Introduction

The Kerne Trick

Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Introduction to Kernel Methods

Fabio A. González Ph.D.

Depto. de Ing. de Sistemas e Industrial Universidad Nacional de Colombia, Bogotá

September 11, 2015

Introduction

The Kerne

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Outline

- 1 Introduction Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm Primal linear regression Dual linear regression
- 5 Kernel Functions

  Mathematical characterisation

  Visualizing kernels in input space
- 6 Kernel Algorithms
- **7** Kernels in Complex Structured Data

A Kernel Pattern Analysis Algorithn

Kernel Functions

Algorithms

Kernels in

Kernels in Complex Structured Data

# Outline

1 Introduction

Motivation

2 The Kernel Trick

Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm Primal linear regression Dual linear regression
- **5** Kernel Functions

Mathematical characterisation Visualizing kernels in input space

- 6 Kernel Algorithms
- 7 Kernels in Complex Structured Data

Introduction

# The Kern

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Algorithms

Kernels in

Kernels in Complex Structured Data

# Outline

## Introduction Motivation

- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- Kernel Functions
   Mathematical characterisation
   Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

Introduction

## Motivation

The Kern

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

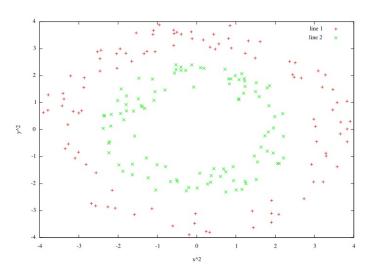
Kernel Function

Algorithms

Kernels in Complex Structured Data

## Problem 1

How to separate these two classes using a linear function?



Problem 2

Kernel Algorithms

Kernels in Complex Structured

## How to do symbolic regression?

$$\Sigma = \{A,\,C,\,G,\,T\}$$

$$\begin{array}{ccccc} f: & \Sigma^d & \rightarrow & \mathbb{R} \\ & ACGTA & \mapsto & 10.0 \\ & GTCCA & \mapsto & 11.3 \\ & GGTAC & \mapsto & 1.0 \\ & CCTGA & \mapsto & 4.5 \\ & \vdots & \vdots & \vdots \end{array}$$

Introduction

# The Kernel

space to the featu space Calculating the do

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Functions Kernel

Kernels in Complex Structured

## Outline

1 Introduction
Motivation

## 2 The Kernel Trick

Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm Primal linear regression Dual linear regression
- Kernel Functions
   Mathematical characterisation
   Visualizing kernels in input space
- **6** Kernel Algorithms
- **7** Kernels in Complex Structured Data

Calculating the de product in the feature space

Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Functions

Kernel Algorithms

Kernels in Complex Structured Data

## Outline

1 Introduction
Motivation

2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

Introductio

The Kerne

Mapping the input space to the feature space

Calculating the de product in the feature space

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

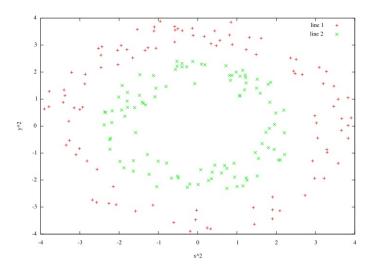
Kernel Functions

Kernel Algorithms

Kernels in Complex Structured Data

## Problem 1

• How to separate these two classes using a linear function?



Calculating the opposition of the feature space

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

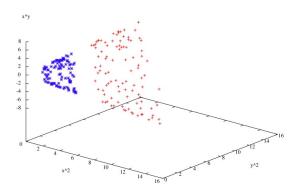
Kernel Algorithms

Kernels in Complex Structured Data

## Solution

• Map to  $\mathbb{R}^3$ :

$$\begin{array}{ccc} \phi: \mathbb{R}^2 & \to & \mathbb{R}^3 \\ (x,y) & \mapsto & (x^2,y^2,xy) \end{array}$$



Calculating the d product in the feature space

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

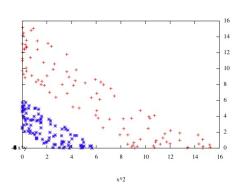
Kernel Algorithms

Kernels in Complex Structured Data

## Solution

• Map to  $\mathbb{R}^3$ :

$$\begin{array}{ccc} \phi: \mathbb{R}^2 & \to & \mathbb{R}^3 \\ (x,y) & \mapsto & (x^2,y^2,xy) \end{array}$$



y^2

Introductio

## The Kerne

Mapping the input space to the feature space

Calculating the d

The Kernel Approach t Machine

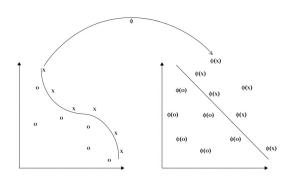
A Kernel Pattern Analysis

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured

# Input space vs. feature space



space to the feature space Calculating the dot

Calculating the de product in the feature space

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured Data

## Outline

1 Introduction
Motivation

2 The Kernel Trick

Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm Primal linear regression Dual linear regression
- Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

