (II)$$

Dimensionality Reduction: d summands $\rightarrow$ 2 summands,
but classification rules of (I) and (II) are equivalent!

Combination of classifiers:
Multiple Classifier: initial (linear) classifier h (in feature space $x_1, \cdots, x_d$) and refinement (linear) classifier g (in feature space α, β)

$\Downarrow$

We are able to improve accuracy without increasing complexity!

Lecture 10:

Margin-based Refinement for Support-Vector-Machine Classification

6th International Conference on
Pattern Recognition Applications and Methods, Porto, Portugal
24.-26. February 2017

OUTLINE

- Introduction
- Approach
- Experimental Results
- Summary

INTRODUCTION

- Classification task
- Supervised / unsupervised learning
- Decision-making procedure
- Taxonomy

Iris setosa *Iris versicolor* *Iris virginica*

© palass.org

Wikipedia says: "…Taxonomy is the practice and science of classification… The word finds its roots in the Greek language τάξις, *taxis* (meaning 'order', 'arrangement') and νόμος, *nomos* ('law' or 'science')…"

INTRODUCTION

Problem: poor generalisation of classification

Known: Input, Information, Data, Features…

Training of Rules

Class, Cluster, Output, Knowledge, Decision,…

© karrierebibel.de

Unknown: Input, Information, Data, Features…

Application of Rules

Class, Cluster, Output, Knowledge, Decision,…

➢ The reason for poor generalisation might be e.g. incomplete training samples

poor generalisation:

© fotolia.com

MOTIVATION

We consider **„poor generalisation“ problem** for **incomplete/small** training samples, because of **at least 4 reasons**:

- scientifically challenging problem
- suitable to not well-balanced samples (high relations to reality)
- industrially attractive for resource-efficient environments
- important for authentication task

PRELIMINARIES

Primary basics for **good generalisation** [1] of a classifier:

- simple models (*Occam's razor* principle)
- regularisation (e.g. large classification margin like SVM)

Secondary basics (methods) might be **good for generalisation** [1] of a classifier, e.g.:

- dimensionality reduction
- multiple classifier
- …

[1] E. Alpaydın, Introduction to Machine Learning, 2nd ed. Cambridge: The MIT Press, 2010.

PRELIMINARIES

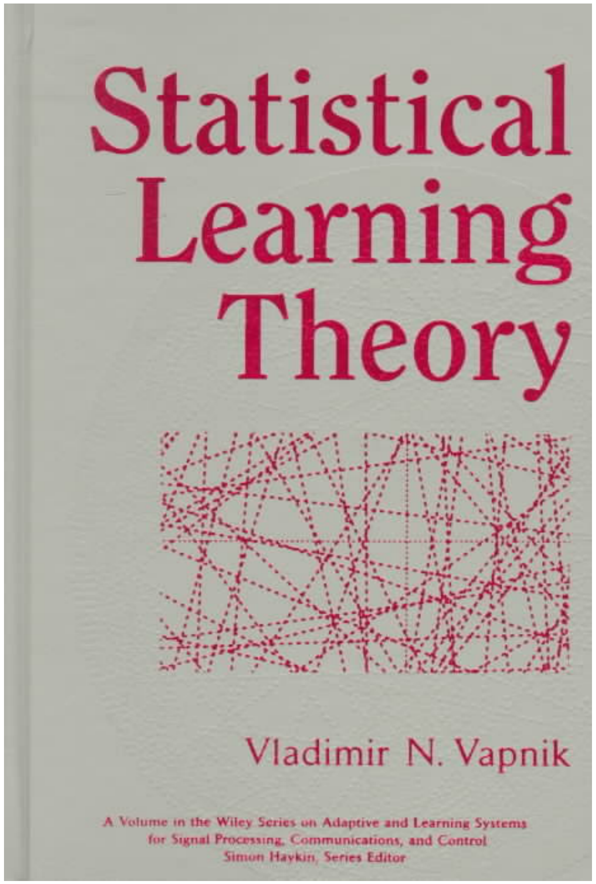
All features are **equally relevant** for the classification boundary calculation, **but**:

Principles of the **feature extraction** [1] **:**

- a helpful feature may be **less relevant** by itself;
- combination of individually irrelevant features **may become relevant**

[1] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing). Secaucus and NJ and USA: Springer-Verlag New York, Inc, 2006.

