

Reduce the Number of Support Vectors by Using Clustering Techniques

Anh Tran Quang ⁽¹⁾, Qianli Zhang, Xing Li

Email ⁽¹⁾: chenguangying@tsinghua.org.cn

Network Research Center
Tsinghua University, China

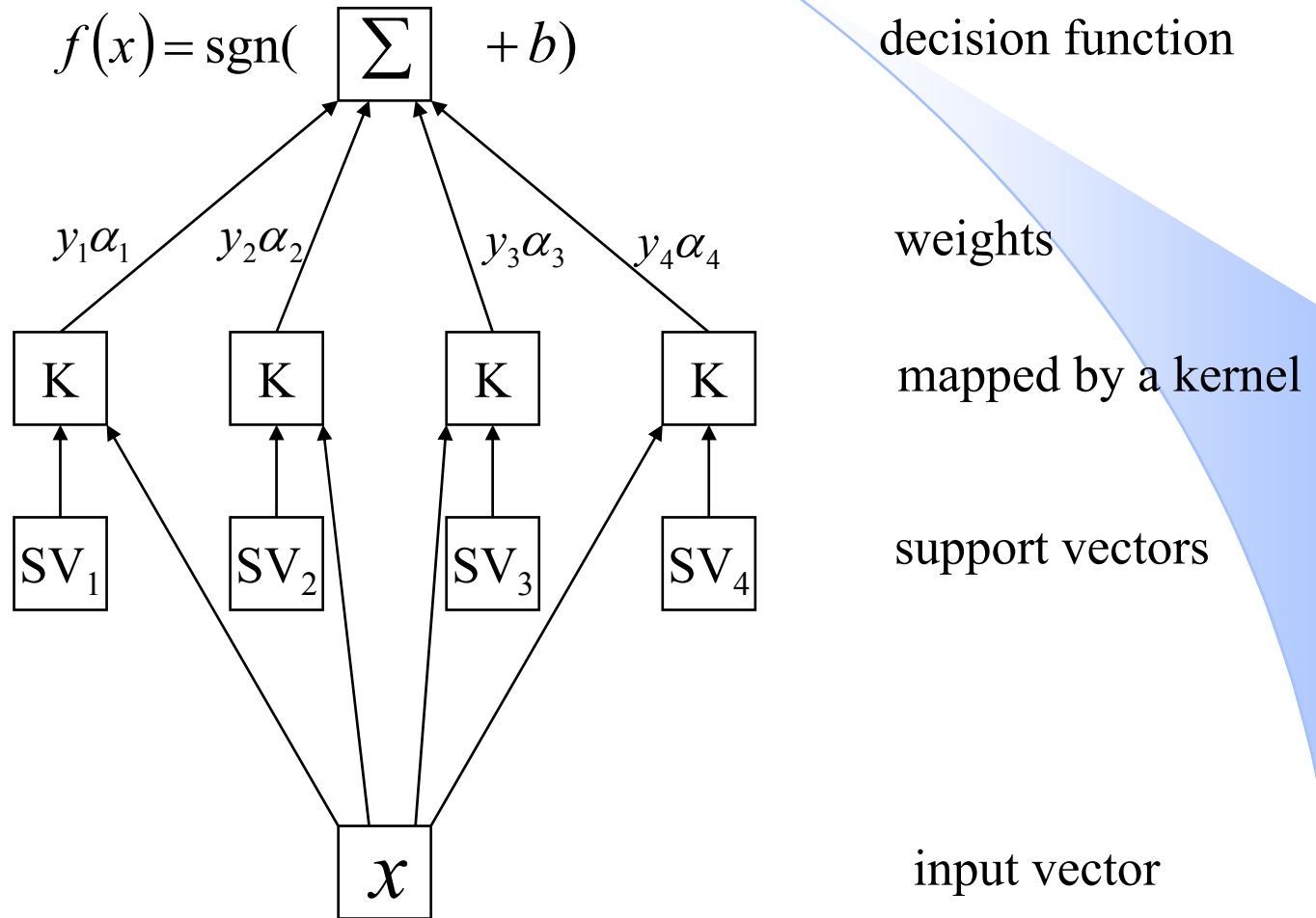
2004-02-11

Presentation at the Academic Symposium of the National Natural Science Foundation of China