#### Introduction

## The Kerne

space to the featu

#### Calculating the dot product in the feature space

The Kernel Approach to Machine

A Kernel Pattern Analysis Algorithm

### Kernel Functions

Kernel Algorithms

Kernels in Complex Structured Data

$$\phi: \mathbb{R}^2 \to \mathbb{R}^3 
(x_1, x_2) \mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

$$\langle \phi(x), \phi(z) \rangle = \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \rangle$$

$$= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1x_2z_1z_2$$

$$= (x_1z_1 + x_2z_2)^2$$

$$= \langle x, z \rangle^2$$

- A function  $k: X \times X \to \mathbb{R}$  such that  $k(x,z) = \langle \phi(x), \phi(z) \rangle$  is called a kernel
- Morale: you don't need to apply  $\phi$  explicitly to calculate the dot product in the feature space!

$$\phi : \mathbb{R}^2 \to \mathbb{R}^3$$
  
 $(x_1, x_2) \mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ 

$$\langle \phi(x), \phi(z) \rangle = \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \rangle$$

$$= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1x_2z_1z_2$$

$$= (x_1z_1 + x_2z_2)^2$$

$$= \langle x, z \rangle^2$$

- A function  $k: X \times X \to \mathbb{R}$  such that  $k(x,z) = \langle \phi(x), \phi(z) \rangle$  is called a kerne
- Morale: you don't need to apply  $\phi$  explicitly to calculate the dot product in the feature space!

Kernels in Complex Structured Data

$$\phi : \mathbb{R}^2 \to \mathbb{R}^3$$
  
 $(x_1, x_2) \mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ 

$$\langle \phi(x), \phi(z) \rangle = \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \rangle$$

$$= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1 x_2 z_1 z_2$$

$$= (x_1 z_1 + x_2 z_2)^2$$

$$= \langle x, z \rangle^2$$

- A function  $k: X \times X \to \mathbb{R}$  such that  $k(x,z) = \langle \phi(x), \phi(z) \rangle$  is called a kernel
- Morale: you don't need to apply  $\phi$  explicitly to calculate the dot product in the feature space!

Kernels in Complex Structured Data

$$\phi : \mathbb{R}^2 \to \mathbb{R}^3$$
  
 $(x_1, x_2) \mapsto (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ 

$$\langle \phi(x), \phi(z) \rangle = \langle (x_1^2, x_2^2, \sqrt{2}x_1x_2), (z_1^2, z_2^2, \sqrt{2}z_1z_2) \rangle$$

$$= x_1^2 z_1^2 + x_2^2 z_2^2 + 2x_1x_2z_1z_2$$

$$= (x_1z_1 + x_2z_2)^2$$

$$= \langle x, z \rangle^2$$

- A function  $k: X \times X \to \mathbb{R}$  such that  $k(x, z) = \langle \phi(x), \phi(z) \rangle$  is called a kernel
- Morale: you don't need to apply  $\phi$  explicitly to calculate the dot product in the feature space!

space to the feat space

Calculating the dot product in the feature space

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured Data

# Kernel induced feature space

The feature space induced by the kernel is not unique:
 The kernel

$$k(x,z) = \langle x, z \rangle^2$$

also calculates the dot product in the four dimensional feature space:

$$\phi: \mathbb{R}^2 \to \mathbb{R}^4$$
  
 $(x_1, x_2) \mapsto (x_1^2, x_2^2, x_1 x_2, x_2 x_1)$ 

• The example can be generalised to  $\mathbb{R}^n$ 

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured Data

# Kernel induced feature space

The feature space induced by the kernel is not unique:
 The kernel

$$k(x,z) = \langle x, z \rangle^2$$

also calculates the dot product in the four dimensional feature space:

$$\phi : \mathbb{R}^2 \to \mathbb{R}^4$$
  
 $(x_1, x_2) \mapsto (x_1^2, x_2^2, x_1 x_2, x_2 x_1)$ 

• The example can be generalised to  $\mathbb{R}^n$ 

Introduction

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithm

Kernels in Complex Structured Data

## Outline

- 1 Introduction
  Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- 6 Kernel Algorithms
- 7 Kernels in Complex Structured Data

The Kernel Approach to Machine Learning

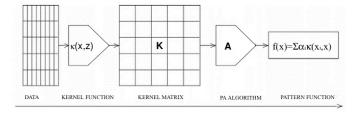
A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## The Process



A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

- Data items are embedded into a vector space called the feature space
- Linear relations are sought among the images of the data items in the feature space
- The pattern analysis algorithm are based only on the pairwise dot products, they do not need the actual coordinates of the embedded points
- The pairwise dot products in the feature space could be efficiently calculated using a kernel function

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

- Data items are embedded into a vector space called the feature space
- Linear relations are sought among the images of the data items in the feature space
- The pattern analysis algorithm are based only on the pairwise dot products, they do not need the actual coordinates of the embedded points
- The pairwise dot products in the feature space could be efficiently calculated using a kernel function

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

- Data items are embedded into a vector space called the feature space
- Linear relations are sought among the images of the data items in the feature space
- The pattern analysis algorithm are based only on the pairwise dot products, they do not need the actual coordinates of the embedded points
- The pairwise dot products in the feature space could be efficiently calculated using a kernel function