SUPPORT-VECTOR-MACHINE CLASSIFICATION



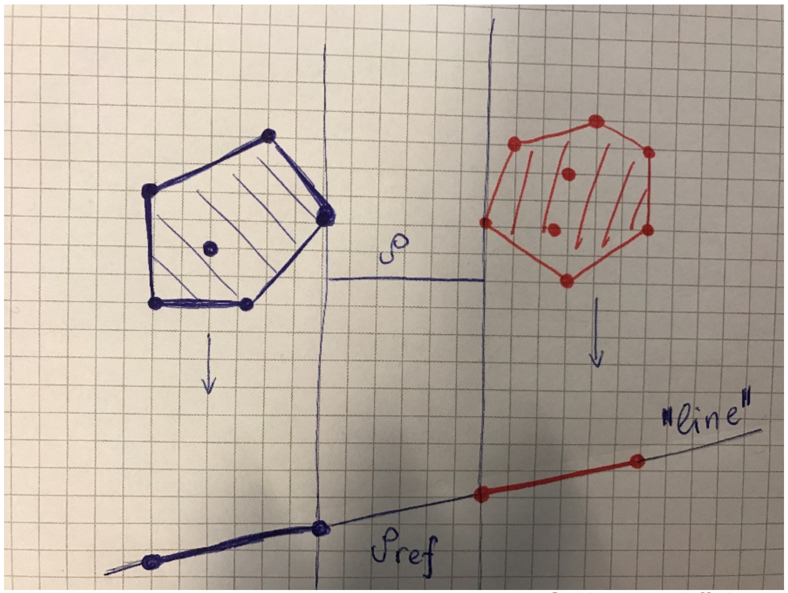
© google

© datascience.columbia.edu

Wikipedia says: "… As of February 2017, he has an ***h-index*** of **115** and, overall, his publications have been cited close to **180,000** times. His book on "Statistical Learning Theory" alone has been cited close to **60,000** times…"

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

„Very simple“ geometrical interpretation:



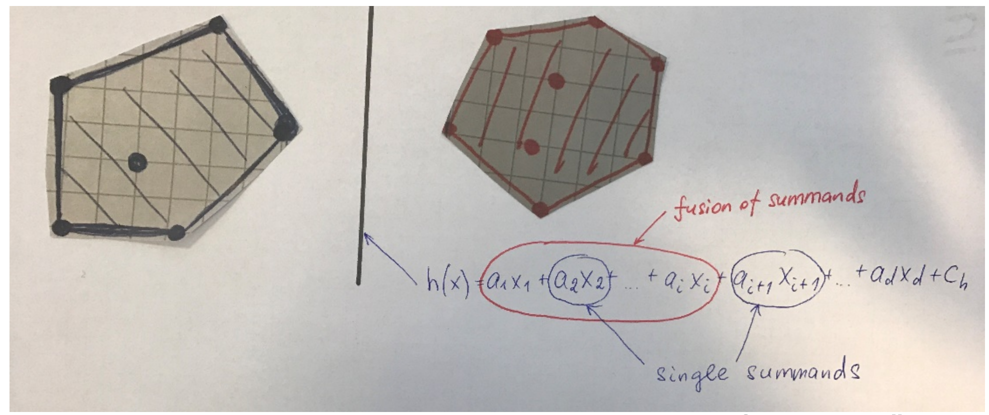
- <u>margin-based refinement</u> means increasing generalisation by margin
- **steepness** (absolute value of slope) of “line” is key factor for $\boldsymbol{\rho_{ref}}$
- <u>infinitely</u> many “lines” !