Contents

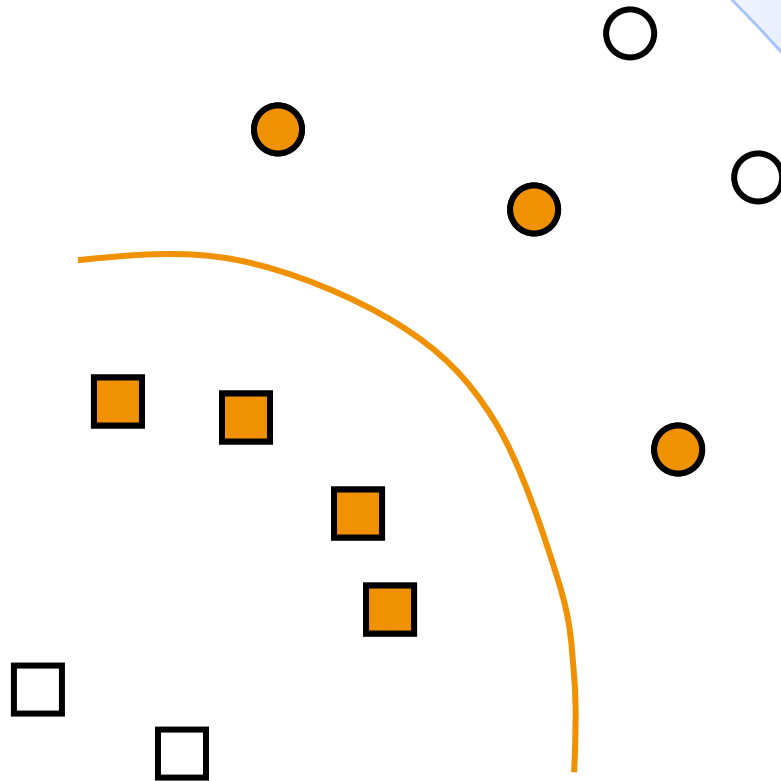
1. The role of support vectors
2. The clustering SVM method
3. Experiments
4. Conclusion