Machine Learning

A Kernel
Pattern

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

- Data items are embedded into a vector space called the feature space
- Linear relations are sought among the images of the data items in the feature space
- The pattern analysis algorithm are based only on the pairwise dot products, they do not need the actual coordinates of the embedded points
- The pairwise dot products in the feature space could be efficiently calculated using a kernel function

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

### A Kernel Pattern Analysis Algorithm

Primal linear regression

Dual linear regress

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

# Outline

- 1 Introduction
  Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm

Primal linear regression Dual linear regression

- 5 Kernel Functions
  Mathematical character
  - Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

### Primal linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Outline

- 1 Introduction
  Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

## Primal linear regression

Dual linear regres

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Problem definition

• Given a training set  $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$  of points  $\mathbf{x}_i \in \mathbb{R}^n$  with corresponding labels  $y_i \in \mathbb{R}$  the problem is to find a real-valued linear function that best interpolates the training set:

$$g(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle = \mathbf{w}' \mathbf{x} = \sum_{i=1}^{n} w_i x_i$$

• If the data points were generated by a function like g(x), it is possible to find the parameters w by solving

$$Xw = y$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{x'}_1 \\ \vdots \\ \mathbf{x'}_l \end{bmatrix}$$

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

Dual linear regres

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Problem definition

• Given a training set  $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\}$  of points  $\mathbf{x}_i \in \mathbb{R}^n$  with corresponding labels  $y_i \in \mathbb{R}$  the problem is to find a real-valued linear function that best interpolates the training set:

$$g(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle = \mathbf{w}' \mathbf{x} = \sum_{i=1}^{n} w_i x_i$$

• If the data points were generated by a function like g(x), it is possible to find the parameters w by solving

$$Xw = y$$

where

$$\mathbf{X} = \left[ \begin{array}{c} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_l \end{array} \right]$$

Introductio

The Kerne

The Kerne Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

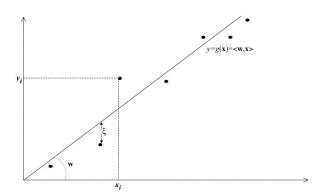
Dual linear regress

Kernel

Kernel

Kernels in Complex Structured

# Graphical representation



Introduction

The Kerne

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

Dual linear regress

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Loss function

Minimize

$$\mathcal{L}(g, S) = \mathcal{L}(\mathbf{w}, S) = \sum_{i=1}^{l} (y_i - g(x_i))^2 = \sum_{i=1}^{l} \xi_i^2$$
$$= \sum_{i=1}^{l} \mathcal{L}(g, (\mathbf{x}_i, y_i))$$

This could be written as

$$\mathcal{L}(w, S) = ||\xi||^2 = (y - Xw)'(y - Xw)$$

Optimization problem:

$$\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, S) = \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})'(\mathbf{y} - \mathbf{X}\mathbf{w})$$

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

Dual linear regres

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Loss function

Minimize

$$\mathcal{L}(g, S) = \mathcal{L}(\mathbf{w}, S) = \sum_{i=1}^{l} (y_i - g(x_i))^2 = \sum_{i=1}^{l} \xi_i^2$$
$$= \sum_{i=1}^{l} \mathcal{L}(g, (\mathbf{x}_i, y_i))$$

• This could be written as

$$\mathcal{L}(w, S) = ||\xi||^2 = (y - Xw)'(y - Xw)$$

Optimization problem:

$$\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, S) = \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})'(\mathbf{y} - \mathbf{X}\mathbf{w})$$

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

Dual linear regres

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Loss function

Minimize

$$\mathcal{L}(g, S) = \mathcal{L}(\mathbf{w}, S) = \sum_{i=1}^{l} (y_i - g(x_i))^2 = \sum_{i=1}^{l} \xi_i^2$$
$$= \sum_{i=1}^{l} \mathcal{L}(g, (\mathbf{x}_i, y_i))$$

This could be written as

$$\mathcal{L}(w, S) = ||\xi||^2 = (y - Xw)'(y - Xw)$$

Optimization problem:

$$\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}, S) = \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})'(\mathbf{y} - \mathbf{X}\mathbf{w})$$

Introduction

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

#### Primal linear regression

Dual linear regress

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured

$$\frac{\partial \mathcal{L}(\mathbf{w},S)}{\partial \mathbf{w}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{w} = 0,$$

therefore

$$X'Xw = X'y,$$

and

$$w = (X'X)^{-1}X'y$$

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Dual linear regression

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

# Outline

1 Introduction
Motivation

2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Primal lin

Dual linear regression

Kernel

Kernel Algorithm

Kernels in Complex Structured

# Dual representation of the problem

• 
$$w = (X'X)^{-1}X'y = X'X(X'X)^{-2}X'y = X'\alpha$$

• So, w is a linear combination of the training samples,  $\mathbf{w} = \sum_{i=1}^{l} \alpha_i \mathbf{x}_i$ .

A Kernel Pattern Analysis

Primal lin

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Dual representation of the problem

• 
$$w = (X'X)^{-1}X'y = X'X(X'X)^{-2}X'y = X'\alpha$$

• So, w is a linear combination of the training samples,  $\mathbf{w} = \sum_{i=1}^{l} \alpha_i \mathbf{x}_i$ .

