

SVM-KM

Seminários 2 - Técnicas Clássicas de Reconhecimento de Padrões

Matheus Bitarães de Novaes - PPGEE - UFMG

Sumário

1. Treinamento de uma SVM
2. Proposta do Artigo
3. Exemplos do Artigo
4. Aplicações

Treinamento de uma SVM

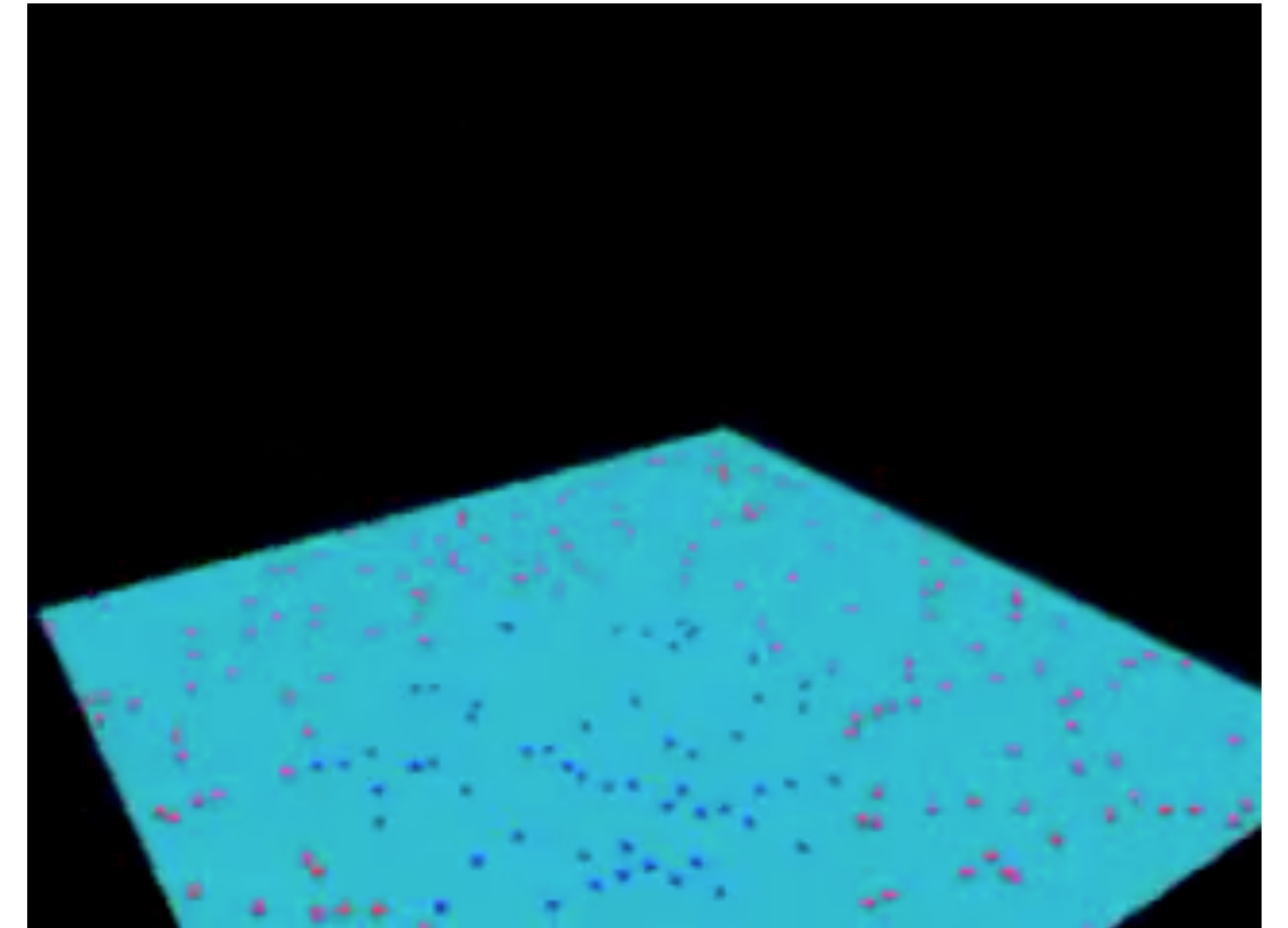
- Encontrar hiperplano ótimo que consiga separar as classes do problema.
- Equação do hiperplano:

$$w^T \psi(x) + b = 0$$

w^T : Vetor de pesos

$\psi(x)$: Mapeia vetor x em um espaço de maior dimensão onde os dados serão linearmente separáveis

b : *bias*



treinamento SVM [5]

- Este hiperplano é encontrado através de uma otimização quadrática.
- Variáveis de folga podem ser adicionadas para que as margens não sejam rígidas.

Treinamento de uma SVM

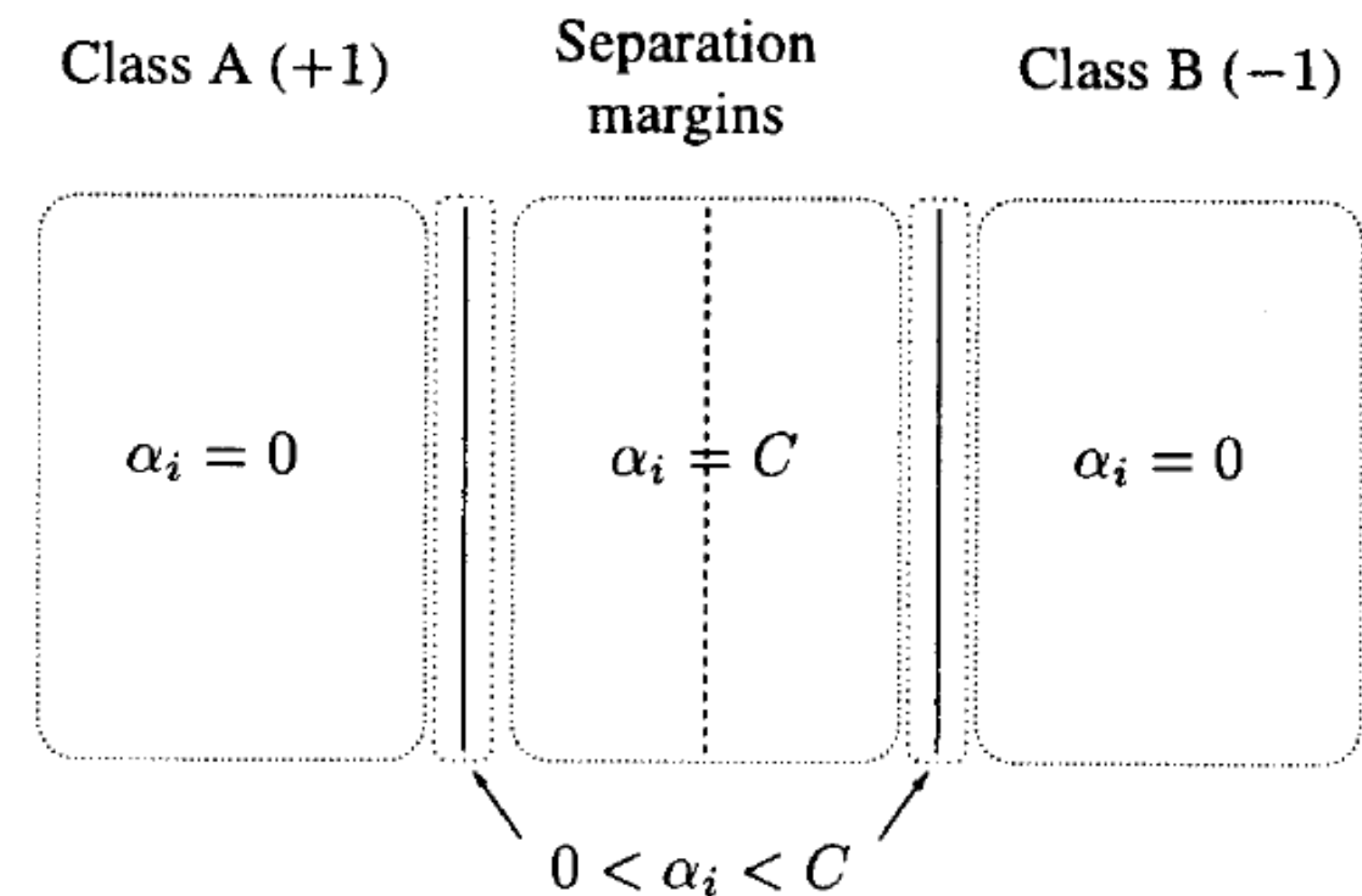
- A função $\psi(x)$ é substituída por funções de Kernel conhecidas, com certos hiper-parâmetros informados *a priori*.
- A formulação final do problema de otimização fica assim:

Dados o parâmetro C e a matriz \mathbf{Z} , encontre Multiplicadores de Lagrange α_i que maximizem a função

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j, \mathbf{Z})$$

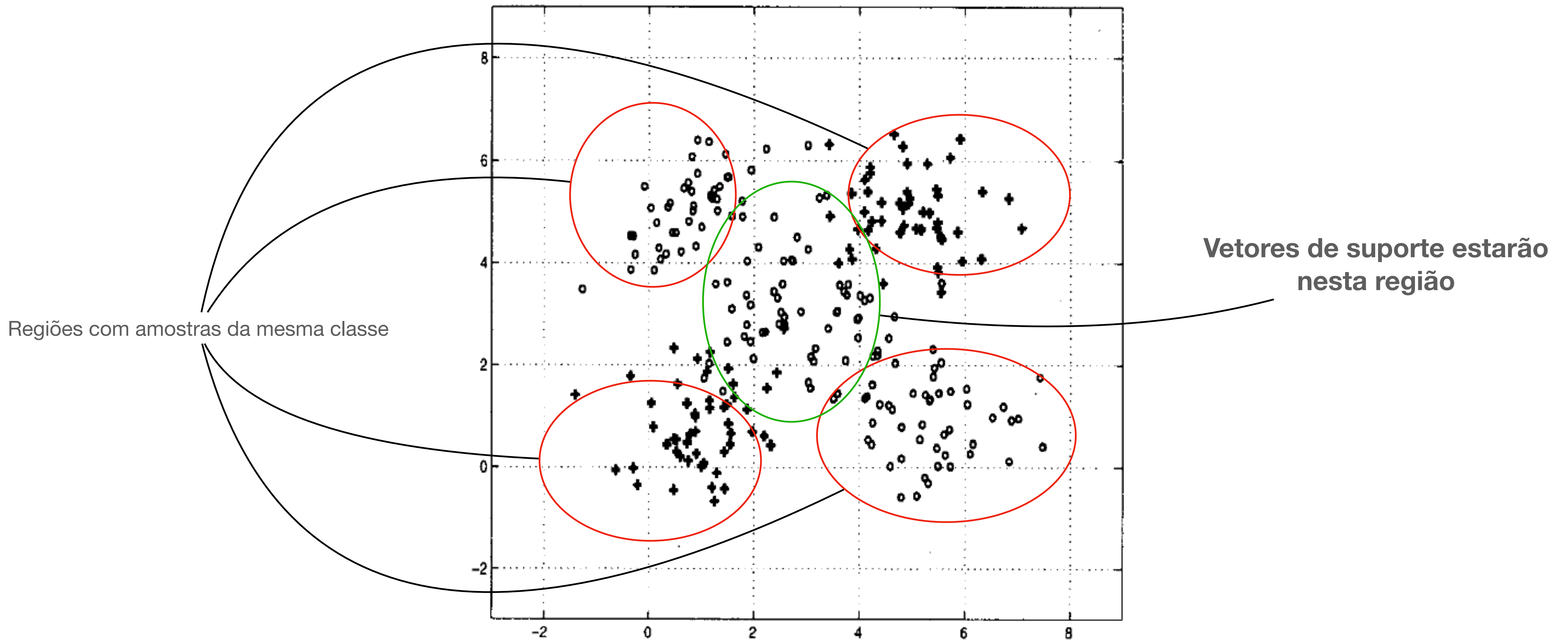
e que satisfaçam às seguintes restrições:

- $\sum_{i=1}^N \alpha_i y_i = 0$
- $0 \leq \alpha_i \leq C$



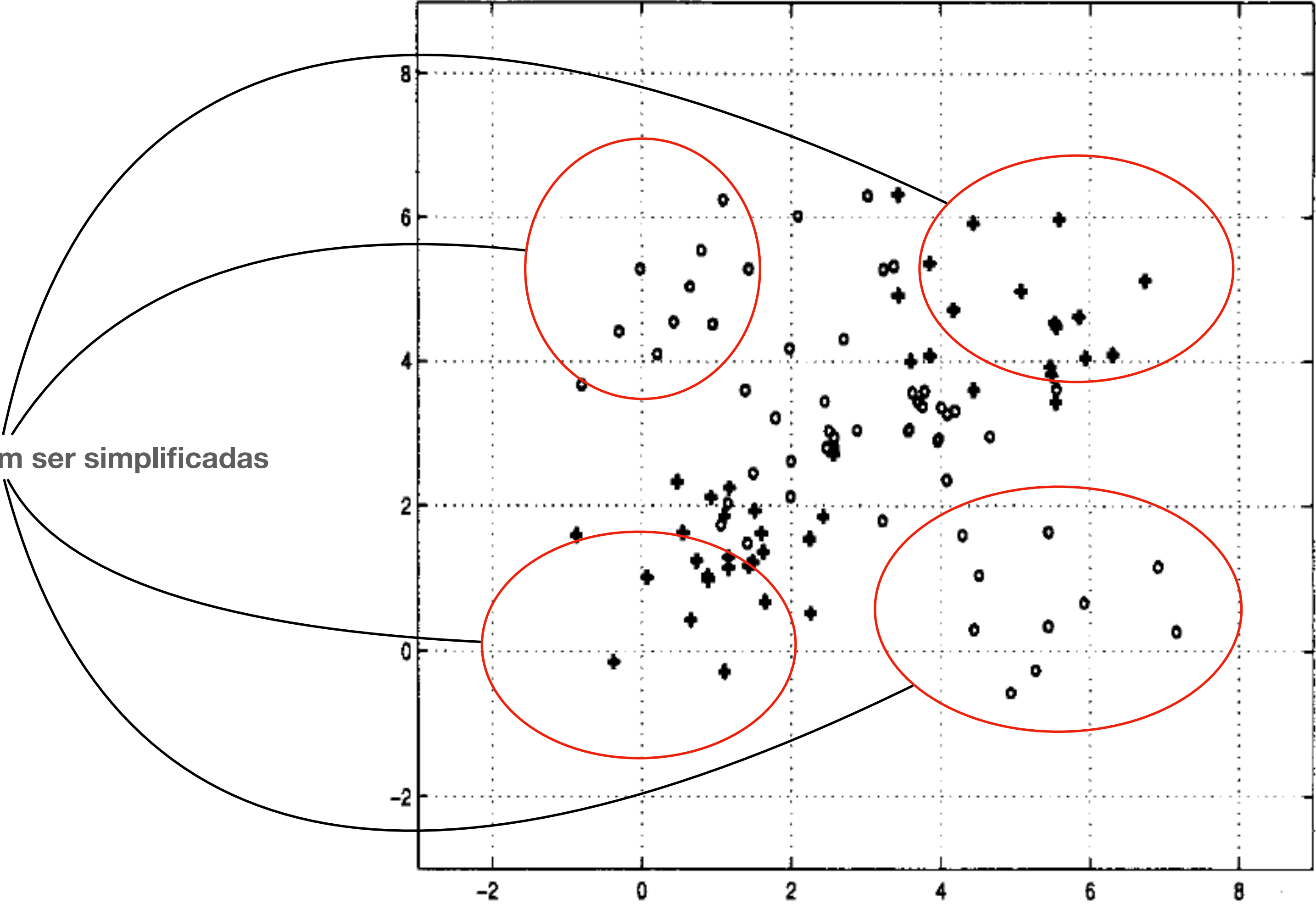
- Dos possíveis vetores a serem gerados, queremos apenas os que estão dentro da margem de separação.
- Se todas as amostras forem consideradas, o espaço de soluções pode conter vetores que não estão dentro da margem de separação.

