Universidade do Minho
Escola de Engenharia
Departamento de Informática

Machine Learning: Support Vector Machines 2017/2018

Paulo Novais, Tiago Pinto

Synthetic Intelligence Lab

Summary

Pattern Analysis

Metalearning

Supervised Learning

Reinforcement Learning

Machine Learning

Unsupervised Learning

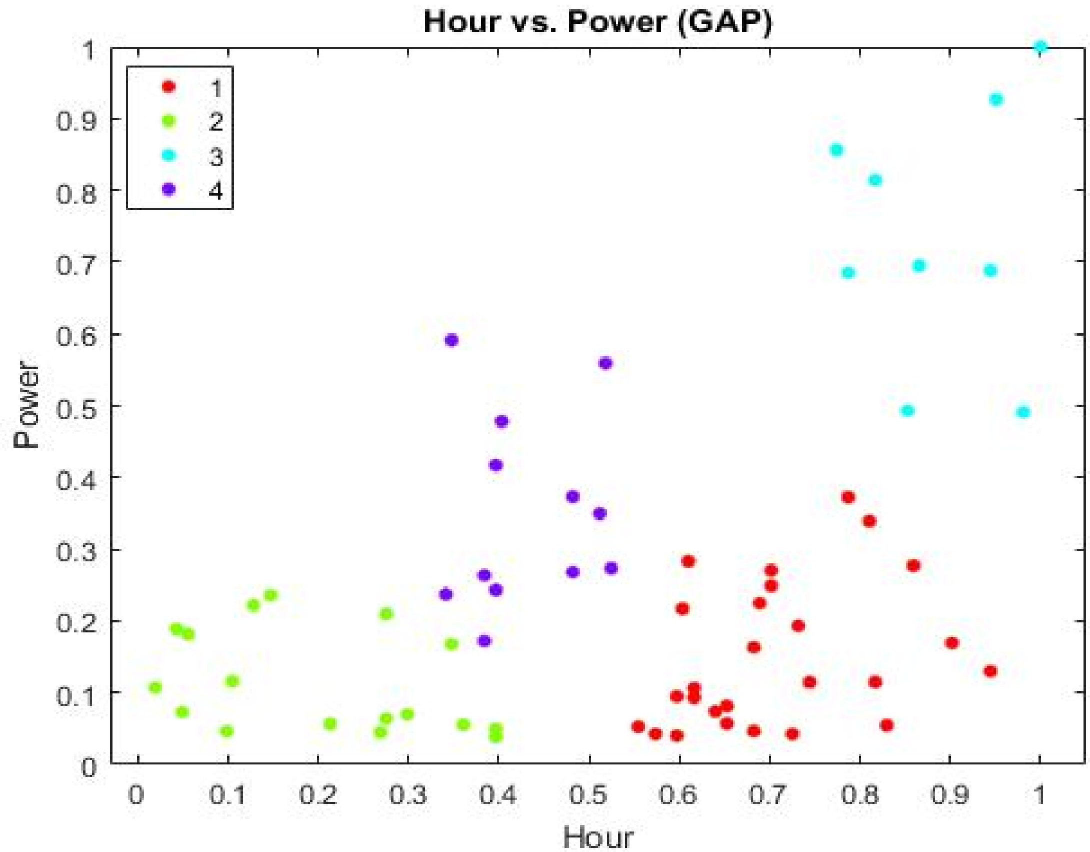
Computational Intelligence

Fuzzy Logic

Data Mining

Cased-Based Reasoning

- Forecasting
- Clustering
- Classification
- Association



ISLab
Synthetic Intelligence Lab

Computational Intelligence

- Artificial Neural Networks
- Fuzzy Logic
- Metaheuristics

Efficiency

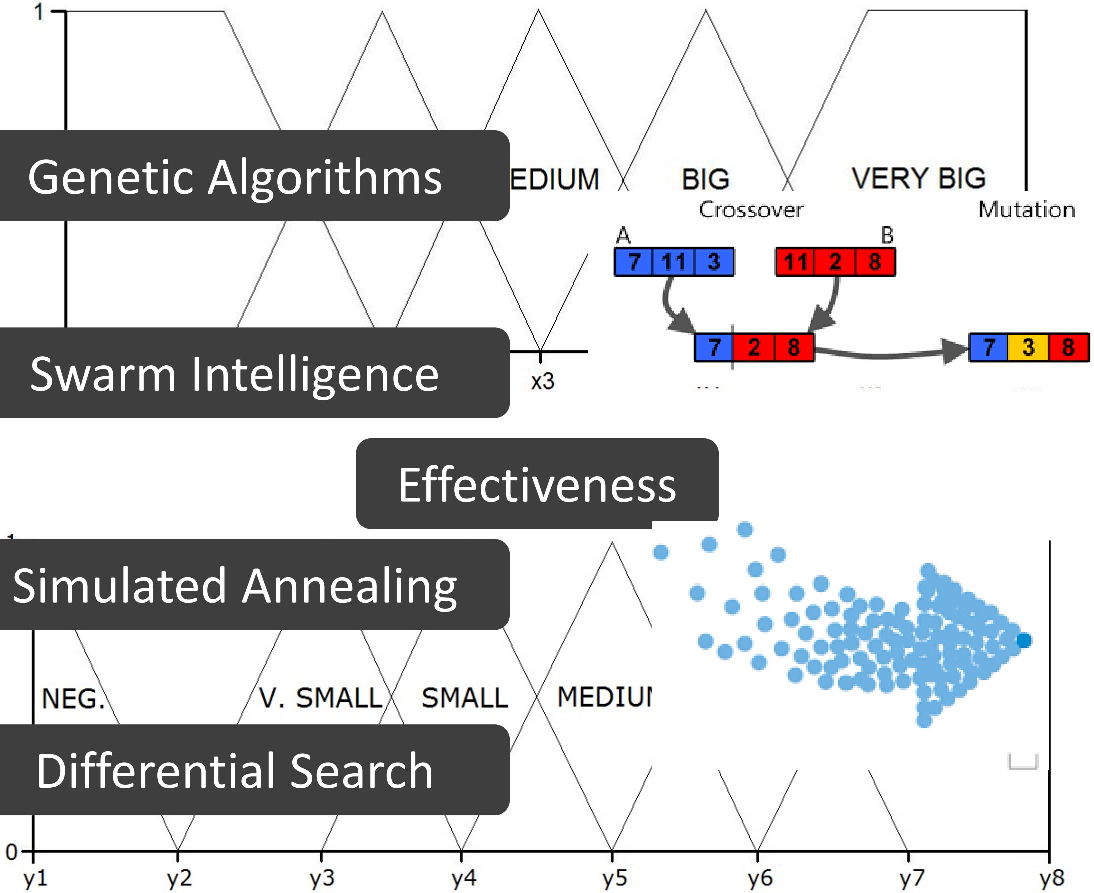
Genetic Algorithms

Swarm Intelligence

Effectiveness

Simulated Annealing

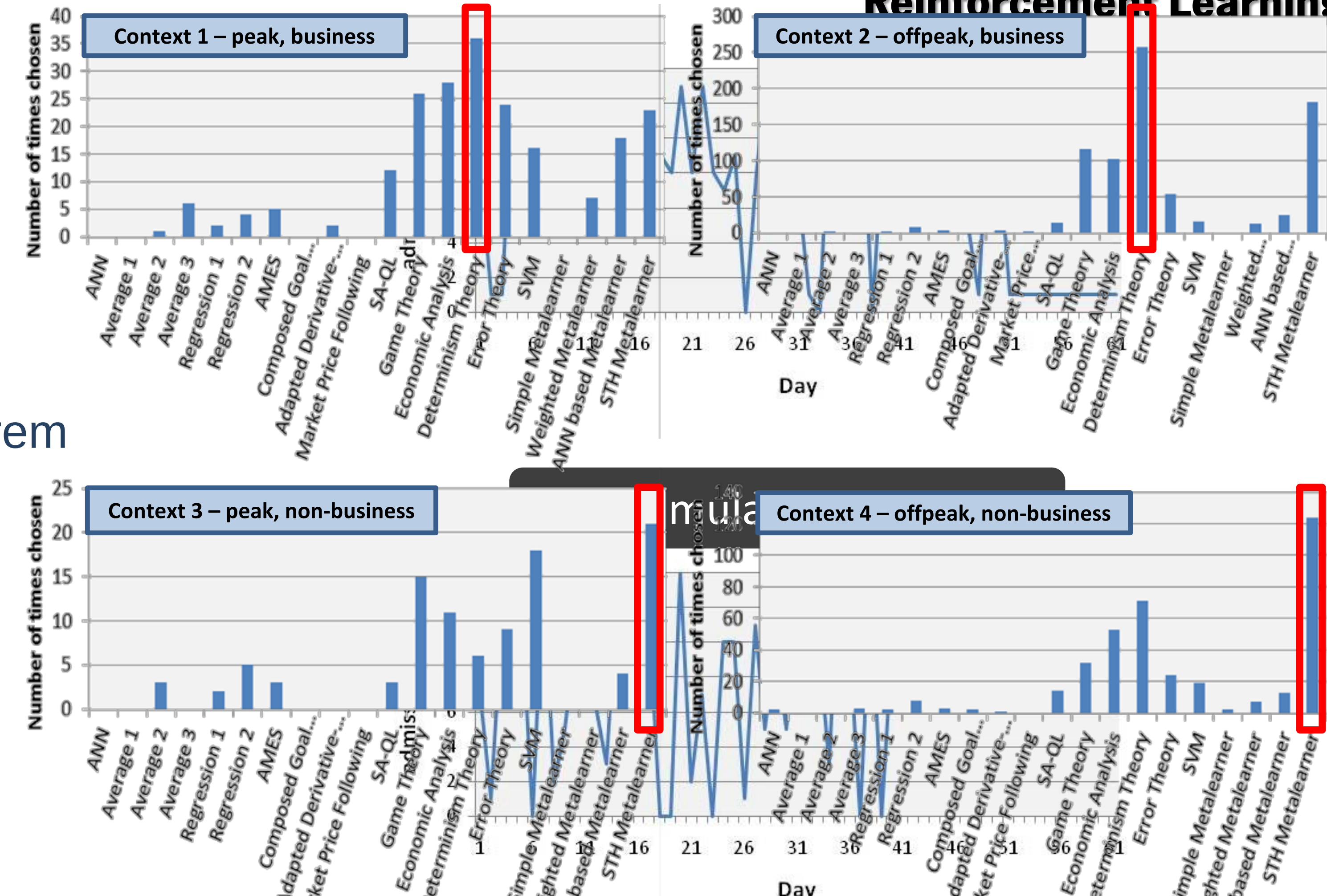
Differential Search



Synthetic Intelligence Lab