Kernels in Complex Structured Data

## Solution

• From the solution of the primal problem:

$$X'Xw = X'y,$$

then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

• then

$$\alpha = (XX')^{-1}y,$$

and

$$g(x) = w'x = \alpha'Xx.$$

Kernels in Complex Structured Data

## Solution

• From the solution of the primal problem:

$$X'Xw = X'y,$$

then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

then

$$\alpha = (XX')^{-1}y,$$

and

$$g(\mathbf{x}) = \mathbf{w}'\mathbf{x} = \alpha'\mathbf{X}\mathbf{x}.$$

Introduction

The Kerne

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Primal linea

Dual linear regression

Kernel

Kernel Algorithm

Kernels in Complex Structured Data

## Solution

• From the solution of the primal problem:

$$X'Xw = X'y,$$

• then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

then

$$\alpha = (XX')^{-1}y,$$

and

$$g(\mathbf{x}) = \mathbf{w}'\mathbf{x} = \alpha'\mathbf{X}\mathbf{x}.$$

Kernels in Complex Structured Data

#### Solution

• From the solution of the primal problem:

$$X'Xw = X'y,$$

• then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

then

$$\alpha = (XX')^{-1}y,$$

and

$$g(\mathbf{x}) = \mathbf{w}'\mathbf{x} = \alpha'\mathbf{X}\mathbf{x}.$$

Primal line

Dual linear regression

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

### Solution

• From the solution of the primal problem:

$$X'Xw = X'y,$$

• then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

then

$$\alpha = (XX')^{-1}y,$$

and

$$g(x) = w'x = \alpha'Xx.$$

Kernels in Complex Structured Data

## Solution

• From the solution of the primal problem:

$$X'Xw=X'y,$$

• then

$$XX'Xw = XX'y,$$

using the dual representation

$$XX'XX'\alpha = XX'y,$$

then

$$\alpha = (XX')^{-1}y,$$

and

$$g(x) = w'x = \alpha'Xx.$$

A Kernel Pattern Analysis Algorithm

Primal linear

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Ridge regression

- If XX' is singular, the pseudo-inverse could be used: to find the w that satisfies X'Xw = X'y with minimal norm.
- Optimisation problem:

$$\min_{\mathbf{w}} \mathcal{L}_{\lambda}(\mathbf{w}, S) = \min_{\mathbf{w}} \lambda \|\mathbf{w}\|^{2} + \sum_{i=1}^{l} (y_{i} - g(x_{i}))^{2},$$

where  $\lambda$  defines the trade-off between norm and loss. This controls the complexity of the model (the process is called *regularization*).

A Kernel Pattern Analysis Algorithm

Algorithm Primal linea

Dual linear regression

Kernel

Kernel Algorithms

Kernels in Complex Structured

## Ridge regression

- If XX' is singular, the pseudo-inverse could be used: to find the w that satisfies X'Xw = X'y with minimal norm.
- Optimisation problem:

$$\min_{\mathbf{w}} \mathcal{L}_{\lambda}(\mathbf{w}, S) = \min_{\mathbf{w}} \lambda \|\mathbf{w}\|^{2} + \sum_{i=1}^{l} (y_{i} - g(x_{i}))^{2},$$

where  $\lambda$  defines the trade-off between norm and loss. This controls the complexity of the model (the process is called *regularization*).

A Kernel Pattern Analysis Algorithm

Primal linear regression

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data • Taking the derivative and making it equal to zero:

$$X'Xw + \lambda w = (X'X + \lambda I_n)w = X'y,$$

where  $I_n$  is an identity matrix of  $n \times n$  dimension,

• then,

$$w = (X'X + \lambda I_n)^{-1}X'y.$$

• In terms of  $\alpha$ :

$$w = \lambda^{-1}X'(y - Xw) = X'\alpha,$$

$$\alpha = \lambda^{-1}(y - Xw) = (XX' + \lambda I_l)^{-1}y.$$

Dual linear regression

Taking the derivative and making it equal to zero:

$$X'Xw + \lambda w = (X'X + \lambda I_n)w = X'y,$$

where  $I_n$  is an identity matrix of  $n \times n$  dimension,

then,

$$w = (X'X + \lambda I_n)^{-1}X'y.$$

• In terms of  $\alpha$ :

$$w = \lambda^{-1}X'(y - Xw) = X'\alpha,$$

$$\alpha = \lambda^{-1}(y - Xw) = (XX' + \lambda I_l)^{-1}y.$$

Kernels in Complex Structured • Taking the derivative and making it equal to zero:

$$X'Xw + \lambda w = (X'X + \lambda I_n)w = X'y,$$

where  $I_n$  is an identity matrix of  $n \times n$  dimension,

then,

$$w = (X'X + \lambda I_n)^{-1}X'y.$$

• In terms of  $\alpha$ :

$$w = \lambda^{-1} X'(y - Xw) = X'\alpha,$$

$$\alpha = \lambda^{-1}(y - Xw) = (XX' + \lambda I_l)^{-1}y.$$

Dual linear regression

Taking the derivative and making it equal to zero:

