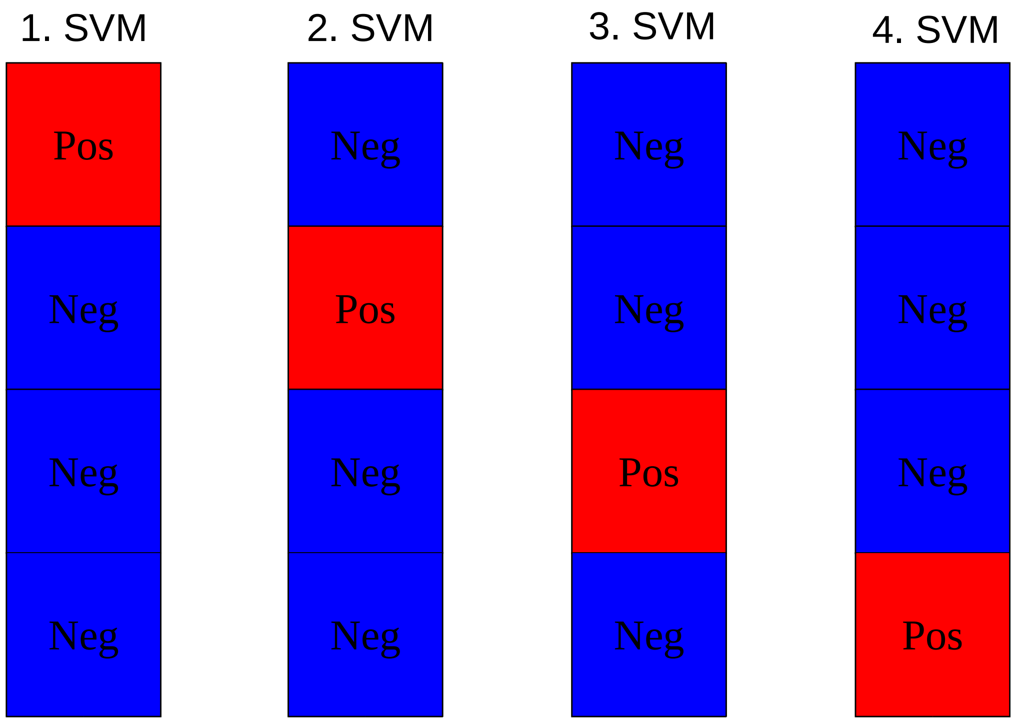
Ein Vergleich von Methoden für Multi-klassen Support Vector Maschinen

- Einführung
- Auf binären Klassifikatoren beruhende Methoden
  - One-Against-All
  - One-Against-One
  - DAGSVM
- Methoden die alle Daten zugleich betrachten (All-Together Methoden)
  - Eine von Vapnik beschriebene Methode
  - Eine Methode von Crammer und Singer
- Experimente und Resultate

Einführung

- SVM wurden ursprünglich für Zweiklassenprobleme entworfen.
- Wie sie effizient zur Multi-Klassen Klassifikation erweitert werden können, ist immer noch ein aktives Forschungsfeld.
- Grundsätzlich gibt es zwei Typen von Ansätzen für Multi-Klassen SVM:
- Eine Methode besteht darin mehrere binäre SVM zu kombinieren.
- Die andere behandelt direkt alle Daten in einer Optimierungsgleichung. (All-together Methoden)

One-Against-All



One-Against-All

$$\underset{w_i, b_i, {}_i}{min} \frac{1}{2}\|w_i\|^2 + C\sum_{j=1}^{l} {}_{ij}$$