Reinforcement Learning

- Q-Learning
- Roth-Erev
- Bayes Theorem



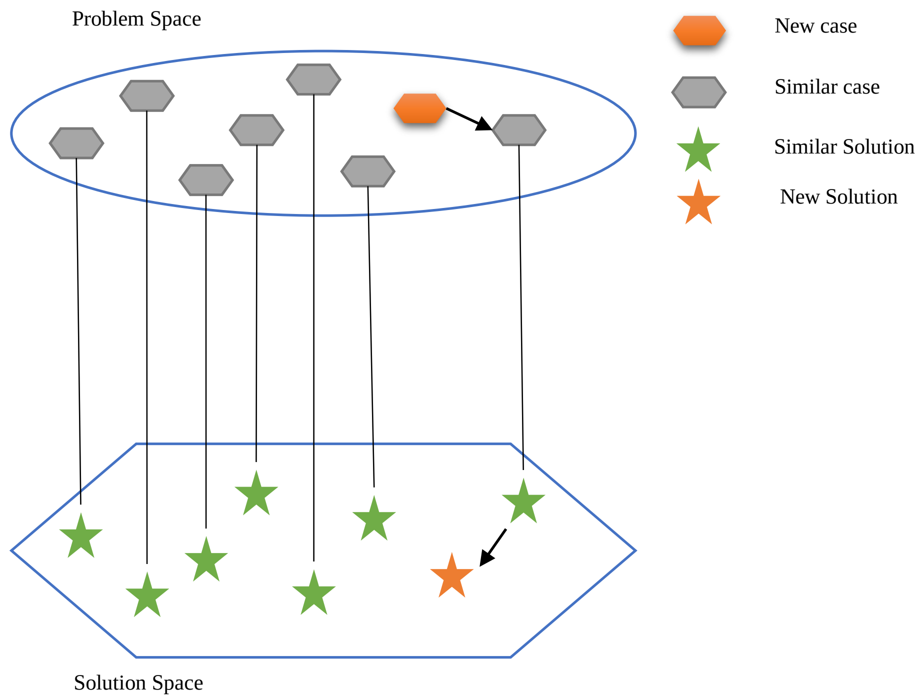
ISLab
Synthetic Intelligence Lab

Case-Based Reasoning

- Retrieve
- Reuse
- Revise
- Retain

Clustering
K-NN
Decision
Trees

Expert
Systems



Synthetic Intelligence Lab

Pattern Analysis

- Most repeated action
- Sequences in the past matching the last few actions. The longer matches to the recent history are attributed an higher importance
- Most repeated sequence along the historic of actions of this player
- Most recent sequence among all the found ones
- History matching

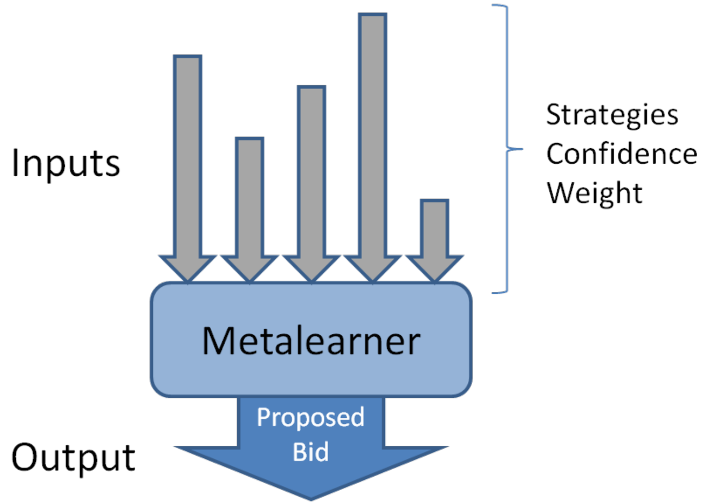
Synthetic Intelligence Lab

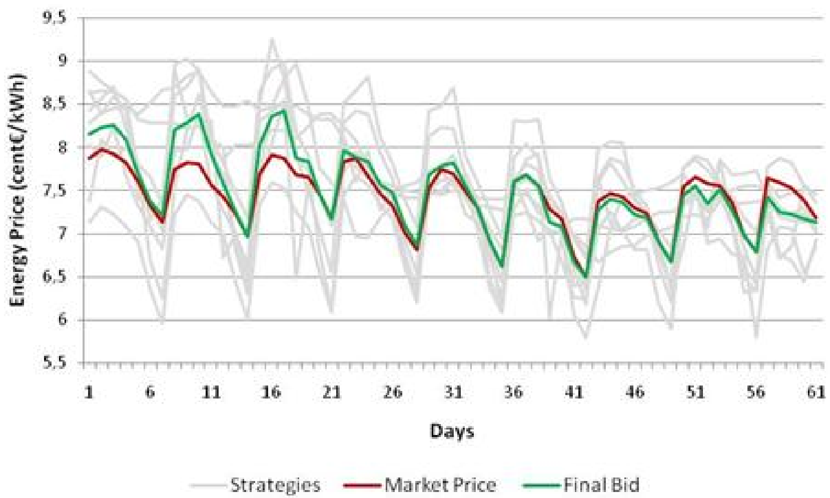
Pattern Analysis

- Second Guessing the predictions
- Third-Guess
- Self Model prediction
- Second-Guess the Self Model prediction