$$X'Xw + \lambda w = (X'X + \lambda I_n)w = X'y,$$

where  $I_n$  is an identity matrix of  $n \times n$  dimension,

then,

$$w = (X'X + \lambda I_n)^{-1}X'y.$$

• In terms of  $\alpha$ :

$$w = \lambda^{-1} X'(y - Xw) = X'\alpha,$$

$$\alpha = \lambda^{-1}(y - Xw) = (XX' + \lambda I_l)^{-1}y.$$

A Kernel Pattern Analysis

Primal linea

Dual linear regression

Kernel

Kernel

Kernels in Complex Structured

#### Prediction function

$$g(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle = \left\langle \sum_{i=1}^{l} \alpha_i \mathbf{x}_i, \mathbf{x} \right\rangle = \sum_{i=1}^{l} \alpha_i \left\langle \mathbf{x}_i, \mathbf{x} \right\rangle$$

Introductio

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Primal linear regression

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

# Ridge regression as a kernel method

• The Gram matrix G = XX' is the matrix of dot products

$$G = XX' = \left[ \begin{array}{c} x'_1 \\ \vdots \\ x'_l \end{array} \right] [x_1 \cdots x_l] = \left[ \begin{array}{cc} \langle x_1, x_1 \rangle & \langle x_1, x_l \rangle \\ \\ \langle x_l, x_1 \rangle & \langle x_l, x_l \rangle \end{array} \right]$$

- G may be replaced by a general kernel matrix, K, with  $k_{ii} = k(x_i, x_i) = \langle \phi(x_i), \phi(x_i) \rangle$
- The  $\alpha$ 's are calculated as:

$$\alpha = (\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \mathbf{y}$$

$$g(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}, \mathbf{x}_i) = y'(\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_l) \end{bmatrix}$$

Introductio

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Primal li

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

# Ridge regression as a kernel method

• The Gram matrix G = XX' is the matrix of dot products

$$G = XX' = \begin{bmatrix} x'_1 \\ \vdots \\ x'_l \end{bmatrix} [x_1 \cdots x_l] = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_l \rangle \\ \langle x_l, x_1 \rangle & \langle x_l, x_l \rangle \end{bmatrix}$$

- G may be replaced by a general kernel matrix, K, with  $k_{ij}=k(x_i,x_i)=<\phi(x_i),\phi(x_i)>$
- The  $\alpha$ 's are calculated as:

$$\alpha = (\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \mathbf{y}$$

$$g(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}, \mathbf{x}_i) = y'(\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_l) \end{bmatrix}$$

Introduction

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Primal li

Dual linear regression

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

# Ridge regression as a kernel method

• The Gram matrix G = XX' is the matrix of dot products

$$G = XX' = \left[ \begin{array}{c} x'_1 \\ \vdots \\ x'_l \end{array} \right] [x_1 \cdots x_l] = \left[ \begin{array}{cc} \langle x_1, x_1 \rangle & \langle x_1, x_l \rangle \\ \\ \langle x_l, x_1 \rangle & \langle x_l, x_l \rangle \end{array} \right]$$

- G may be replaced by a general kernel matrix, K, with  $k_{ij}=k(x_i,x_i)=<\phi(x_i),\phi(x_i)>$
- The  $\alpha$ 's are calculated as:

$$\alpha = (K + \lambda I_l)^{-1} y$$

$$g(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}, \mathbf{x}_i) = y'(\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_l) \end{bmatrix}$$

Dual linear regression

## Ridge regression as a kernel method

• The Gram matrix G = XX' is the matrix of dot products

$$G = XX' = \left[ \begin{array}{c} x'_1 \\ \vdots \\ x'_l \end{array} \right] [x_1 \cdots x_l] = \left[ \begin{array}{cc} \langle x_1, x_1 \rangle & \langle x_1, x_l \rangle \\ \\ \langle x_l, x_1 \rangle & \langle x_l, x_l \rangle \end{array} \right]$$

- G may be replaced by a general kernel matrix, K, with  $k_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_i) \rangle$
- The α's are calculated as:

$$\alpha = (K + \lambda I_l)^{-1} y$$

$$g(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}, \mathbf{x}_i) = y'(\mathbf{K} + \lambda \mathbf{I}_l)^{-1} \begin{bmatrix} k(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ k(\mathbf{x}, \mathbf{x}_l) \end{bmatrix}$$

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

#### Kernel Functions

Mathematical characterisation
Visualizing kernels i

Kernel Algorithms

Kernels ir Complex Structure Data

### Outline

1 Introduction
Motivation

2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- **6** Kernel Functions

Mathematical characterisation Visualizing kernels in input space

- 6 Kernel Algorithms
- Mernels in Complex Structured Data

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

### Mathematical characterisation

Visualizing kernel

Kernel Algorithms

Kernels ir Complex Structure Data

## Outline

- 1 Introduction
  Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

#### Characterisation

#### Theorem

(Mercer's Theorem) A function

$$k: X \times X \to \mathbb{R}$$
,

which is either continuous or has a countable domain, can be decomposed

$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle$$

into a feature map  $\phi$  into a Hilbert space F applied to both its arguments followed by the evaluation of the inner product in Fif and only if it satisfies the finitely positive semi-definite property.

