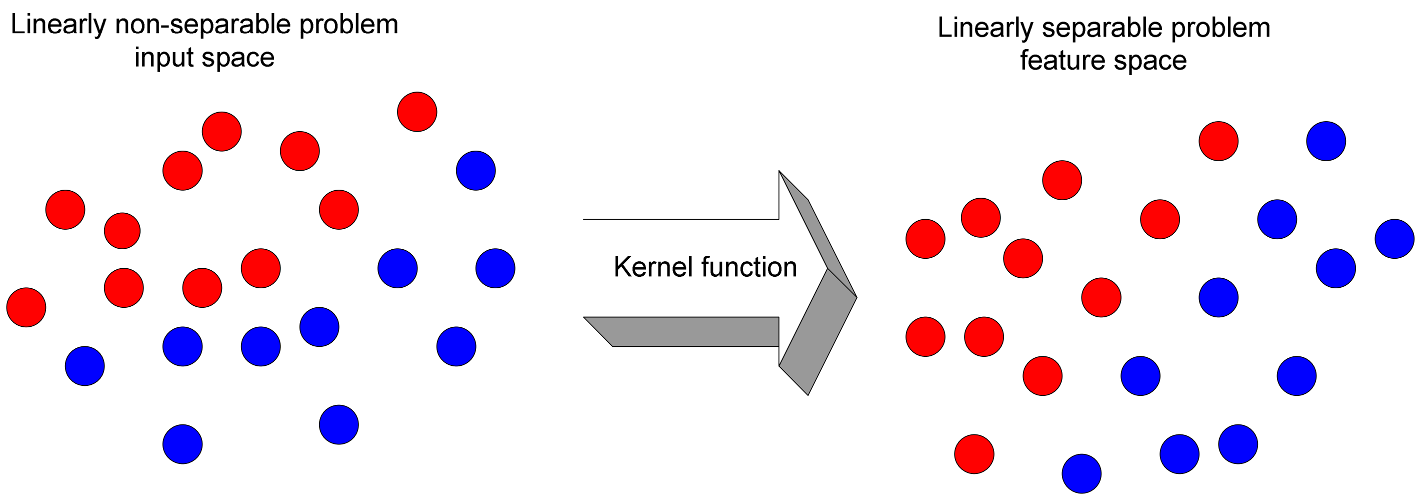
Machine Learning: Basic Principles

Sparse Kernel Machines

20070226 | Janne Argillander, jargilla@cc.hut.fi

Contents

Contents

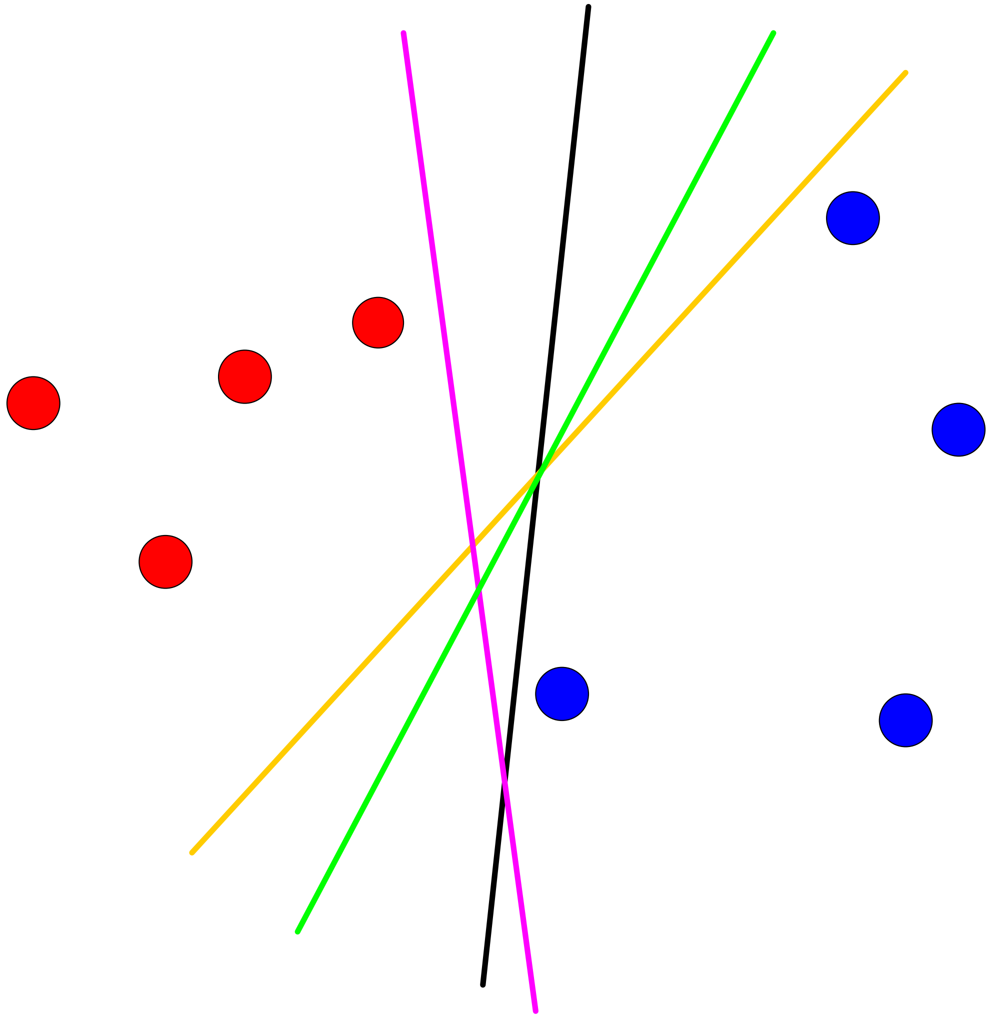


- Argorithms based on the non-linear kernels are nice, but because the kernel function $k(\mathbf{x}_n,\mathbf{x}_m)$ has to be evaluated for possible pairs of the training points the computational complexity is high.

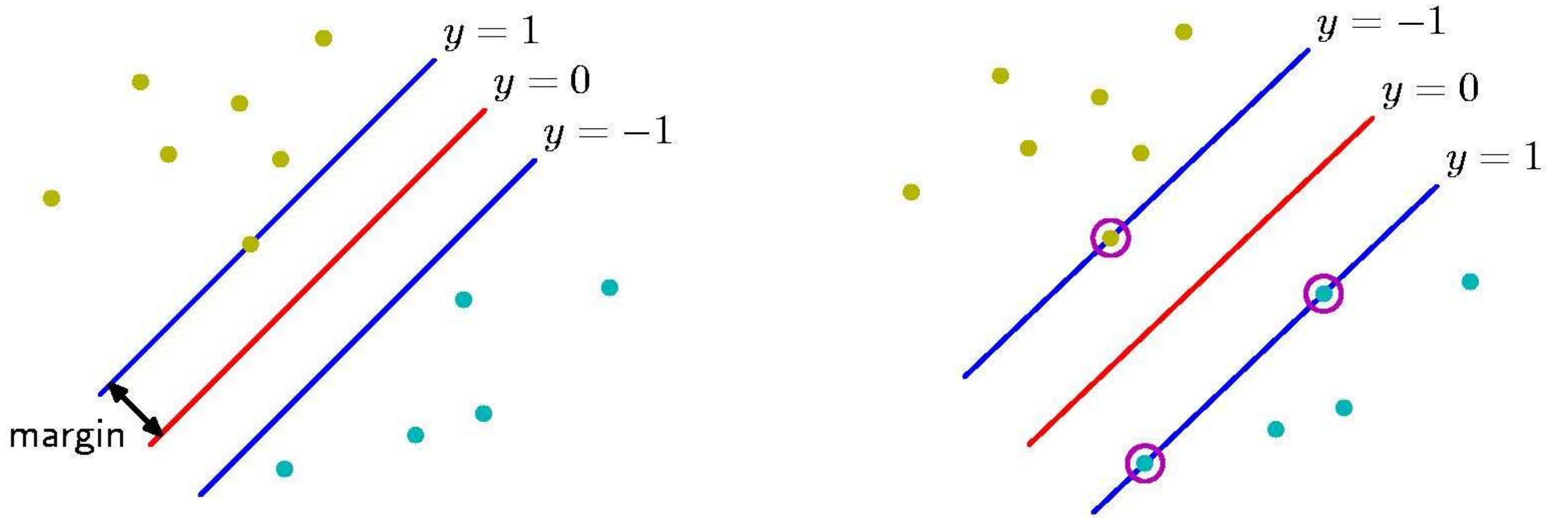
  → Some kind of ”reduction” has to be done

- We say solution to be *sparse* when only the subset of the training data points are used for evaluation.