- why is it might work (especially well) for incomplete/small samples?
- how to select the right „line“?

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

Margin-based refinement considered as a special case of:

Combinatorial Refinement (ComRef)[1]



© Helene Dörksen

- a helpful summand may be **less relevant** by itself
- a **fusion** of individually irrelevant **summands may become relevant**

[1]) H. Dörksen and V. Lohweg, “Combinatorial Refinement of Feature Weighting for Linear Classification,” in 19th IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA 2014), 2014.

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

$$h(x) = \boxed{a_1 x_1 + a_2 x_2 + \cdots + a_i\, x_i} + \boxed{a_{i+1} x_{i+1} + \cdots + a_d x_d} + c_h \qquad (I)$$

$$\underbrace{\text{fusion of summands } 1, \cdots, i}_{\sum_{j=1}^{i} a_j x_j} \qquad \underbrace{\text{fusion of summands } i+1, \cdots, d}_{\sum_{j=i+1}^{d} a_j x_j}$$

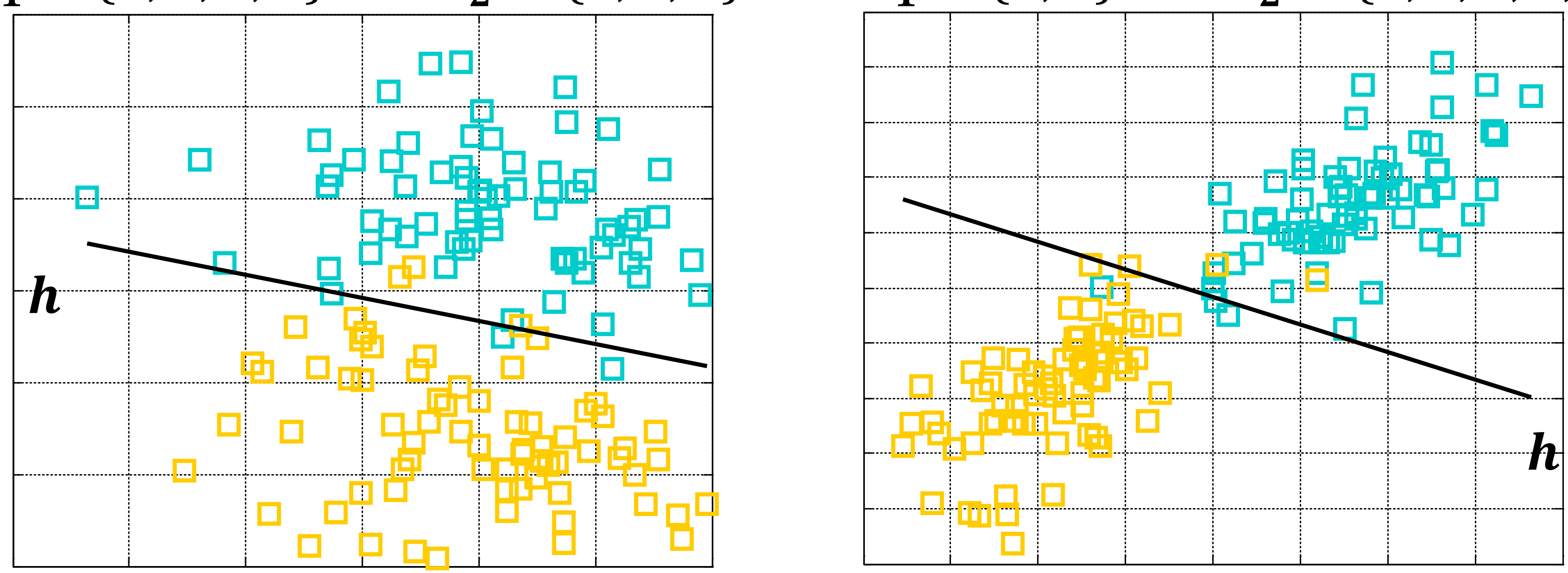
$$h(x) = \boxed{\sum_{j \in I_1} a_j\, x_j} + \boxed{\sum_{j \in I_2} a_j x_j} + c_h \qquad (II)$$

Dimensionality Reduction: d summands $\rightarrow$ 2 summands,
but classification rules of (I) and (II) are equivalent!

