# Support Vector Machines

24 August 2018

Torben Tvedebrink
tvede@math.aau.dk

Data Science using R

AALBORG UNIVERSITY
DENMARK

# Support Vector Machines



The separable situation

Slack variable

The non-separable situation

Kernel trick

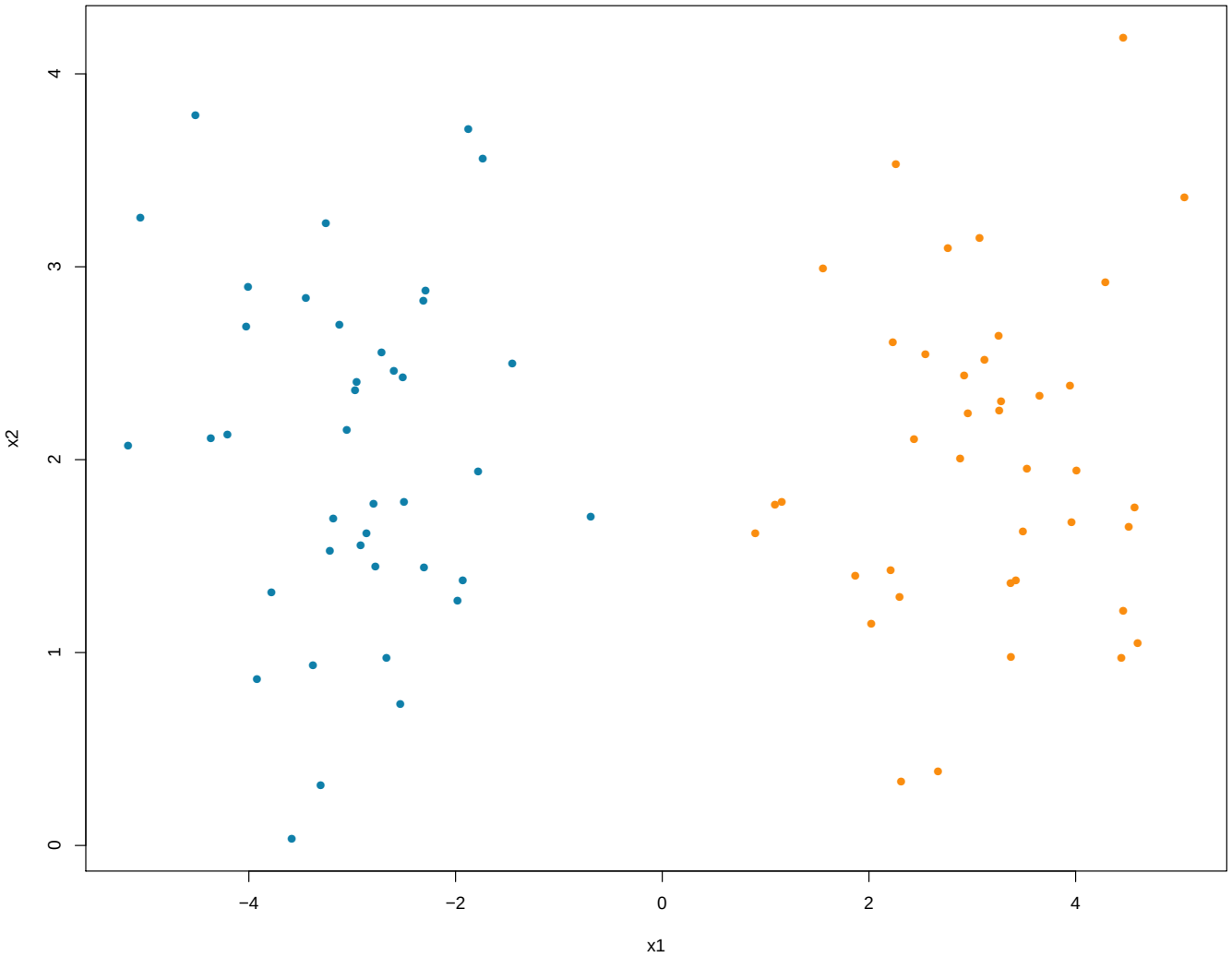
Support Vector Machines (SVM) is a type classifier that has proven itself quite efficient in many different areas of application.

We start by considering the simplest case...



