Support Vector Machines

EECE 580B

Lecture 1

January 26, 2010

Jan Kodovský, Jessica Fridrich

BINGHAMTON UNIVERSITY

State University of New York

Course Information

Lectures: TR 2:50 pm – 4:15 pm, LH – 12

Instructors: Jessica Fridrich (fridrich@binghamton.edu)

office hours: Monday 1pm – 3pm, office EB – Q16

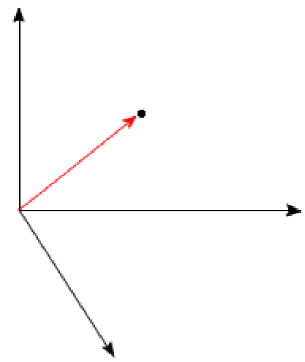
Jan Kodovský (jan.kodovsky@binghamton.edu)

office hours: Wednesday 1pm – 3pm, office LSG – 606

Webpage: http://dde.binghamton.edu/kodovsky/svm

Prerequisites

- Working knowledge of linear algebra and calculus.

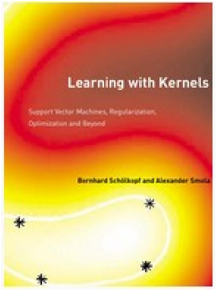


- Basics of programming in Matlab

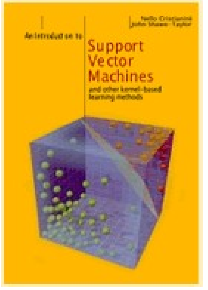
    Matlab tutorials available online

Course Materials

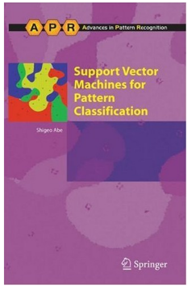
- Learning with Kernels
  - Authors: B. Schölkopf, A. J. Smola
  - Webpage: http://www.learning-with-kernels.org
  - Several chapters available online!
  - BU Library (Q325.5 .S32 2002) – on reserve



- Support Vector Machines and other kernel-based learning methods
  - Authors: N. Cristianini, J. Shawe-Taylor
  - Webpage: http://www.support-vector.net



- Support Vector Machines for Pattern Classification
  - Author: S. Abe
  - BU Library (QA76.9.T48 A23 2005) – on reserve



Other Online Resources

- SVM gateways
  - http://www.support-vector-machines.org
  - http://www.kernel-machines.org (seems to be better maintained)
  - Links to many online tutorials / books / papers / lectures / software
- A Tutorial on Support Vector Machines for Pattern Recognition (1998) link
  - Christopher J.C. Burges
- A Tutorial on ν-Support Vector Machines (2005) link
  - P.H. Chen, C.J. Lin, B. Schölkopf
- Kernel Methods in Machine Learning (2008) link
  - T. Hofmann, B. Schölkopf, A. Smola
- Videolectures on SVMs link
  - More than 40 SVM-related videolectures, slides available
  - C.J. Lin, B. Schölkopf, A. Smola, J. Shawe-Taylor, C. Campbell

Grading System

- Homework Assignments
  - 65% of the final grade
  - 8 – 10 in total
  - Variable weights
  - No assignments accepted after due date!

- Final exam
  - 35% of the final grade
  - Take home

Students are expected to work **individually**

Academic Honesty

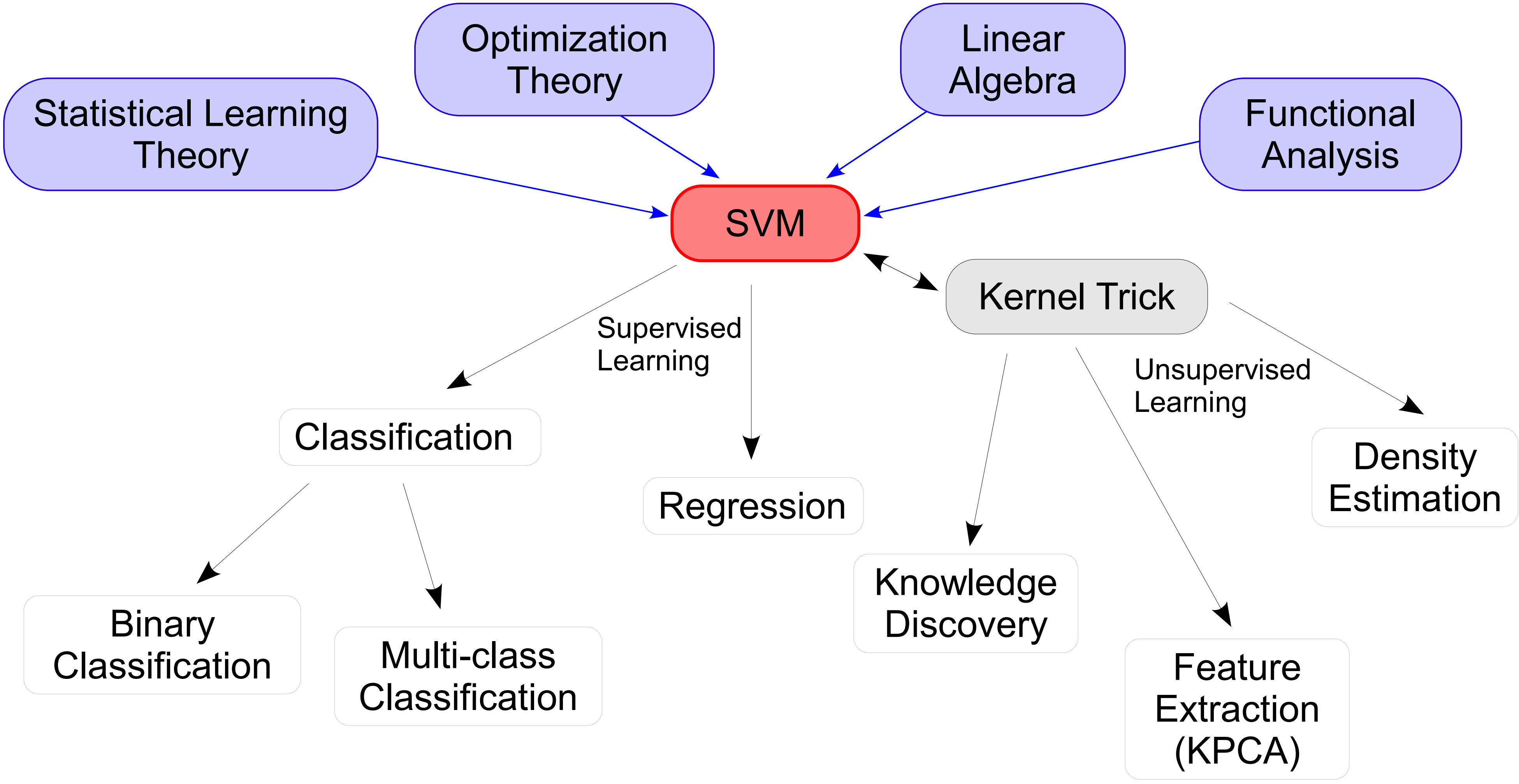
- Student Academic Honesty Code (Binghamton University)
- Student Academic Honesty Code (Watson School)