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

SVM for classes ☐ and ☐ of ***SEEDS*** [1)] : $\boldsymbol{h(x)} = \sum_{i=1}^{d} a_i x_i + c_h \quad (d = 7)$

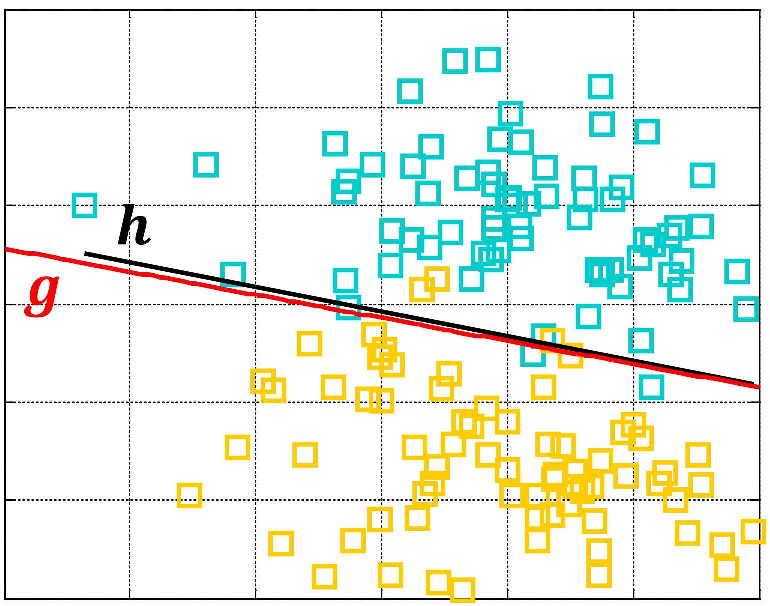
$I_1 = \{1, 2, 3, 7\}$ and $I_2 = \{4, 5, 6\}$ $\qquad$ $I_1 = \{4, 7\}$ and $I_2 = \{1, 2, 3, 5, 6\}$



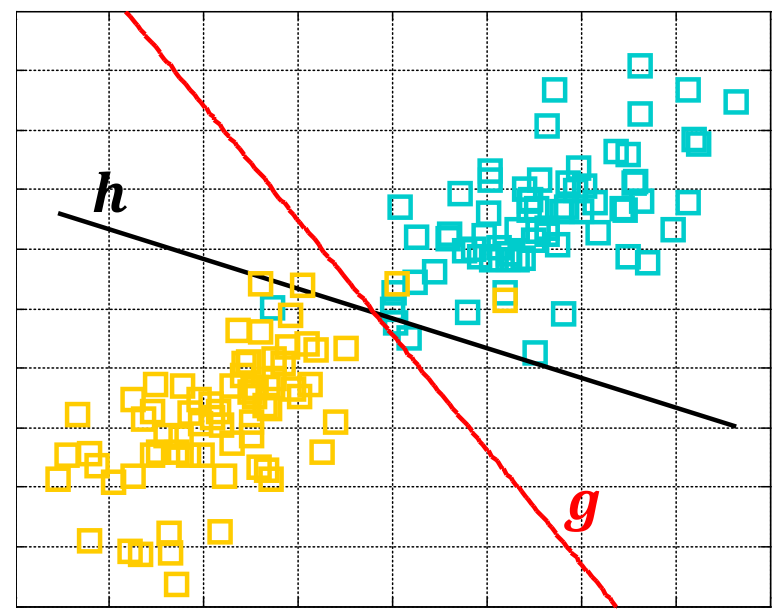
classification rate of $\boldsymbol{h}$ for both reductions: $\mathbf{95.71\%}$