Torben Tvedebrink
tvede@math.aau.dk

# Example
## Separation of two classer

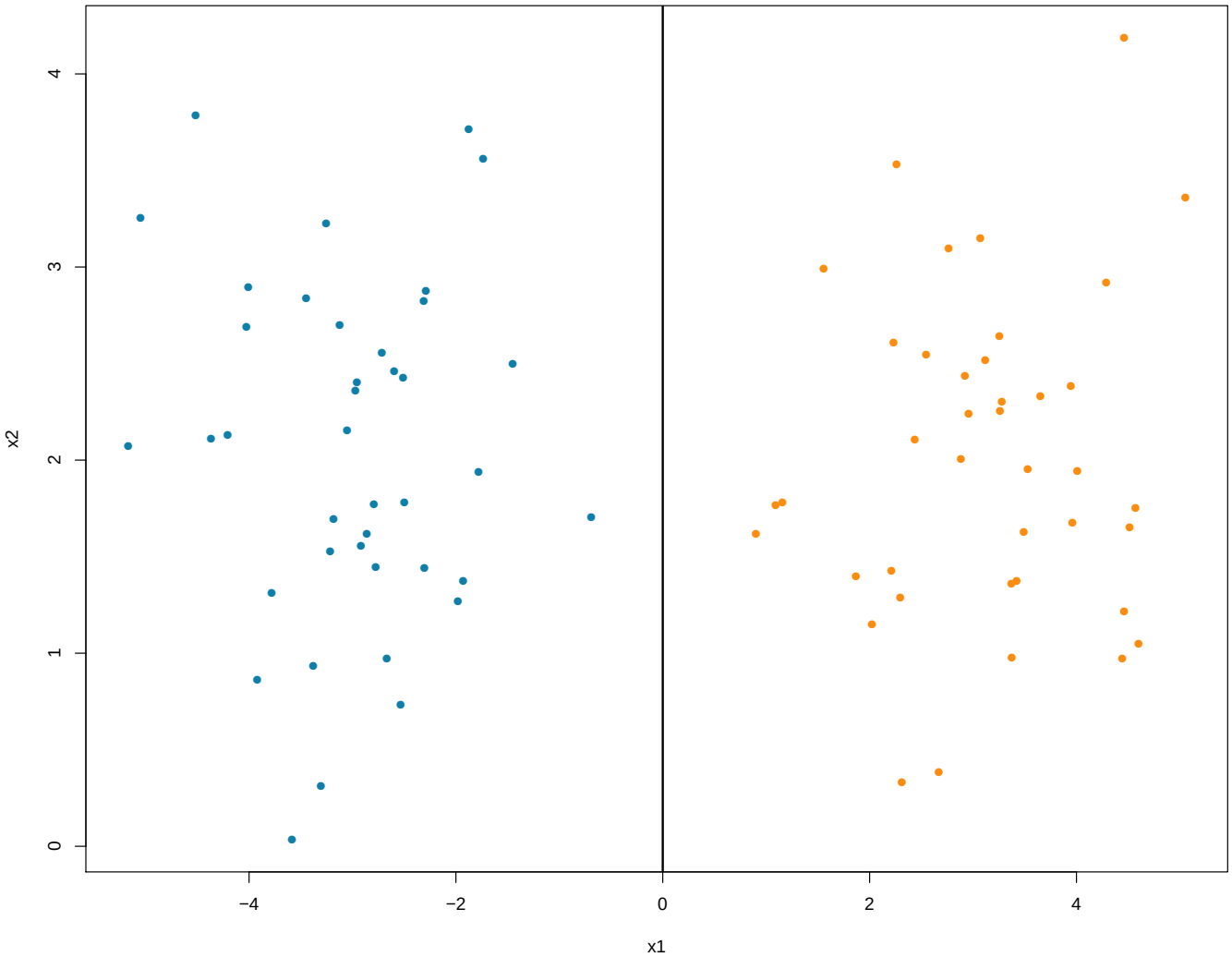





2 The separable situation
Slack variable
The non-separable situation
Kernel trick


Torben Tvedebrink
tvede@math.aau.dk

# Example
## Separation of two classer – One possible choise

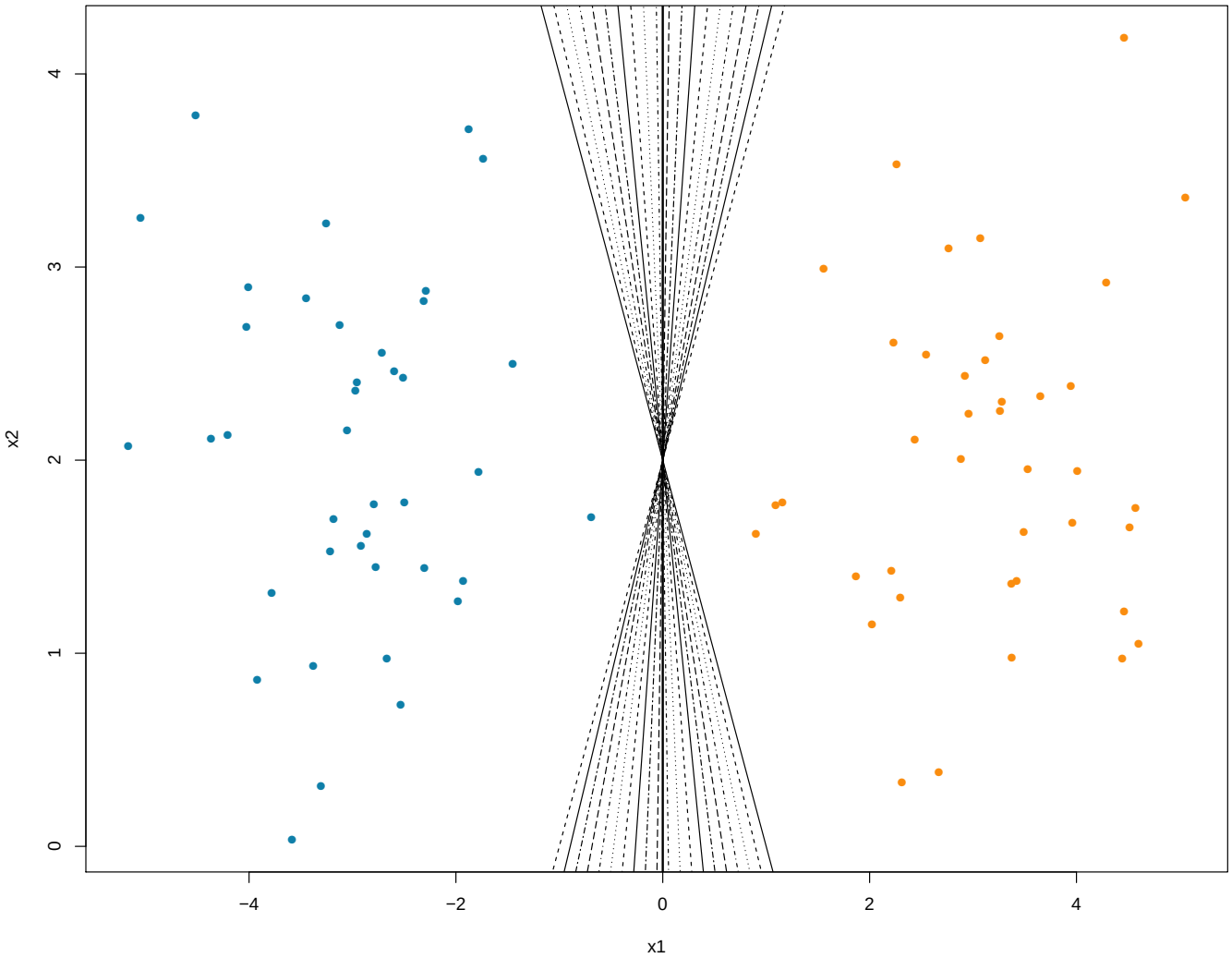





2 The separable situation
Slack variable
The non-separable situation
Kernel trick


Torben Tvedebrink
tvede@math.aau.dk

# Example
## Separation of two classer – Infinitely many options








Torben Tvedebrink
tvede@math.aau.dk

# The optimal hyper-plane



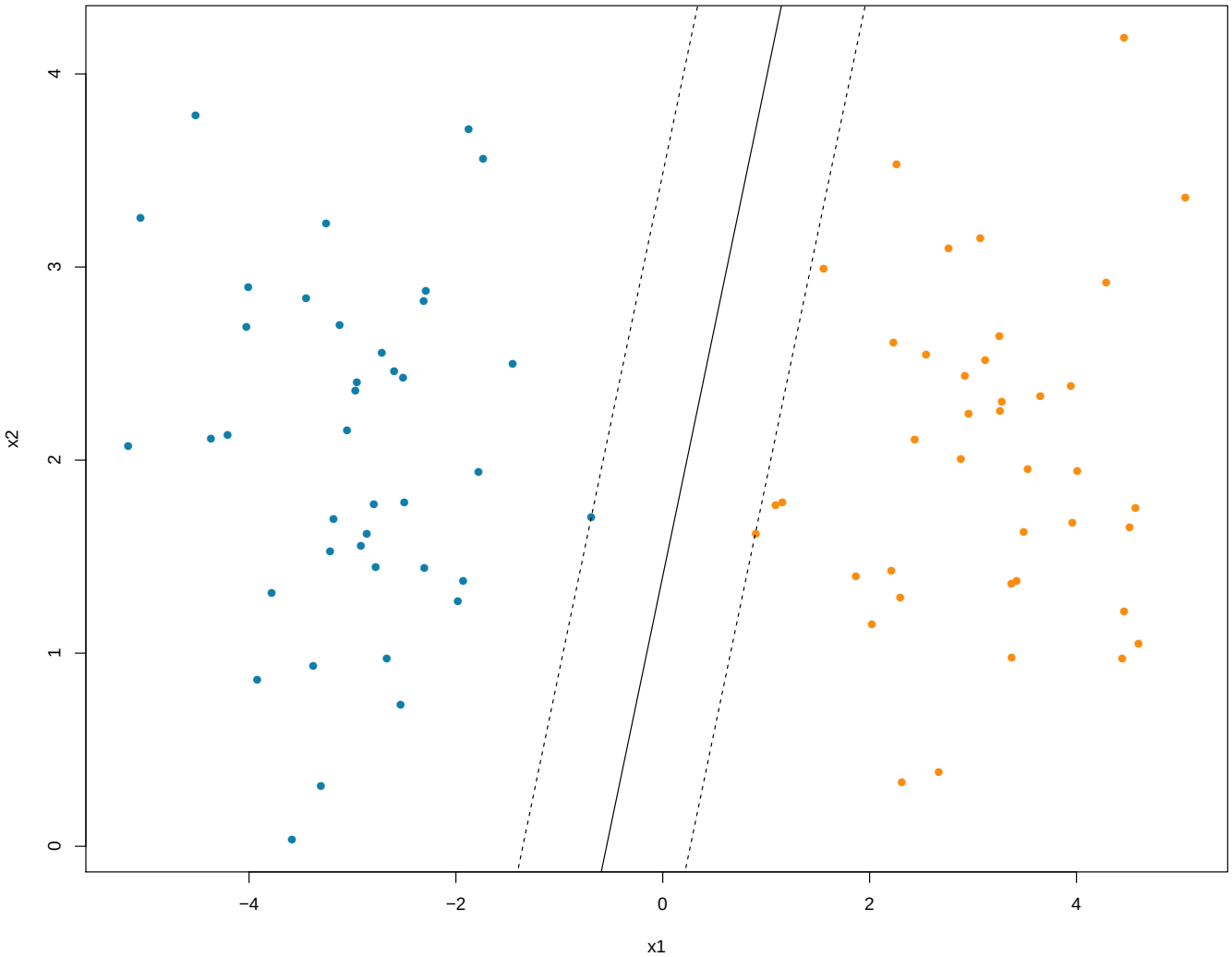
Can be uniquely defined by demanding:

a) Separates the data into two disjoint classes

b) Will have the maximal distance to the nearest points from each class


Torben Tvedebrink
tvede@math.aau.dk

# Example
## The optimal hyper-plane





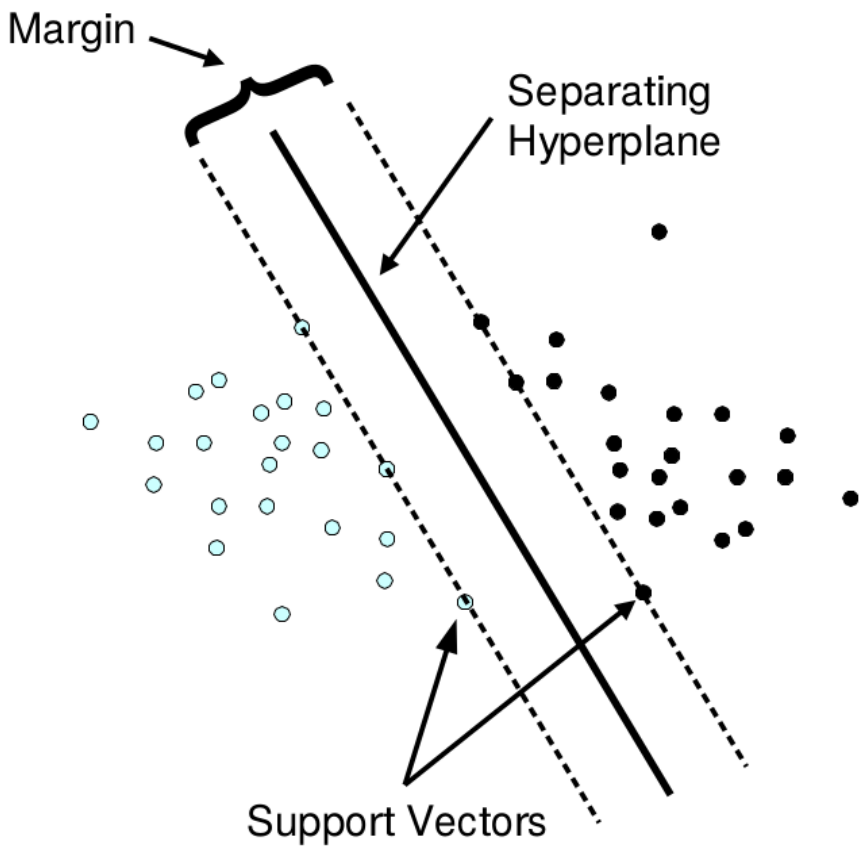

4 The separable situation
Slack variable
The non-separable situation
Kernel trick


Torben Tvedebrink
tvede@math.aau.dk

# Support Vectors

From the way the hyper-plane is computed, only the points closest to the plane have an influence on the expression:

– The **support vectors**!