Metalearning

- Simple
- Weighted Metalearner
- Artificial Neural Network
- Reinforcement Learning





Support Vector Machines

- Support Vector Machines (SVM) have showed robustness in many application domains
  - pattern recognition, image recognition, classification and regression analysis, text categorization, medical science, classification of proteins, weather forecast, wind speed prediction, energy prices forecast, among other practical applications

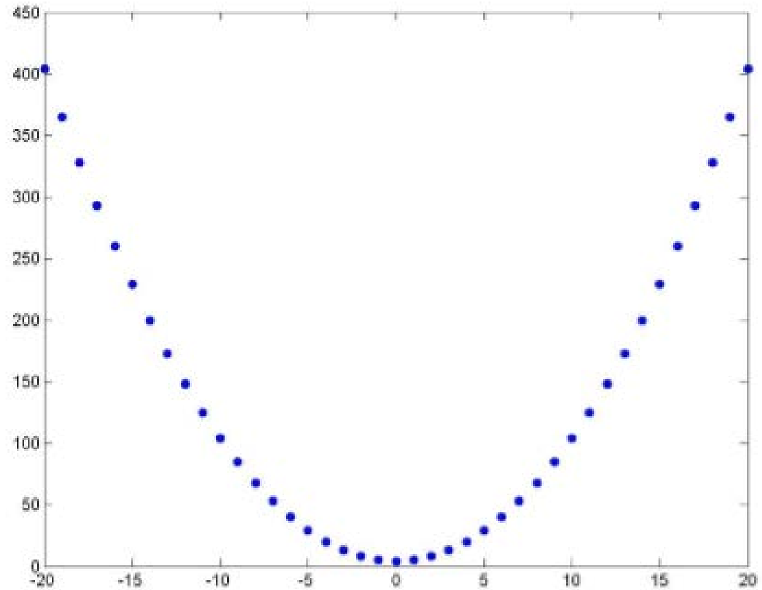
- Often concentrating on convex problems

- Allowing many linear algebra techniques to be used in a non-linear way

ISLab
Synthetic Intelligence Lab

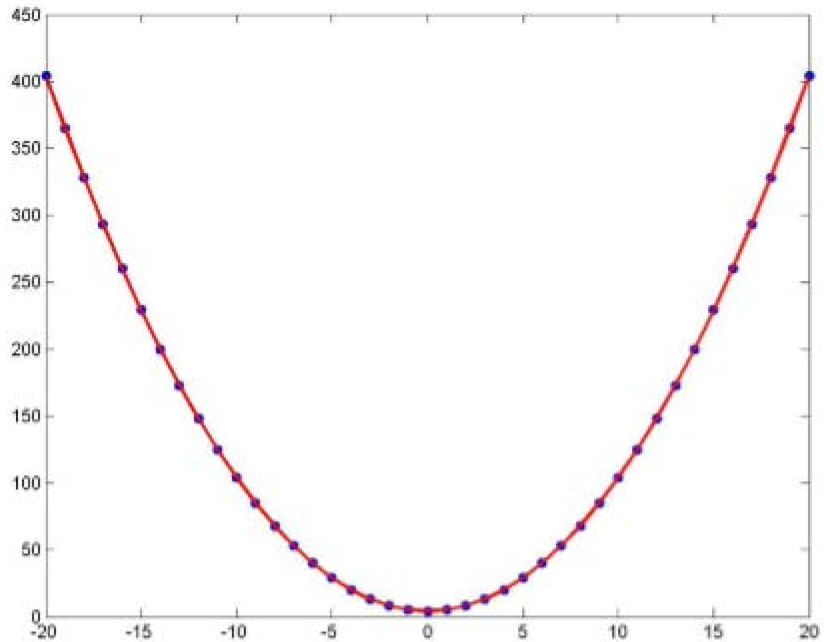
- SVM vs Artificial Neural Networks
  - Advantages
    - Spend fewer resources and half the time of artificial neural networks
    - Less training data
    - Better control of overfitting
  - Disadvantages
    - Lesser capability to deal with co-related data series
    - Weaker for more complex data structures

- Predict y given x



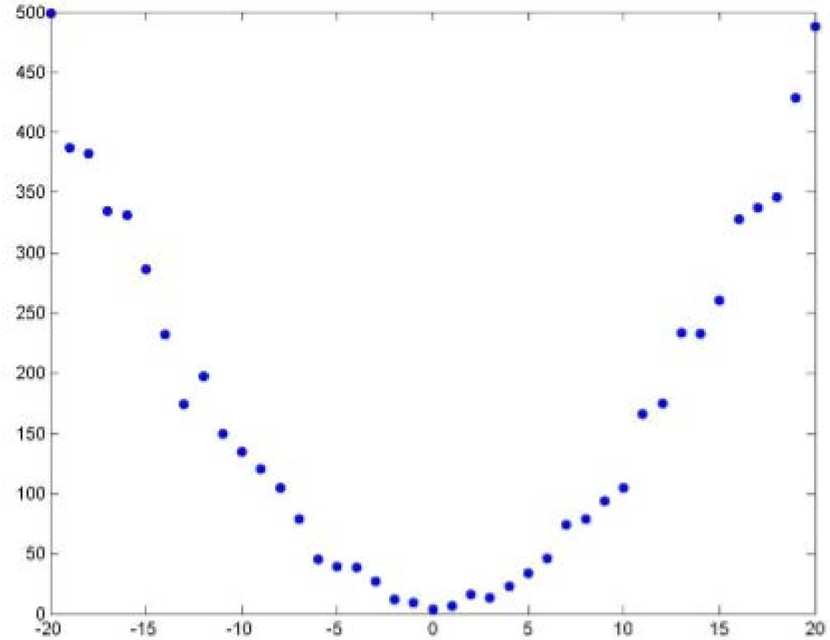
- Try to fit a function to describe this

- Easy!



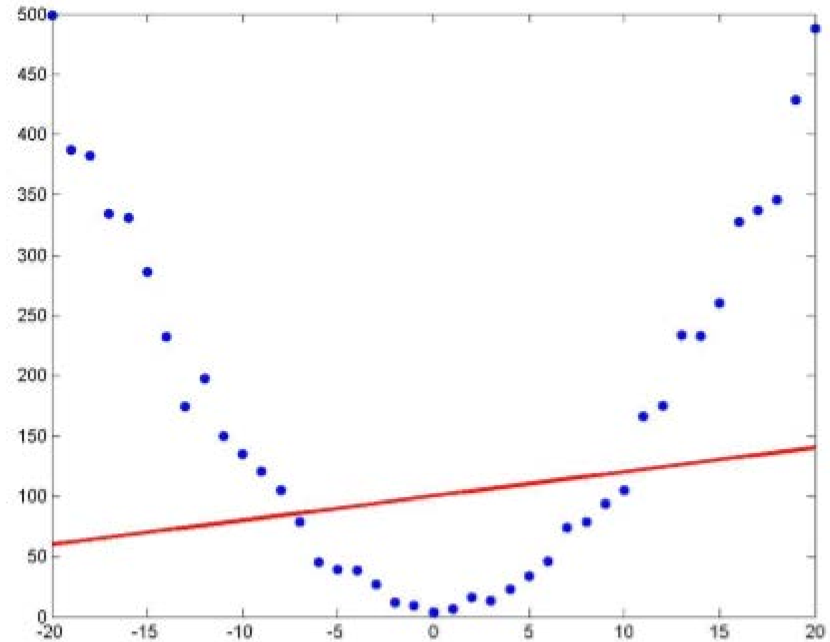
- For a new point we will be correct

- What if we add some noise?



- In real data we cannot “see” the function

- We could assume the relationship to be linear



- How wrong are we?
- How do we know which parameters are the best?