The role of support vectors



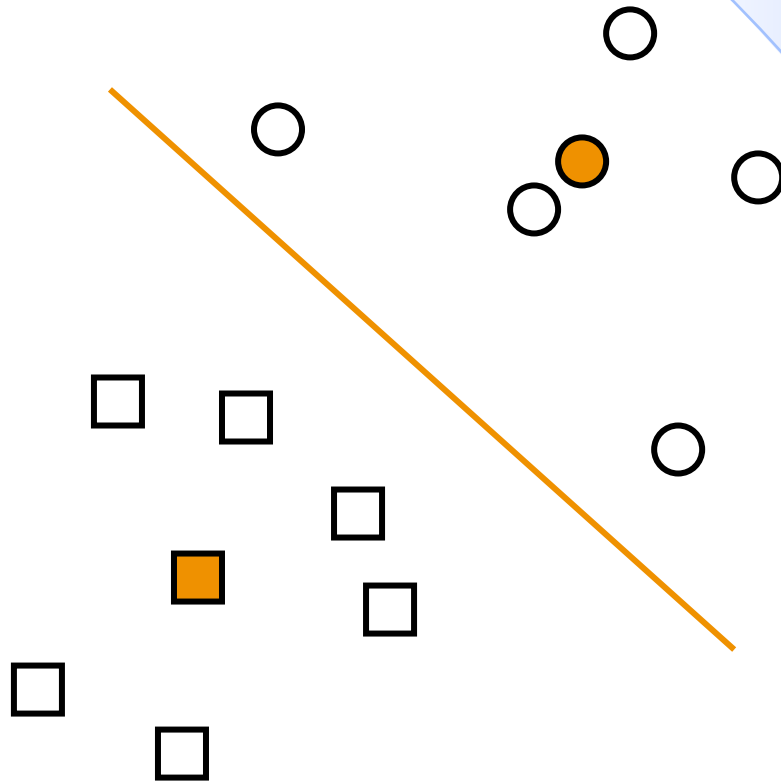
The clustering SVM method

- Support vector machines



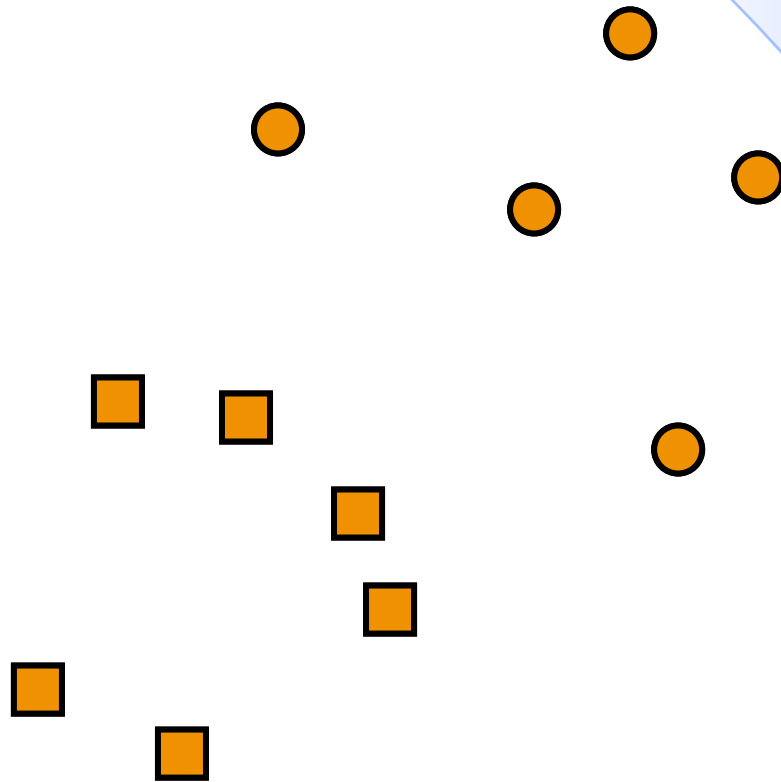
The clustering SVM

- Rocchio



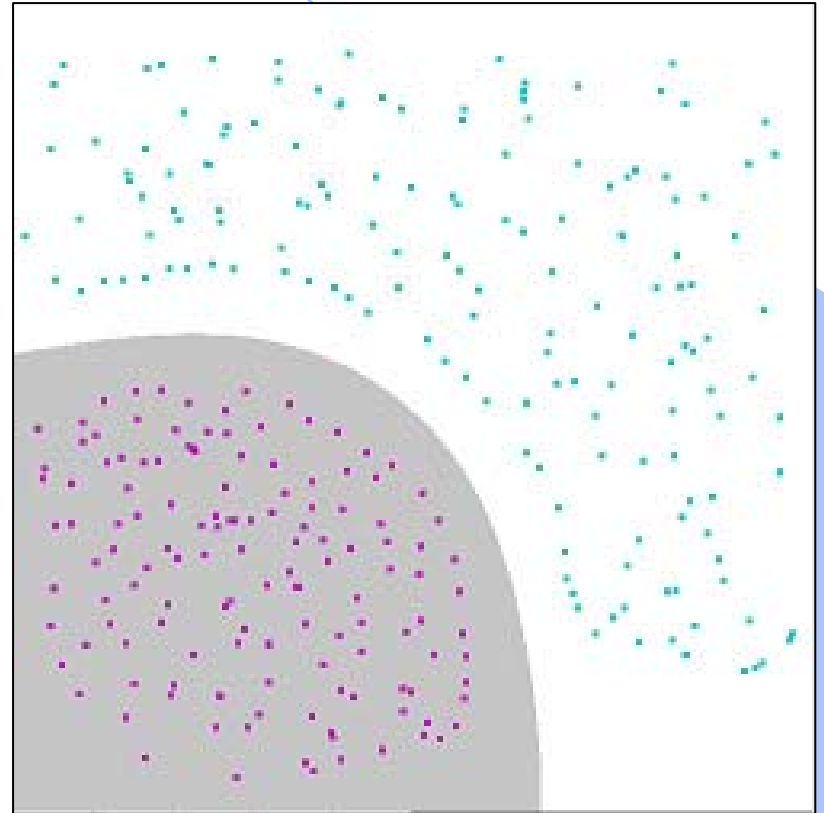
The clustering SVM