Torben Tvedebrink
tvede@math.aau.dk

# Optimal for the test data?

In the previous figure, the optimal hyper-plane was highly influenced by the right-most blue observation.





Torben Tvedebrink
tvede@math.aau.dk

# Optimal for the test data?
## Slack variable



The separable situation
6 Slack variable
The non-separable situation
Kernel trick

In the previous figure, the optimal hyper-plane was highly influenced by the right-most blue observation.

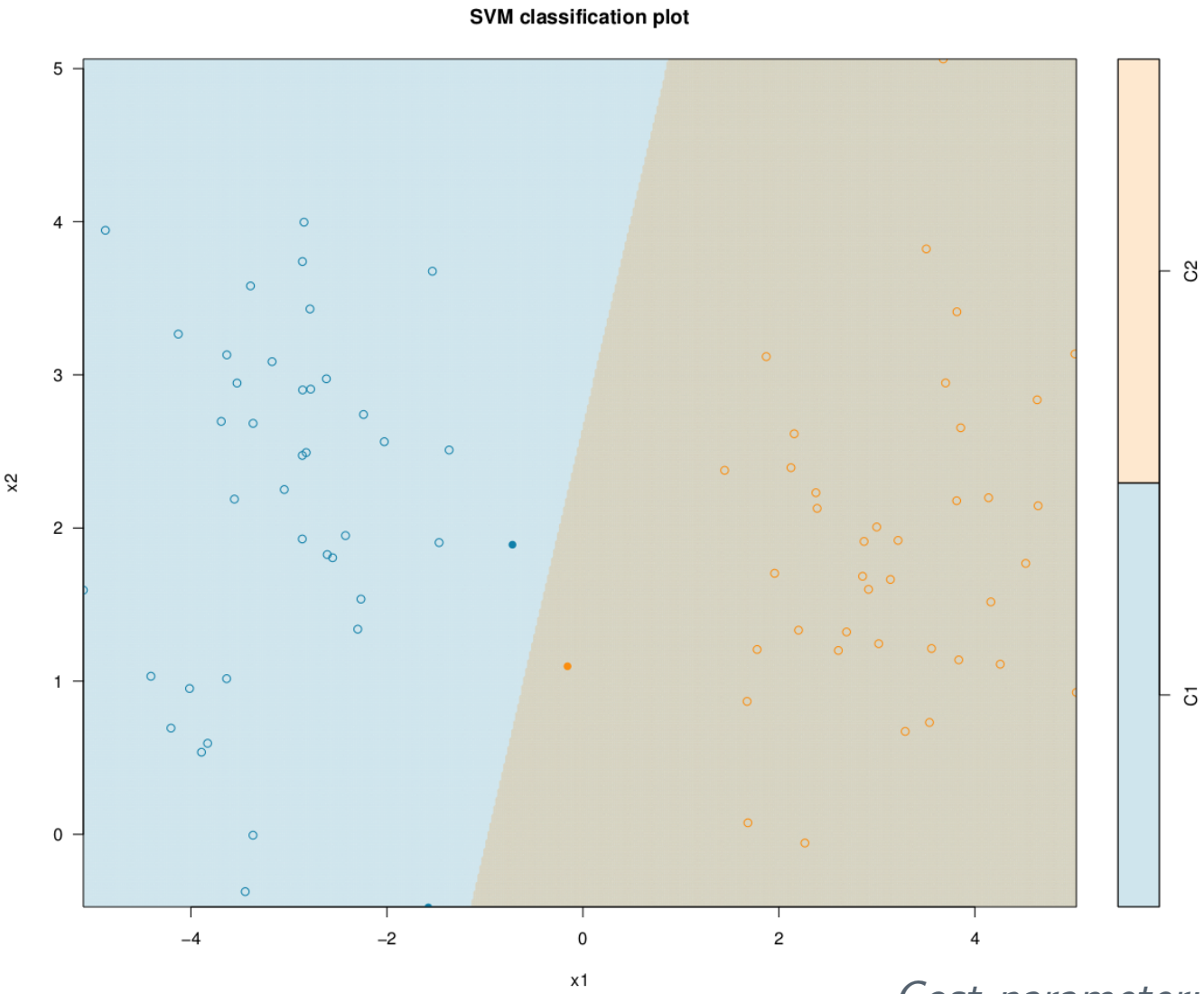
By using so-called *slack variables* we may find a hyper-plane with different properties – hopefully with lower test errors.

The purpose of *slack variables* are to allow some observations to be on the *wrong* side of the hyper-plane den.

*"A few rotten apples in the basket"*


Torben Tvedebrink
tvede@math.aau.dk

# Example
## No *slack variables*



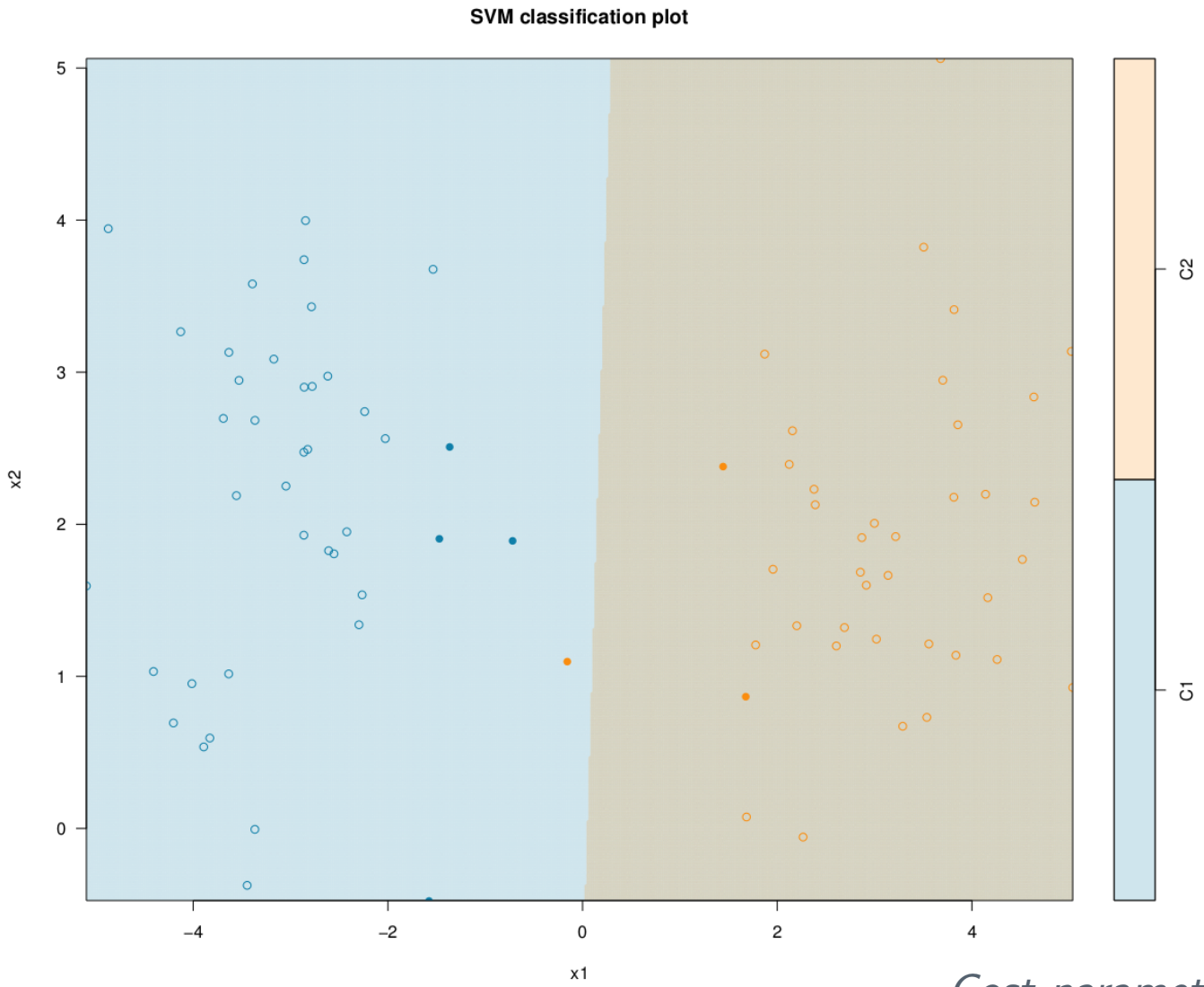

*Cost-parameter:* $10^{10}$





Torben Tvedebrink
tvede@math.aau.dk

# Example
## Using *slack variables*




*Cost-parameter:* 1





Torben Tvedebrink
tvede@math.aau.dk

# The non-separable situation




In cases where there exists no separating hyper-plane it is necessary to use slack variable – otherwise there is no solution.

The slack variables is used to measure how far the mis-classified observations are some the hyper-plane.