- First instance of academic dishonesty:
  - No credit for the assignment / exam on which the offense was committed
  - Reduction in course grade by one letter grade
  - Record of offense will be reported to university administration

- Second instance of academic dishonesty:
  - Failure of course
  - Further consequences outside the class (e.g., suspension)

Support Vector Machines (SVMs)

(Big Picture)



Support Vector Machines (SVMs)

- Advantages:
  - High performance
  - Controllable generalization ability
  - Optimization with no local minima
  - Robustness to outliers / high dimension
  - Kernelization of other dot-product based algorithms

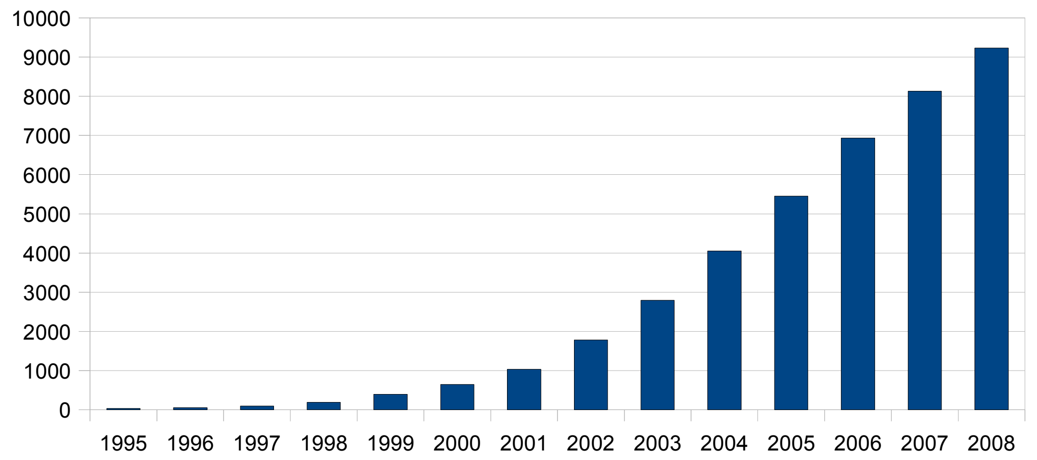
- Disadvantages:
  - Non-trivial extension to multi-class classification
  - Problem with proper selection of the kernel function parameters
  - Slow training for larger problems

Historical Remarks

- 1930s (R. A. Fisher)
  - Dependency estimation = estimating a finite number of parameters
- 1960s (F. Rosenblatt)
  - First model of learning machine – Perceptron
- 1960s – 1980s (V. Vapnik, A. Chervonenkis)
  - Building of the complex Statistical Learning Theory
  - Structural risk minimization inductive principle
- 1980s (D. E. Rumelhart, G. E. Hinton, R. J. Williams)
  - Second boom of neural networks (backpropagation algorithm)
- 1990s (V. Vapnik)
  - Support Vector Machines introduced
  - Rapidly growing community of researchers

Number of Related Publications

(Engineering, Computer Science, and Mathematics)