- k-nearest neighbor



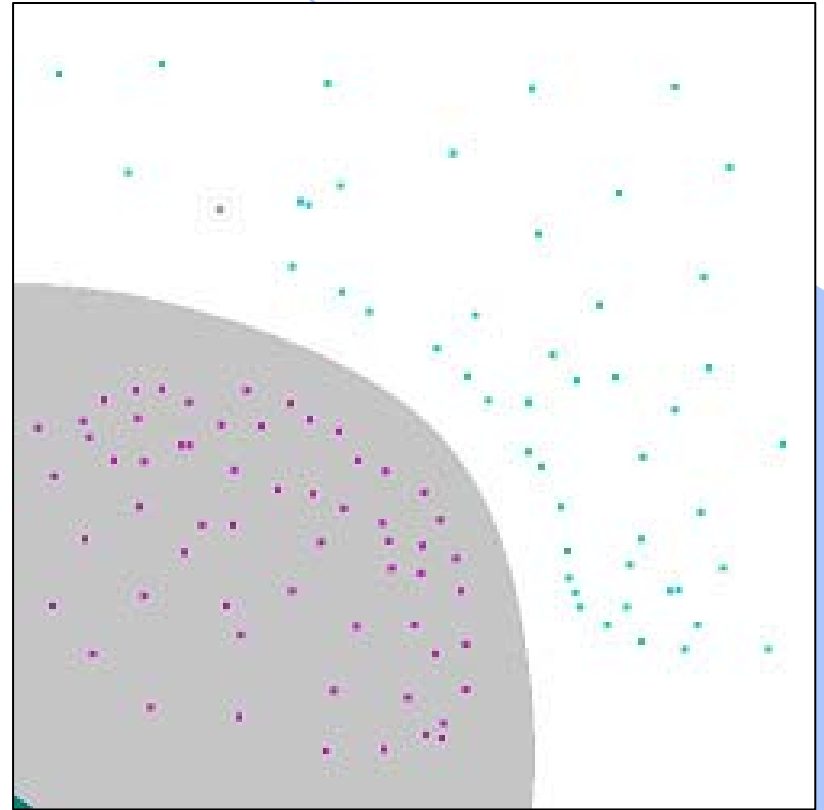
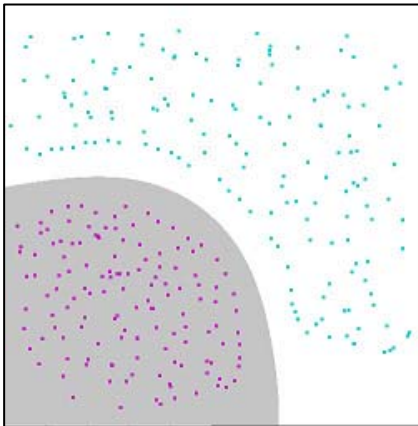
Experiment

- # of data = 292
- # of SVs = 32



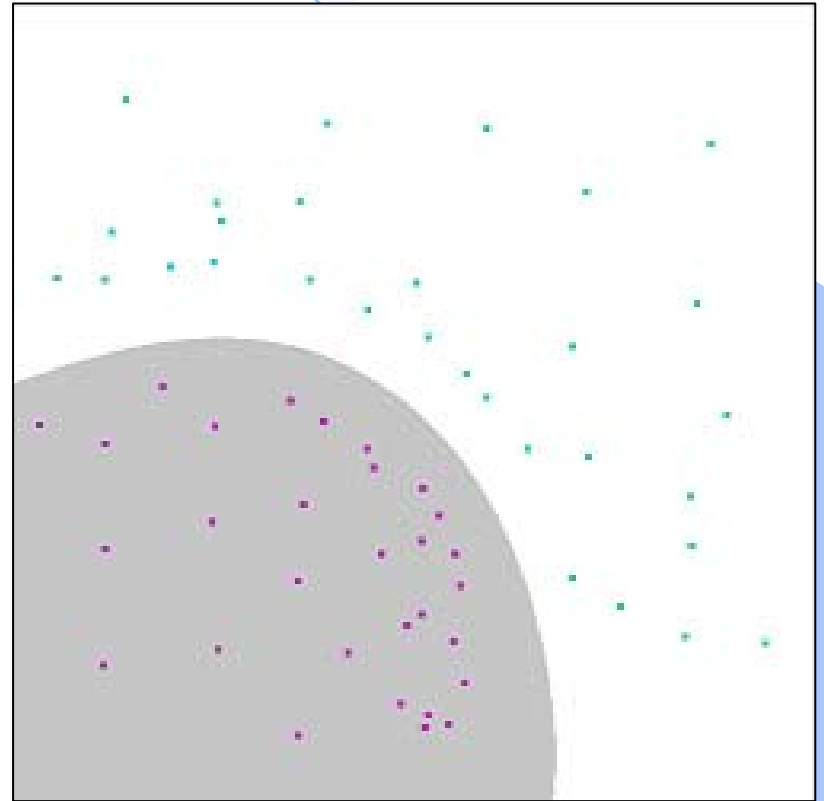
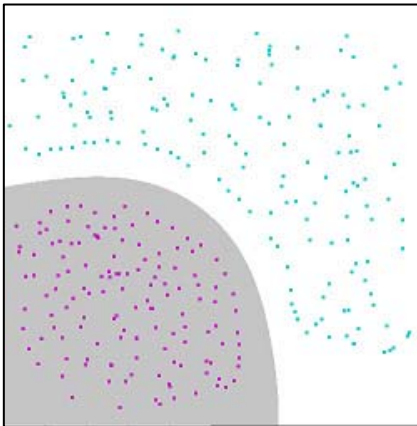
Experiment