- We can measure the square loss

$$L(\mathbf{y}, f(\mathbf{x})) = \sum_{i=1}^{\ell} (y_i - \hat{y_i})^2$$

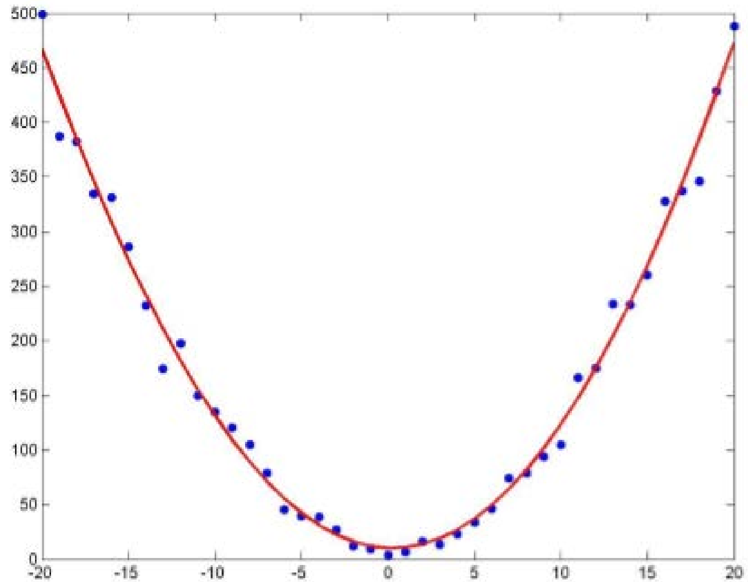
- This suggests an algorithm

$$\min_{\mathbf{w}} \sum_{i=1}^{\ell} (((\mathbf{w} \cdot \mathbf{x}) + c) - y_i)^2$$

- Linear is still bad
- Increase parameters, and consider a quadratic function

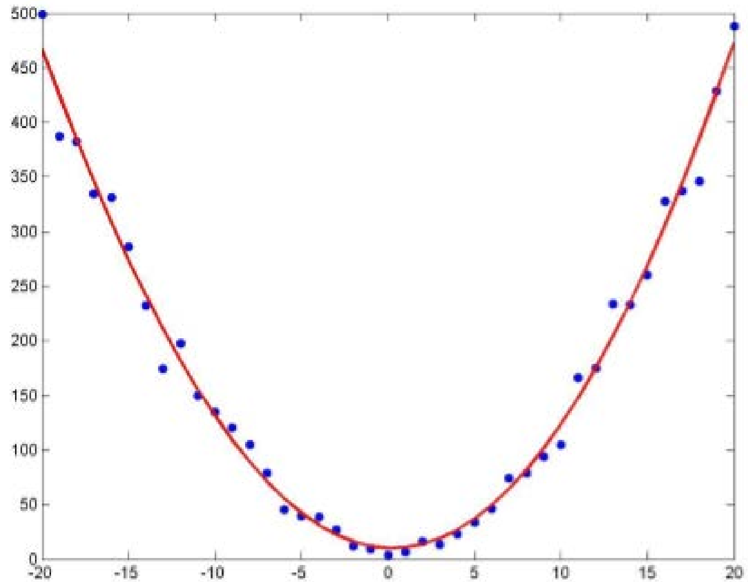
$$f(\mathbf{x}) = w_2\mathbf{x}^2 + w_1\mathbf{x} + c$$

- This is better



- Linear is still bad
- Increase parameters, and consider a quadratic function

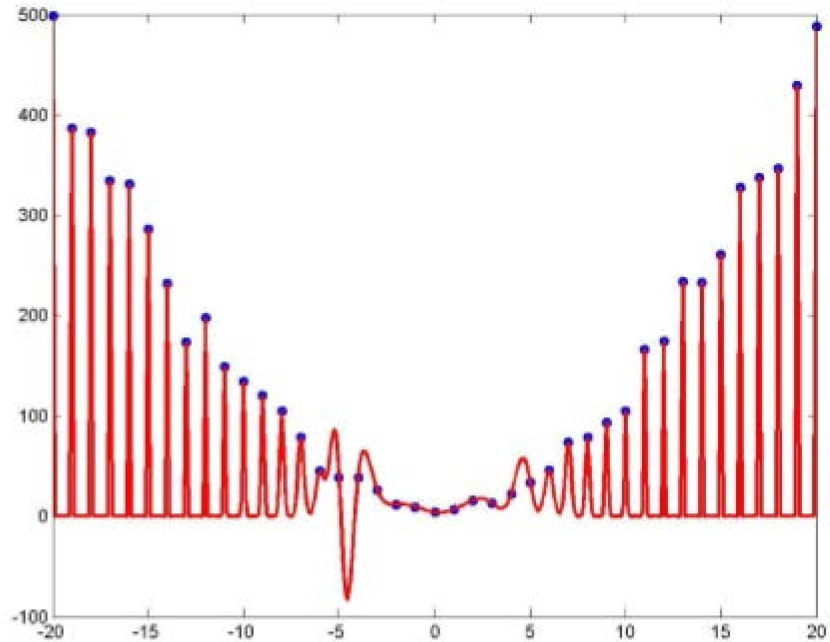
$$f(\mathbf{x}) = w_2\mathbf{x}^2 + w_1\mathbf{x} + c$$



- This is better
- But still has some loss
- Increase parameters!

Support Vector Machines

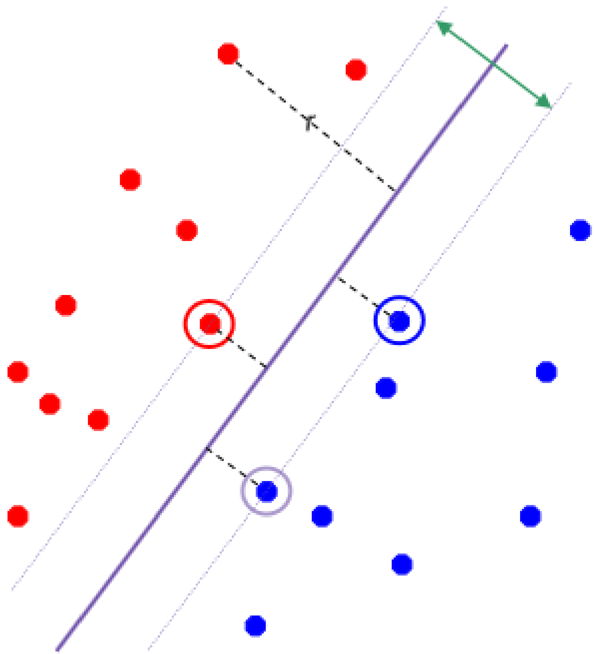
- Zero error on training set!
- What about a test example?
- Too specific = overfitting
- Empirical risk minimization: overfits if you follow it blindly



- How to avoid?
- Need to balance training error and capacity of a function
  - Too complex a function will overfit
  - Not enough complexity will not generalize well either
- General strategy:
  - Minimize some combination of a regularizing term (to control capacity) and the loss on the training set

$$\min_{\mathbf{w}} R(\mathbf{w}) + L(\mathbf{w}, X, \mathbf{y})$$

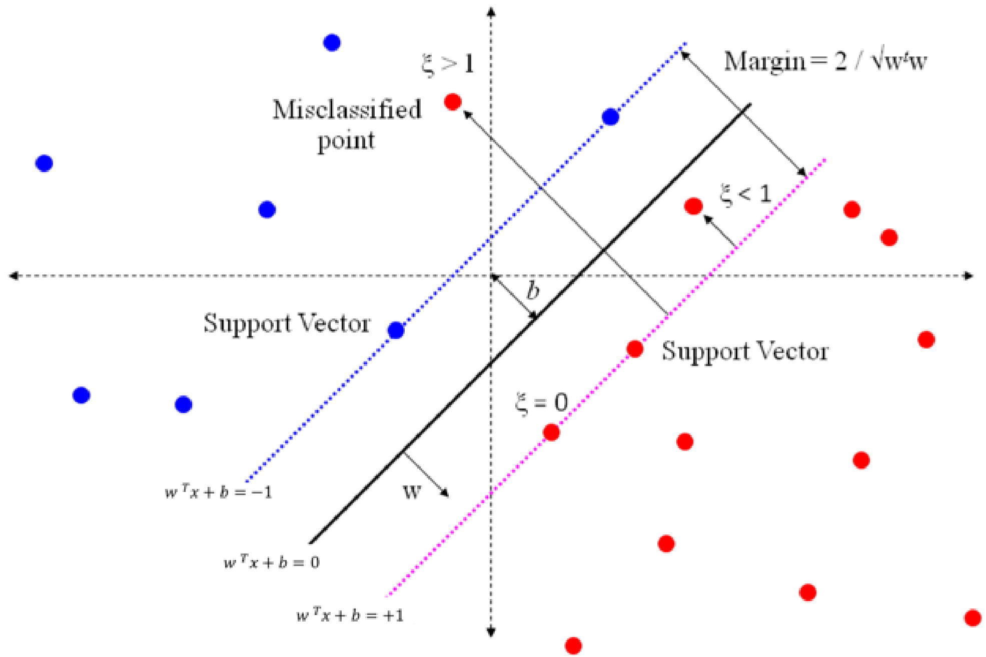
- SVM Linears
- Examples closest to the hyperplane are **support vectors**
- The margin ρ is the distance between support vectors