$$w_i x_j + b_i \geqslant 1 - {}_{ij}, y_j = i,$$

$$w_i x_j + b_i \leqslant -1 + {}_{ij}, y_j \neq i,$$

$${}_{ij} \geqslant 0, j = 1,\ldots,l$$

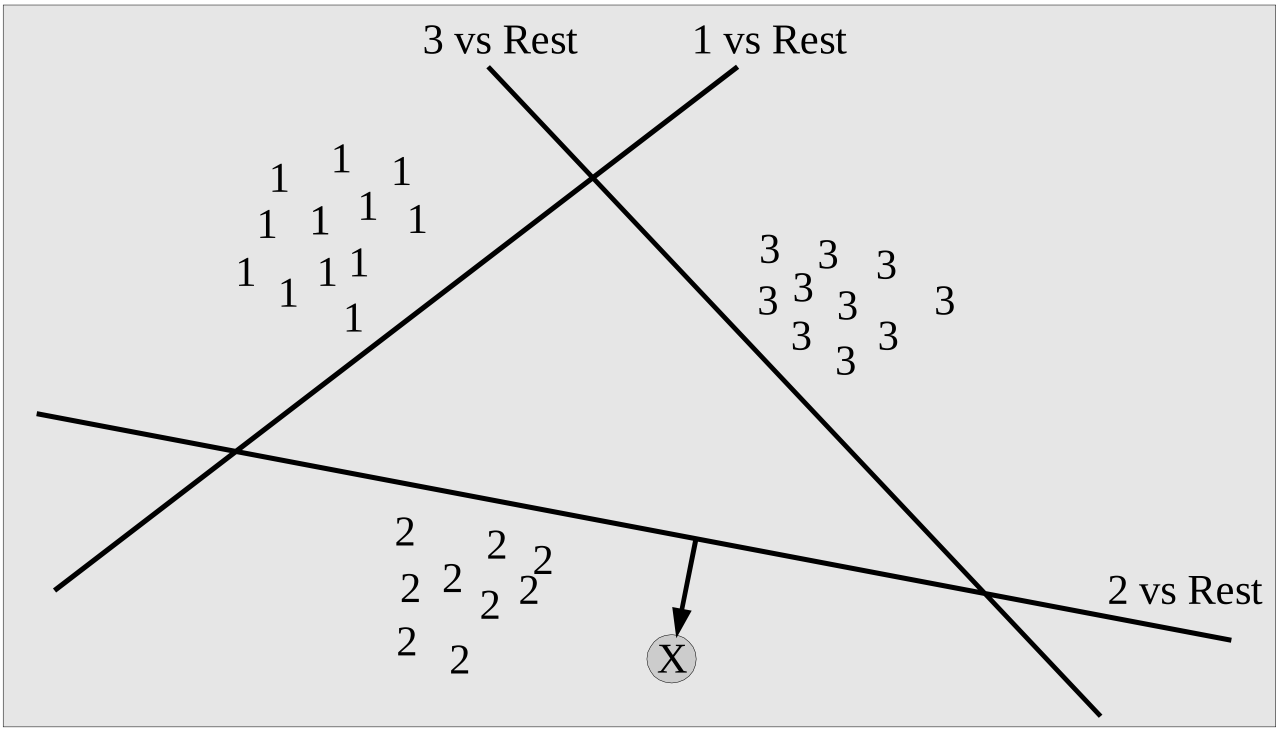
$$w_1 x + b_1$$

$$\ldots$$

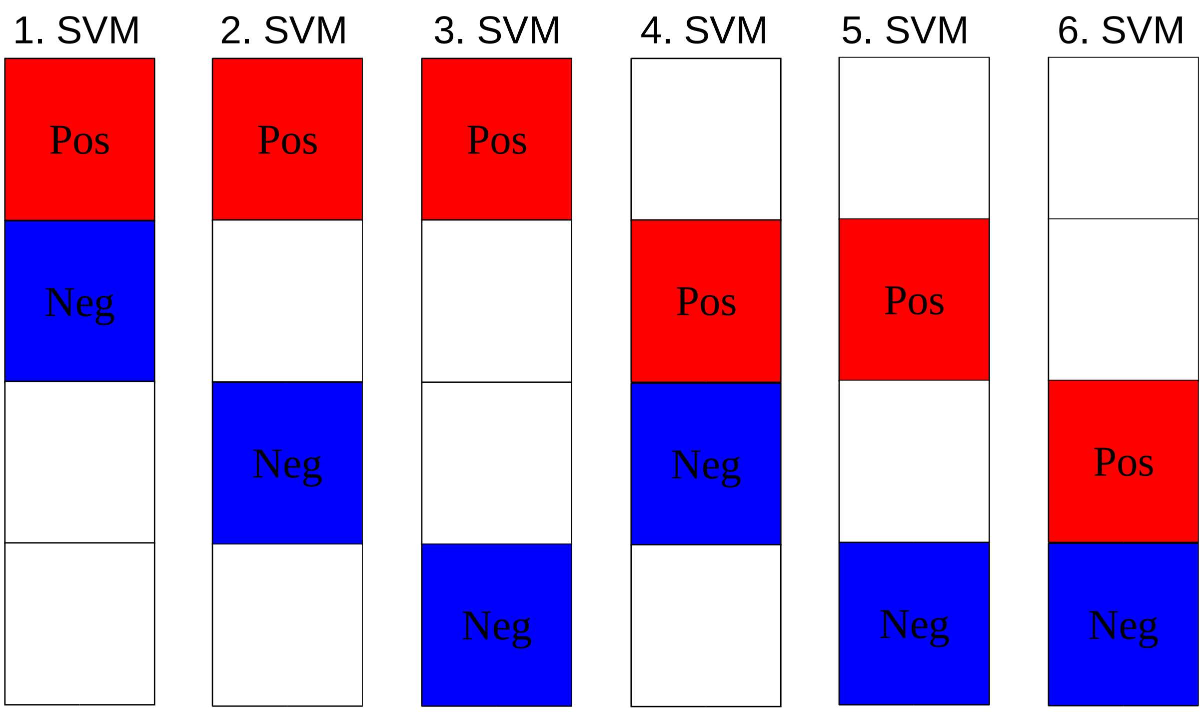
$$w_k x + b_k$$

$$x \equiv argmax_{i=1,\ldots,k}(w_i x + b_i)$$

One-Against-All



One-Against-One



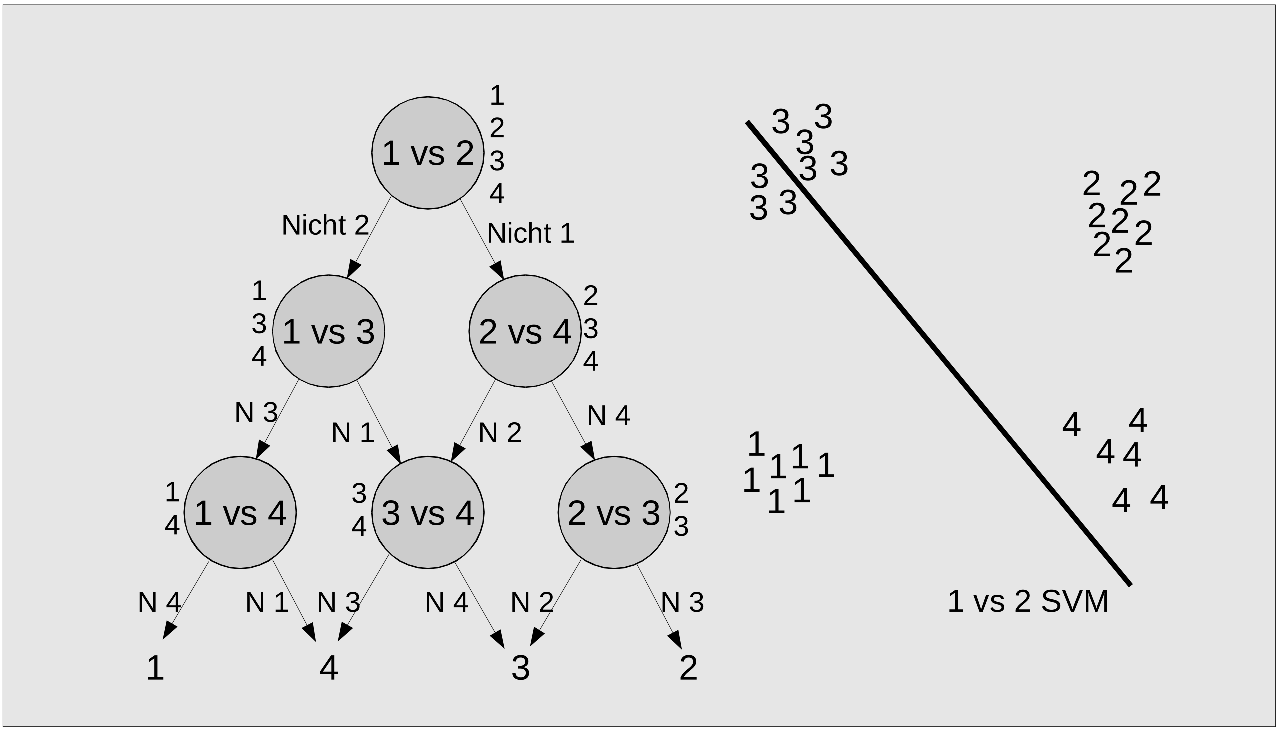
One-Against-One

- Es werden k(k-1)/2 SVMs erzeugt.
- Zum Testen wird die “Max Wins” Strategie benutzt.
- Ein Test sign(w_{ij} x + b_{ij}) entscheidet zwischen den Klassen i und j
- Für jede Klasse wird die Anzahl der für sie positiven Tests gezählt.
- Die Klasse die die meisten Tests für sich entschieden hat, wird als Ergebnis ausgegeben.

Directed Acyclic Graph SVM (DAGSVM)

- Die DAGSVM Trainings Phase ist die selbe wie bei der One-Agains-One Methode.
- Zum Testen wird ein binärer azyklischer Graph benutzt.
- Der Graph enthält k(k-1)/2 interne Knoten und k Randknoten.
- Jeder interne Knoten steht für eine SVM.
- Vorteil: Kürzere Testzeiten im Vergleich zu One-Against-One

Directed Acyclic Graph SVM (DAGSVM)



Methode von Vapnik

- Die Idee ist der des One-Against-All Ansatzes ähnlich.
- Es werden k zwei Klassen Regeln erzeugt.
- Die Regeln werden, im Gegensatz zu One-Against-All, durch das Lösen einer Gleichung erzeugt.
- Die Entscheidungsfunktion ist wieder identisch zu One-Against-All.

Methode von Vapnik

$$\min_{w,b,} \frac{1}{2}\sum_{m=1}^{k} \|w_i\|^2 + C\sum_{i=1}^{l}\sum_{m\neq y_i} {}_{mi}$$