Kernels in Complex Structured

#### Some kernel functions

- $k(x, z) = p(k_1(x, z))$ . p a polynomial with positive coefficients.
- $k(x, z) = \exp(k_1(x, z))$ .
- $k(x, z) = \exp(-\|x z\|^2/(2\sigma^2))$ . Gaussian kernel.
- $k(x, z) = k_1(x, z)k_2(x, z)$

Kernels ir Complex Structure

#### Some kernel functions

- $k(x, z) = p(k_1(x, z))$ . p a polynomial with positive coefficients.
- $k(x, z) = \exp(k_1(x, z))$ .
- $k(x, z) = \exp(-\|x z\|^2/(2\sigma^2))$ . Gaussian kernel.
- $k(x, z) = k_1(x, z)k_2(x, z)$

A Kernel Pattern Analysis Algorithm

Kernel Function

Mathematical characterisation

Visualizing kernel

Kernel Algorithms

Kernels in Complex Structured

#### Some kernel functions

- $k(x, z) = p(k_1(x, z))$ . p a polynomial with positive coefficients.
- $k(x, z) = \exp(k_1(x, z))$ .
- $k(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} \mathbf{z}\|^2/(2\sigma^2))$ . Gaussian kernel.
- $k(x, z) = k_1(x, z)k_2(x, z)$

#### Some kernel functions

- $k(x, z) = p(k_1(x, z))$ . p a polynomial with positive coefficients.
- $k(x, z) = \exp(k_1(x, z)).$
- $k(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} \mathbf{z}\|^2/(2\sigma^2))$ . Gaussian kernel.
- $\bullet k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) k_2(\mathbf{x}, \mathbf{z})$

A Kernel Pattern Analysis Algorithm

Kernel Function

Mathematical

characterisation
Visualizing kernels

Kernel Algorithm

Kernels in Complex Structured

# Embeddings corresponding to kernels

- It is possible to calculate the feature space induced by a kernel (Mercer's Theorem)
- This can be done in a constructive way
- The feature space can even be of infinite dimension.

A Kernel Pattern Analysis Algorithm

Kernel Functions

Mathematical characterisation

Visualizing kernels input space

Kernel Algorithm

Kernels in Complex Structure

# Embeddings corresponding to kernels

- It is possible to calculate the feature space induced by a kernel (Mercer's Theorem)
- This can be done in a constructive way
- The feature space can even be of infinite dimension.

A Kernel Pattern Analysis Algorithm

Kernel Functions

Functions Mathematical

characterisation Visualizing kernel

Kernel Algorithms

Kernels in Complex Structured

# Embeddings corresponding to kernels

- It is possible to calculate the feature space induced by a kernel (Mercer's Theorem)
- This can be done in a constructive way
- The feature space can even be of infinite dimension.

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Mathematical characterisation

characterisation Visualizing kernels in input space

Kernel Algorithm

Kernels in Complex Structure Data

#### Outline

1 Introduction
Motivation

2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space

- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- **5** Kernel Functions

Mathematical characterisation Visualizing kernels in input space

- 6 Kernel Algorithms
- 7 Kernels in Complex Structured Data

Kernels in Complex Structured Data

#### How to visualize?

- Choose a point in input space  $p_0$
- Calculate the distance from another point x to  $p_0$  in the feature space:

$$\begin{aligned} \|\phi(p_0) - \phi(x)\|_F^2 &= \langle \phi(p_0) - \phi(x), \phi(p_0) - \phi(x) \rangle_F \\ &= \langle \phi(p_0), \phi(p_0) \rangle_F + \langle \phi(x), \phi(x) \rangle \\ &- 2 \langle \phi(p_0), \phi(x) \rangle_F \\ &= k(p_0, p_0) + k(x, x) - 2k(p_0, x) \end{aligned}$$

• Plot 
$$f(x) = \|\phi(p_0) - \phi(x)\|_F^2$$

A Kernel Pattern Analysis Algorithm

Kernel Functions

Mathematic

Visualizing kernels in input space

Kernel Algorithms

Kernels in Complex Structured Data

#### How to visualize?

- Choose a point in input space  $p_0$
- Calculate the distance from another point x to  $p_0$  in the feature space:

$$\begin{split} \|\phi(p_0) - \phi(x)\|_F^2 &= \langle \phi(p_0) - \phi(x), \phi(p_0) - \phi(x) \rangle_F \\ &= \langle \phi(p_0), \phi(p_0) \rangle_F + \langle \phi(x), \phi(x) \rangle_F \\ &- 2 \langle \phi(p_0), \phi(x) \rangle_F \\ &= k(p_0, p_0) + k(x, x) - 2k(p_0, x) \end{split}$$