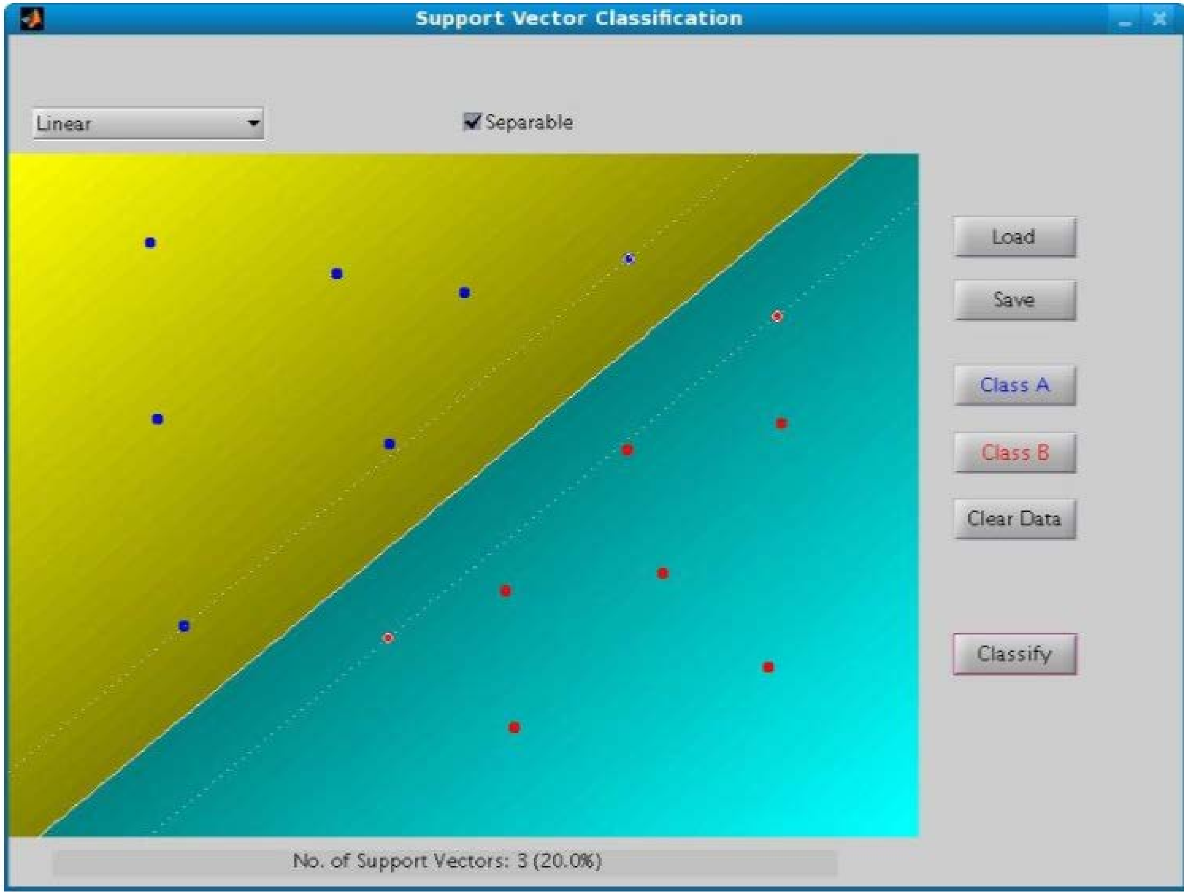
Synthetic Intelligence Lab

Support Vector Machines

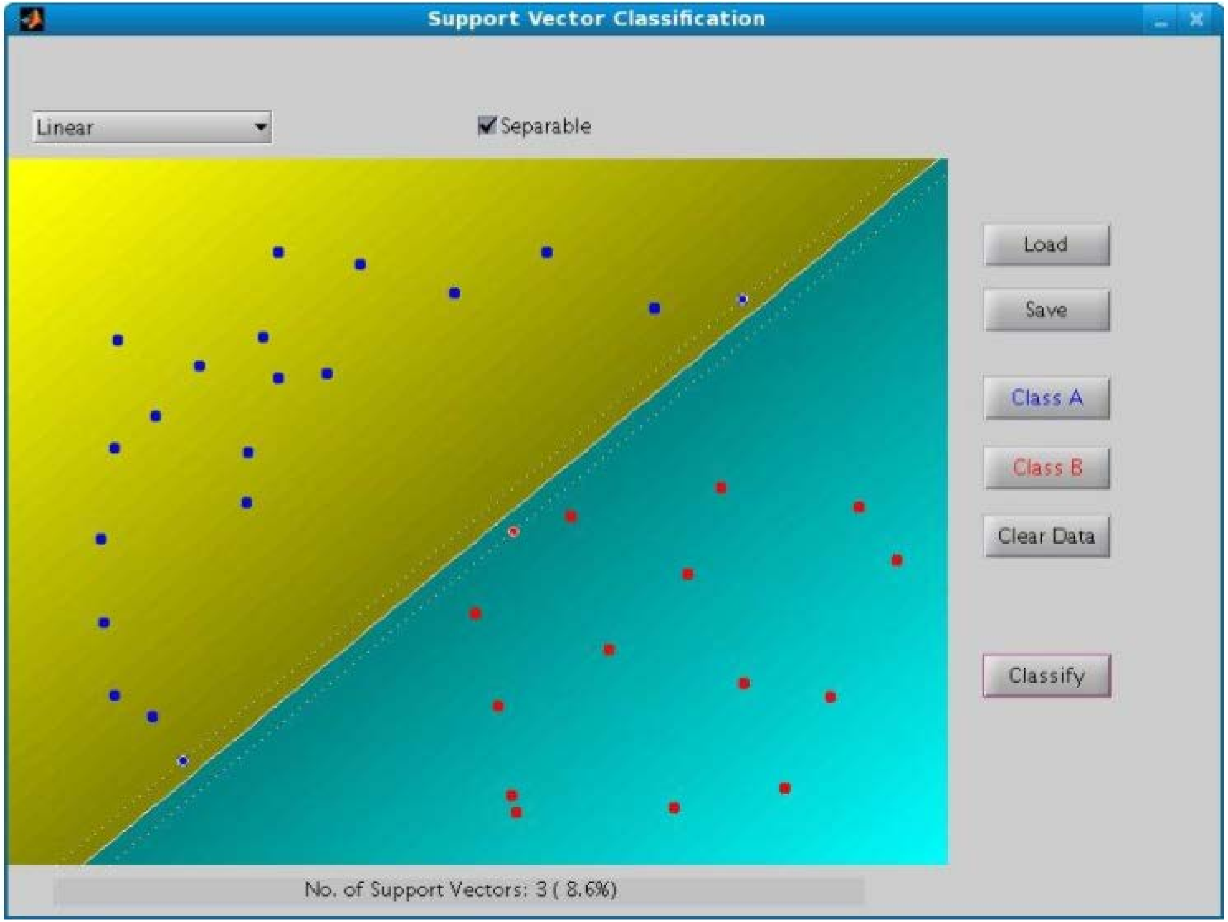
- If the data is not separable
- Slack variables are introduces for difficult examples
- Avoid overfitting



