

Support Vector Machines

Krydsvalidering, Bootstrap, LDA og QDA

Lecture 3

Torben Tvedebrink
tvede@math.aau.dk

Institut for Matematiske Fag



AALBORG UNIVERSITY
DENMARK

Self-study opgaver

Louise og Rasmus



SVM
CV, Bootstrap
LDA, QDA

1 Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

The floor is yours!



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

2 Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

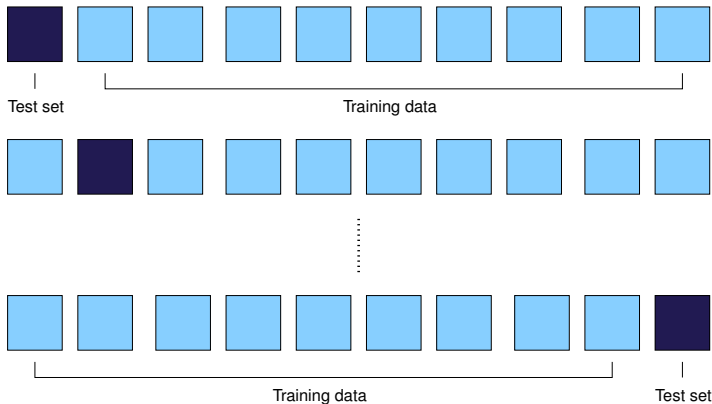
Opgaver

Krydsvalidering og bootstrap

En effektiv og simpel og effektiv måde til at estimere *generaliseringsfejlen* for en statistisk model/metode er vha. **krydsvalidering**.

K -fold krydsvalidering går ud på at imitere processen hvor vi har adgang til nye testdata, ved at vi opdeler data i K dele og successivt bruger de $K - 1$ dele som træningsdata og den resterende del til testdata.

10-fold krydsvalidering



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

4 Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

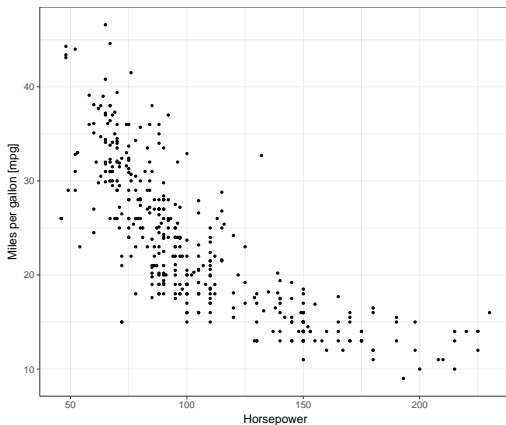
Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel



Vi kan fx. lave 10-fold krydsvalidering til at sige noget om passende valg af kompleksitet for en given model.



SVM

CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

5 Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

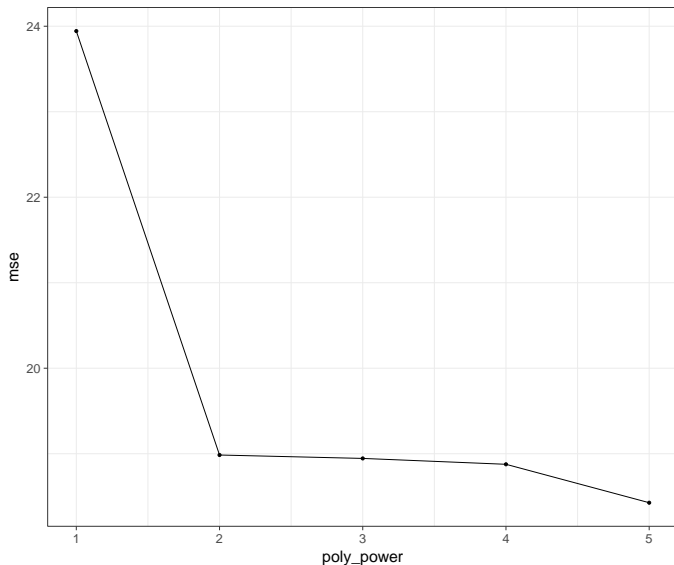
Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