$$w_{y_i}x_i + b_{y_i} \geqslant w_m x_i + b_m + 2 - {}_{mi},$$

$${}_{ij} \geqslant 0,\ j=1,\dots,l,\ m\in\{1,\dots,k\}\setminus y_i$$

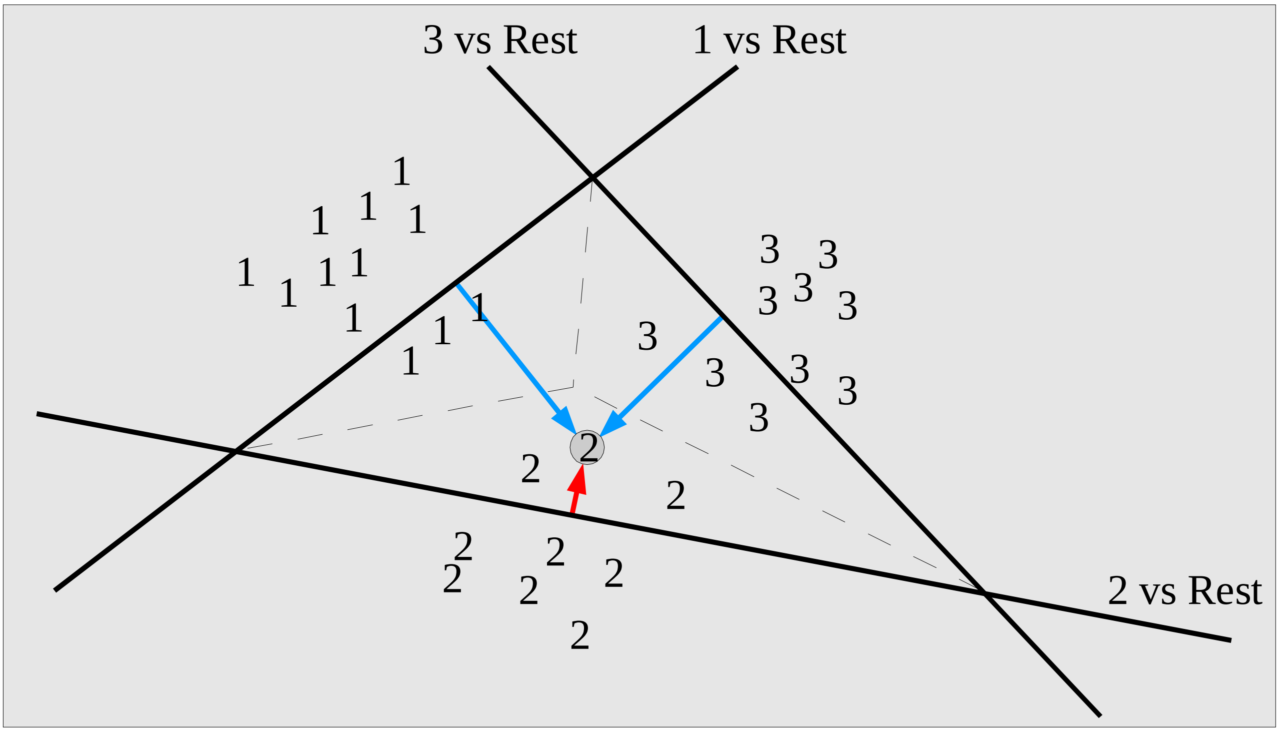
$$w_1 x + b_1$$

$$\dots$$

$$w_k x + b_k$$

$$x \equiv argmax_{m=1,\dots,k}(w_m x + b_m)$$

Methode von Vapnik



Methode von Vapnik

- Zur Implementation wurde eine Decomposition Methode verwendet.
-

Methode von Crammer und Singer