Proposta do Artigo



Proposta do Artigo

Estas regiões podem ser simplificadas



Proposta do Artigo

SVM-KM: speeding SVMs learning with a priori cluster selection and k-means [1]

- 1) Definir um número de k de *agrupamentos* para serem detectados pelo método de *k-means* (depende do número de vetores de treinamento e da dispersão dos dados).
- 2) Rodar o algoritmo de *k-means* e verificar os k pontos centrais encontrados.
 - 1) Para os agrupamentos com apenas um tipo de classe: substituir todos os vetores deste agrupamento pelo seu centro.
 - 2) Para os agrupamentos com mais de um tipo de classe: Considerar todos os vetores
- 3) Prosseguir o treinamento da SVM com os vetores resultantes.

Exemplos do Artigo

- Dois experimentos com dados não linearmente separáveis.
- Cada um repetido 100 vezes.
- Considerados os valores médios dos principais parâmetros.
- Simulações em *Pentium 400MHZ* com 128mb de memória rodando Linux.

Exemplos do Artigo

- Dois experimentos com dados não linearmente separáveis.
- Cada um repetido 100 vezes.
- Considerados os valores médios dos principais parâmetros.
- Simulações em *Pentium 400MHZ* com 128mb de memória rodando Linux.



Computador do Milhão:

- Intel Celeron de 1 GHz
- 128mb de memória
- Windows XP

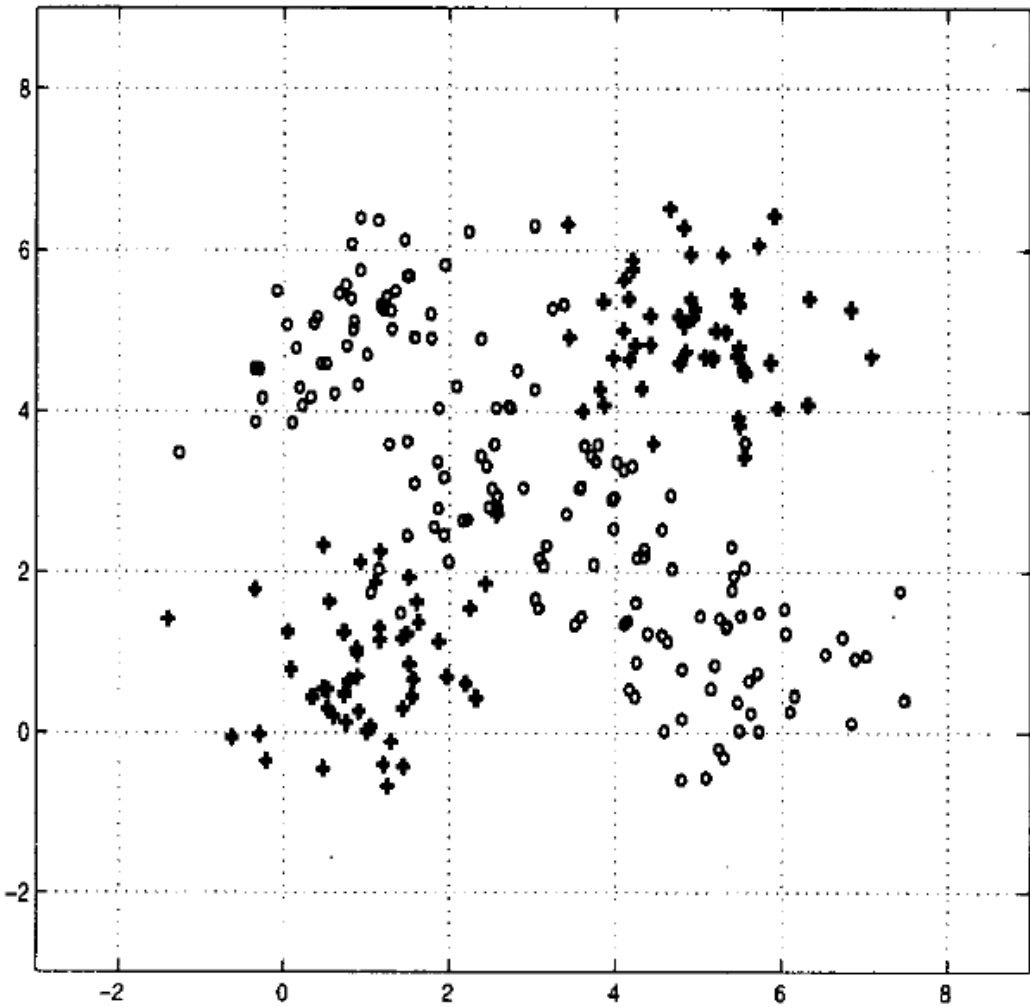
Exemplos do Artigo

Primeiro Experimento

MSE: Mean Square Error
TV: training vectors
SV: support vectors

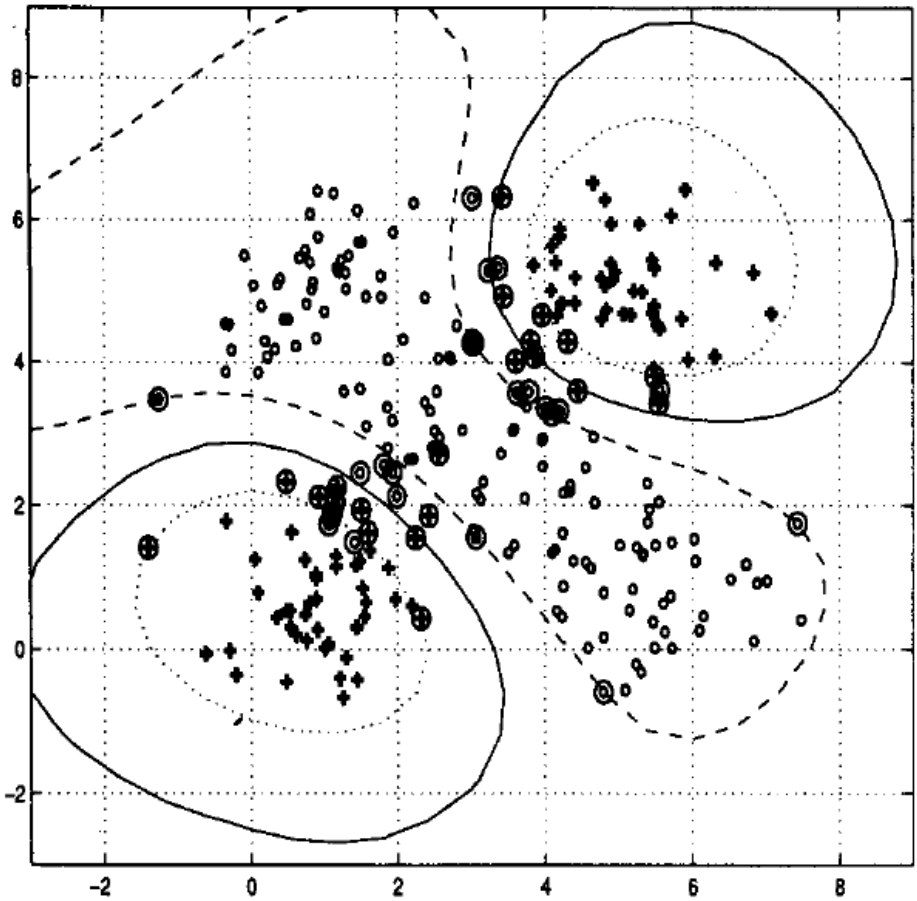
	SMO	SVM-KM (<i>K</i> -MEANS+SMO)
CPU time (s)	2.24	0.71 (0.073+0.64)
$ \mathbf{w} ^2$	46.74	35.95
MSE	3.23	2.88
Num. of TV	250	94.72
Num. of SV	46.61	41.17
Success (%)	93.98%	93.72%