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

classification rate of $\boldsymbol{h}$ is $\mathbf{95.71\%}$



classification rate of $\boldsymbol{g}$: $\mathbf{96.43\%}$

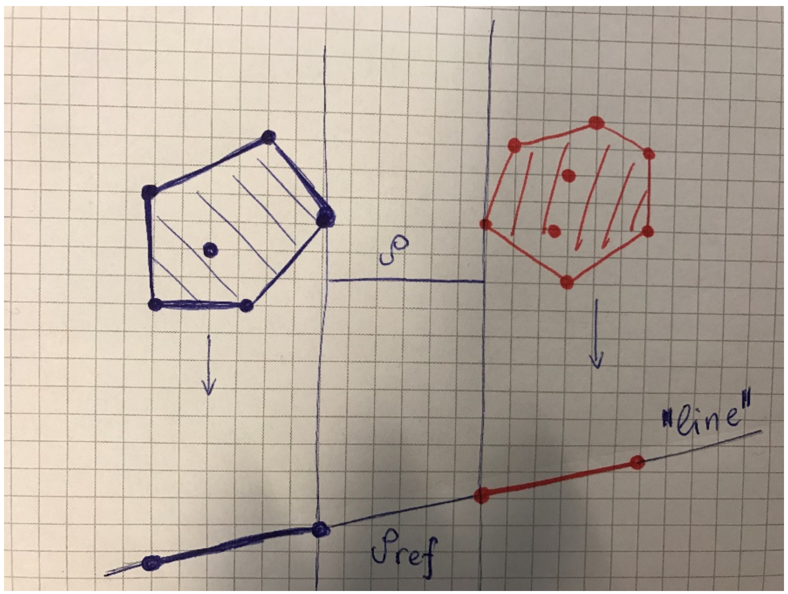


classification rate of $\boldsymbol{g}$: $\mathbf{97.85\%}$

- ComRef is able to increase generalisation ability
- Number of fusions of summands is finite!

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

$$\rho = \frac{2}{\|\mathbf{a}\|} = \frac{2}{\sqrt{a_1^2 + \cdots + a_d^2}}.$$



- We choose candidates for „lines“ by ComRef‘s fusions of summands

- Refinement corresponds to:

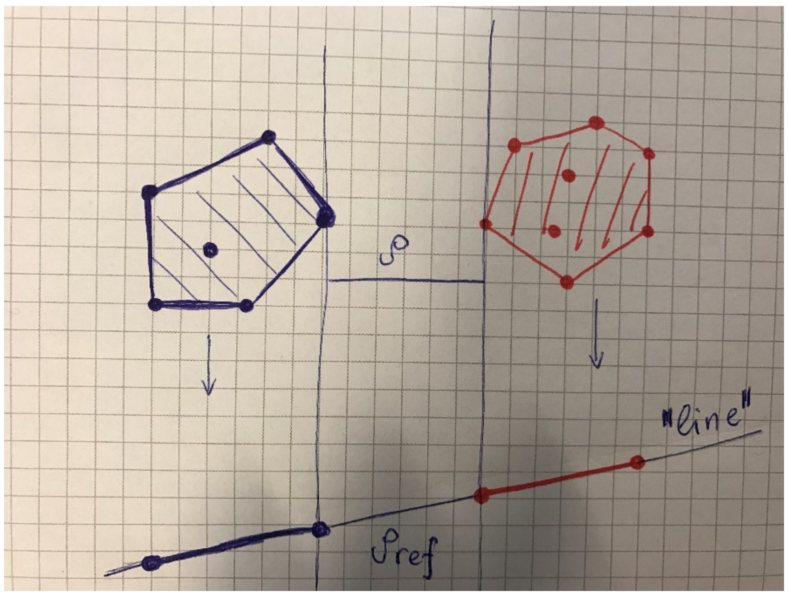
$$b_1 \sum_{i \in I_1} a_i x_i + \cdots + b_j \sum_{i \in I_j} a_i x_i - \tilde{c} = 0$$

$$\rho_{ref} = \frac{2}{\sqrt{b_1^2 \sum_{i \in I_1} a_i^2 + \cdots + b_j^2 \sum_{i \in I_j} a_i^2}}$$