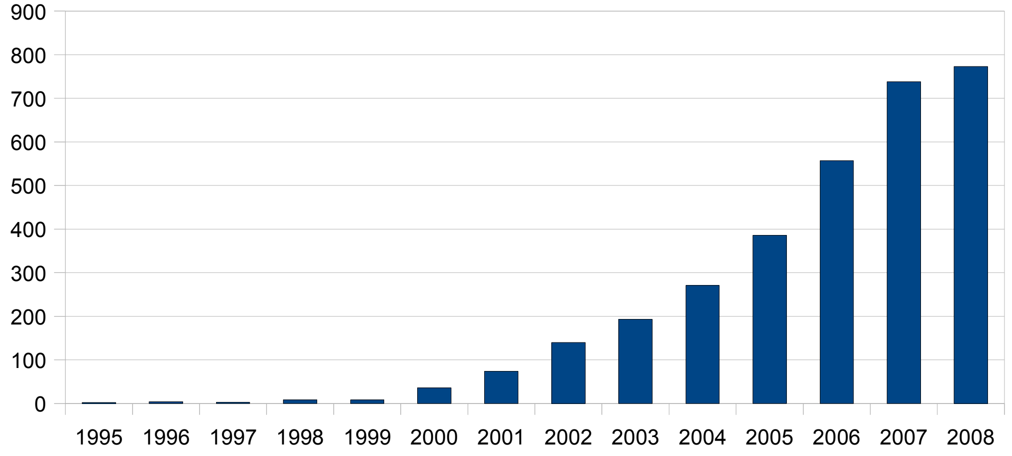


- Text categorization with SVMs (1998)
- SVMs for speaker and language recognition (2006)
- Road-sign detection and recognition based on SVMs (2007)
- Face Recognition using total margin-based adaptive fuzzy SVMs (2007)
- SVM for classification of voltage disturbances (2007)

source: Google scholar

Number of Related Publications

(Biology, Life Sciences, and Environmental Science)

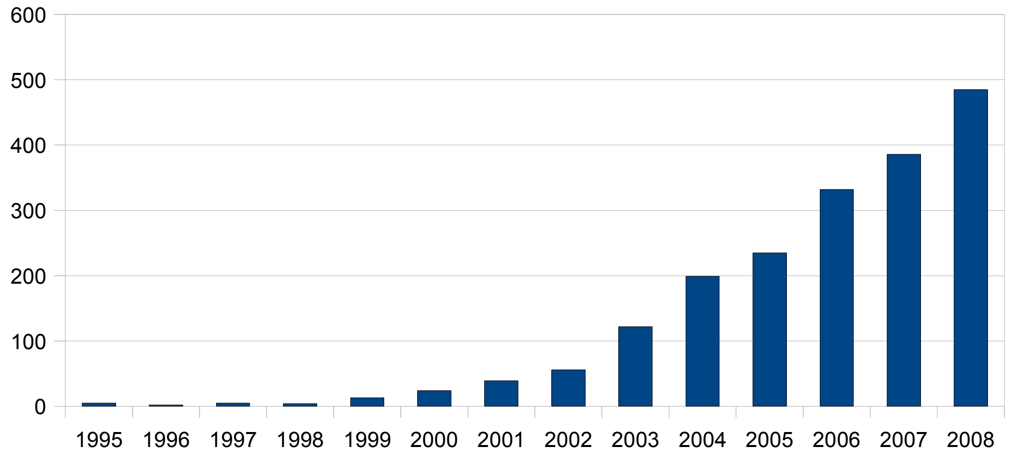


- SVMs for predicting HIV protease cleavage (2002)
- SVMs for cancer diagnosis from the blood concentration (2002)
- SVMs for predicting DNA binding proteins from amino sequences (2003)
- SVMs for predicting distribution of Sudden Oak Death in California (2004)

source: Google scholar

Number of Related Publications

(Social Sciences, Arts and Humanities)

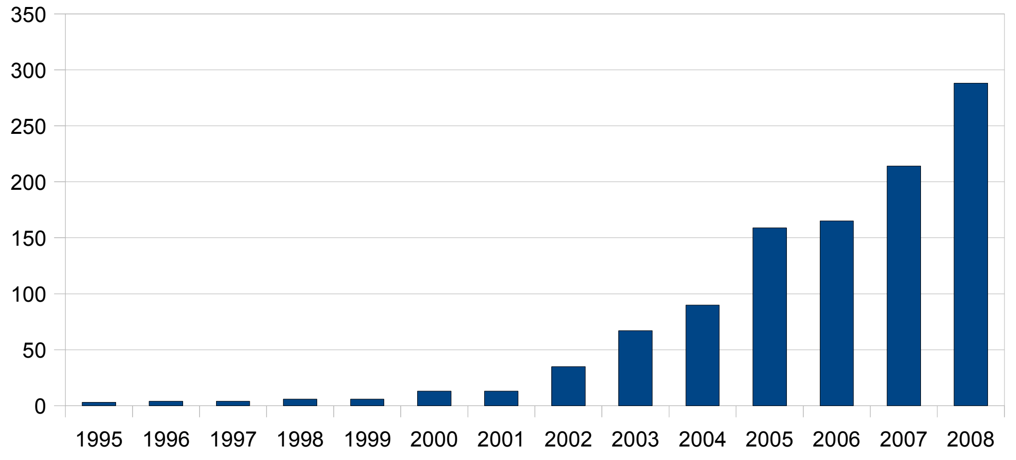


- Detecting emotion in music (2003)
- Recognizing expressions of commonsense psychology in English text (2003)
- Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression (2008)