Dados de treinamento



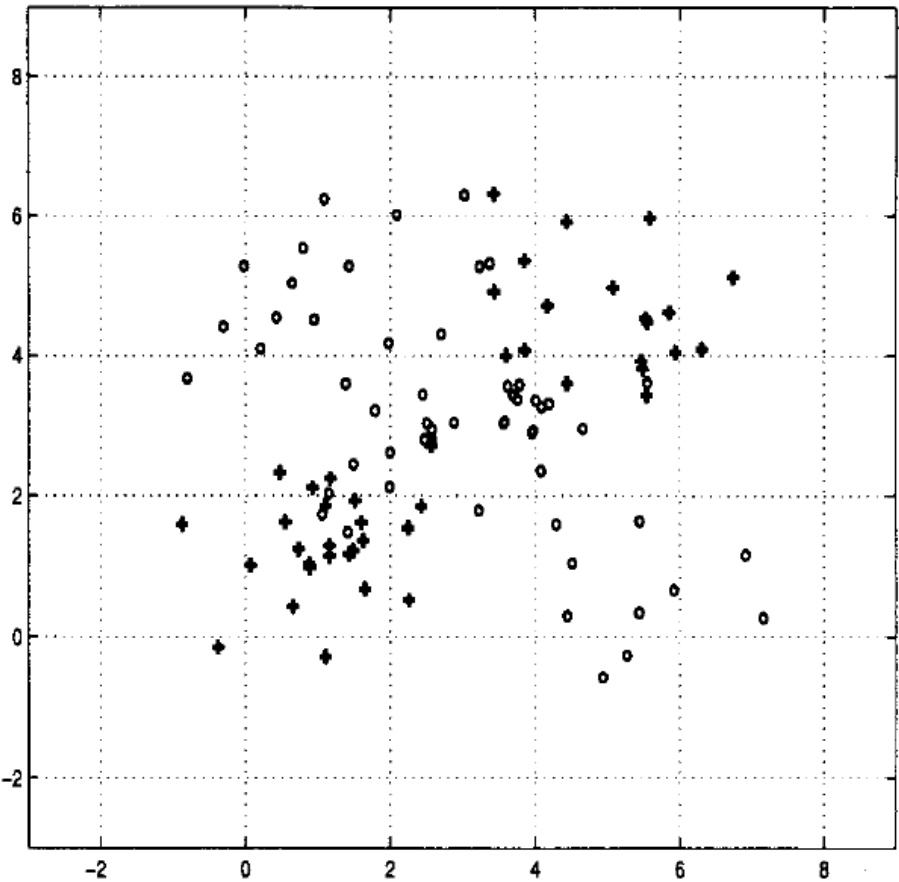
Treinamento SVM

Fronteiras de Decisão

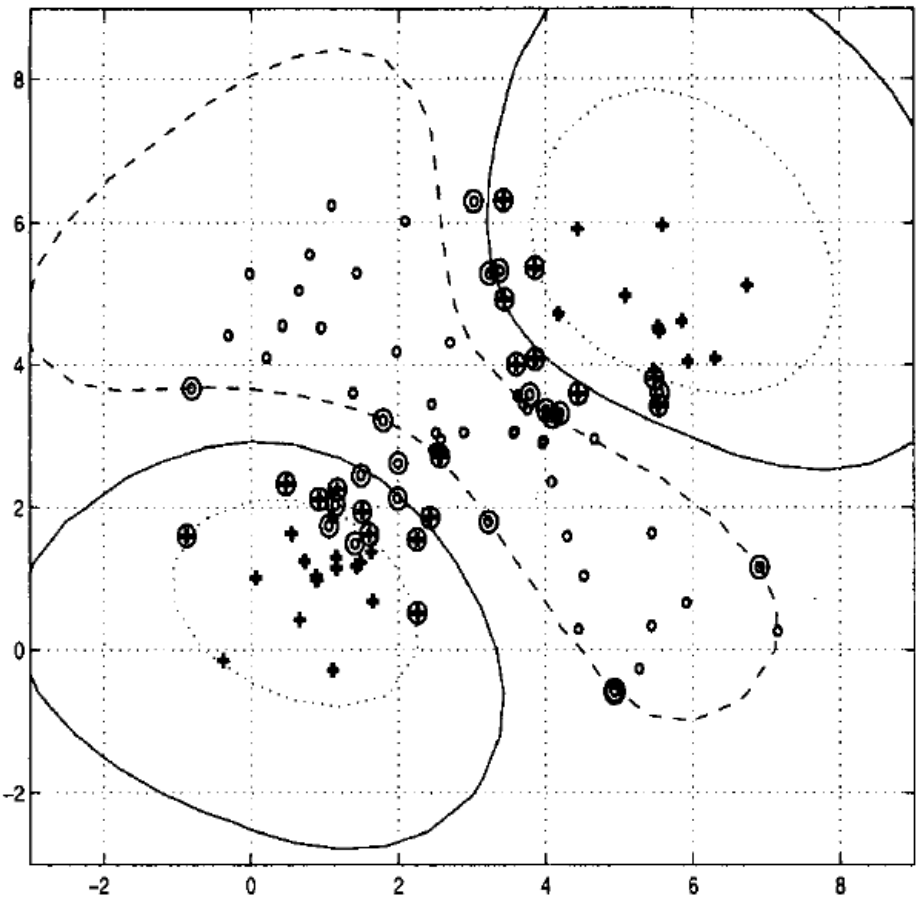


SVM

Aplicação do *k-means*



Treinamento SVM



SVM-KM

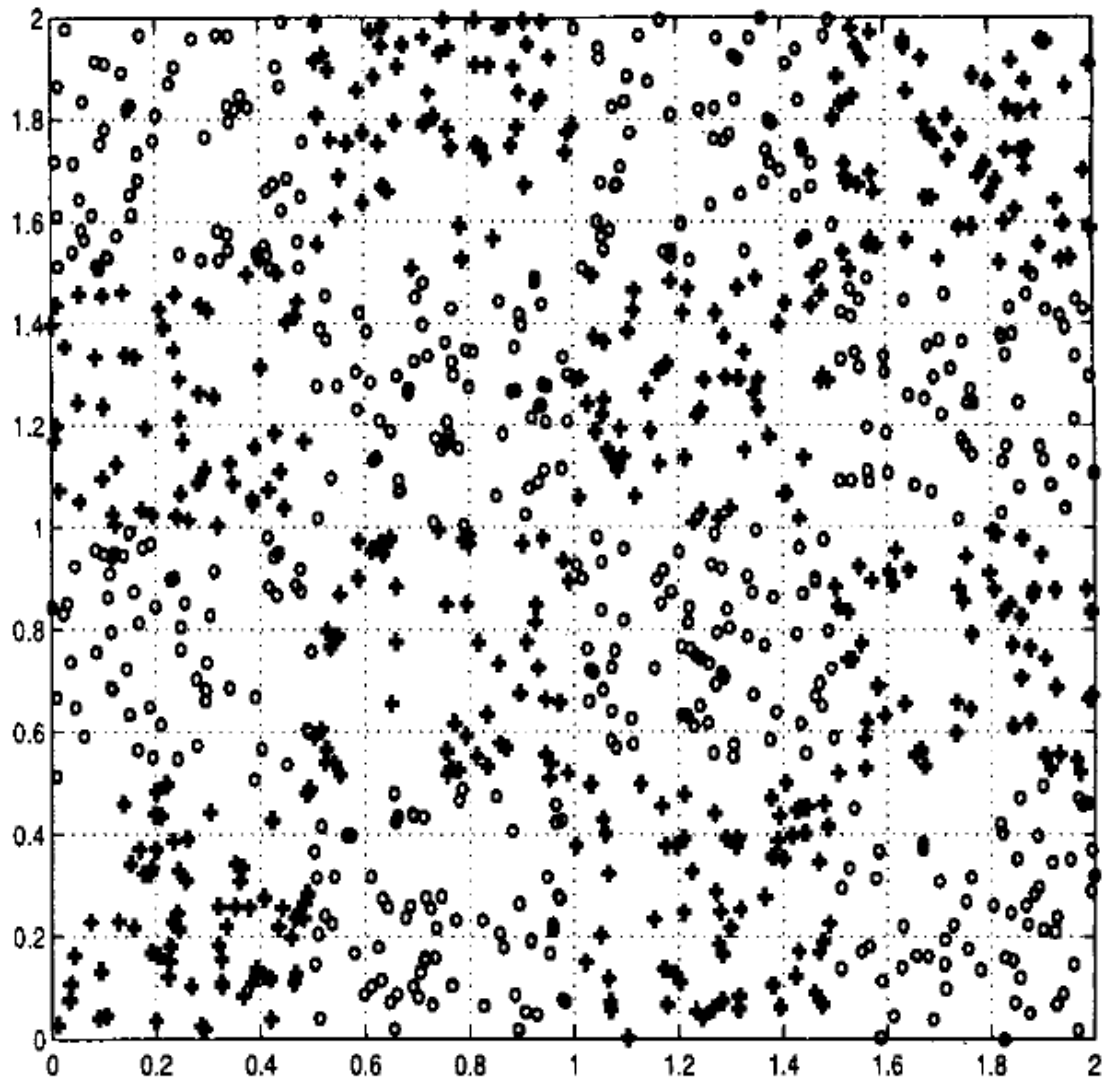
Exemplos do Artigo

Segundo Experimento