Support Vector Machines



Support Vector Machines



ISLab
Synthetic Intelligence Lab

Support Vector Machines

- SVMs can also be applied to non-linear problems (i.e. problems that can not split the training data by a hyper-plane)
- You can make any non-linear problem into linear
- The concept is to map the training data, from its original space for a larger space called feature space
- Kernel Trick

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle$$

- kernel function return the inner product ⟨Φ(x),Φ(x^′)⟩ between the images of two data points x, x' in the feature space Φ(x),Φ(x^′),

- Kernel functions

- Linear:

$$K(x, x') = x \cdot x'$$

- Gaussian Radial Basis Function (RBF):

$$K(x, x') = \exp(-\gamma \|x - x'\|^2),$$

$$\text{where } \gamma = \frac{1}{2\sigma^2}$$

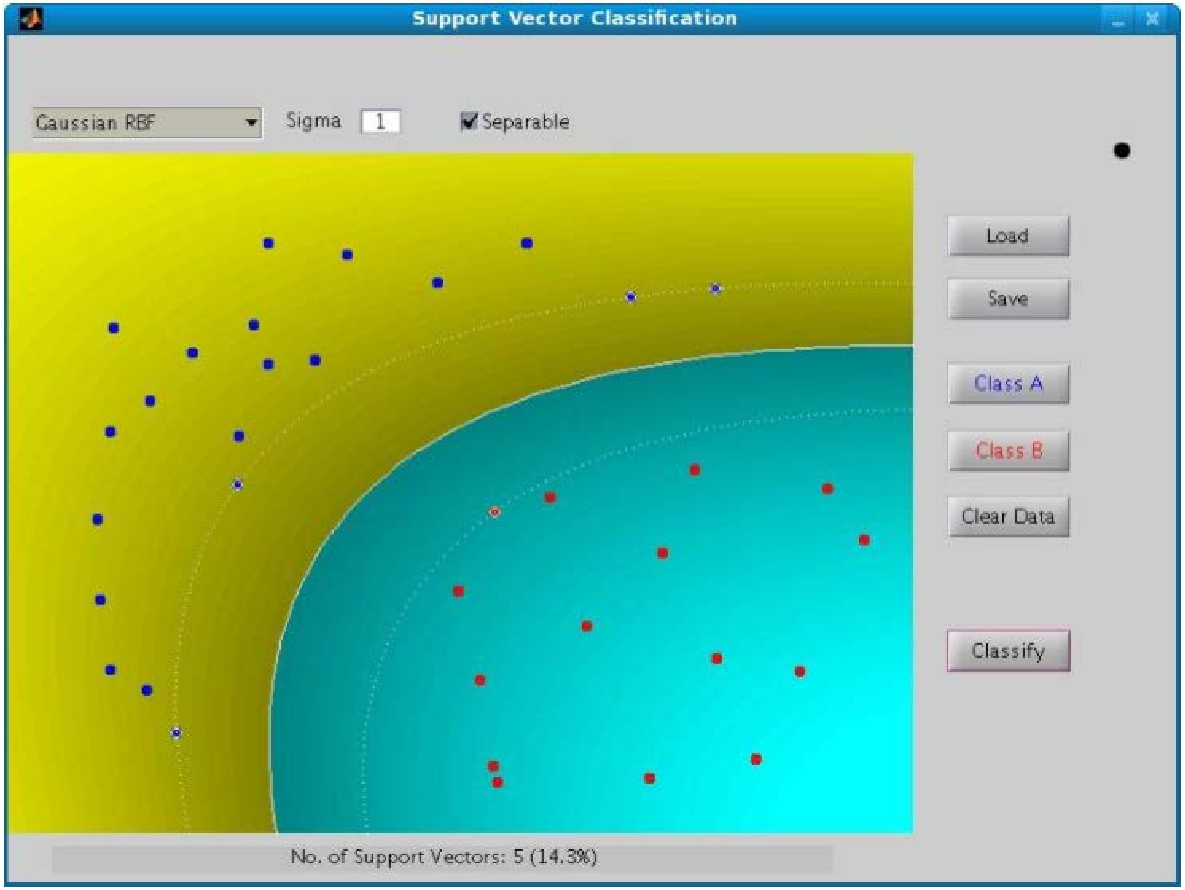
- Polynomial:

$$K(x, x') = (x^T x' + c)^d$$

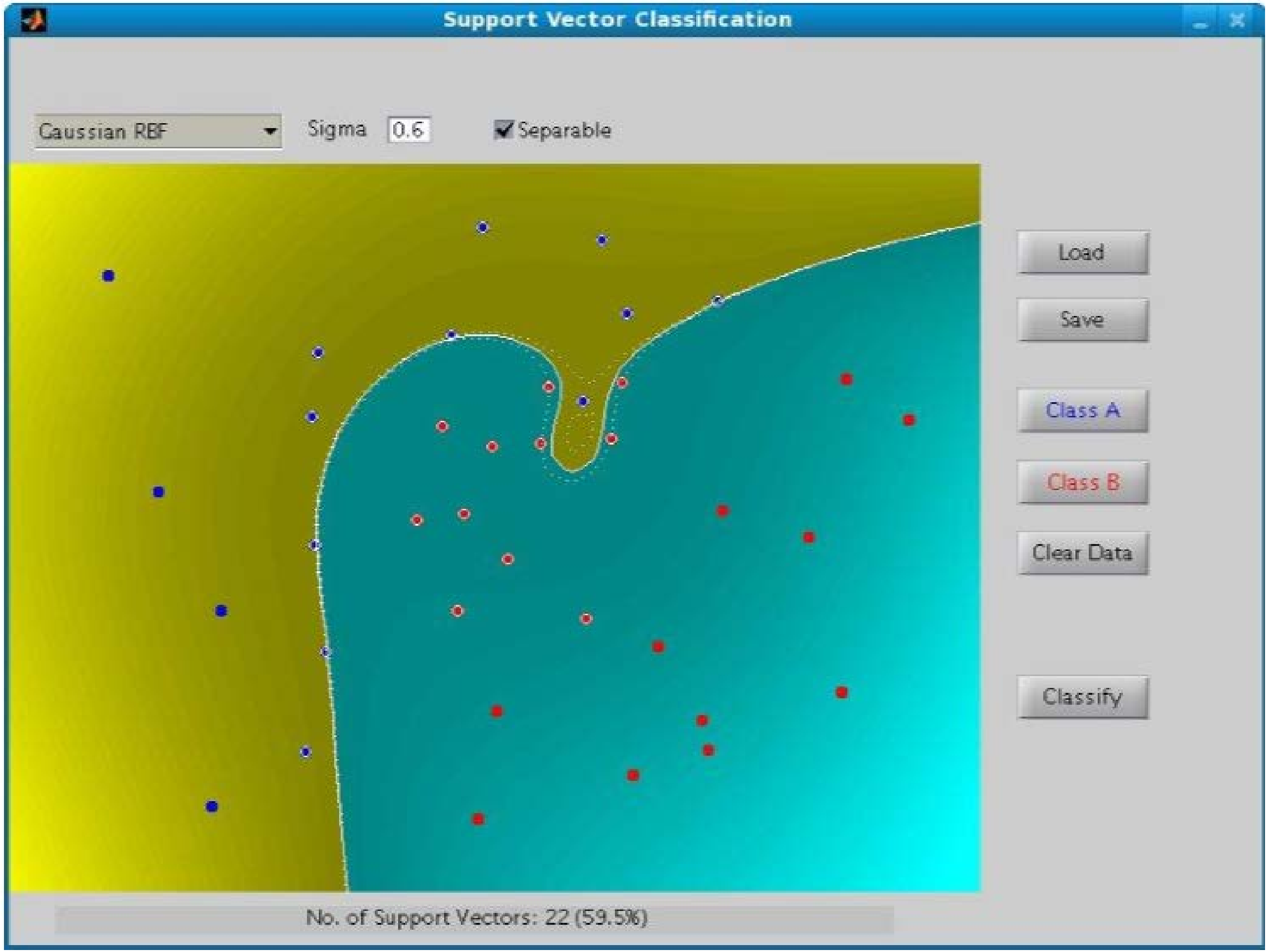
- Sigmoidal:

$$K(x, x') = \tanh(k x^T x' - \delta)$$

Support Vector Machines

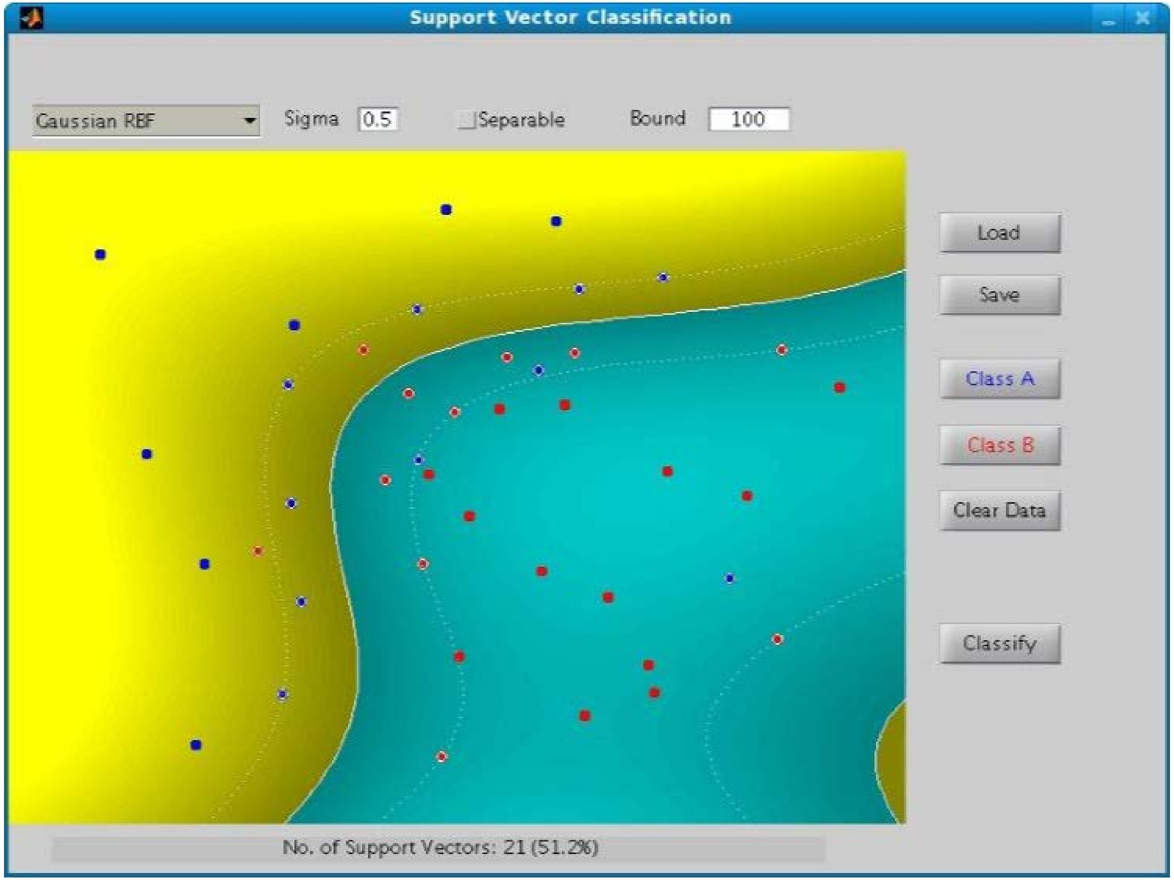


Support Vector Machines



ISLab
Synthetic Intelligence Lab

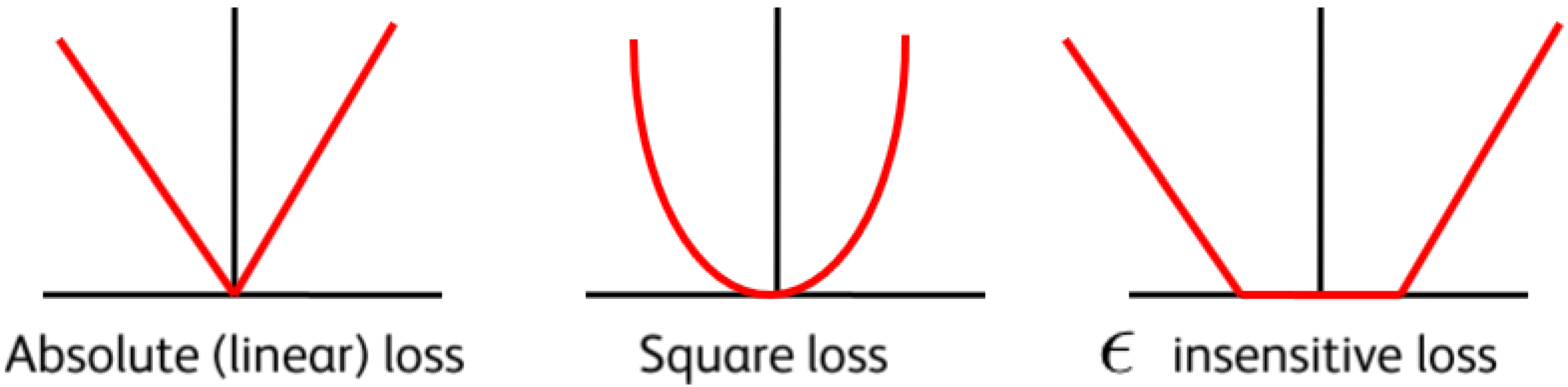
Support Vector Machines



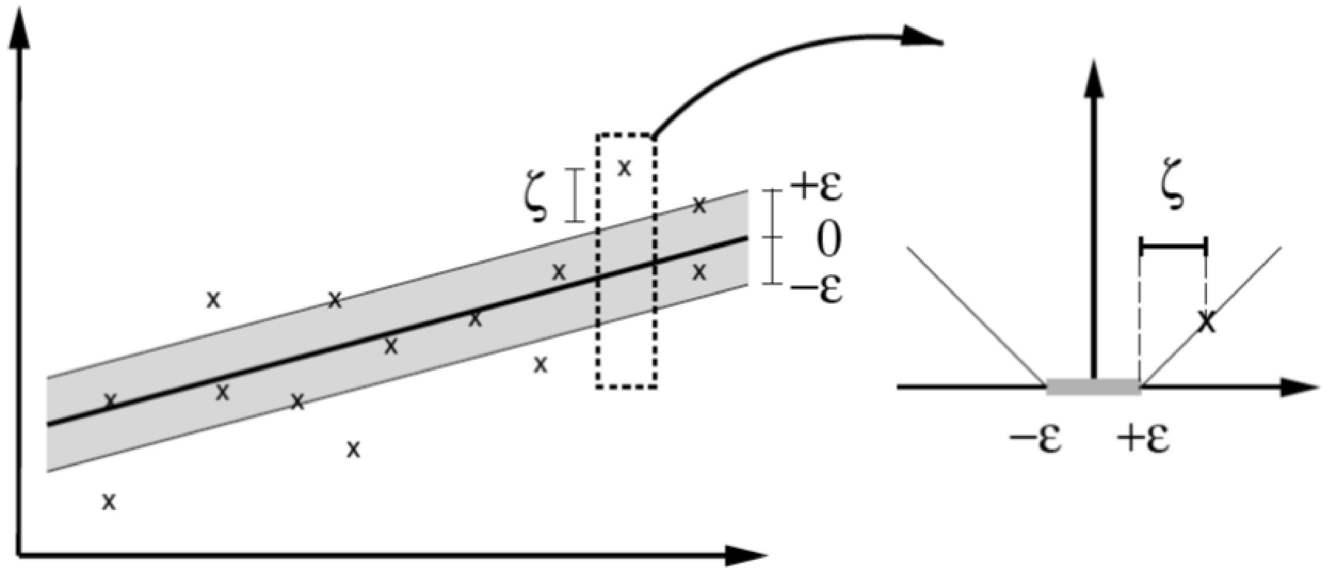
Support Vector Machines

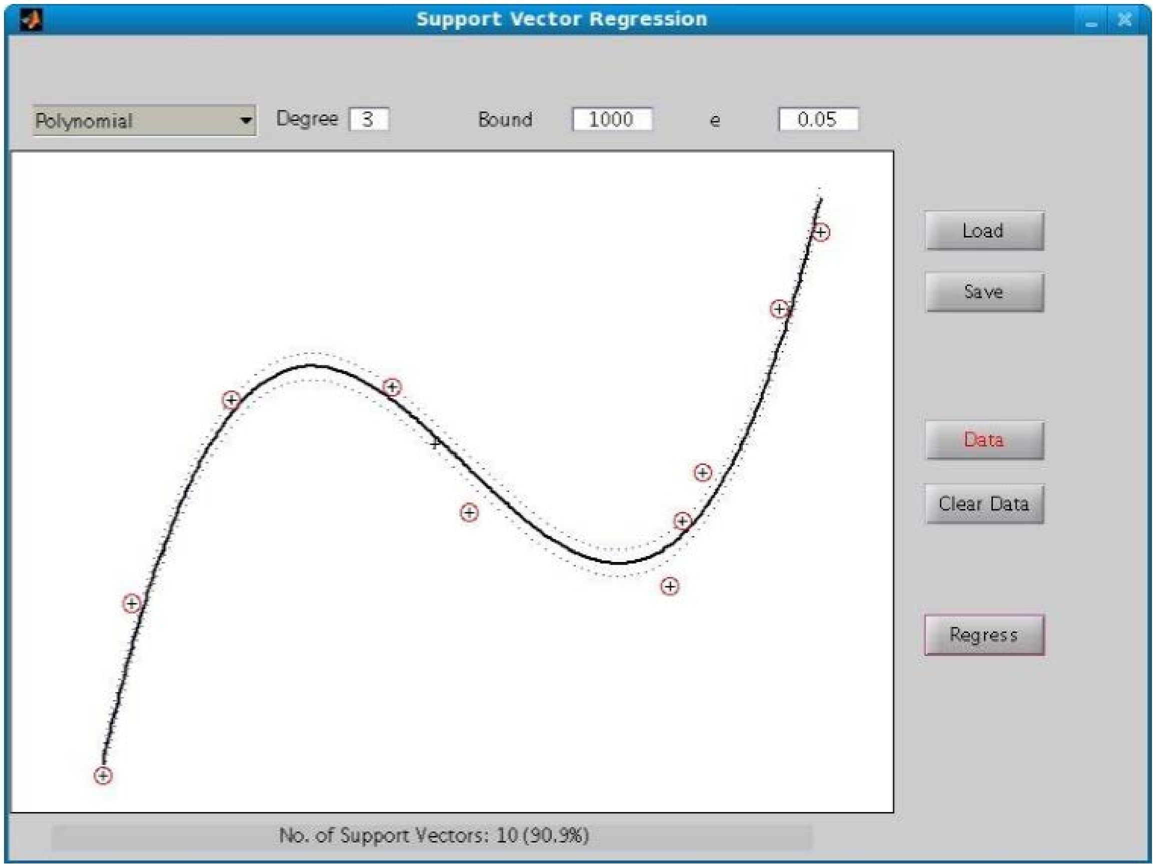
- SVM for classification
- Advantages
  - Unique global minimum
  - Use of kernel functions for non-linear data
  - Support Vector define the max-margin boundary
  - Deals with the curse of dimensionality
- Disadvantages
  - Parameters and kernel must be chosen
  - Can be difficult to interpret non-linear decision rules