source: Google scholar

Number of Related Publications

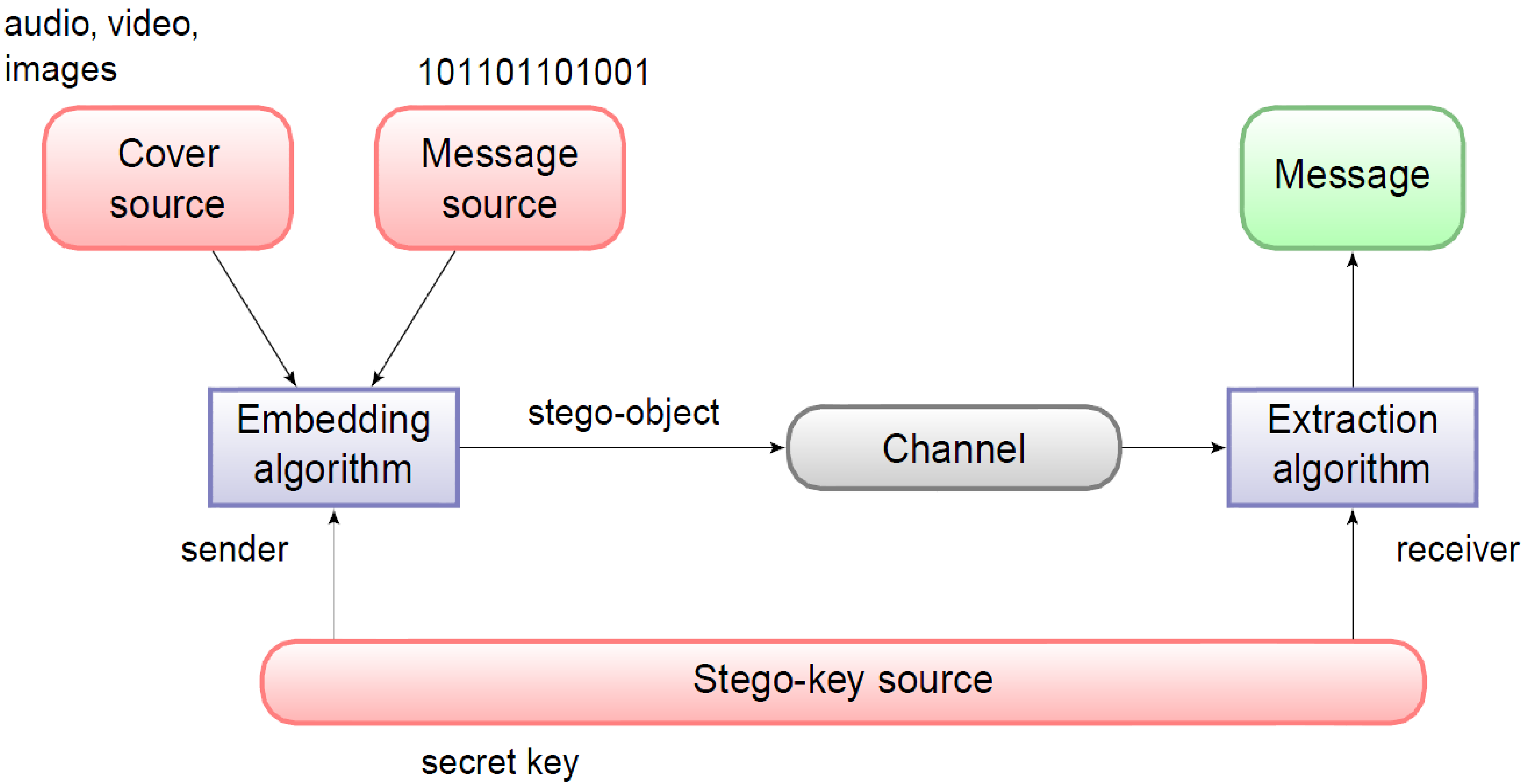
(Business, Administration, Finance, and Economics)



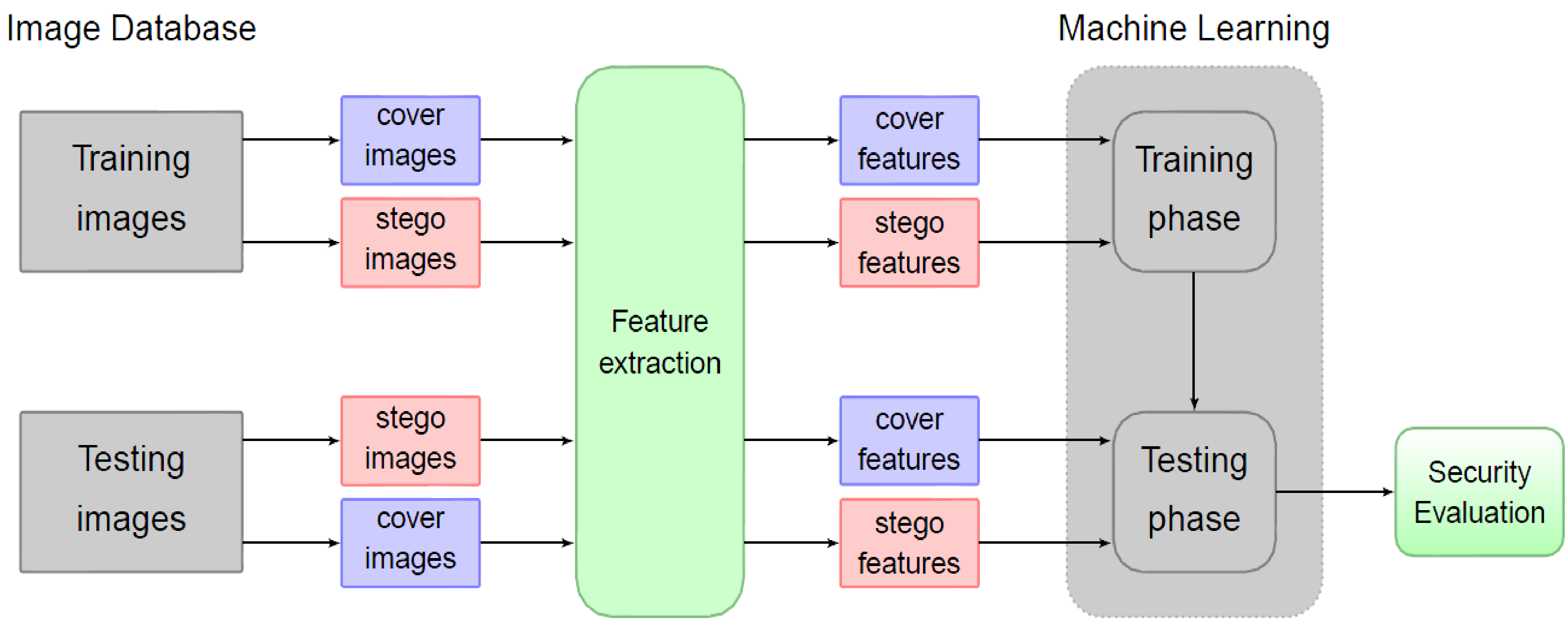
- Tourism demand modelling and forecasting (2001)
- Modified SVMs in financial time series forecasting (2003)
- Adapting SVM methods for horserace odds prediction (2007)