• Plot 
$$f(x) = \|\phi(p_0) - \phi(x)\|_F^2$$

A Kernel Pattern Analysis Algorithm

Kernel Functions

Mathematic

Visualizing kernels in input space

Kernel Algorithms

Kernels in Complex Structured Data

#### How to visualize?

- Choose a point in input space  $p_0$
- Calculate the distance from another point x to  $p_0$  in the feature space:

$$\begin{split} \|\phi(p_0) - \phi(x)\|_F^2 &= \langle \phi(p_0) - \phi(x), \phi(p_0) - \phi(x) \rangle_F \\ &= \langle \phi(p_0), \phi(p_0) \rangle_F + \langle \phi(x), \phi(x) \rangle_F \\ &- 2 \langle \phi(p_0), \phi(x) \rangle_F \\ &= k(p_0, p_0) + k(x, x) - 2k(p_0, x) \end{split}$$

• Plot  $f(x) = \|\phi(p_0) - \phi(x)\|_F^2$ 

The Kernel Approach to Machine

A Kernel Pattern Analysis Algorithm

Kernel Function

Mathemati

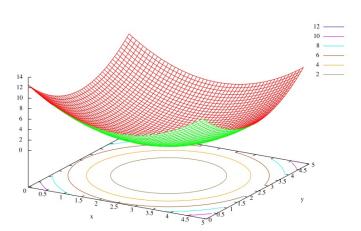
Visualizing kernels in

input space
Kernel

Kernels ir Complex Structure

## Identity kernel

$$k(x,z) = \langle x, z \rangle$$



A Kernel Pattern Analysis Algorithm

Kernel Function

Mathematic

characterisation

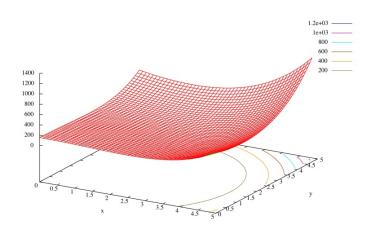
Visualizing kernels in input space

Kernel

Kernels in Complex Structure

# Quadratic kernel (1)

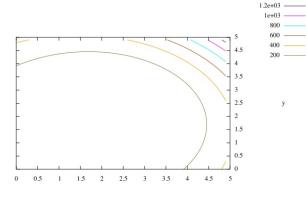
$$k(x,z) = \langle x, z \rangle^2$$



Visualizing kernels in input space

# Identity kernel (2)

$$k(x,z) = \langle x, z \rangle^2$$



800 600

200

y

The Kernel Approach to Machine

A Kernel Pattern Analysis Algorithm

Kernel Function

Mathematic

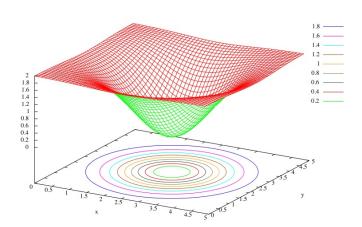
Visualizing kernels in input space

Kernel Algorithm

Kernels in Complex Structure

## Gaussian kernel

$$k(\mathbf{x}, \mathbf{z}) = e^{-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}}$$



#### F. González

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

#### Kernel Algorithms

Kernels in Complex Structured

### Outline

- 1 Introduction
  Motivation
- 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm
  Primal linear regression
  Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- 6 Kernel Algorithms
- 7 Kernels in Complex Structured Data

Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Means
- Distances
- Projections
- Covariance

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

- Means
- Distances
- Projections
- Covariance

The Kerne

The Kernel Approach t Machine Learning

Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Means
- Distances
- Projections
  - Covariance

F. González

Introductio

The Kerne Trick

The Kernel Approach t Machine Learning

Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Means
- Distances
- Projections
- Covariance

Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured

- Support Vector Machines
- Support Vector Regression
- Kernel Fisher Discriminant
- Kernel Perceptron

#### F. González

Introductio

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Support Vector Machines
- Support Vector Regression
- Kernel Fisher Discriminant
- Kernel Perceptron

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

- Support Vector Machines
- Support Vector Regression
- Kernel Fisher Discriminant
- Kernel Perceptron

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Support Vector Machines
- Support Vector Regression
- Kernel Fisher Discriminant
- Kernel Perceptron

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

- Kernel PCA
- Kernel CCA
- Kernel k-means
- Kernel SOM

Introductio

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Kernel PCA
- Kernel CCA
- Kernel k-means
- Kernel SOM

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured

- Kernel PCA
- Kernel CCA
- Kernel k-means
- Kernel SOM

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

- Kernel PCA
- Kernel CCA
- Kernel k-means
- Kernel SOM

#### F. González

Introduction

The Kerne Trick

The Kernel Approach t Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithm

Kernels in Complex Structured

## Outline

- 1 Introduction
  Motivation
  - 2 The Kernel Trick Mapping the input space to the feature space Calculating the dot product in the feature space
- 3 The Kernel Approach to Machine Learning
- 4 A Kernel Pattern Analysis Algorithm Primal linear regression Dual linear regression
- 6 Kernel Functions Mathematical characterisation Visualizing kernels in input space
- **6** Kernel Algorithms
- 7 Kernels in Complex Structured Data

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithm

Kernels in Complex Structured Data

## Kernels in complex structured data

- Since kernel methods do not require an attribute-based representation of objects, it is possible to perform learning over complex structured data (or unstructured data)
- We only need to define a dot product operation (similarity, dissimilarity measure)
- Examples:
  - Strings
  - Texts
  - Trees
  - Graphs