5 Krydsvalidering
Bootstrap

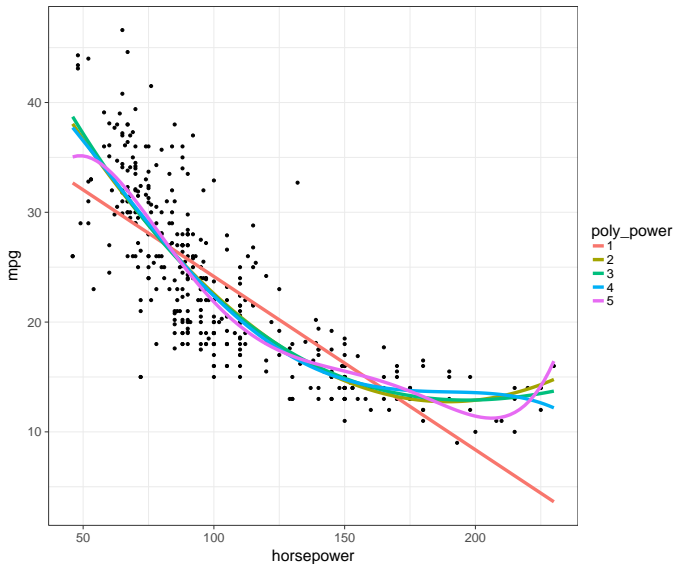
Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

5

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

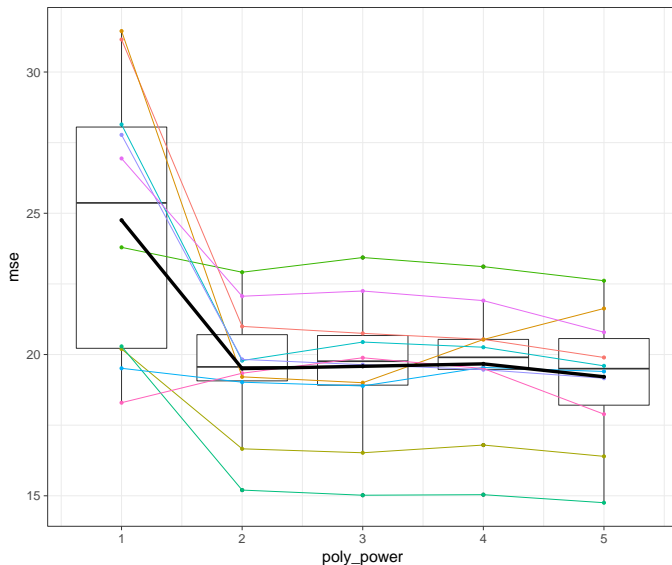
Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

5 Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Idéen bag bootstrap minder om krydsvalidering, idet vi benytter re-sampling af vores data til at estimere standard errors og bias af parameter estimerer.

Normalt noteres en bootstrap sample med $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$, hvor x_i^* er trukket med tilbagelægning fra x_1, \dots, x_n .

Baseret på \mathbf{x}^* kan vi således estimere de ukendte parametre θ og opnå et estimat $\hat{\theta}^*$. Dette kan vi gøre N gange, hvorved vi har $\hat{\theta}_1^*, \dots, \hat{\theta}_N^*$ estimerer af θ , hver baseret et bootstrap sample \mathbf{x}_i^* , $i = 1, \dots, N$.

Vi kan således sige noget om variabiliteten af $\hat{\theta}^*$ omkring $\hat{\theta}$, idet begge er funktioner af de observerede data (og re-samplinger af disse).

Ved en hver form for inferens er vi interesserede i at sige noget variationen af $\hat{\theta}$ omkring den sande værdi θ .

Ved bootstrap kan vi tilgå denne information ved at estimere $\text{sd}(\hat{\theta})$ ved $(N - 1)^{-1} \sum_{j=1}^N (\hat{\theta} - \hat{\theta}_j^*)^2$.

Bootstrap interval



SVM
CV, *Bootstrap*
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

8

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Ud fra vores bootstrap estimatater $\theta_1^*, \dots, \theta_N^*$ kan vi forme et *bootstrap interval* for estimatet af θ .

Fx. hvis $N = 100$ er $[\theta_{(5)}^*, \theta_{(95)}^*]$ et approksimativt 90%-konfidens interval for θ , hvor $\theta_{(1)}^* \leq \theta_{(2)}^* \leq \dots \leq \theta_{(N)}^*$ er ordnet efter størrelse.