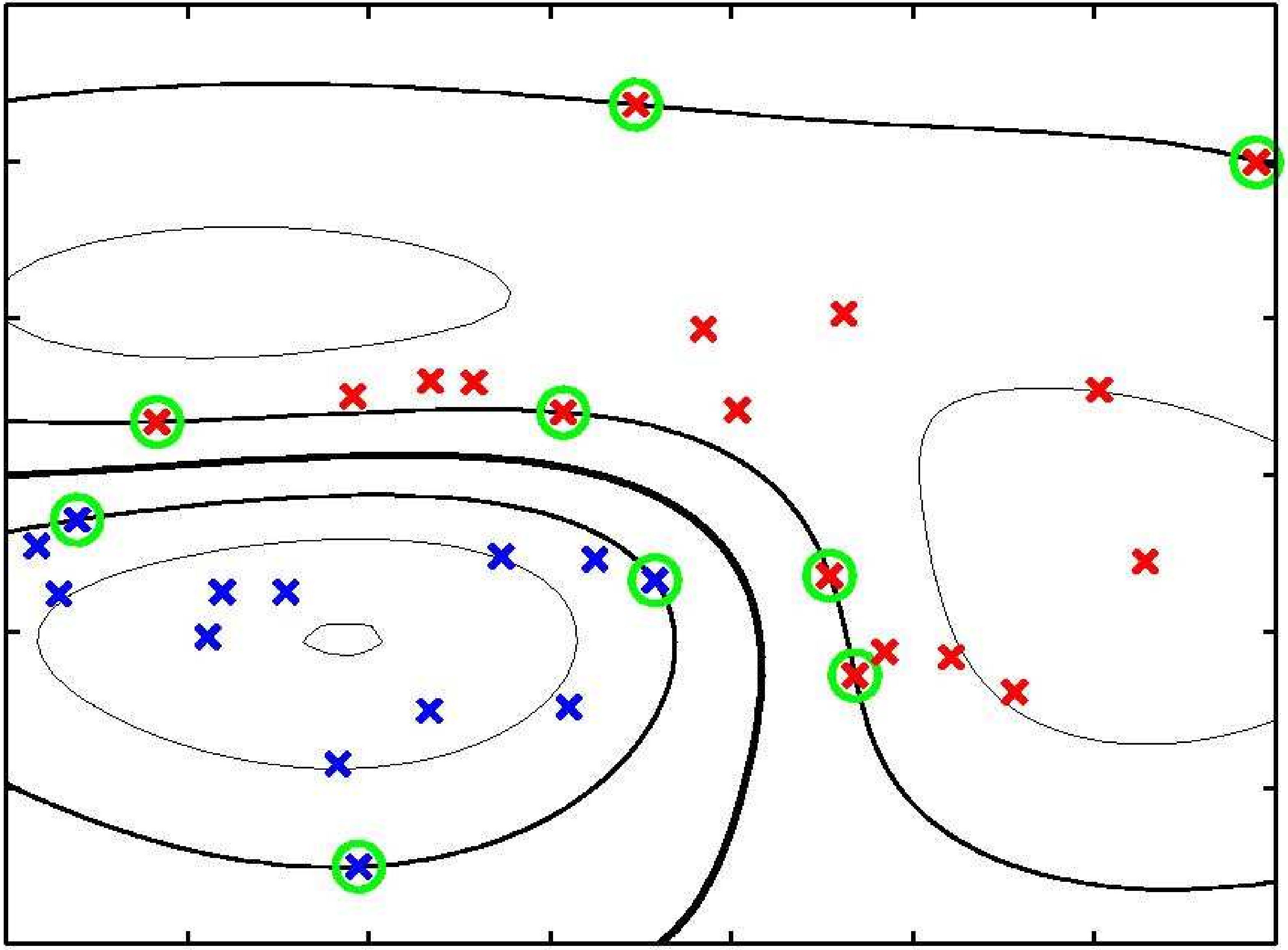
Which boundary should we pick?