- SVM for Regression
- Loss functions
    - Least squares is an empirical risk minimization algorithm that minimizes square loss
    - Many different loss function can be considered

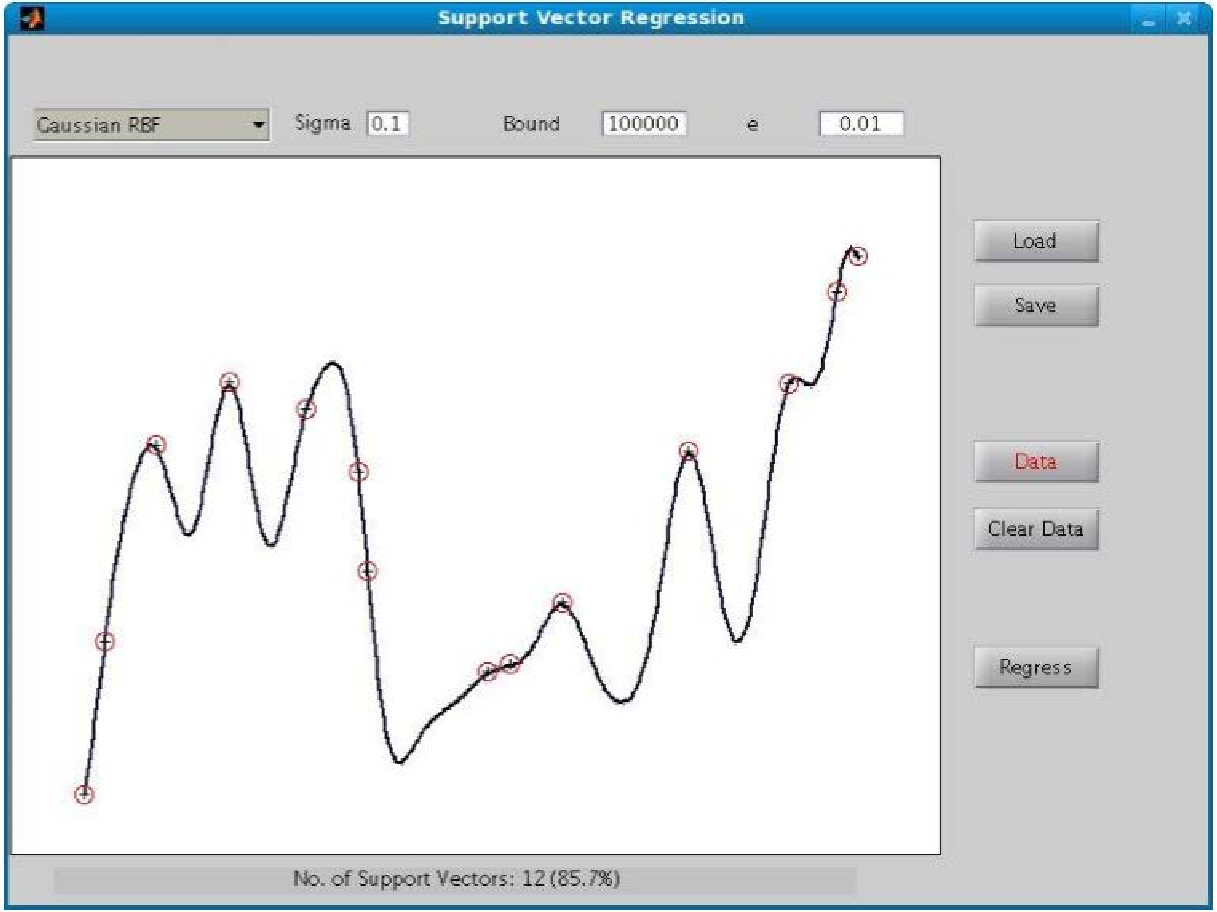


- Epsilon insensitive loss




Support Vector Regression
Polynomial
Degree
3
Bound
1000
e
0.05
Load
Save
Data
Clear Data
Regress
No. of Support Vectors: 10 (90.9%)

Support Vector Machines



Support Vector Machines

- Considerations
  - Loss function
  - Kernel function
  - Dimensionality / non-linearity
  - Parametrization

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Machine Learning: Support Vector Machines 2017/2018

Paulo Novais, Tiago Pinto