Antag at X og Y repræsenterer to afkastet fra to investerings aktiver. Vi ønsker at minimere investerings risikoen, hvilket svarer til at minimere variansen af $\alpha X + (1 - \alpha)Y$, hvor α angiver andelen investeret i X .

Vi kan vise (*opgave*) at det optimale α angives som

$$\alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}},$$

hvor $\sigma_X^2 = \text{Var}(X)$, $\sigma_Y^2 = \text{Var}(Y)$ og $\sigma_{XY} = \text{Cov}(X, Y)$.

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

9

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

Opgaver

Eksempel (fortsat)



Generelt er σ_X^2 , σ_Y^2 og σ_{XY} ukendte og estimeres derfor fra data.

Lad os antage at vi kender de sande værdier. Vi kan således simulere data fra den sande fordeling og se variabiliteten i estimatet for α .

I det følgende simulerer vi 1000 datasæt med 100 observationer i hver. Den første simulation benytter vi efterfølgende til at lave et bootstrap af (ligeledes 1000 bootstrap samples).

SVM
CV, *Bootstrap*
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

10

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

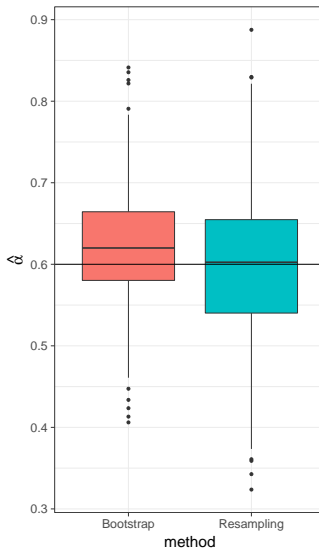
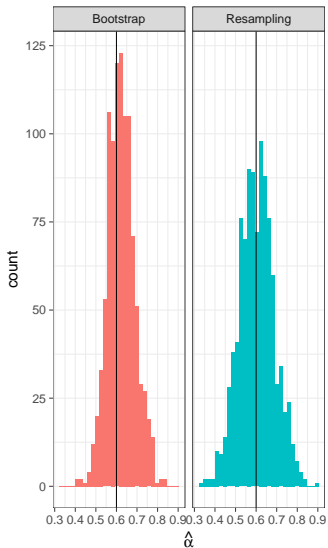
Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

40

Simuleret vs. bootstrap



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

Faldgruber ved bootstrap



Der er dog forskellige setups hvor bootstrap ikke kan anvendes, fx.

- ▶ Tidsrækker
- ▶ Og generelt hvis parameteren under inspektion modellere strukturer som ikke overholdes af bootstrap samplingen.
- ▶ Bootstrap kan ikke bruges til modelvalidering idet vi har stort overlap med data (ca. 2/3 dele af originale data er i bootstrap samples) – afgørende forskel mellem bootstrap og krydsvalidering.
- ▶ Se evt. mere på Thomas Lumley's blog:
<http://notstatschat.tumblr.com/post/156650638586/when-the-bootstrap-doesnt-work>

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

13 Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Diskriminant analyse

SVM
CV, *Bootstrap*
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

14 Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

I klassifikation ønsker vi som regel at bestemme klassen for en given observation vha. en sandsynlighed, fx.

$P(Y = k \mid \mathbf{x})$.

En Bayes klassifier predikterer naturligt observationens klasse som den mest sandsynlige gruppe givet X ,

$$G(\mathbf{x}) = \arg \max_{k \in \{1, \dots, K\}} P(Y = k \mid \mathbf{x}),$$

hvor K er antallet af mulige tilstande (klasser).

Ser vi på $P(Y = k | \mathbf{x})$ har vi fra Bayes sætning at

$$P(Y = k | \mathbf{x}) = \frac{f_k(\mathbf{x})P(Y = k)}{\sum_{l=1}^K f_l(\mathbf{x})P(Y = l)},$$

hvor $P(Y = k) = \pi_k$ er prior sandsynligheden for klasse k og f_k er fordelingen for \mathbf{x} givet klasse k .

Jf. formlen er $P(Y = k | \mathbf{x})$ stort set ensbetydende med $f_k(\mathbf{x})$, idet prior informationen gerne skulle være svag (mest information i data) og nævneren er normaliserings konstant.

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

15 **Diskriminant analyse**

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

Opgaver

Hvis vi antager at $f_k(\mathbf{x})$ er multivariat normal fordelt har vi således

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right\},$$

hvor Σ_k og μ_k er hhv. kovarians matrix og middelværdi for klassen k .