MSE: Mean Square Error
TV: training vectors
SV: support vectors

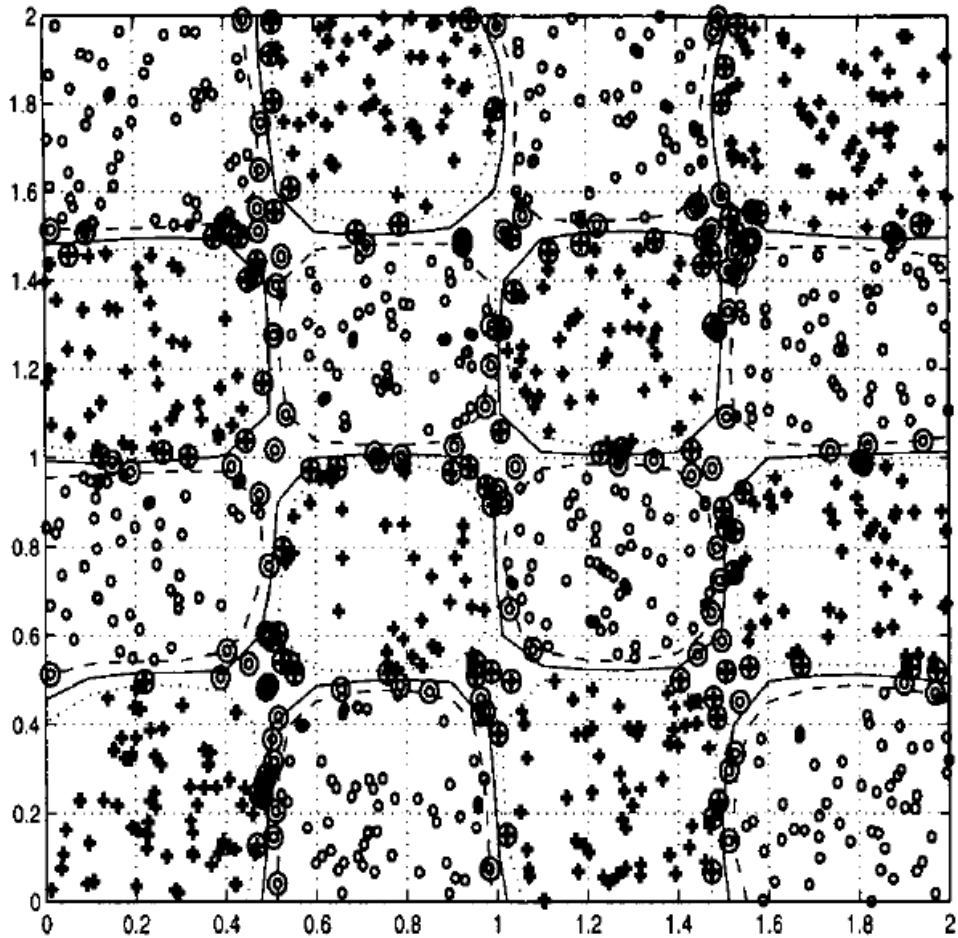
	SMO	SVM-KM (<i>K</i> -MEANS+SMO)
CPU time (s)	32.21	13.89 (0.89+13.00)
$ \mathbf{w} ^2$	1562.2	1349.9
MSE	1.78	1.99
Num. of TV	1000	497.15
Num. of SV	172.19	159.69
Success (%)	95.96%	94.92%