- $k = 60$
- # of data = 120
- # of SVs = 25



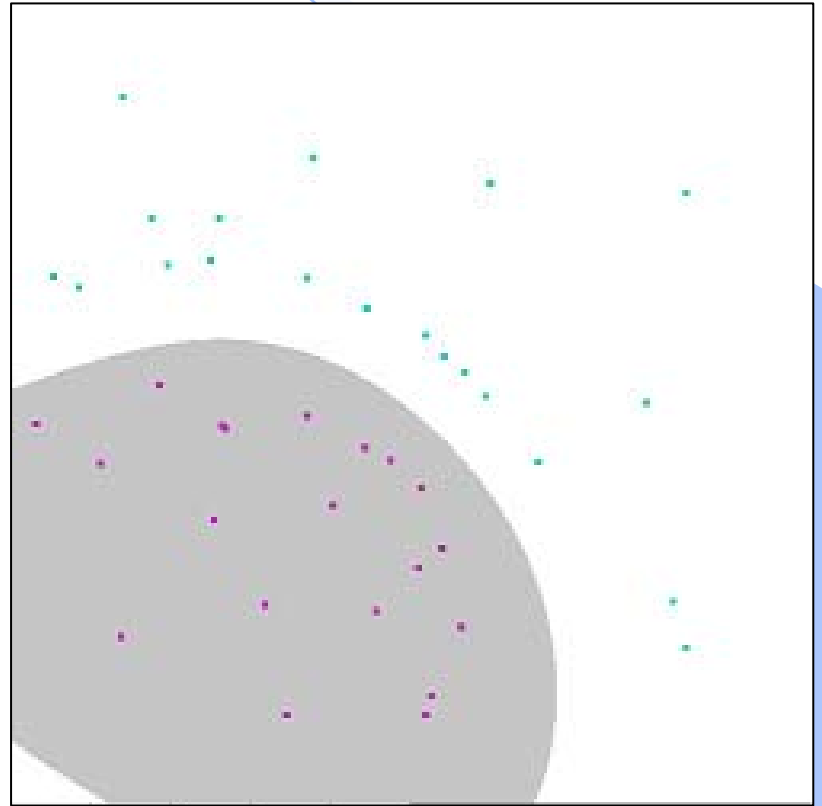
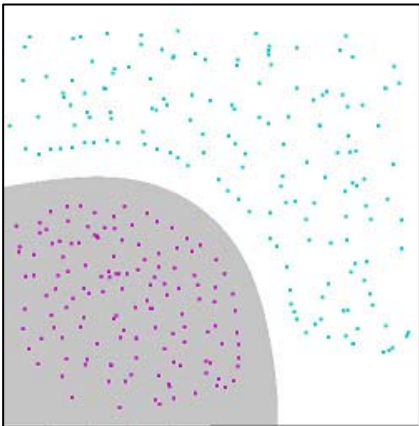
Experiment