Lineær Diskriminant Analyse



Lineær Diskriminant Analyse (eller Fisher's diskriminant analyse, opfundet af af R.A. Fisher i 1936) laver den antagelse at $\Sigma_k = \Sigma$ for alle klasser k .

Hvad betyder dette?

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

17 Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

Lineær Diskriminant Analyse (eller Fisher's diskriminant analyse, opfundet af af R.A. Fisher i 1936) laver den antagelse at $\Sigma_k = \Sigma$ for alle klasser k .

Hvad betyder dette?

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ (\mathbf{x} - \mu_k)^\top \Sigma^{-1} (\mathbf{x} - \mu_k) \right\},$$

Tager vi log har vi

$$\log f_k(\mathbf{x}) \propto -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \left\{ \mathbf{x}^\top \Sigma^{-1} \mathbf{x} + \mu_k^\top \Sigma^{-1} \mu_k - 2 \mathbf{x}^\top \Sigma^{-1} \mu_k \right\}$$

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

17 Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation

Slack variable

Den ikke-separable situation

Kernel trick

Opgaver

Ser vi således på log-ratioet af posterior sandsynligheden for en klasse mod en anden, fx. k i forhold til l , får vi således

$$\begin{aligned}\log \frac{P(Y = k | \mathbf{x})}{P(Y = l | \mathbf{x})} &= \log \frac{f_k(\mathbf{x})\pi_k}{f_l(\mathbf{x})\pi_l} \\ &= \log \frac{\pi_k}{\pi_l} + \log f_k(\mathbf{x}) - \log f_l(\mathbf{x}) \\ &= \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k - \mu_l)^\top \Sigma (\mu_k + \mu_l) \\ &\quad + \mathbf{x}^\top \Sigma (\mu_k - \mu_l),\end{aligned}$$

Som altså er *lineær* i \mathbf{x} – heraf navnet.

Self-study opgaver

Krydsvalidering og
bootstrapKrydsvalidering
BootstrapDiskriminant
analyse18 Lineær (LDA)
Kvadratisk (QDA)Support Vector
MachinesDen separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

LDA - estimering af parametre



SVM
CV, Bootstrap
LDA, QDA

Givet data kan vi estimere π_k , μ_k og Σ for $k = 1, \dots, K$.

Dette gøres naturligt ved (bias korrigerede) MLE

$$\hat{\pi}_k = N_k / N$$

$$\hat{\mu}_k = N_k^{-1} \sum_{i: Y_i=k} \mathbf{x}_i$$

$$\hat{\Sigma} = (N - K)^{-1} \sum_{k=1}^K \sum_{i: Y_i=k} (\mathbf{x}_i - \hat{\mu}_k)(\mathbf{x}_i - \hat{\mu}_k)^\top$$