Contents



- Support Vector Machine (SVM) approaches the problem stated in the previous slide through the concept of the *margin*
- The margin is the smallest distance between the decision boundary and any of the training samples
- The points with the smallest margin define the decision boundary and are called as *support vectors*

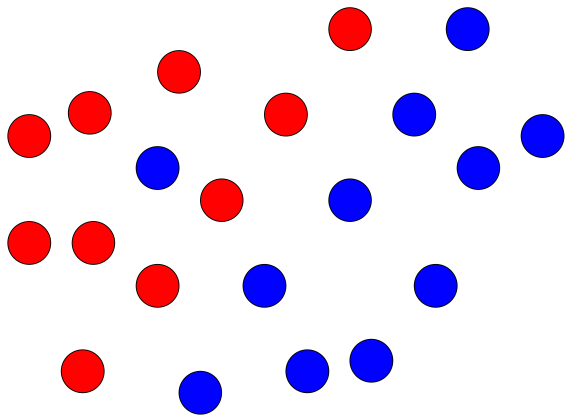


Properties of SVM

- SVM does **not** provide posterior probabilities. The output value can be seen as a distance from the decision boundary to the test point (measured in the feature space)
  - Considering the retrieval task (most probable first), how do we sort the results?
  - Can we combine two different methods or models?
- Can be used for classifying, regression and to data reduction
  - In the classifying task the samples are usually annotated as a positive or negative examples. SVM accepts also examples without annotations.
- SVM is fundamentally a two-class classifier
  - What if we have more classes?

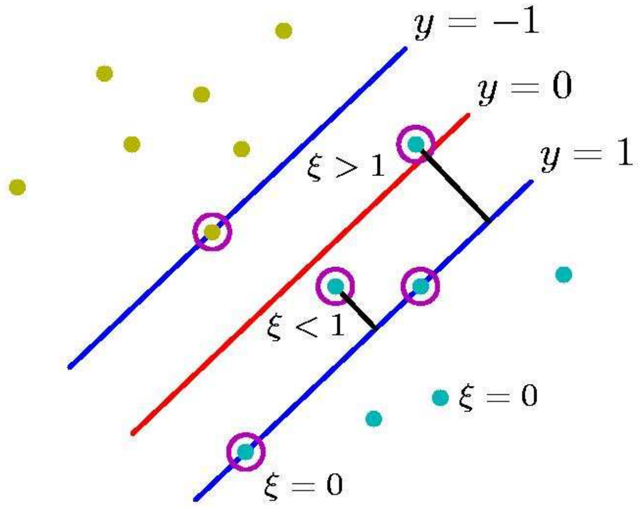
Properties of SVM

- The naive solution is slow $O(N^3)$, but with the help of lagrange multipliers task is only from $O(N)$ to $O(N^2)$
- Handles non-línearly separable data set in the linearly separable feature space.
  - This space is defined by the nonlinear kernel function K.
  - Dimensionality of the feature space can be high
- Traditional SVM does not like outliers



Contents

Overlapping class distributions



- Penalty that increases with the distance (*slack variables*)
- *v*-SVM
  - ☺ Method that can handle data sets with few outliers
  - ☹ Parameters have to be estimated
    - Cross-validation over the hold-out set

Contents

20070226 | Janne Argillander, jargilla@cc.hut.fi

SVM is fundamendally a two-class classifier

Let us consider problem with K classes

- One-versus-rest approach (1-BGM)
  - Training: C_k marked as positive, rest K-1 classes as negative

    → Training sets will be inbalanced (if we had symmetry now it is lost)