- $k = 30$
- # of data = 60
- # of SVs = 23



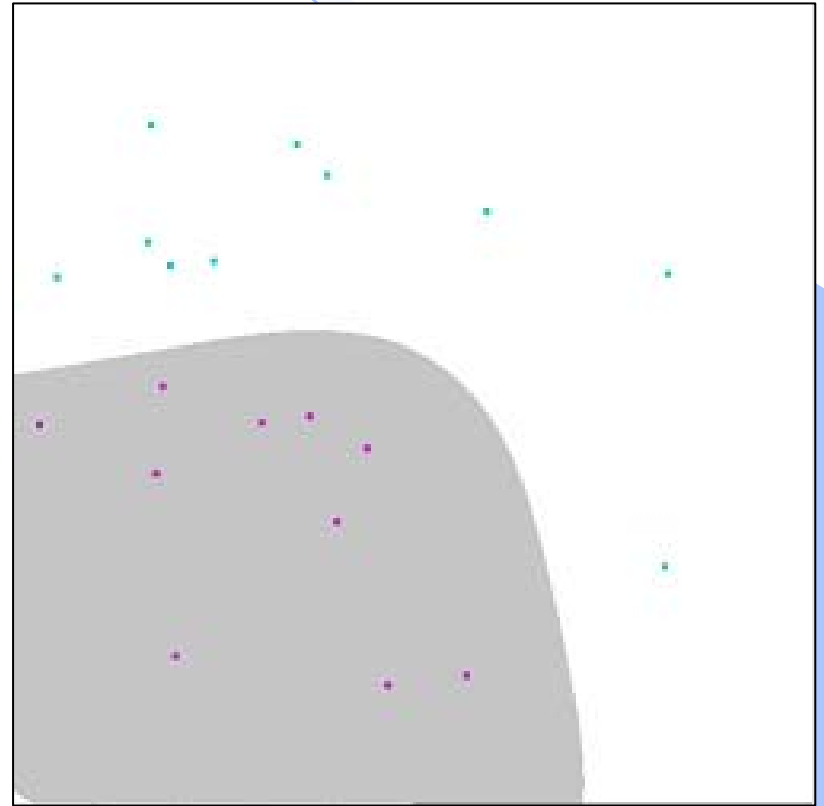
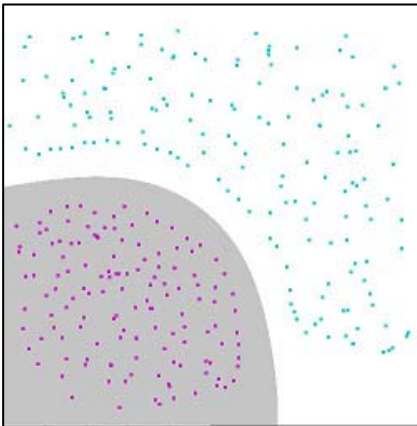
Experiment

- $k = 20$
- # of data = 40
- # of SVs = 20



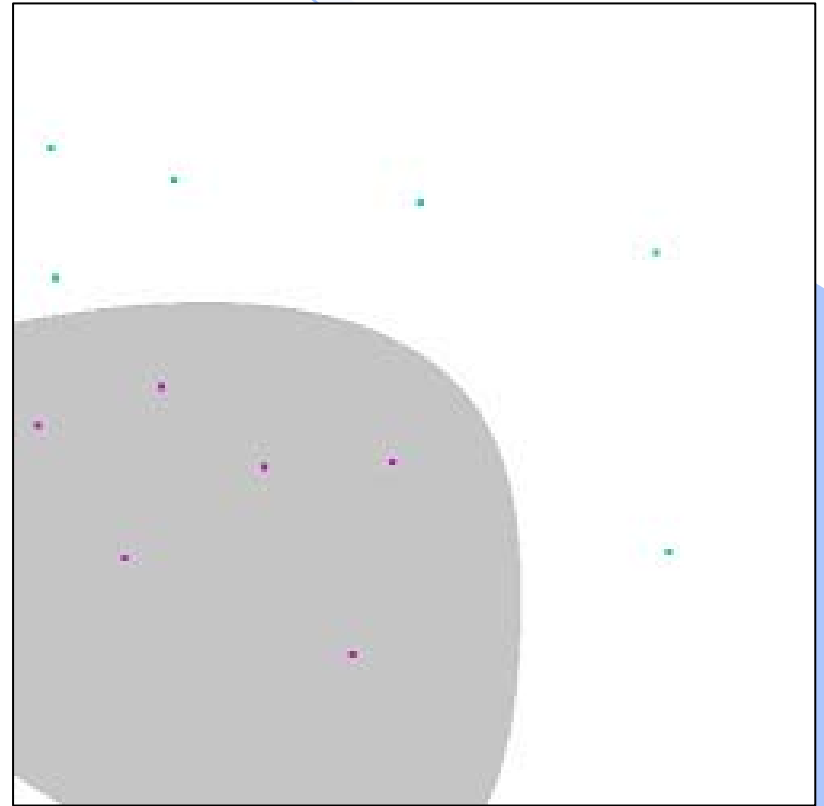
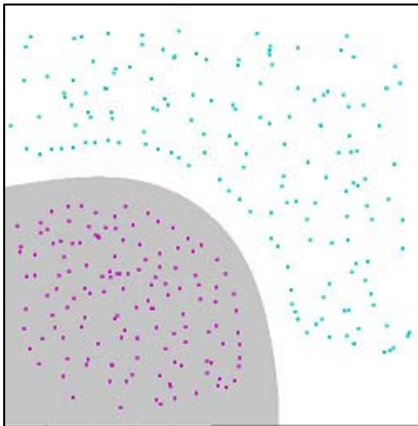
Experiment