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

19 Lineær (LDA)
Kvadratisk (QDA)

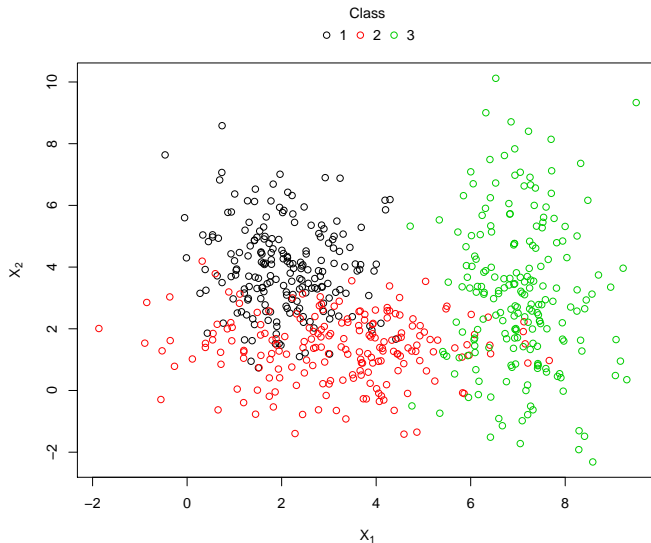
Support Vector
Machines

Den separable
situation

Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

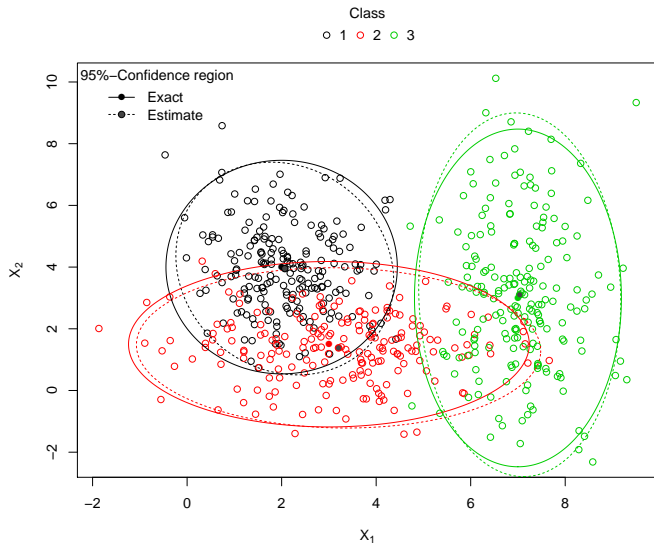
20 Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

20

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

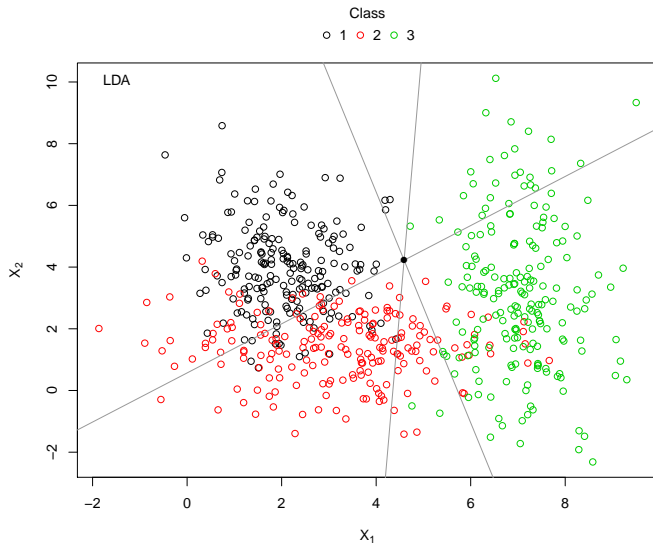
Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

20 Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

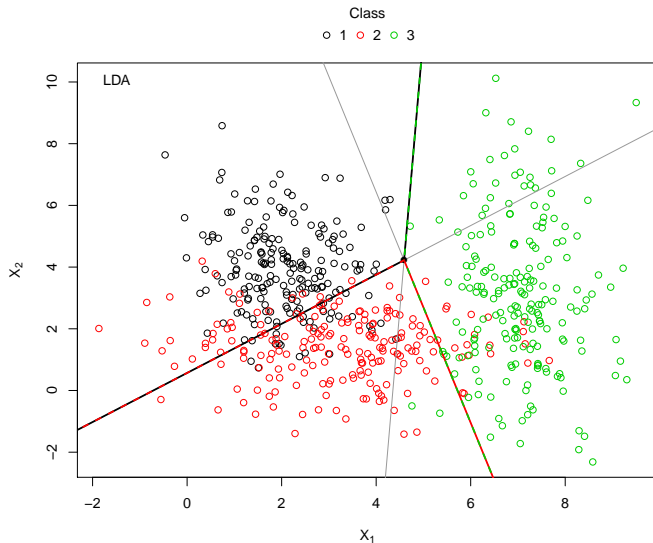
Den separable
situation

Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

20 Lineær (LDA)
Kvadratisk (QDA)

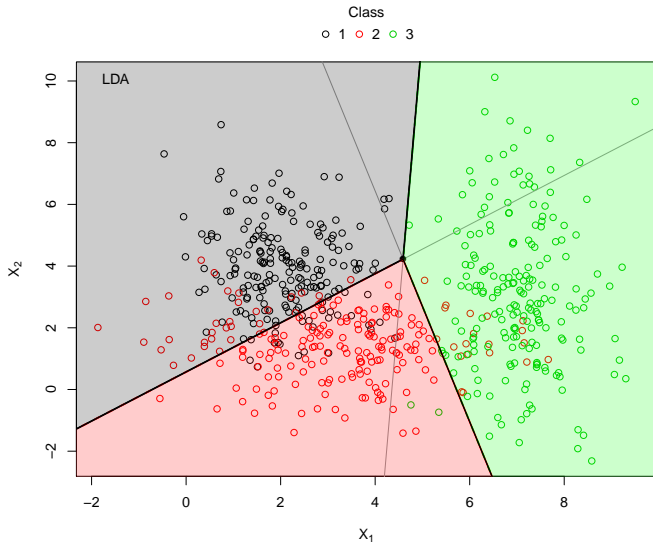
Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