Dados de treinamento



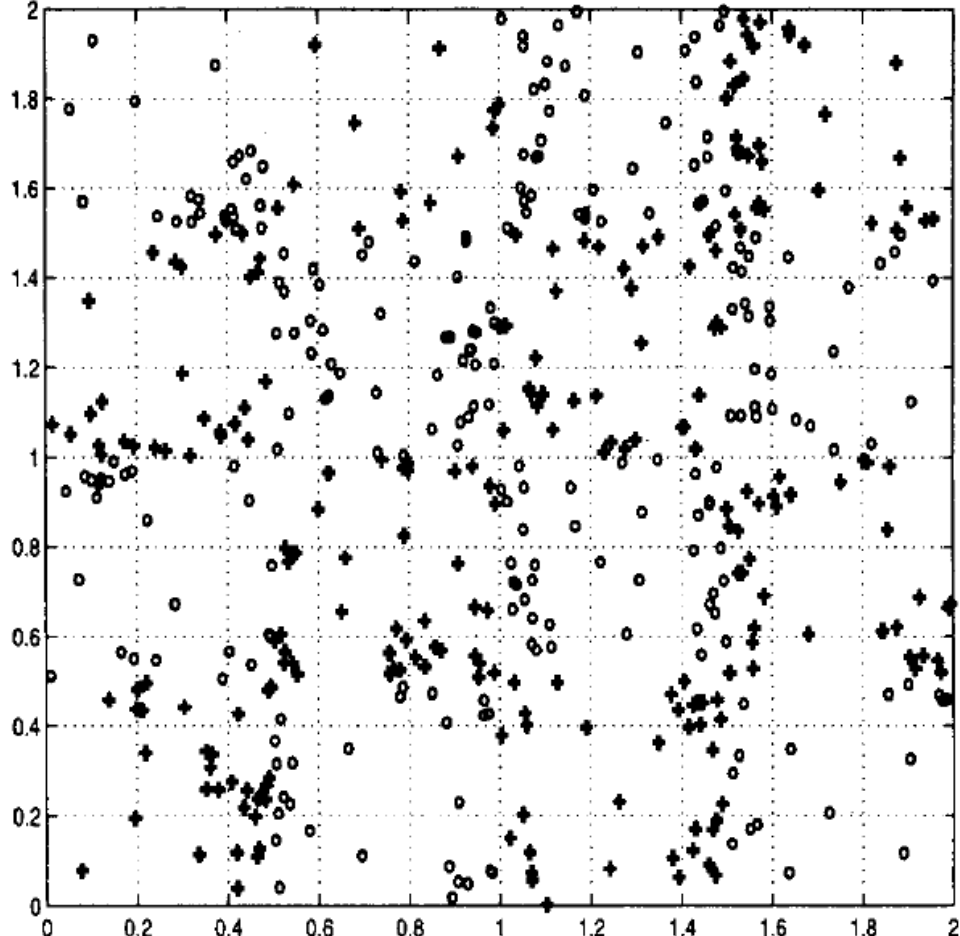
Treinamento SVM

Fronteiras de Decisão

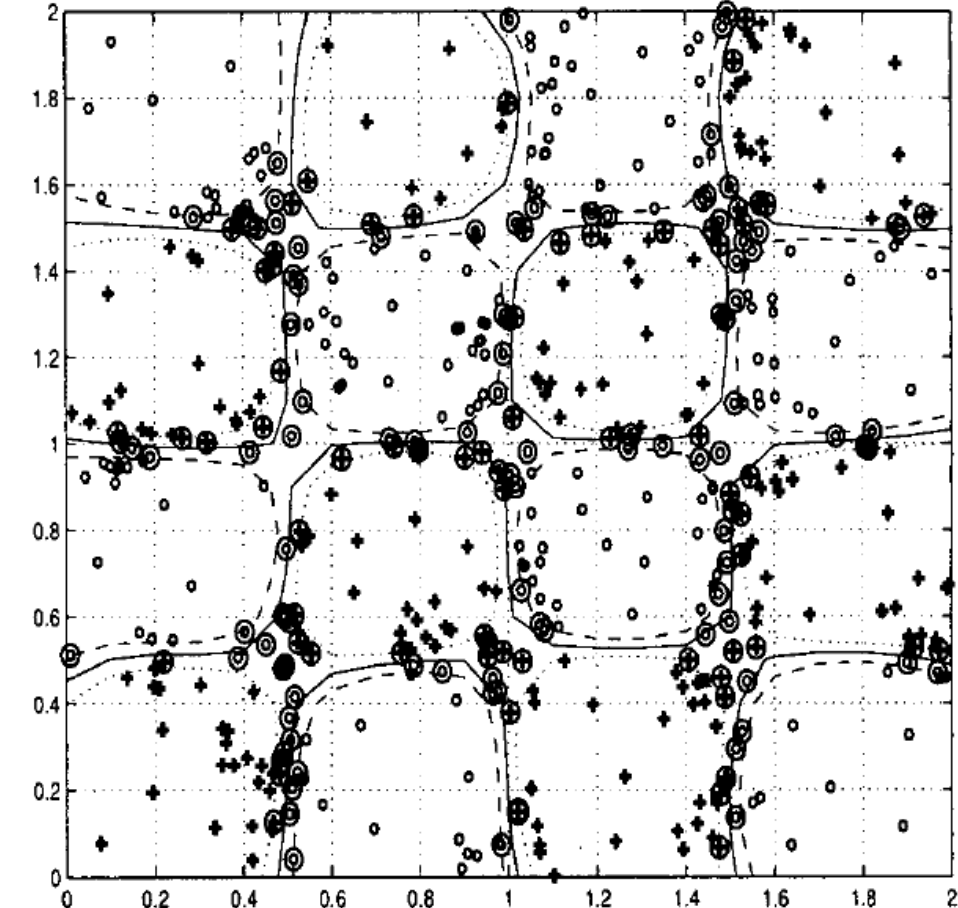


SVM

Aplicação do *k-means*



Treinamento SVM



SVM-KM

Aplicações

(2015) Moisture damage evaluation in SBS and lime modified asphalt using AFM and artificial intelligence [2]

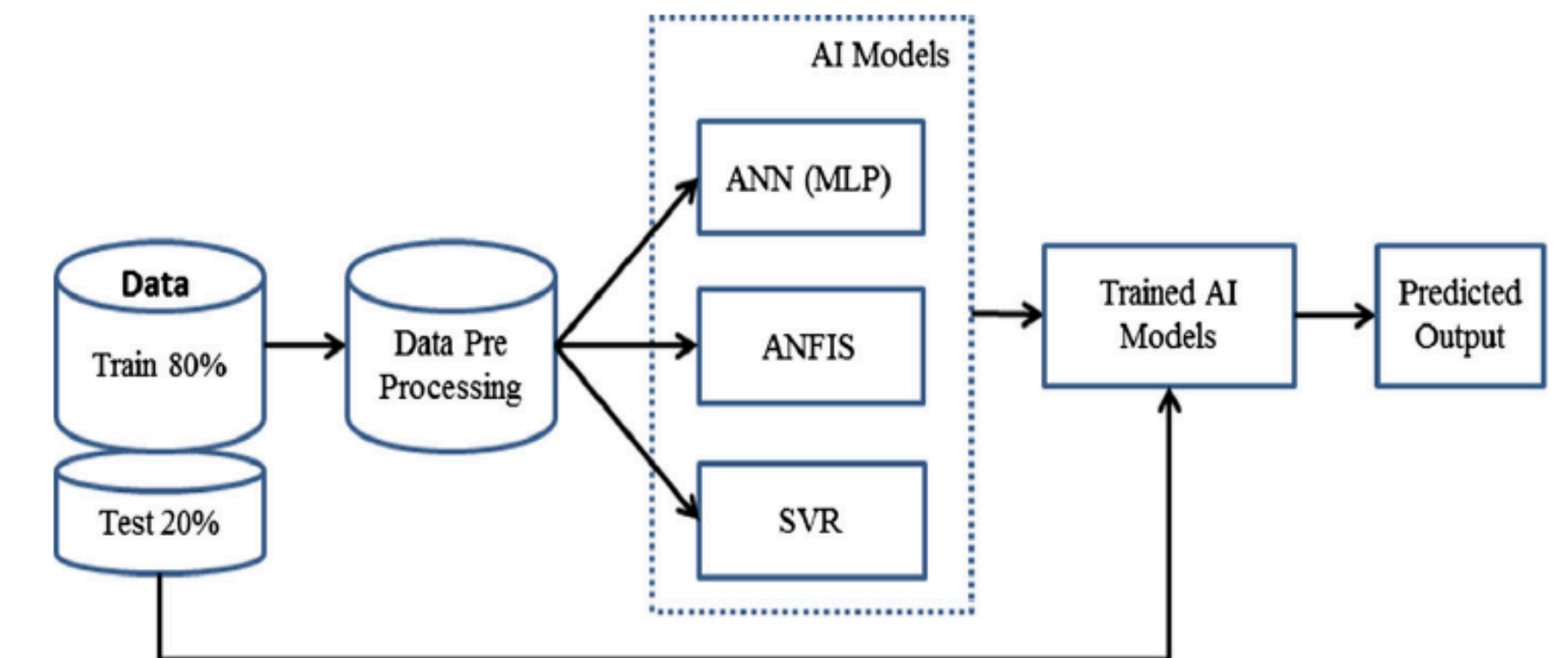
- Predição de danos por umidade asfaltos modificados com cal e polímero.
- Modelos para determinar a vulnerabilidade dos tipos de asfalto em condições secas e molhadas.

Table 2 Performance measures of the ANN and ANFIS models for wet sample

Statistical parameter	SVM-KM		MLP		ANFIS_by_FCM	
	Train (80 %)	Test (20 %)	Train (80 %)	Test (20 %)	Train (80 %)	Test (20 %)
RMSE	40.1193	63.2925	14.7293	57.7385	34.9144	36.4708
CC	0.9426	0.8046	0.9895	0.7858	0.9376	0.9309
MAPE	18.8938	17.6330	7.3598	15.9455	17.1568	15.8106
MAE	32.6608	46.2858	11.6650	43.5721	27.4170	30.8518
MRE	0.1889	0.1763	0.0736	0.1595	0.0233	0.0379
Time	0.3120	0.0000	0.8736	0.0624	0.7098	0.1501
SD	6.0030	9.6309	0.4915	11.5508	0.3425	0.0600

Table 3 Performance measures of the ANN and ANFIS models for dry sample

Statistical parameter	SVM-KM		MLP		ANFIS by FCM	
	Train (80 %)	Test (20 %)	Train (80 %)	Test (20 %)	Train (80 %)	Test (20 %)
RMSE	38.2850	40.3204	28.8451	33.0028	37.3467	36.0526
CC	0.8531	0.9505	0.9211	0.8096	0.8529	0.7223
MAPE	19.8008	21.6526	16.7708	17.5623	23.2327	24.1009
MAE	17.8519	31.6254	21.0612	24.9595	30.0750	30.6259
MRE	0.1980	0.2165	0.1677	0.1756	0.0986	0.3672
Time	0.3276	0.0158	7.4724	0.0312	0.1905	0.4873
SD	2.6754	4.5292	2.8053	5.6481	0.2446	0.3602



CC: Coeficiente de correlação
 RMSE: Erro quadrático médio
 MAPE: Erro médio percentual
 MAE: Erro médio relativo
 SD: Desvio padrão

Aplicações

(2014) Prediction of non-hydrocarbon gas components in separator by using Hybrid Computational Intelligence models [3]

- Predição de presença de gases não-hidrocarbonetos em um separador para redução de custo em processos com petróleo.
- SVM-KM é utilizada como uma das abordagens testadas
- Modelos híbridos também foram testados (incluindo Algoritmos Genéticos).