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

$$\rho = \frac{2}{\|\mathbf{a}\|} = \frac{2}{\sqrt{a_1^2 + \cdots + a_d^2}}.$$

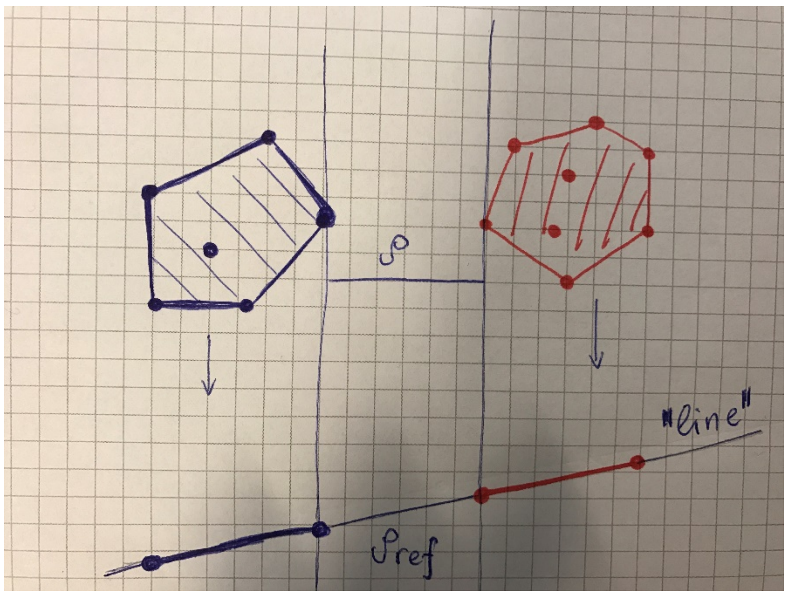
$$\rho_{ref} = \frac{2}{\sqrt{b_1^2 \sum_{i \in I_1} a_i^2 + \cdots + b_j^2 \sum_{i \in I_j} a_i^2}}$$



? **Nominator**

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE

$$\rho = \frac{2}{\|\mathbf{a}\|} = \frac{2}{\sqrt{a_1^2 + \cdots + a_d^2}}.$$



© Helene Dörksen

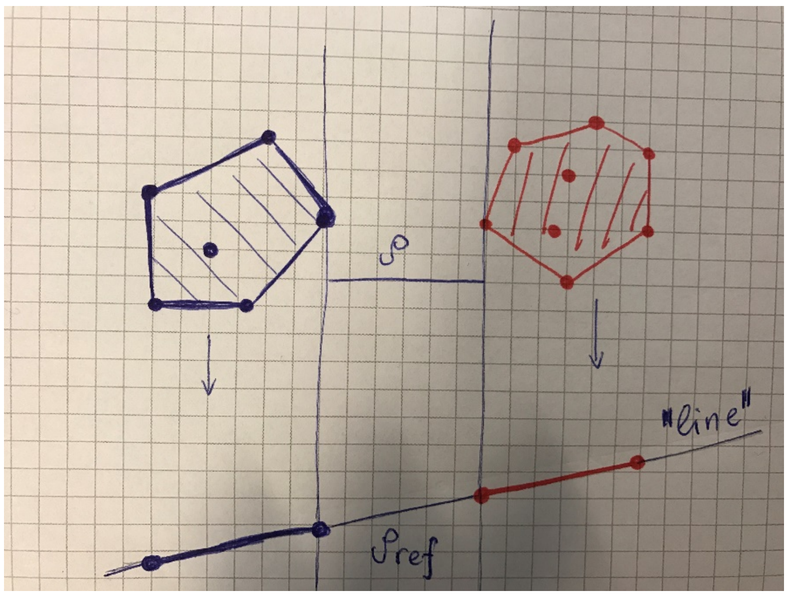
$$\rho_{ref} = \frac{2}{\sqrt{b_1^2 \sum_{i \in I_1} a_i^2 + \cdots + b_j^2 \sum_{i \in I_j} a_i^2}}$$

Nominator needs to be as small as possible!