The separable situation
Slack variable
8 The non-separable situation
Kernel trick

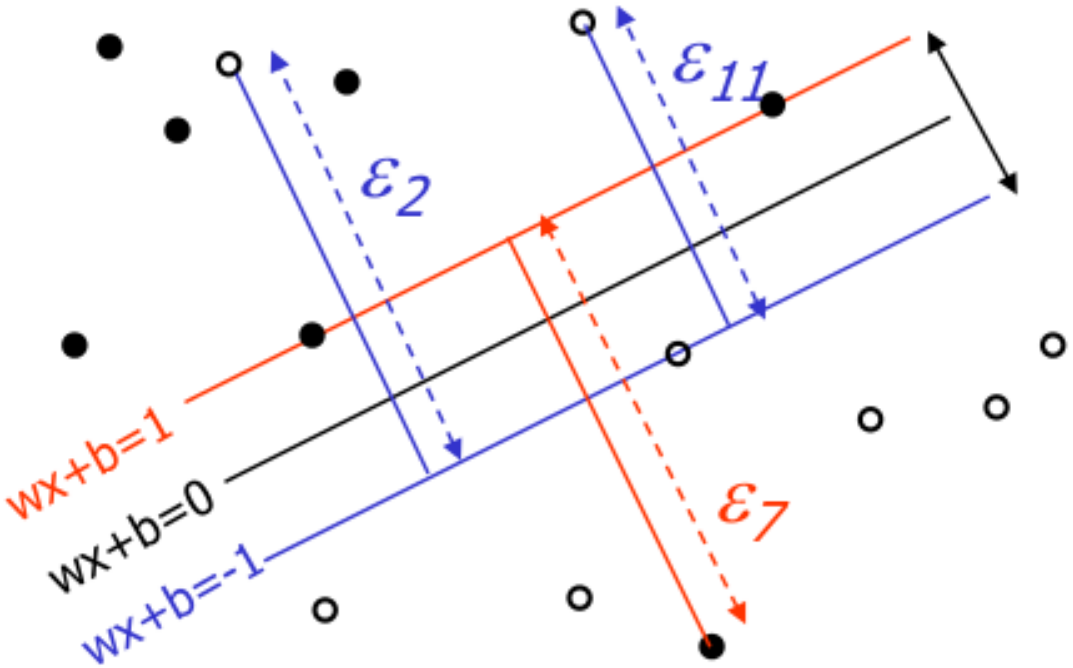

Torben Tvedebrink
tvede@math.aau.dk

# The non-separable situation



In cases where there exists no separating hyper-plane it is necessary to use slack variable – otherwise there is no solution.

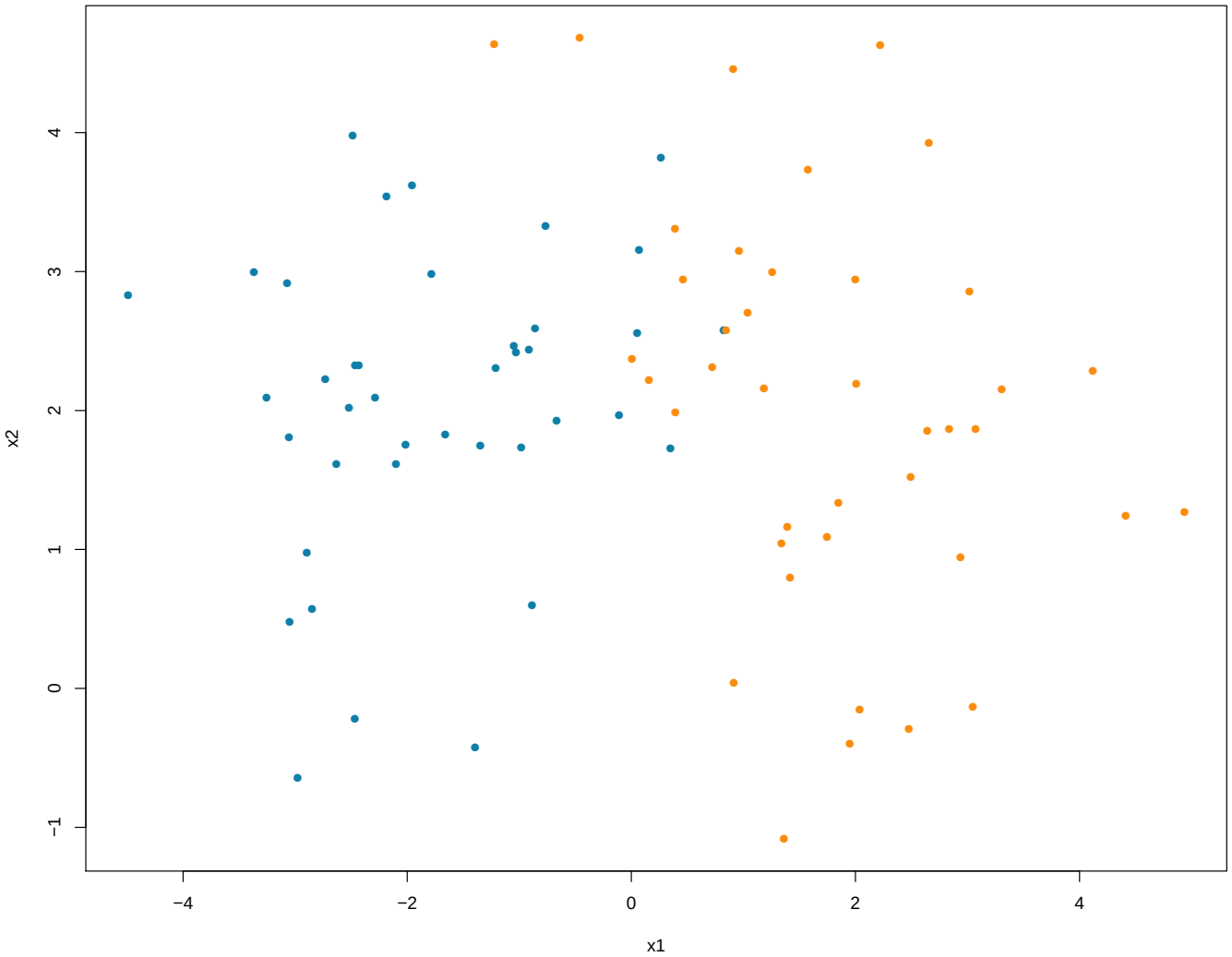
The slack variables is used to measure how far the mis-classified observations are some the hyper-plane.