source: Google scholar

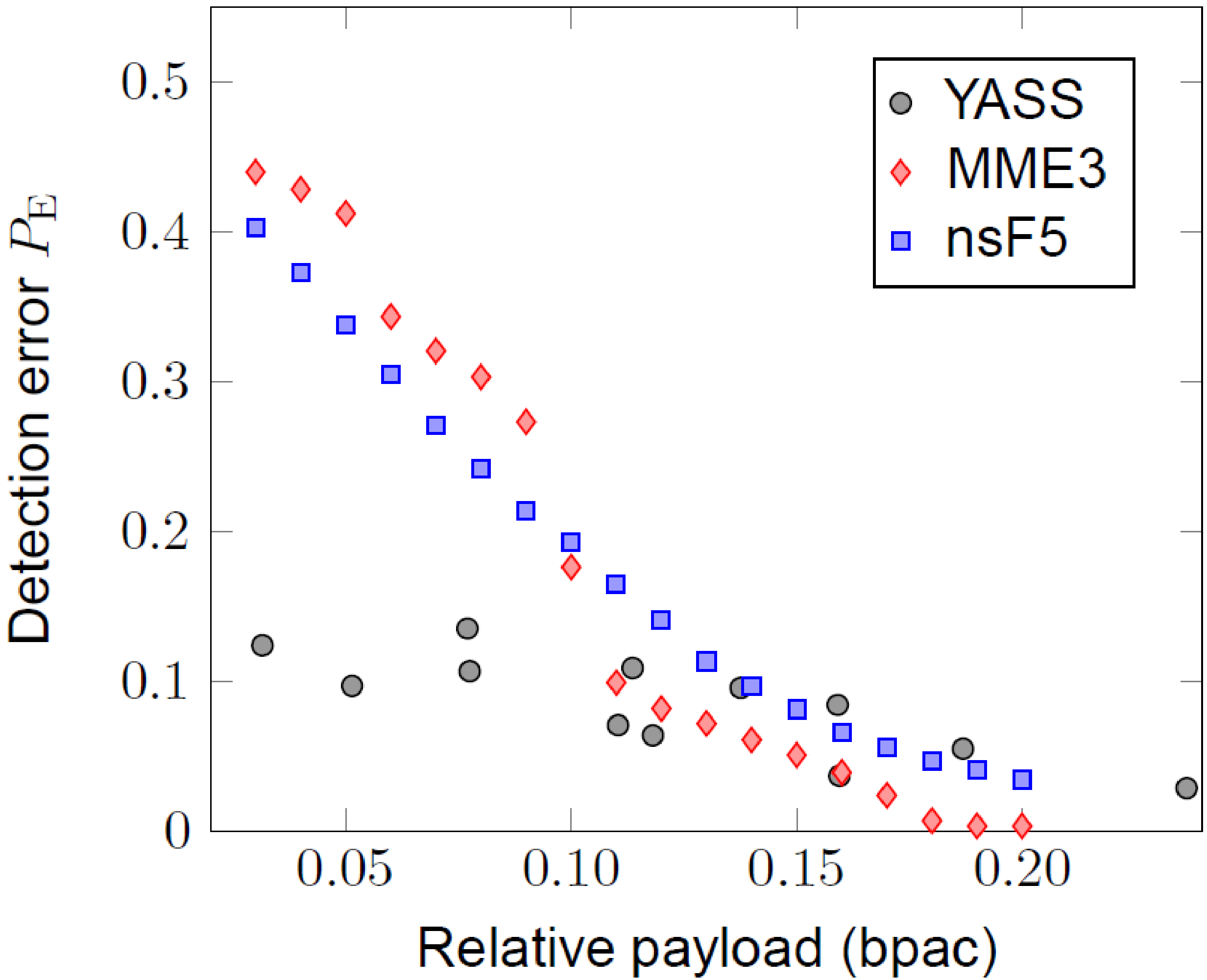
What is Steganography



What is Steganalysis



What is Steganalysis



SVM Software

- LibSVM
  - Authors: Chih-Chung Chang, Chih-Jen Lin
  - http://www.csie.ntu.edu.tw/~cjlin/libsvm
  - Integrated library for SVC, SVR, density estimation, multi-class classification
  - Sources in C++ and Java, interfaces for Python, R, MATLAB, Perl, and more
- Other
  - SVM^{light} – Author: Thorsten Joachims (Cornell University)
  - MATLAB – Bioinformatics Toolbox – Statistical Learning routines
- Extensive lists of SVM implementations / applets / packages / toolboxes
  - http://www.kernel-machines.org
  - http://www.support-vector-machines.org