Table 4 Performance of CI models on test data

Test	PR-EOS		MLP			SVR			ANFIS		
	CC	RMSE	CC	<i>p</i> value	RMSE	CC	<i>p</i> value	RMSE	CC	<i>p</i> value	RMSE
N ₂	0.9586	0.7402	0.9403	0.0001	1.5183	0.9843	0.0000	1.2947	0.9764	0.0000	0.6331
CO ₂	0.9989	0.3114	0.9669	0.0000	1.2819	0.9419	0.0000	2.0379	0.9693	0.0000	1.1854
H ₂ S	0.6800	2.0037	0.9160	0.0019	1.2807	0.9149	0.0002	2.8769	0.7828	0.0044	1.0078

CC: Coeficiente de correlação

p value: *p*-valor da correlação

RMSE: Erro quadrático médio

Aplicações

(2004) Cluster-based support vector machines in text-independent speaker identification [4]

- SVM-KM para identificação de orador.
- Outra abordagem de agrupamento é proposta.

Método proposto no artigo

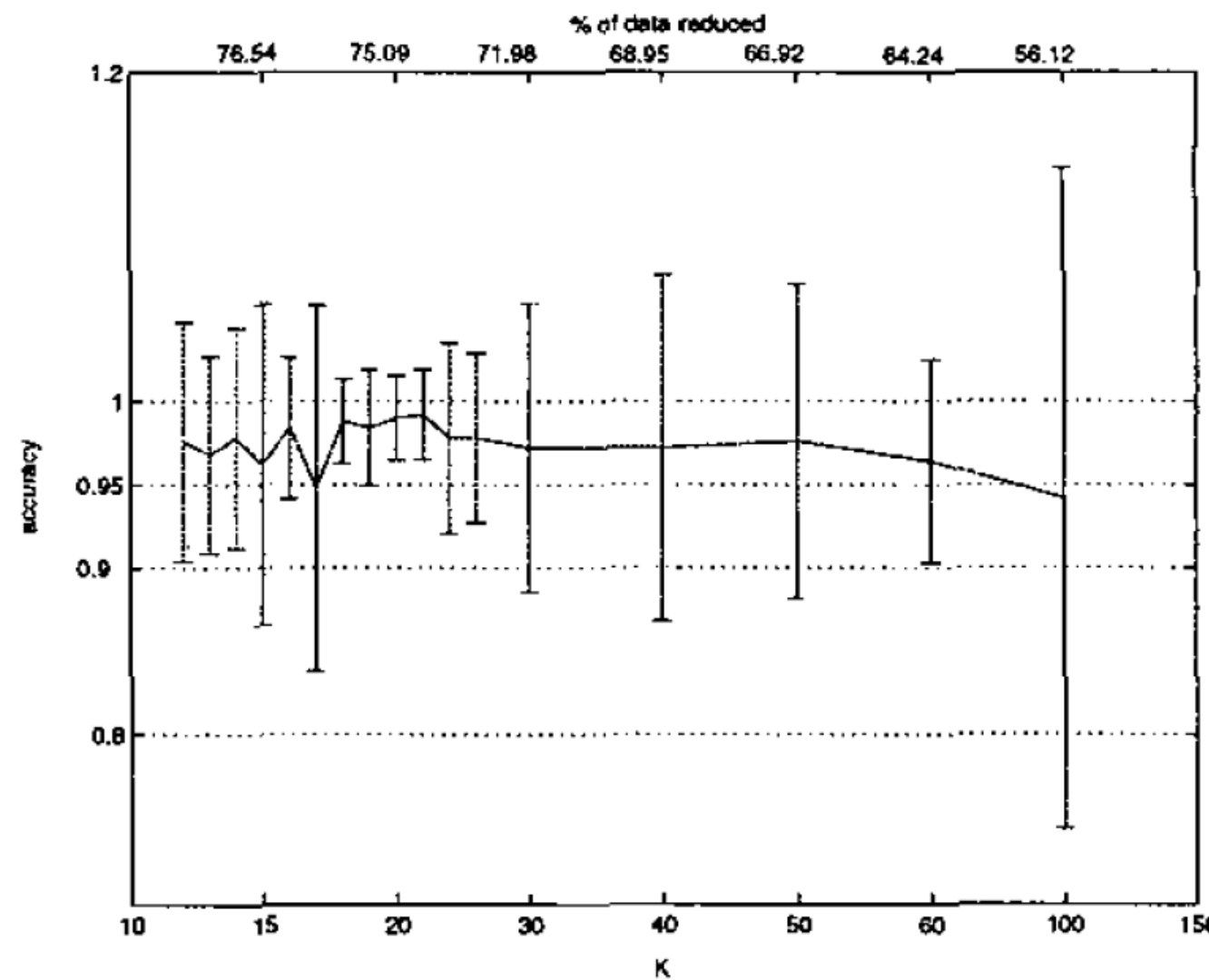


Fig. 7. The accuracy curve of our method with the reduction ratio T to be 0.7, and $k=12 \sim 100$ for 20 speakers. The Standard deviation are very small when $18 \leq k \leq 21$.