- $k = 10$
- # of data = 20
- # of SVs = 14



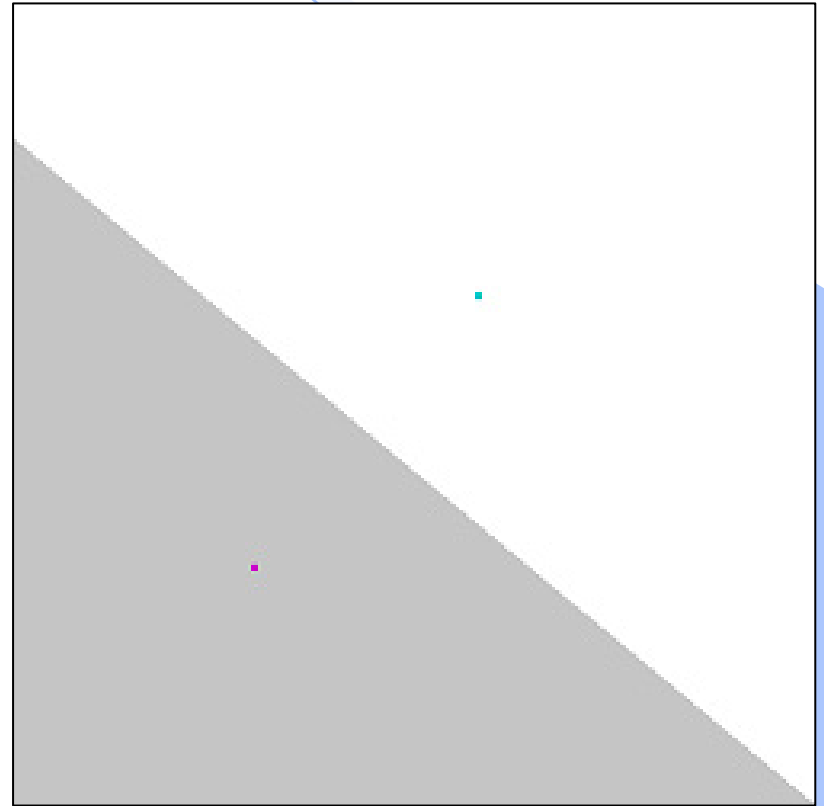
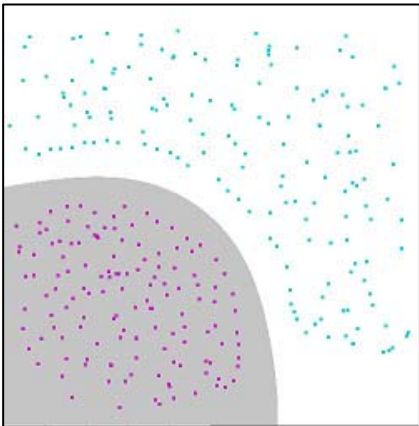
Experiment

- $k = 10$
- # of data = 20
- # of SVs = 14



Experiment

- $k = 1$
- # of data = 2
- # of SVs = 2



5. Conclusion

- Control tradeoff between # of SVs and performance
- $2 * k$ is upper bound of # of SVs

Thank you !

Email ⁽¹⁾: chenguangying@tsinghua.org.cn
Network Research Center
Tsinghua University, China

2004-02-11