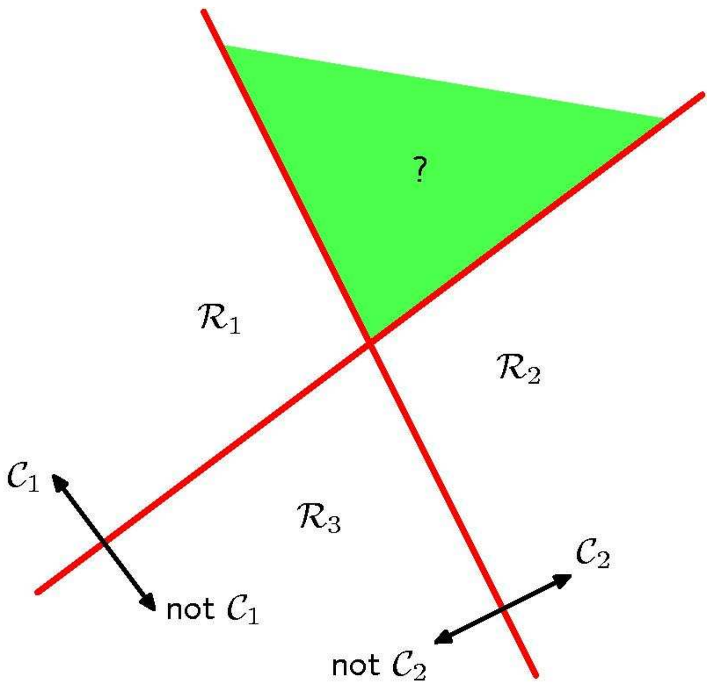
    (10 'faces', 100000 non faces)

    → What if we had similar classes?

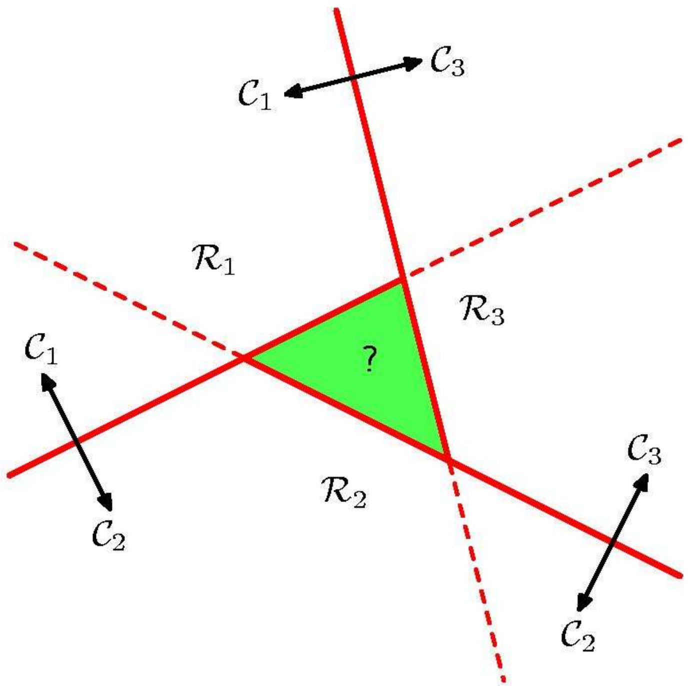
    → Can lead to very very slow models, $O(KN^2)$ – $O(K^2N^2)$

- One-versus-one approach
  - Training: 2-Class SVMs on all possible pair of classes

Multiclass = problems?



One-versus-rest



One-versus-one

HINT: Generally multiclass classification is an open issue. → Solve it and become famous!