Course Objectives

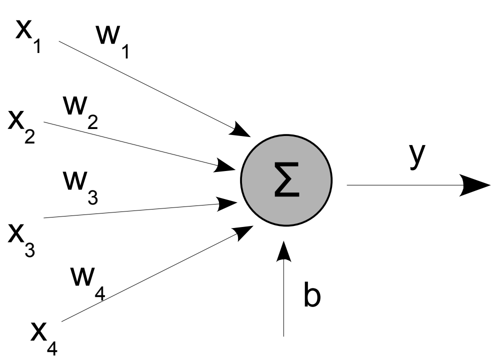
- Understand the core concepts SVMs are built on
- Gain practical experience with using SVM for classification problems
- Implement your own SVM machine (in Matlab)
- Be aware of potential issues when using SVMs
- Be able to use publicly available SVM libraries (and understand them)

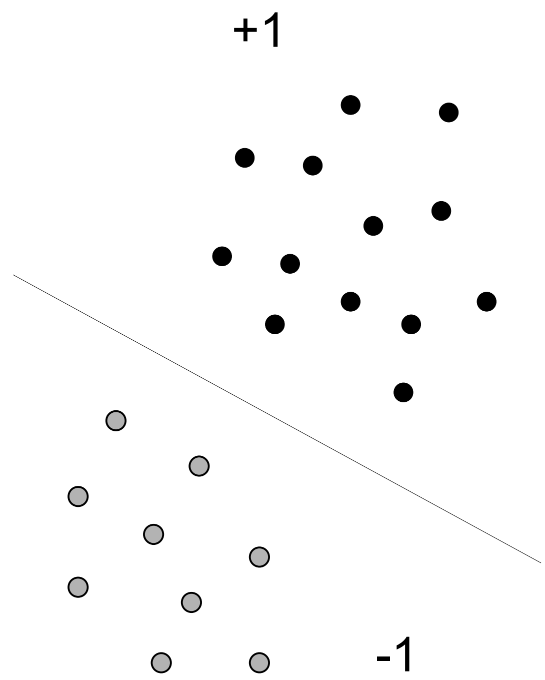
Class Outline

Introduction to classification

- First linear learning algorithms
- Perceptron
- Maximum margin classifier

Lecture 2 – 4



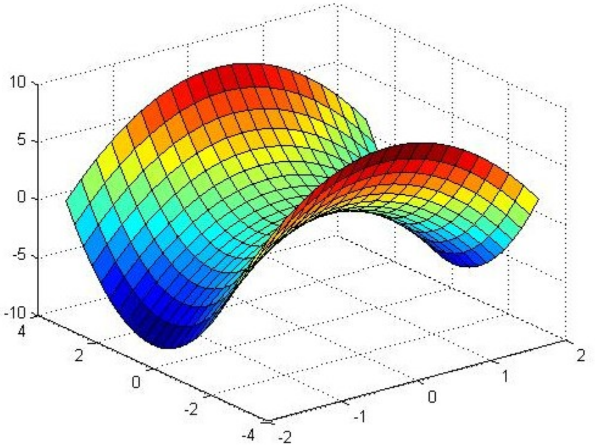


Class Outline

Optimization Theory

- Lagrangian Theory
- Duality
- KKT conditions

Lecture 5 – 6



minimize $\frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x}$

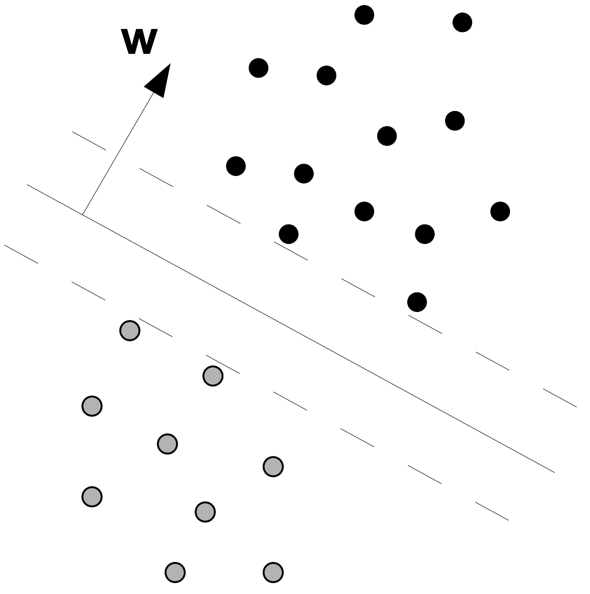
subject to $A\mathbf{x} \leq \mathbf{b}$

Class Outline

Simplest version of SVM

- Hard-margin SVM
- Linearly separable data
- Dual formulation

Lecture 7



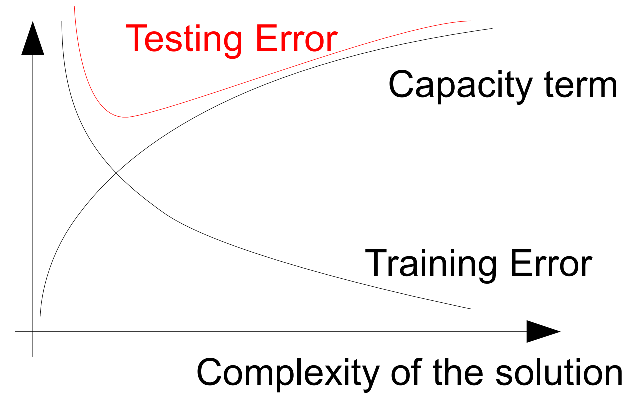
$$\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2$$