If for each $k = 1, \cdots, j$ holds $b_k^2 \leq 1$

then $\rho_{ref} \geq \rho$.

MARGIN-BASED REFINEMENT FOR SUPPORT-VECTOR-MACHINE



© Helene Dörksen

- We define „MIN/MAX RULE“ [1] to overcome combinatorial nature of ComRef

Min/Max Rules for Margin-based Refinement

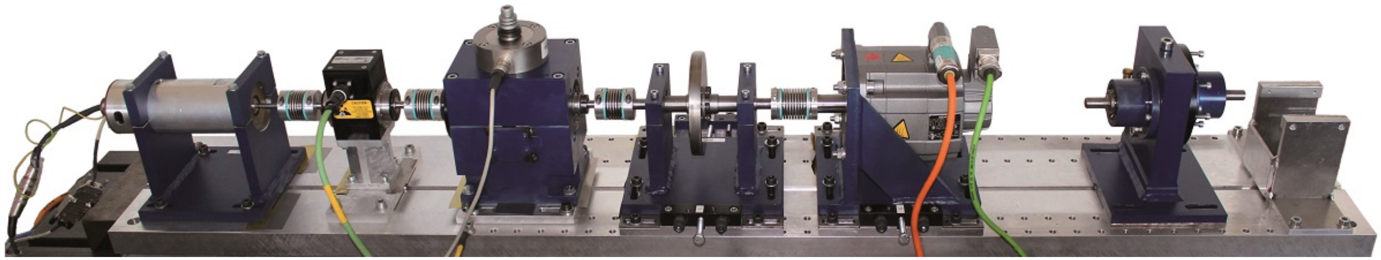
i. Compute initial *SVM* hyperplane $h(\mathbf{x})$

ii. Find set I_k of two indices from $\{1,\cdots,d\}$ such that $\sum_{i\in I_k} a_i^2$ is minimal/maximal, i.e. choose from $|a_1|,\cdots,|a_d|$ two with mininimal/maximal values.

iii. Recalculate *SVM* refinement $g(\mathbf{u})$ for $\mathbf{u} = (a_1x_1,\cdots,\sum_{i\in I_k} a_ix_i,\cdots,a_dx_d)$.

[1] Dörksen, H.; Lohweg, V.: Margin-based Refinement for Support-Vector-Machine Classification. Proceedings of 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM2017), Porto, Portugal, 2017