20070226 | Janne Argillander, jargilla@cc.hut.fi

Contents

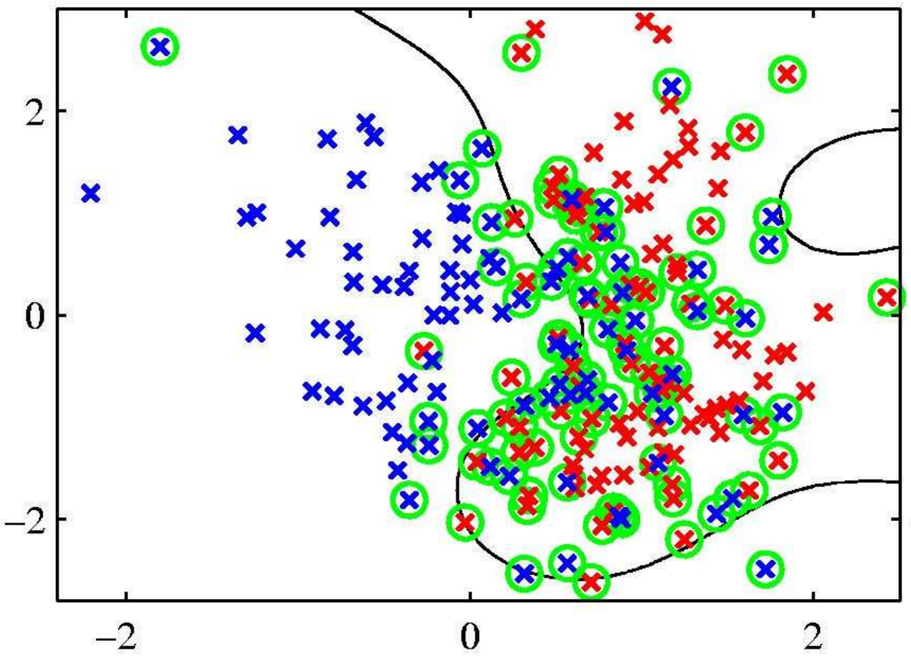
- Introduction
- Maximum margin classifier
- Overlapping class distributions
- Multiclass classifier
- Relevance Vector Machines
- Outroduction

On the way to even more sparse solutions

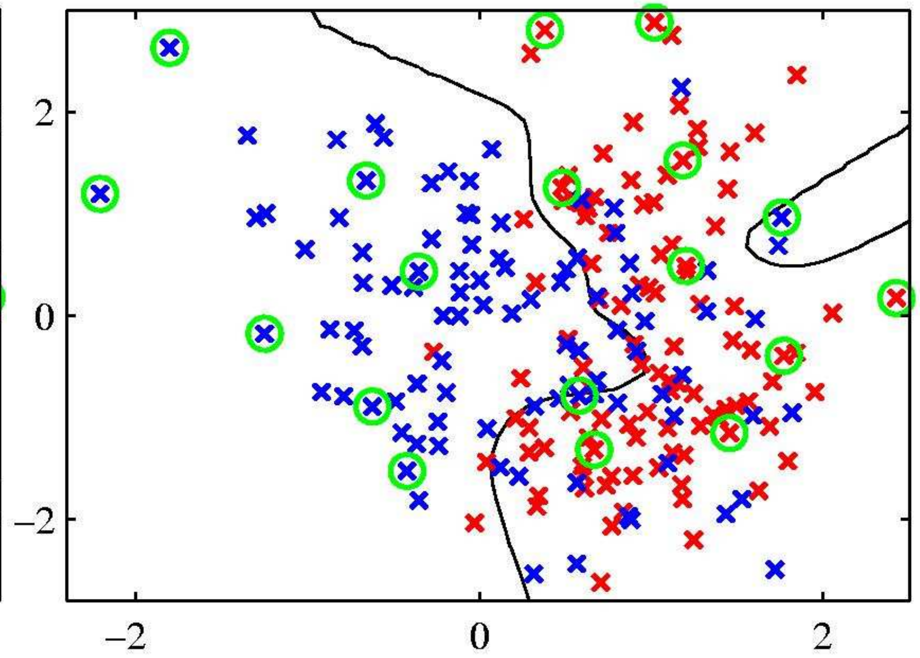
- Some limitations of the SVM
  - Outputs of an SVM represent decisions
  - Slack variables are found by using a hold-out method
  - Positive definite kernels
  - SVM is just not cool anymore?