Torben Tvedebrink
tvede@math.aau.dk

# Example
## Non-separable case

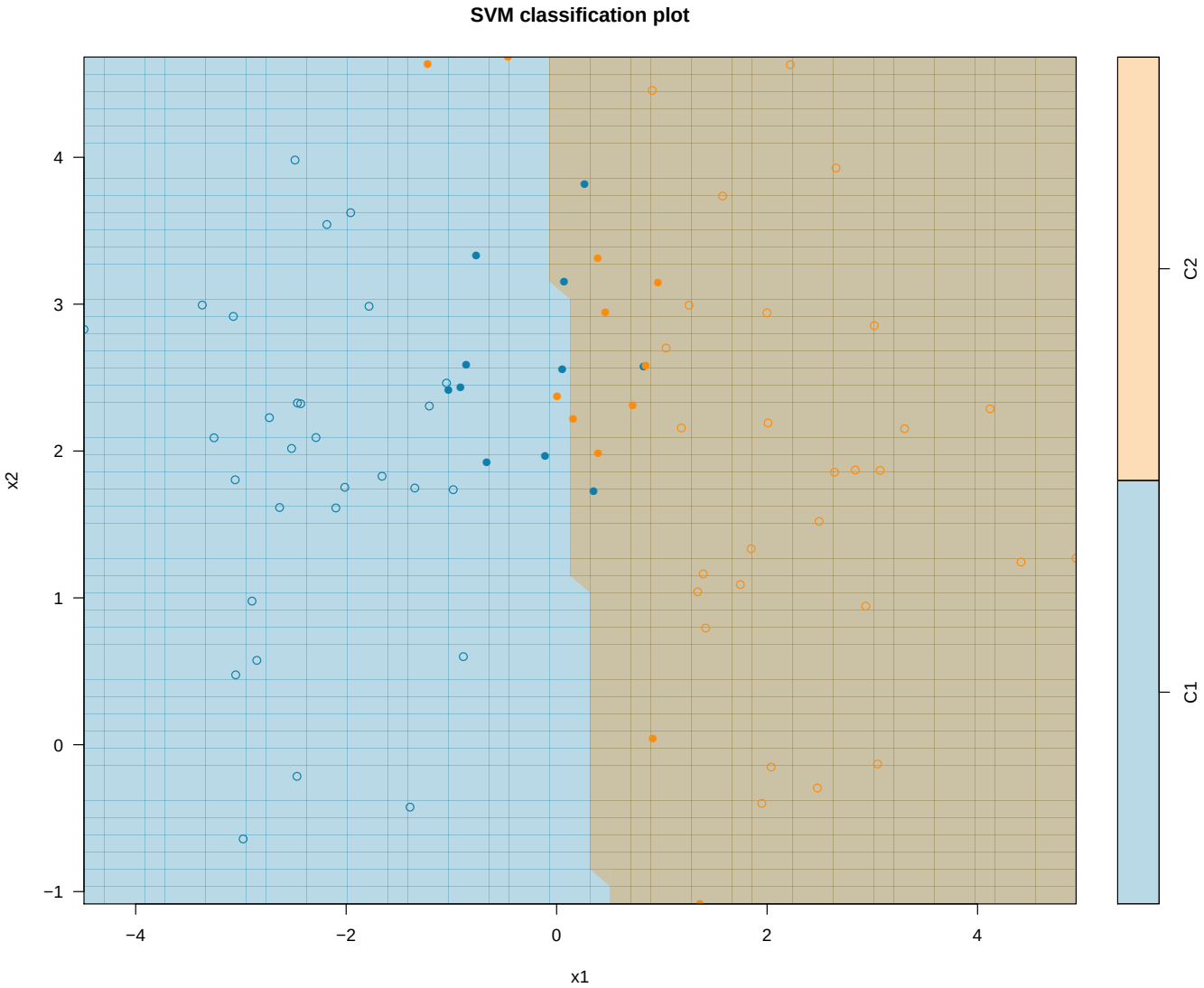







Torben Tvedebrink
tvede@math.aau.dk

# Example
## Non-separable case

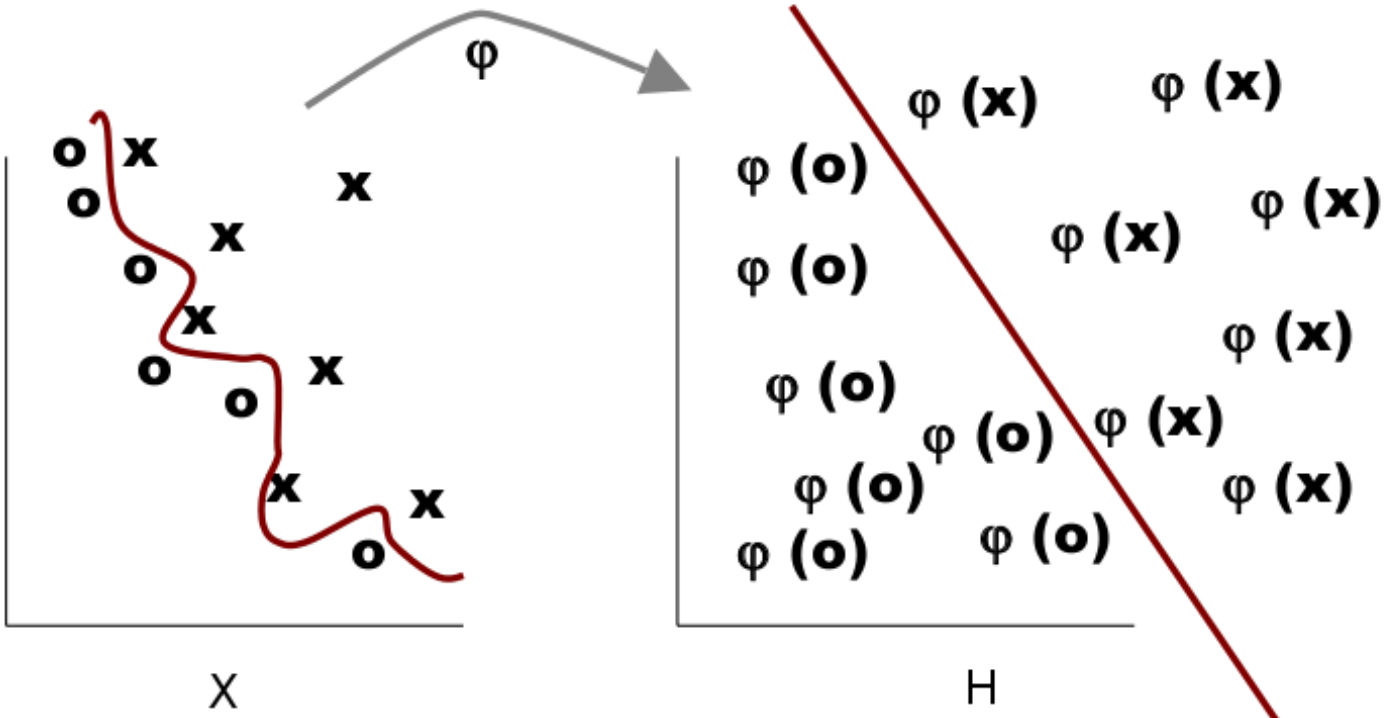







Torben Tvedebrink
tvede@math.aau.dk

# Extend/map data into a higher dimension



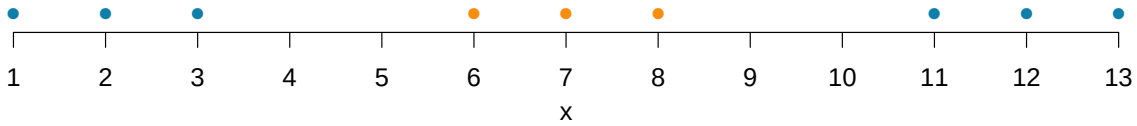





Torben Tvedebrink
tvede@math.aau.dk