Class Outline

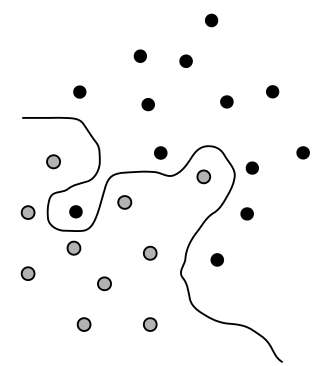
Statistical Learning Theory

- Why maximum margin?
- Structural risk minimization
- Generalization properties

Lecture 8



Testing Error < Training Error + Capacity term

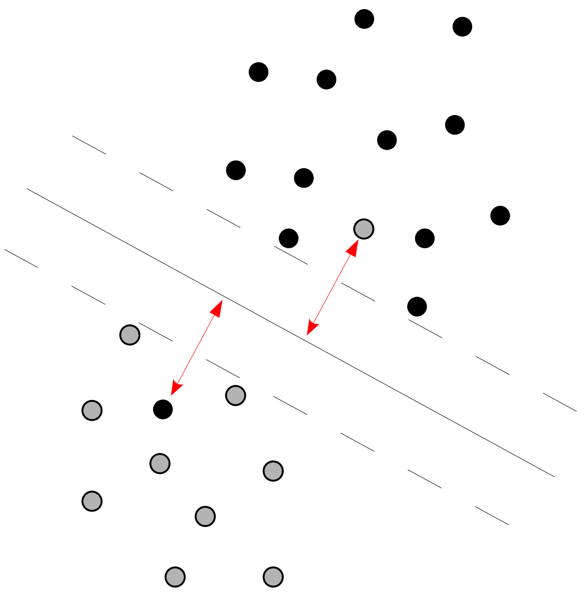


Class Outline

Soft Margin SVMs

- Generalization
- Linearly non-separable data
- Penalization of errors

Lecture 9 – 10



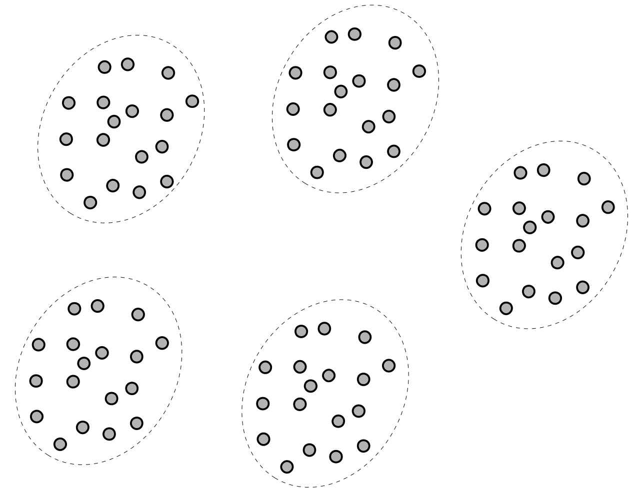
$$\text{minimize} \quad \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \xi_i$$