- Das Prinzip ist das selbe, wie bei der Methode von Vapnik.
- Im Gegensatz zur Methode von Vapnik werden hier nur *l* Lockerungsvariablen eingesetzt. *l* entspricht der Anzahl der Trainingsdatensätze.
- Diese Methode verzichtet ausserdem auf den Koeffizienten b_i .

$$\min_{w_m,\ \ _i} \frac{1}{2} \sum_{m=1}^{k} \|w_m\|^2 + C \sum_{i=1}^{l} \quad _i$$

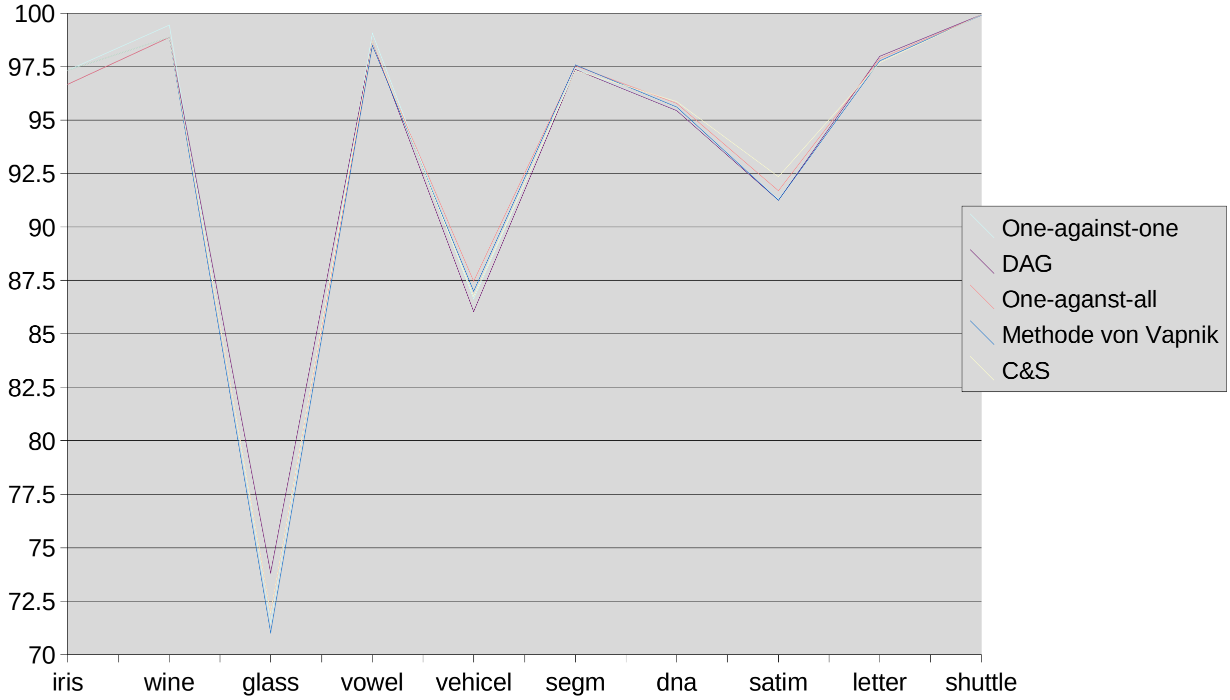
$$w_{y_i} x_i - w_m x_i \geqslant e_{mi} - \quad _i , i = 1, \ldots, l ,$$