# Simple examples of the concept
## Mapping into a higher dimension

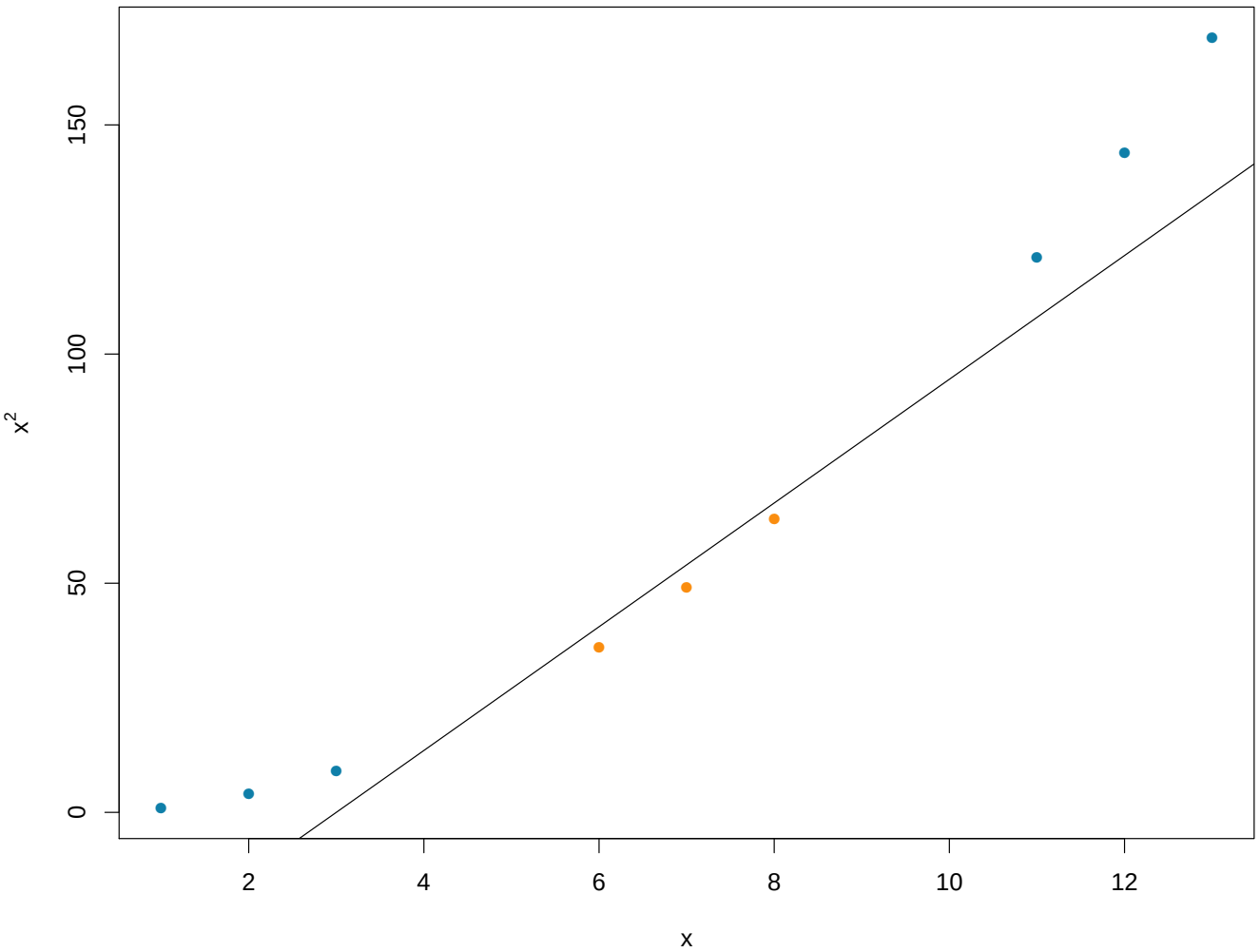






Torben Tvedebrink
tvede@math.aau.dk

# Simple examples of the concept
## Mapping into a higher dimension

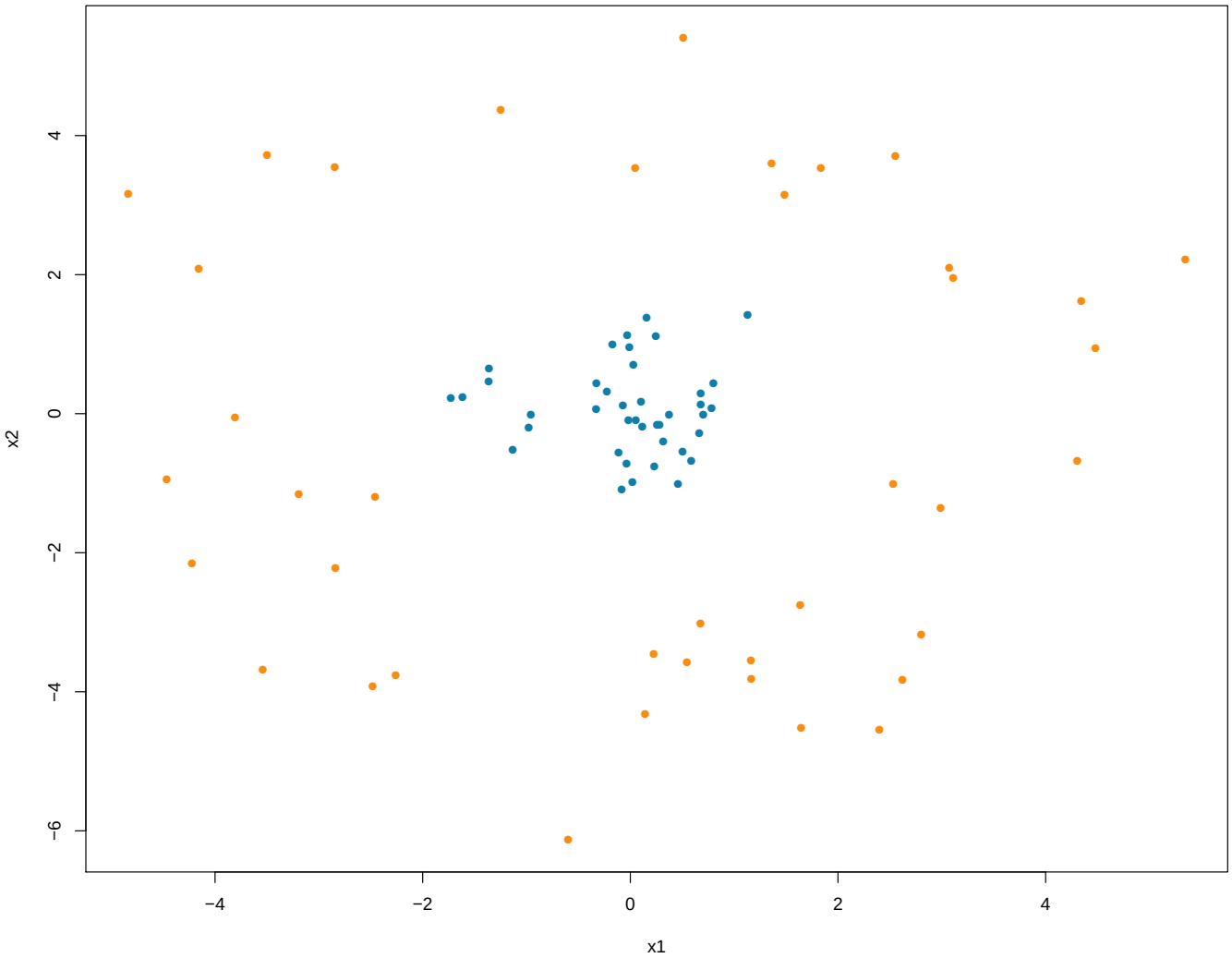







Torben Tvedebrink
tvede@math.aau.dk

# Simple examples of the concept
## Mapping into a higher dimension



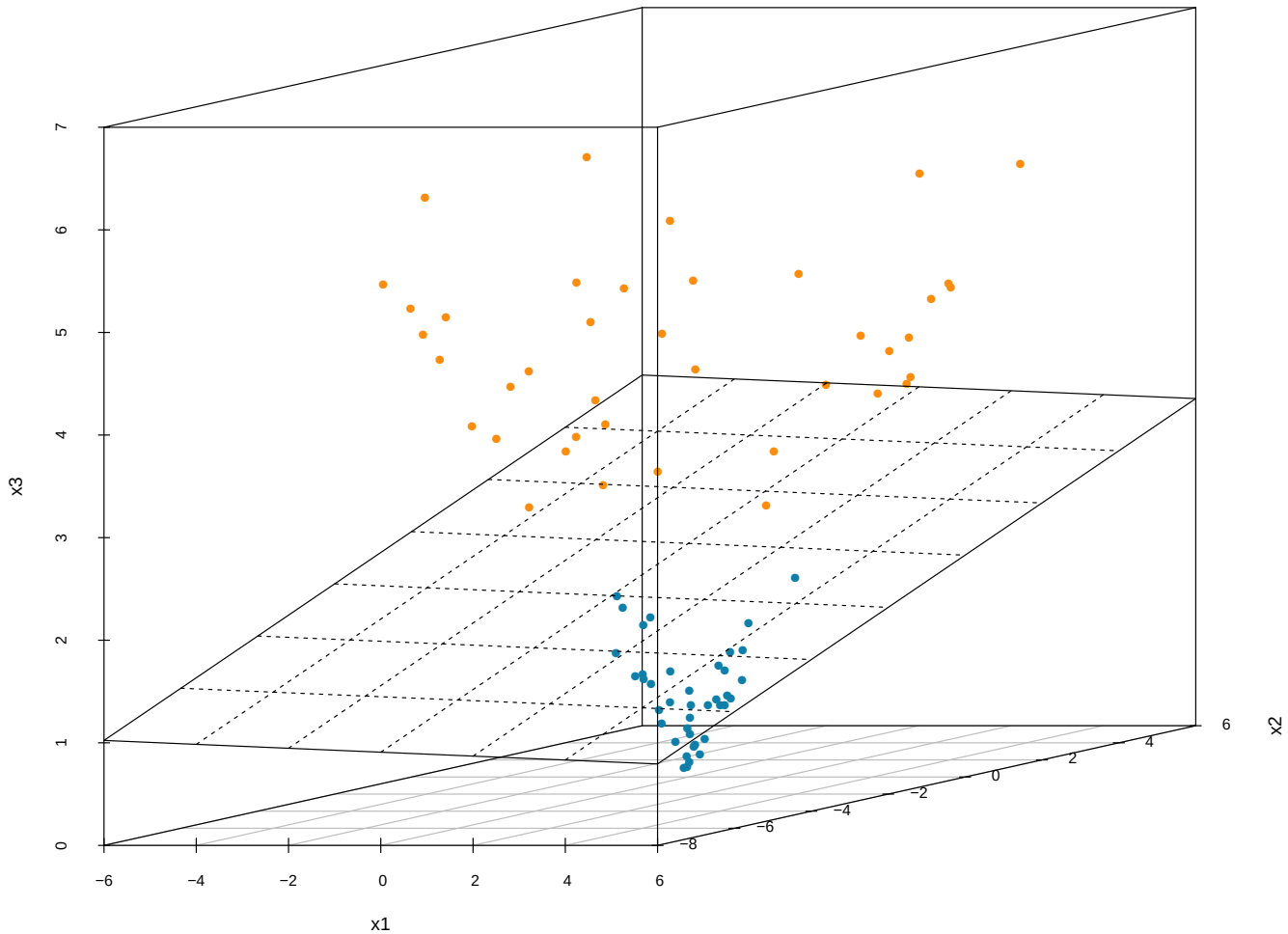



The separable situation
Slack variable
The non-separable situation
11 Kernel trick