RESULTS

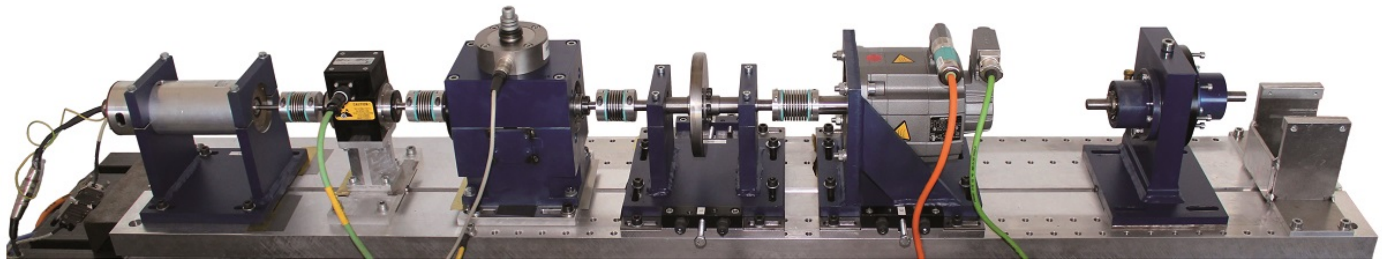


© inIT

real-world industrial application: **Motor Drive Diagnosis**

- condition monitoring of running motors in industry
- two classes considered: *intact* and *anomaly*
- typical anomalies: ball bearings, axle displacement or inclination of gear-wheels
- tests on two samples:
  Motor-I (M-I) 52 features and 5,318 objects (total)
  Motor-II (M-II) 72 features and 10,638 objects (total)
- due to different environmental conditions (e.g. stable/unstable position of the motor, air temperature/humidity, life/running time, etc.), real <u>completeness/balance</u> of sample is assumed to be <u>unknown</u>

RESULTS



real-world industrial application: **Motor Drive Diagnosis**

Benchmarking:

K-fold	*SVM*	*Ref*MIN	t_{K-1}MIN	ρ_{ref}MIN
M-I ($K = 30$)	92.26	93.88	7.69	2.80
M-I ($K = 100$)	88.77	90.37	8.85	2.67
M-II ($K = 30$)	96.00	97.08	15.25	2.84
M-II ($K = 100$)	94.39	95.45	10.24	2.83

- *K-fold cv paired t test* with K = 30 and K = 100
- $\boldsymbol{t_{K-1}}$-statistic for comparing of two classification algorithms is key value resp. generalisation: threshold $\boldsymbol{t_{29} = 2.05}$ or higher

RESULTS

UCI Samples

Benchmarking:

dataset	# features	# objects	*SVM*	*Ref* MIN	t_{K-1}MIN	ρ_{ref}MIN
CNAE (classes 6 vs. 7)*	299	240	79.35	86.52	2.89	1.89
CNAE (classes 7 vs. 9)	333	240	86.75	92.96	2.92	1.70
Ecoli*	6	220	95.22	96.47	2.52	2.17
Heart	13	270	74.89	77.44	3.91	2.49
Promoters	57	212	74.64	78.32	5.45	2.30
Seeds	7	140	91.66	93.49	5.81	2.54
Splice (classes *E* vs. *I*)*	60	1535	84.66	87.13	7.04	2.16
Splice (classes *I* vs. *N*)*	60	2423	81.72	84.54	6.25	2.09

- *K-fold cv paired t test* with K = 10

- t_{K-1}-statistic for comparing of two classification algorithms is key value resp. generalisation: threshold $t_9 = 2.26$ or higher

RESULTS

UCI Samples (*Fertility* and *Hepatitis* are not well-balanced originally, other samples balanced artificially)

Benchmarking:

dataset	# features	# objects	balance	$F_{SVM}(T^+ / T^-)$	$F_{RefMIN}(T^+ / T^-)$
Fertility	9	100	7.3	0.12 / 0.82	0.16 / 0.83
Hepatitis	16	80	5.2	0.32 / 0.84	0.35 / 0.85
Heart	13	150	4.0	0.47 / 0.84	0.51 / 0.84
Mammography	5	628	4.6	0.67 / 0.91	0.69 / 0.92
Pima	8	331	4.3	0.45 / 0.83	0.49 / 0.84
Promoters	57	120	7.6	0.04 / 0.93	0.14 / 0.93
Vertebral	6	244	6.2	0.40 / 0.86	0.44 / 0.86

- balance of sample as number of objects of one class divided by number of objects of other

- ***F***-measures are scores based on true positives and false negatives for each class

SUMMARY

- Margin-based refinement is based on dimensionality reduction
- „MIN/MAX RULE“ allows fast calculation (linear time complexity)
- Margin-based refinement with „MIN/MAX RULE“ is able to improve generalisation of classification
- Approach works especially well on incomplete/small and not well-balanced samples

Homework: Exercises and Labs

for the next week prepare practical exercises and labs from
Exercises Lec 10 (you will find it in the download area)

Online-Exercises coming now

We will discuss **Exercises Lec 9**