$$wenn\ y_i = m\ dann\ e_{mi} = 0, sonst\ e_{mi} = -1$$

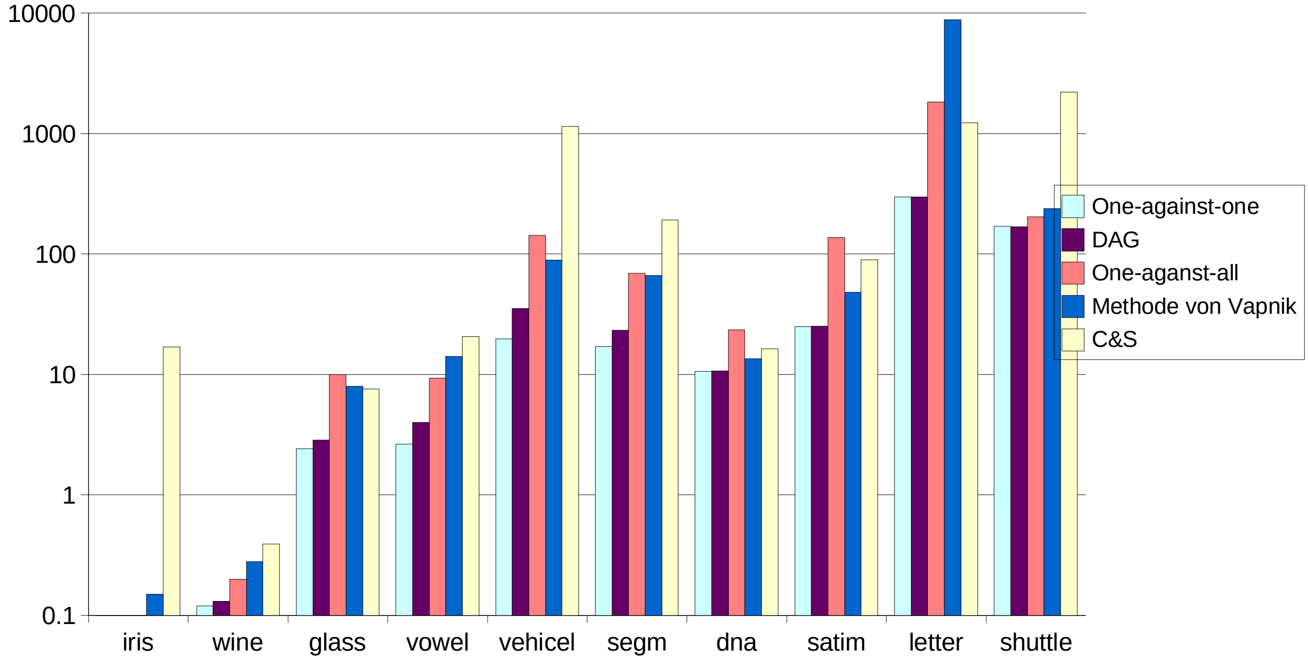
Experimente und Resultate

- Quellen: UCI Repository und Statlog Collection
- Für jedes Problem werden die bestmöglichen Parameter gesucht.
- $\gamma = [2^4, 2^3, 2^2, \ldots, 2^{-10}]$ und $C = [2^{12}, 2^{11}, 2^{10}, \ldots, 2^{-2}]$
- Das entspricht 15x15 = 255 Kombinationen.
- Als Stopkriterium für den Optimierungsalgorithmus wurde in den Tests eine KKT Verletzung von $< 10^{-3}$ verwendet.

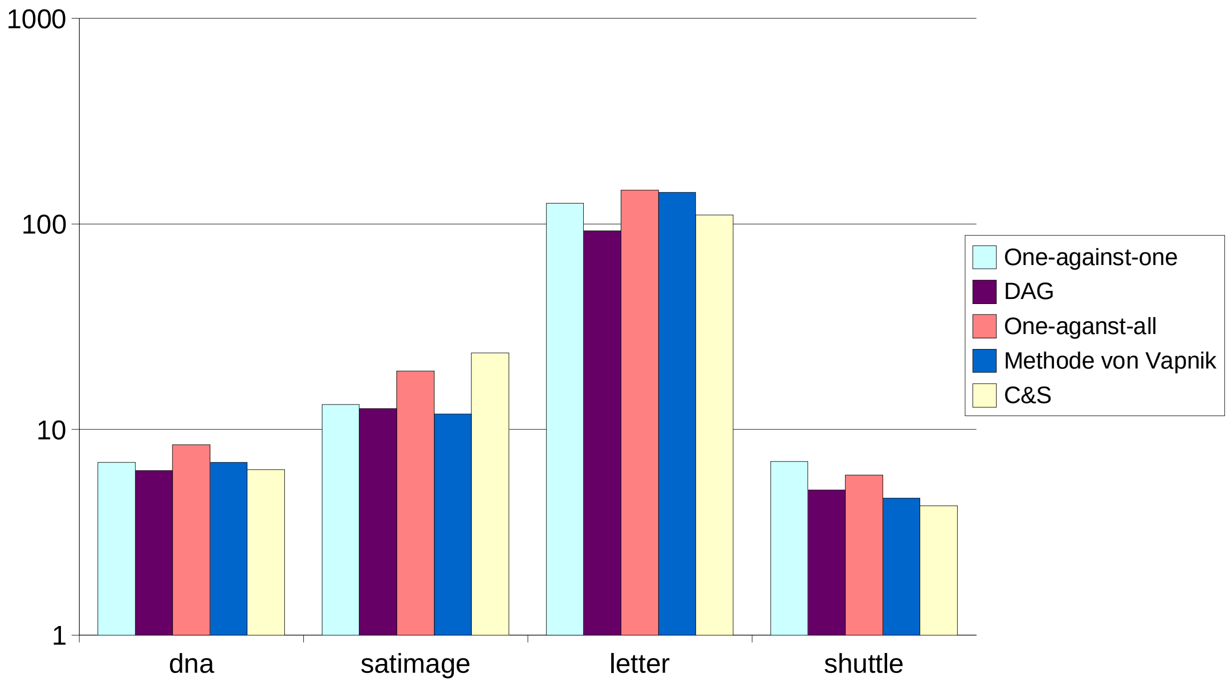
Vergleich der Klassifikationsgenauigkeit (in Prozentangaben)



Training Zeit (Logarithmisch Dargestellt)



Test Zeit (Logarithmisch Dargestellt)



Quellenverweise

[1] C.W.Hsu, C.J. Lin. A Comparison of Methods for Multi-class Support Vector Machines. National Taiwan University, Department of Computer Science and Information Engineering.

[2] Blake, C.L. & Merz, C.J. (1998). UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science.

[3] J. C. Platt, N. Cristiani, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. In *Advances in Neural Information Processing Systems*, volume 12, pages 547-553. MIT Press, 2000.

[4] V.Vapnik. *Statistical Learning Theory*. Wiley, New York, NY, 1998.