Torben Tvedebrink
tvede@math.aau.dk

# Simple examples of the concept
## Mapping into a higher dimension

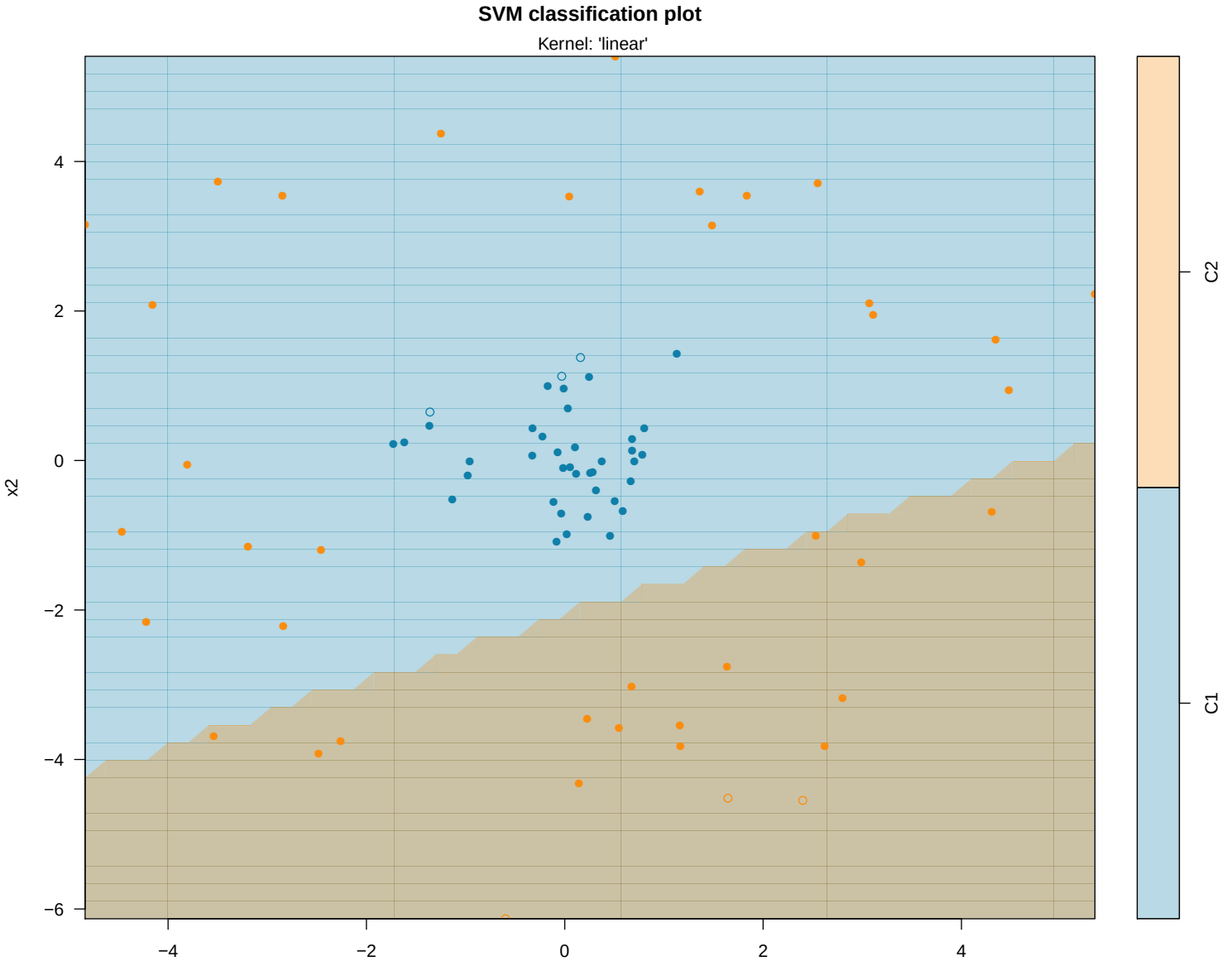







Torben Tvedebrink
tvede@math.aau.dk

# Eksempel
## Cont'f - mapping into higher dimension

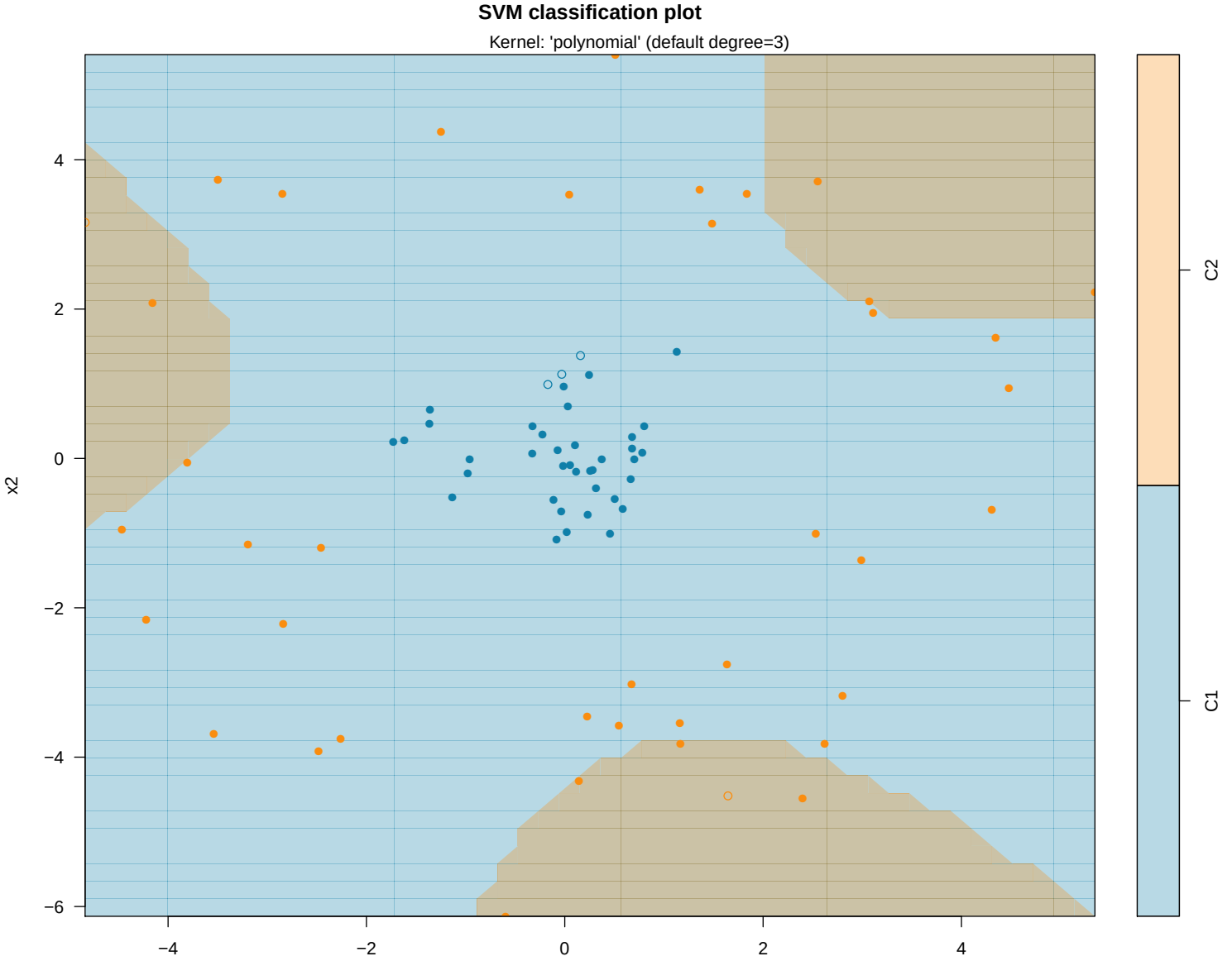







Torben Tvedebrink
tvede@math.aau.dk

# Eksempel
## Cont'f - mapping into higher dimension

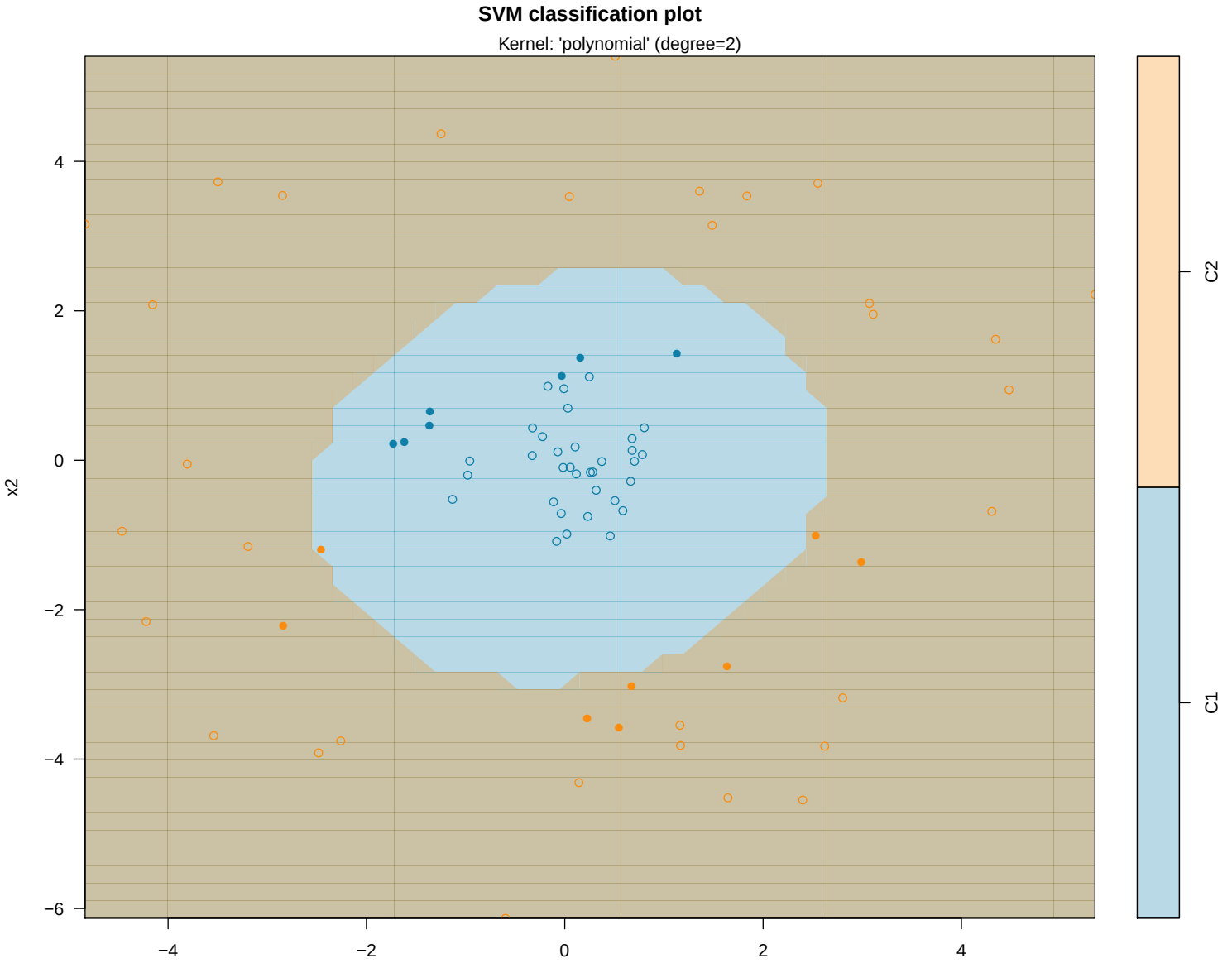







Torben Tvedebrink
tvede@math.aau.dk

# Eksempel
## Cont'f - mapping into higher dimension

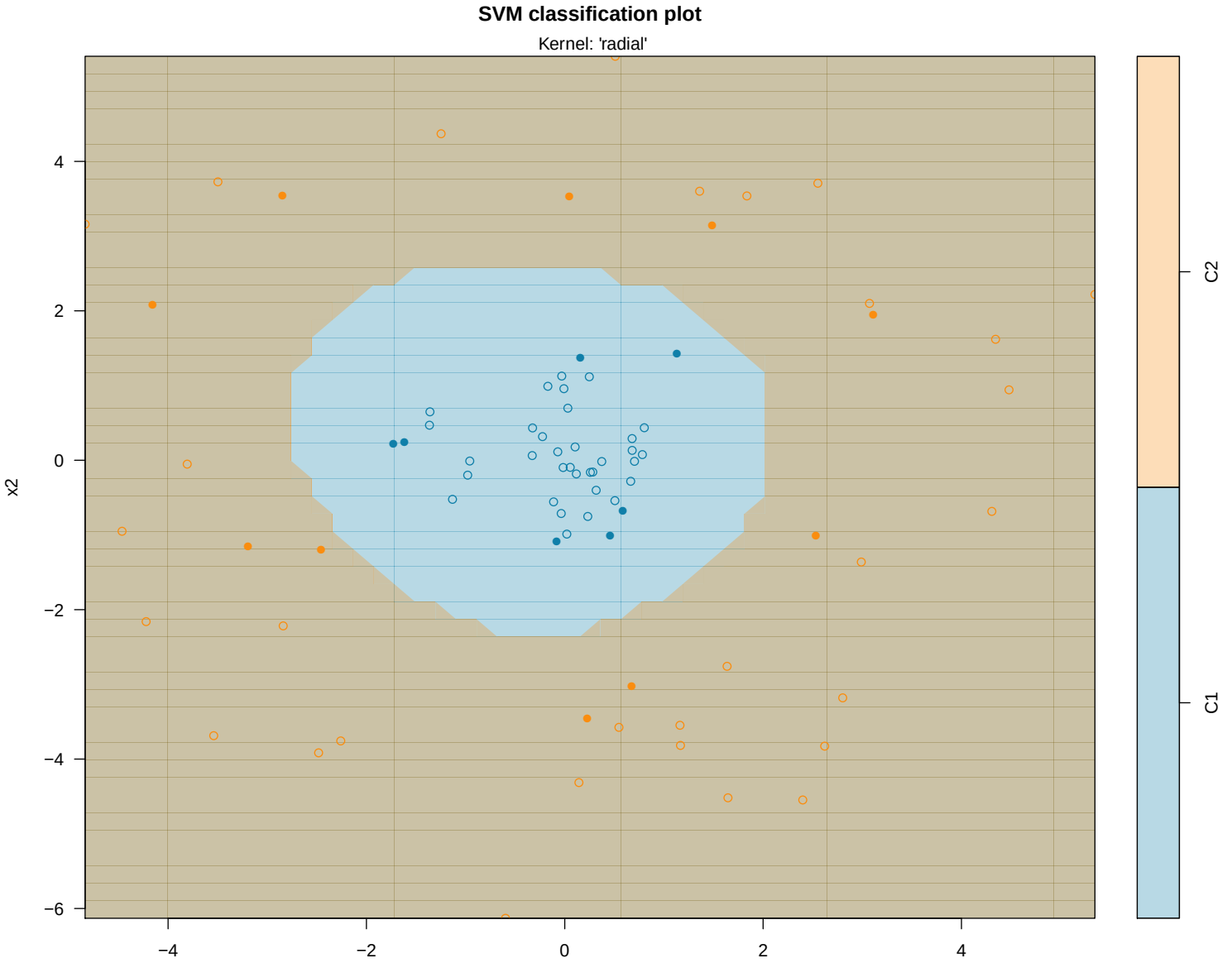







Torben Tvedebrink
tvede@math.aau.dk

# Eksempel
## Cont'f - mapping into higher dimension

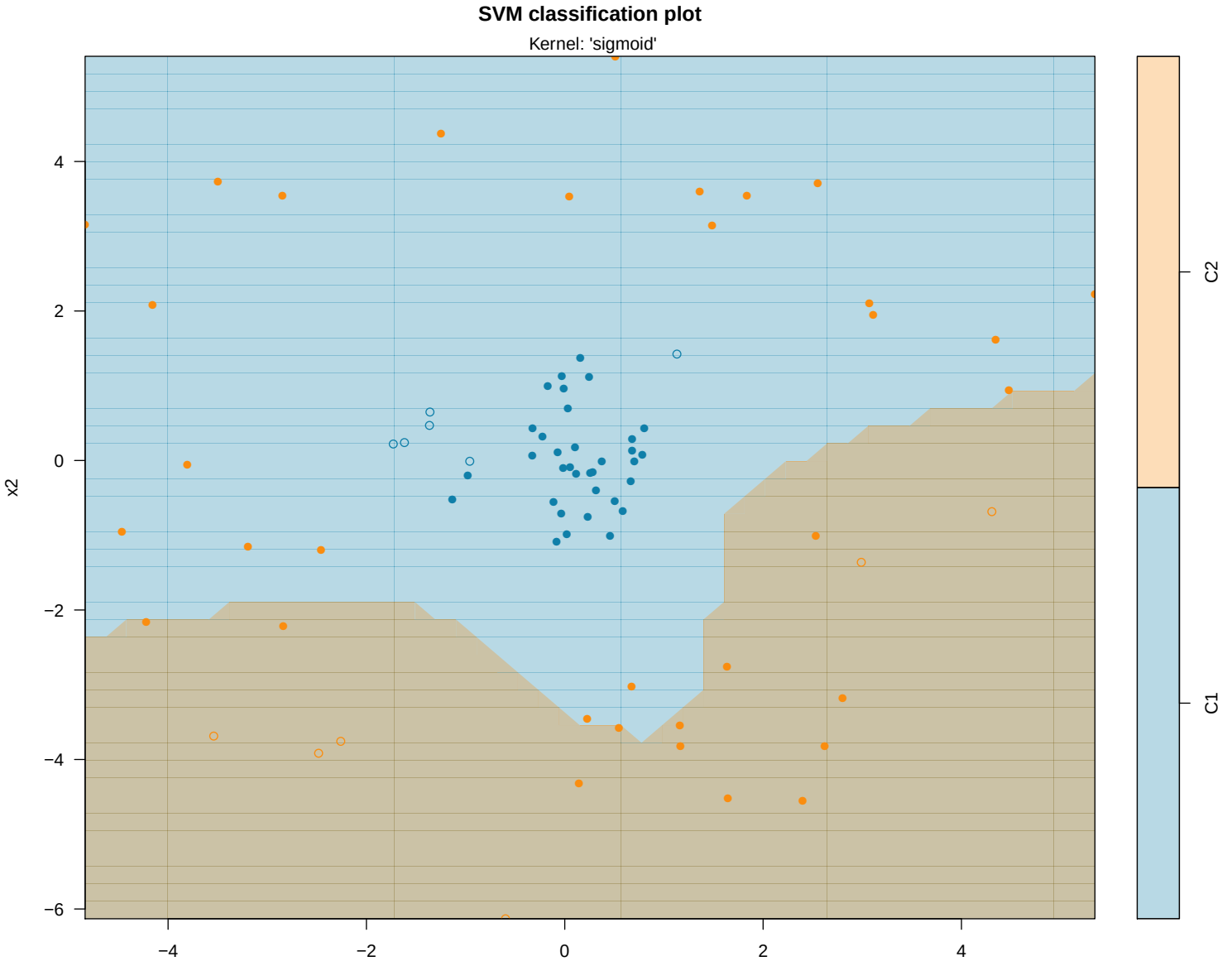







Torben Tvedebrink
tvede@math.aau.dk

# Eksempel
## Cont'f - mapping into higher dimension








Torben Tvedebrink
tvede@math.aau.dk

# Some frequently used kernels

| Name | Expression | **R** notes |
|---|---|---|
| `linear` | $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ | `kernel="linear"` |
| `polynomial` | $(\gamma \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + r)^p$ | $p$ (`degree`) , $\gamma$ (`gamma`) can $r$ (`coef0`) can be tuned |
| `radial` | $\exp\left(-\gamma \lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert^2\right)$ | $\gamma$ (`gamma`) can be tuned |
| *Is also called Gaussian – default in* **R** | | |