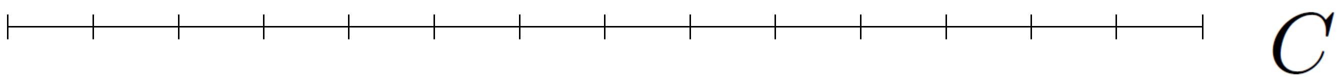
Class Outline

Practical Considerations

- Grid-search
- Cross-validation
- ROC curve

Lecture 11



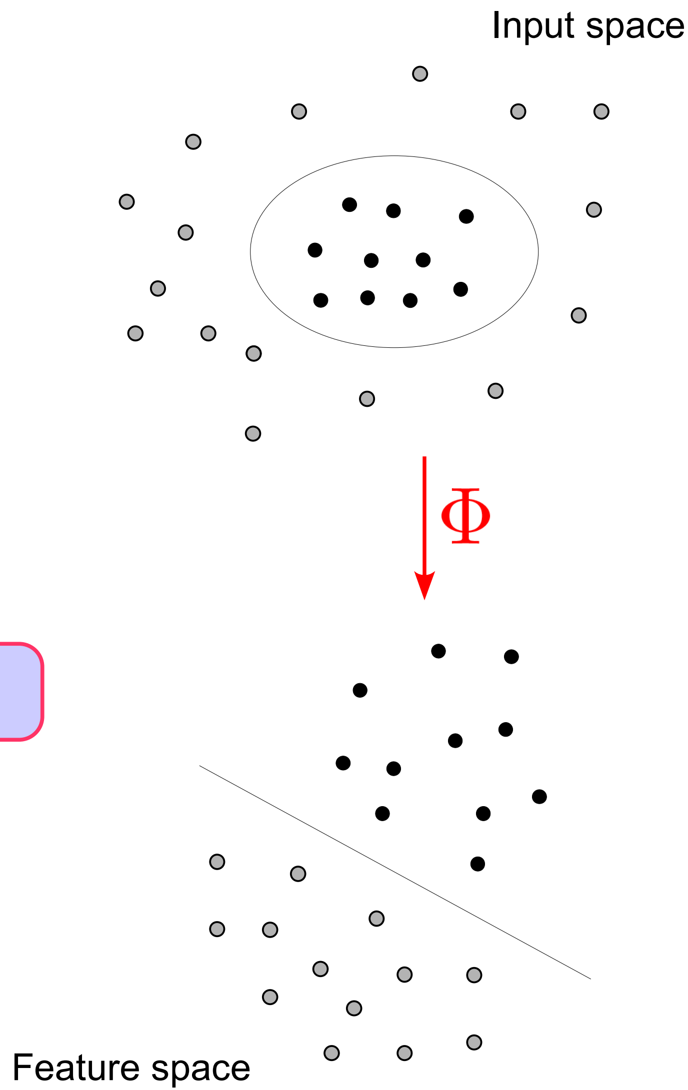


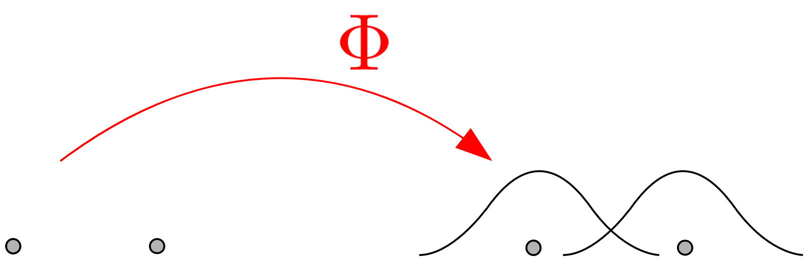
Class Outline

Kernel Trick

- Feature spaces
- Kernels
- Non-linear SVM

Lecture 12 – 13





Class Outline

Implementation Issues

- Training / Testing
- Stopping conditions
- Sequential Minimal Optimization

Lecture 14 – 15

```
Procedure takeStep(i1,i2)
   if (i1 == i2) return 0
   alph1 = Lagrange multipl
   y1 = target[i1]
   E1 = SVM output on point
   s = y1*y2
   Compute L, H
   if (L==H)
       return 0
   k11 = kernel(point[i1],p
   k12 = kernel(point[i1],p
   k22 = kernel(point[i2],p
   eta = 2*k12-k11-k22
   if (eta<0)
   {
       a2 = alph2 - y2*(E1-
       if (a2 < L) a2 = L
       else if (a2 > H) a2
   }
   else
   {
       Lobj = objective fun
       ...
```

Class Outline

Theoretical Foundations

- We don't need to know Φ !
- Mercer's Theorem
- Reproducing Kernel Hilbert Spaces

Lecture 16 – 17

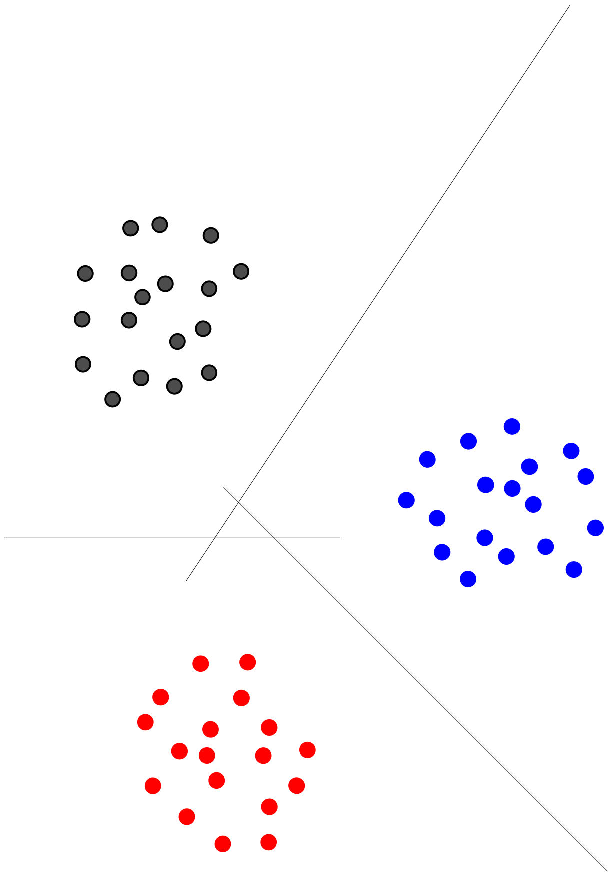
$$K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$$

Class Outline

Multi-classification

- One-against-All SVMs
- Pairwise SVMs
- All-at-Once SVMs

Lecture 18 – 20

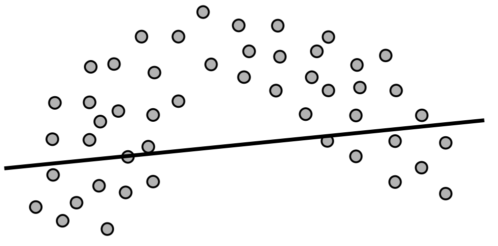


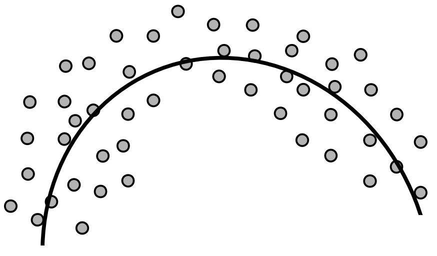
Class Outline

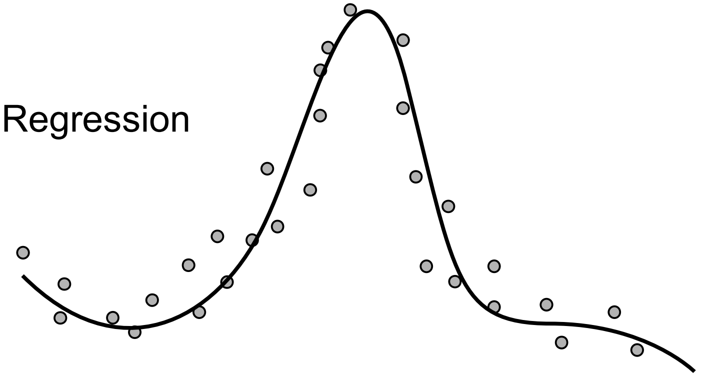
Advanced Topics

- Support vector regression
- Kernel PCA

Lecture 21 – 24







Class Outline

Application to Steganalysis

- Introduction to steganography
- Blind steganalysis

Lecture 25 – 27

cover

?

stego