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithm

Kernels in Complex Structured Data

## Kernels in complex structured data

- Since kernel methods do not require an attribute-based representation of objects, it is possible to perform learning over complex structured data (or unstructured data)
- We only need to define a dot product operation (similarity, dissimilarity measure)
- Examples:
  - Strings
  - Texts
  - Trees
  - Graphs

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithm

Kernels in Complex Structured Data

## Kernels in complex structured data

- Since kernel methods do not require an attribute-based representation of objects, it is possible to perform learning over complex structured data (or unstructured data)
- We only need to define a dot product operation (similarity, dissimilarity measure)
- Examples:
  - Strings
  - Texts
  - Trees
  - Graphs

Introduction

The Kerne Trick

The Kernel Approach to Machine Learning

A Kernel Pattern Analysis Algorithm

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured How to do symbolic regression?

$$\Sigma = \{A,\,C,\,G,\,T\}$$

$$\begin{array}{ccccc} f: & \Sigma^d & \rightarrow & \mathbb{R} \\ & ACGTA & \mapsto & 10.0 \\ & GTCCA & \mapsto & 11.3 \\ & GGTAC & \mapsto & 1.0 \\ & CCTGA & \mapsto & 4.5 \\ & \vdots & \vdots & \vdots \end{array}$$

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured Data

## Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - a function that measures string similarity
  - higher value for similar strings, smaller value for different strings

• 
$$k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n cqual(s_i, t_i)$$
  
 $cqual(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$ 

- k(ACTAG, CCTCG) = ?
- Is it a kernel?

### Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - a function that measures string similarity
  - higher value for similar strings, smaller value for different strings
  - $k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n cqual(s_i, t_i)$ •  $equal(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$
- k(ACTAG, CCTCG) = ?
- Is it a kernel?

Kernels in Complex Structured Data

### Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - · a function that measures string similarity
  - higher value for similar strings, smaller value for different strings

• 
$$k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n equal(s_i, t_i)$$
  
 $equal(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$ 

- k(ACTAG, CCTCG) = ?
- Is it a kernel?

Kernel Function

Kernel Algorithms

Kernels in Complex Structured Data

### Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - · a function that measures string similarity
  - higher value for similar strings, smaller value for different strings

• 
$$k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n equal(s_i, t_i)$$
  
 $equal(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$ 

- k(ACTAG, CCTCG) = ?
- Is it a kernel?

Kernels in Complex Structured Data

### Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - a function that measures string similarity
  - higher value for similar strings, smaller value for different strings

• 
$$k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n equal(s_i, t_i)$$
  
 $equal(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$ 

- k(ACTAG, CCTCG) = ?
- Is it a kernel?

Kernels in Complex Structured Data

### Solution

$$k: \Sigma^d \times \Sigma^d \to \mathbb{R}$$

- Use the kernel along with a kernel learning regression algorithm to find the regression function
- What is a good candidate for k?
  - a function that measures string similarity
  - higher value for similar strings, smaller value for different strings

• 
$$k(s_1 \dots s_d, t_1 \dots t_d) = \sum_{i=1}^n equal(s_i, t_i)$$
  
 $equal(s_i, t_i) = \begin{cases} 1 & \text{if } s_i = t_i \\ 0 & \text{otherwise} \end{cases}$ 

- k(ACTAG, CCTCG) = ?
- Is it a kernel?

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel Algorithm

Kernels in Complex Structured

## Induced Feature Space

What is the feature space induced by k?

$$\phi: \Sigma^{a} \to \mathbb{R}^{4a}$$

$$s_{1} \dots s_{d} \mapsto (x_{1}^{1}, \dots, x_{4}^{1}, x_{1}^{2}, \dots, x_{4}^{2}, \dots, x_{1}^{d}, \dots, x_{4}^{d})$$

$$(x_{1}^{j}, \dots, x_{4}^{j}) = \begin{cases} (1, 0, 0, 0) & \text{if } s_{j} = 'A' \\ (0, 1, 0, 0) & \text{if } s_{j} = 'C' \\ (0, 0, 1, 0) & \text{if } s_{j} = 'C' \end{cases}$$

Kernel Functions

Kernel Algorithms

Kernels in Complex Structured

## Induced Feature Space

- What is the feature space induced by k?
- •

$$\phi: \Sigma^{d} \to \mathbb{R}^{4d}$$

$$s_{1} \dots s_{d} \mapsto (x_{1}^{1}, \dots, x_{4}^{1}, x_{1}^{2}, \dots, x_{4}^{2}, \dots, x_{1}^{d}, \dots, x_{4}^{d})$$

$$(x_{1}^{j}, \dots, x_{4}^{j}) = \begin{cases} (1, 0, 0, 0) & \text{if } s_{j} = 'A' \\ (0, 1, 0, 0) & \text{if } s_{j} = 'C' \\ (0, 0, 1, 0) & \text{if } s_{j} = 'G' \\ (0, 0, 0, 1) & \text{if } s_{j} = 'T' \end{cases}$$

A Kernel Pattern Analysis Algorithm

Kernel Function

Kernel

Kernels in Complex Structured

### References



Shawe-Taylor, J. and Cristianini, N. 2004 Kernel Methods for Pattern Analysis. Cambridge University Press.