- Relevance Vector Machine (RVM) has derived from the SVM
  - Shares many (good) characteristics of the SVM
  - No restriction to positive definite kernels
  - No cross-validation needed
  - Number of relevance vectors << Number of support vectors → faster, sparser
  - When training the RVM model we have to calculate inverse of M x M matrix (where M is amount of basis functions) → $O(M^3)$ → RVM training takes more time

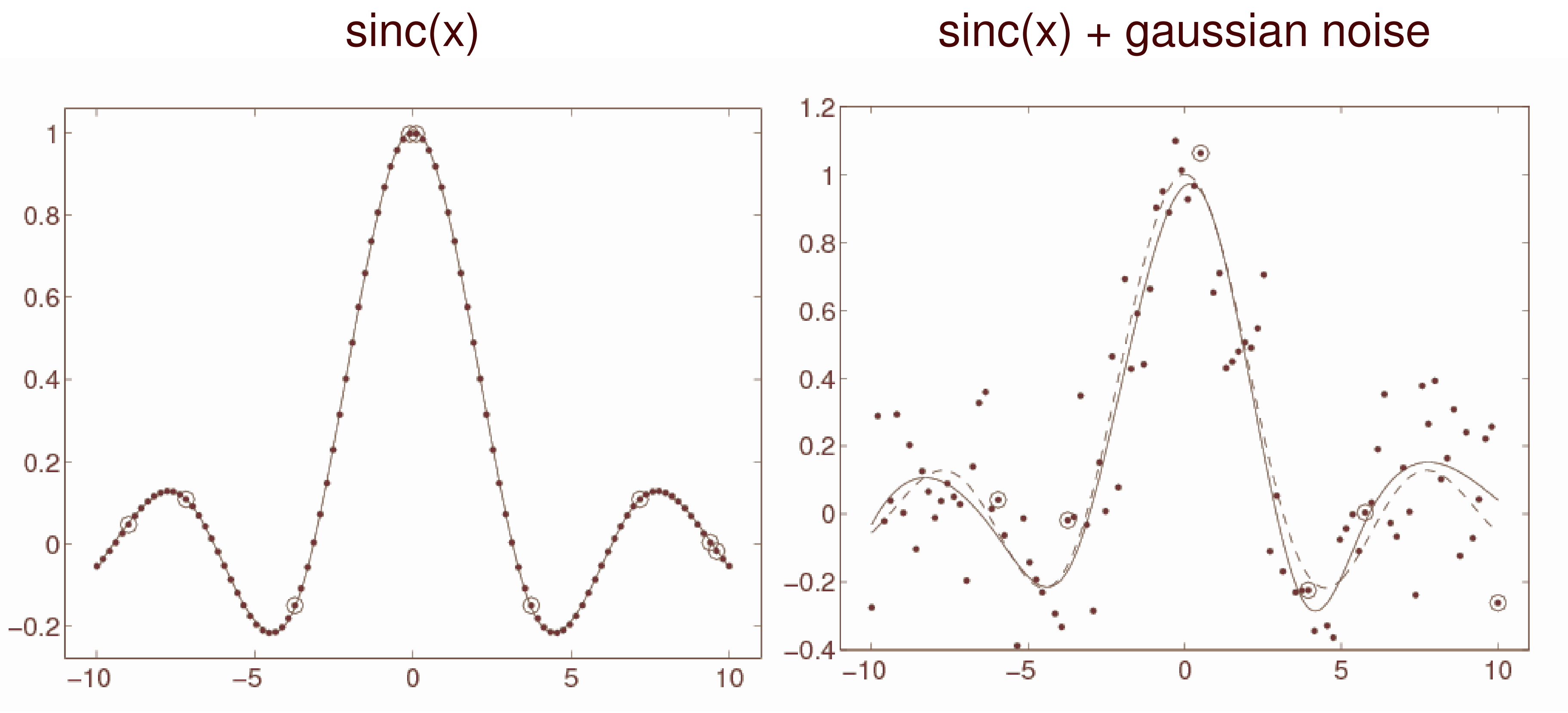
v-SVM



RVM



Function approximation (RVM)



Tipping, M. E. (2000). The Relevance Vector Machine

Contents

Software packages and good to check

- Command line tool

  SVM^{LIGHT} http://svmlight.joachims.org/
  SVMTorch Multi III http://www.torch.ch/
  - one-versus-rest approach

- Do it yourself

  LIBSVM http://www.csie.ntu.edu.tw/~cjlin/libsvm/

- Interfaces to LIBSVM
  - Matlab, R, Weka. Labview,...

- http://www.kernel-machines.org/

- Google