| `sigmoid` | $\tanh(\gamma \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + r)$ | $\gamma$ (`gamma`) and $r$ (`coef0`) can be tuned |





Torben Tvedebrink
tvede@math.aau.dk

# Some frequently used kernels



| Name | Expression | **R** notes |
|---|---|---|
| `linear` | $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ | `kernel="linear"` |
| `polynomial` | $(\gamma \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + r)^p$ | $p$ (`degree`) , $\gamma$ (`gamma`) can $r$ (`coef0`) can be tuned |
| `radial` | $\exp\left(-\gamma \lVert \boldsymbol{x}_i - \boldsymbol{x}_j \rVert^2\right)$ | $\gamma$ (`gamma`) can be tuned |
| *Is also called Gaussian – default in* **R** | | |
| `sigmoid` | $\tanh(\gamma \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + r)$ | $\gamma$ (`gamma`) and $r$ (`coef0`) can be tuned |

The functionen `tune.svm` can be used to 'grid' over selected values of the tuning parameters.


Torben Tvedebrink
tvede@math.aau.dk

# Other SVM-packages in **R**

| Library | Primary function | features |
|---|---|---|
| `kernlab` | `ksvm()` | More kernels and custom kernels |
| `klaR` | `svmlight()` | Requires systen install of `svmlight` |
| `svmpath` | `svmpath()` | Avanceret metode til parameter valg - egne kernels kan bruges |





Torben Tvedebrink
tvede@math.aau.dk