SVM-KM

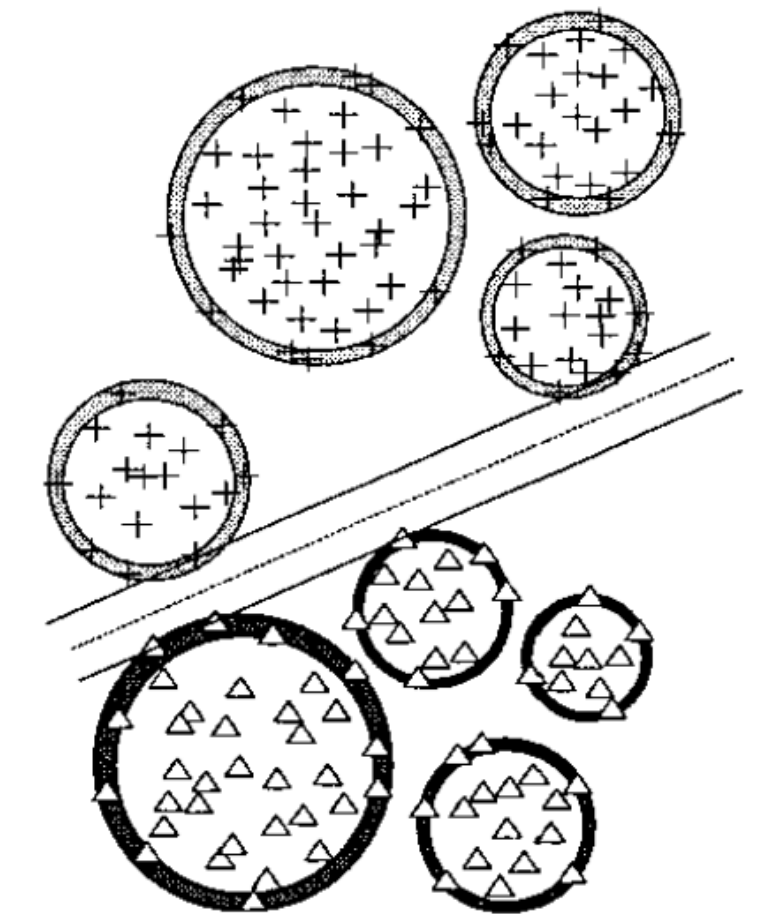
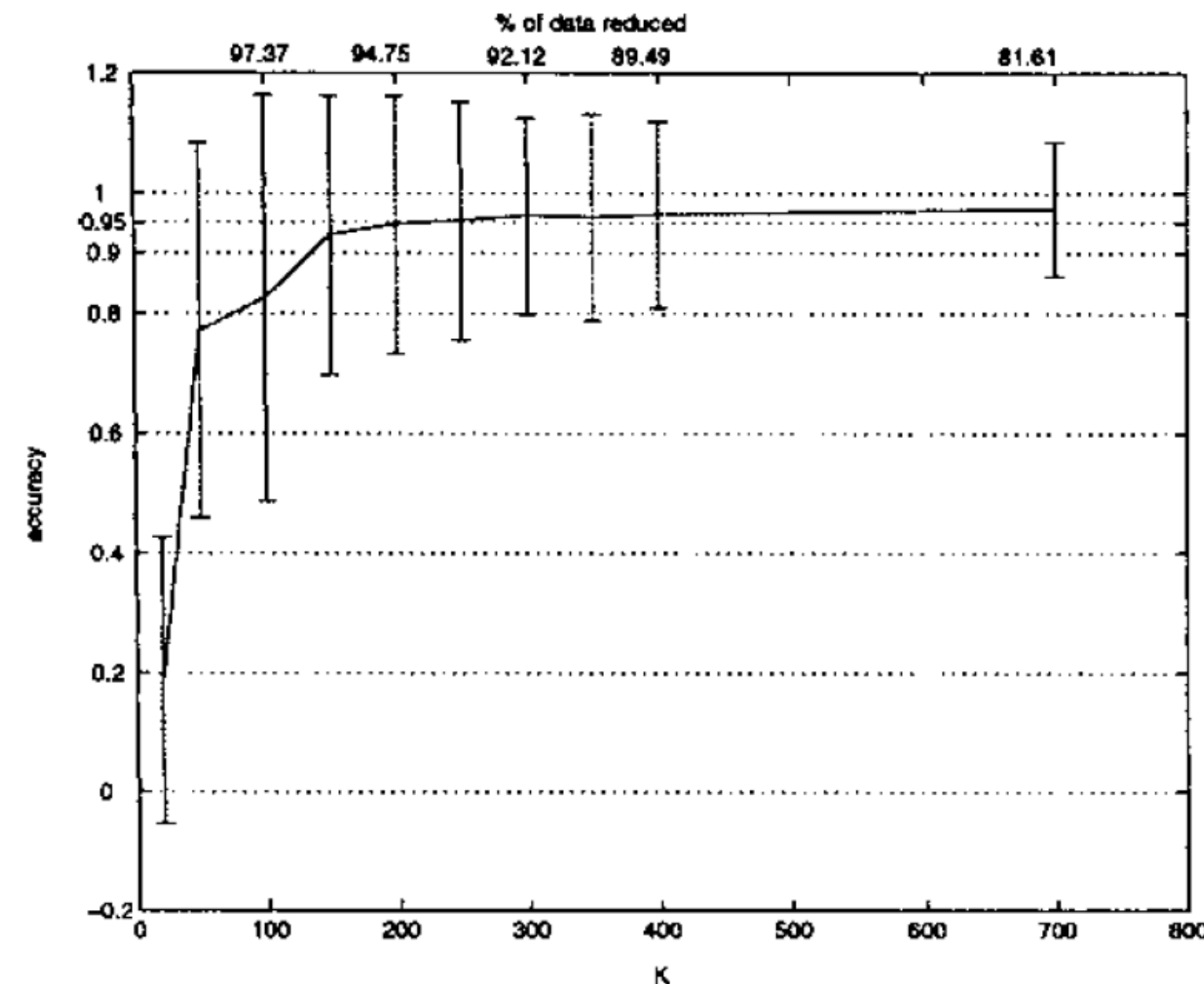


Fig. 3. The original training data with two classes and four clusters each.

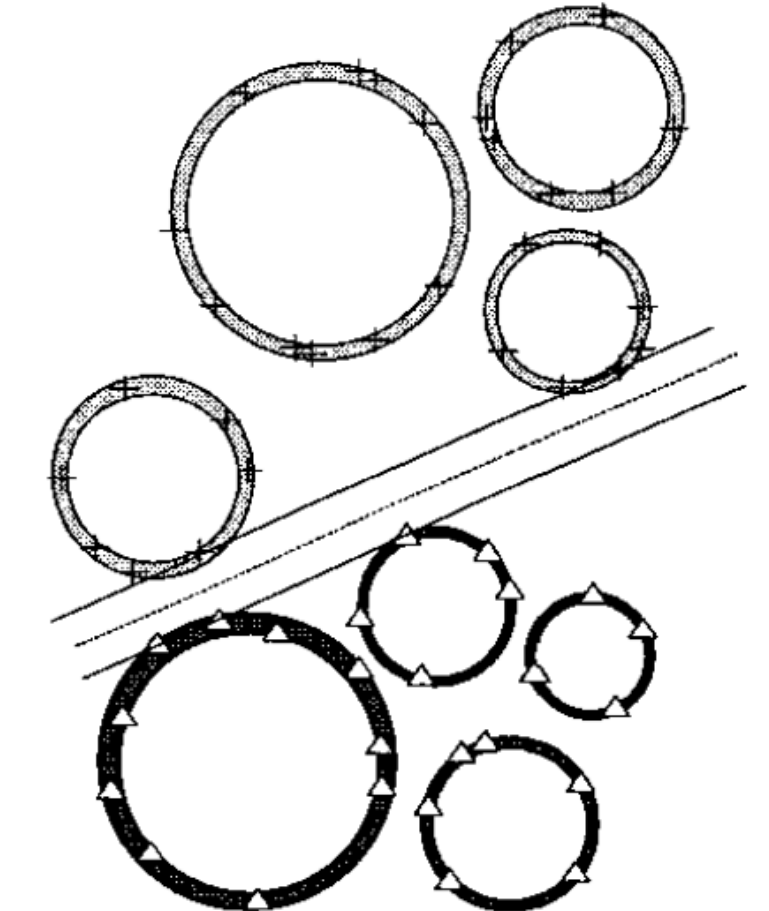


Fig. 4. The training data after selecting

OBRIGADO

Referências

- [1] M. Barros de Almeida, A. de Padua Braga and J. P. Braga, "SVM-KM: speeding SVMs learning with a priori cluster selection and k-means," *Proceedings. Vol.1. Sixth Brazilian Symposium on Neural Networks*, 2000, pp. 162-167, doi: 10.1109/SBRN.2000.889732.
- [2] Arifuzzaman, M., Islam, M.S. & Hossain, M.I. Moisture damage evaluation in SBS and lime modified asphalt using AFM and artificial intelligence. *Neural Comput & Applic* 28, 125–134 (2017). <https://doi.org/10.1007/s00521-015-2041-6>
- [3] Helmy, T., Hossain, M.I., Abdulraheem, A. *et al.* Prediction of non-hydrocarbon gas components in separator by using Hybrid Computational Intelligence models. *Neural Comput & Applic* 28, 635–649 (2017). <https://doi.org/10.1007/s00521-015-2088-4>
- [4] Sheng-Yu Sun, C. L. Tseng, Y. H. Chen, S. C. Chuang and H. C. Fu, "Cluster-based support vector machines in text-independent speaker identification," *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, 2004, pp. 729-734, doi: 10.1109/IJCNN.2004.1380008.
- [5] <https://gfycat.com/cluelessdefinitiveblackandtancoonhound-howto-style-mathmatics-polynomial>