Eksempel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

20 Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk

I det tilfælde hvor $K = 2$ kan det vises (opgave) at vi klassificere en observation \mathbf{x} som klasse 1 hvis

$$\mathbf{x}^\top \hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) > \frac{1}{2}(\hat{\mu}_1 - \hat{\mu}_2)^\top \hat{\Sigma}^{-1}(\hat{\mu}_1 + \hat{\mu}_2) - \log \frac{N_1}{N_2}$$

Kvadratisk Diskriminant Analyse (QDA)



SVM
CV, Bootstrap
LDA, QDA

Den bekvemme antagelse om $\Sigma_k = \Sigma$ for alle k gjorde at leddet $\mathbf{x}^\top \Sigma_k^{-1} \mathbf{x}$ var det samme for alle k , hvilket derfor forkortede ud i $\log P(Y = k | \mathbf{x}) / P(Y = l | \mathbf{x})$.

Vores udtryk for $\log P(Y = k | \mathbf{x})$ i tilfældet med variende kovarians matricer for vi derfor

$$-\frac{1}{2} \left\{ \log |\Sigma_k| + (\mathbf{x} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right\} + \log \pi_k,$$

som er kvadratisk i \mathbf{x} – heraf *kvadratisk diskriminant analyse*.

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

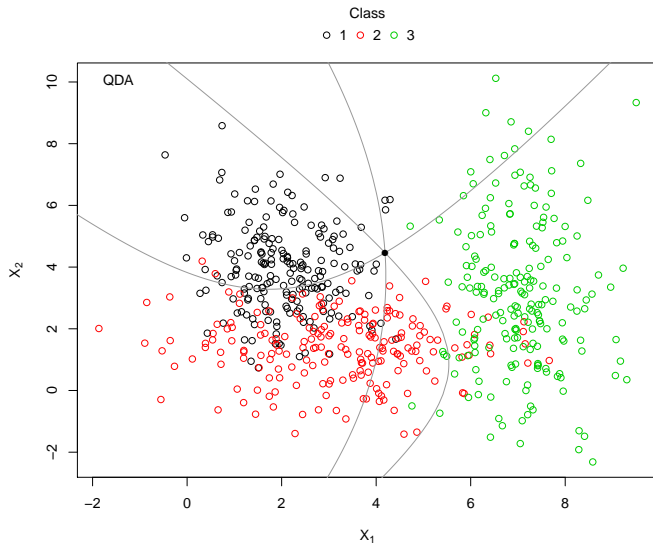
Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

Opgaver

22

40

Eksempel (fortsat)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

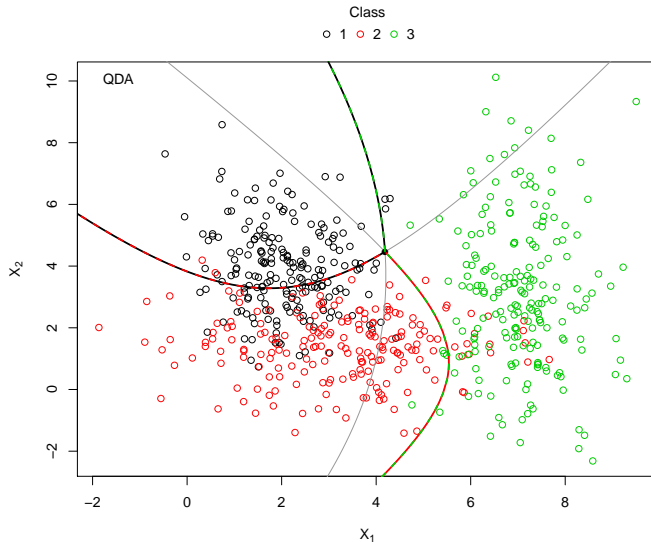
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel (fortsat)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

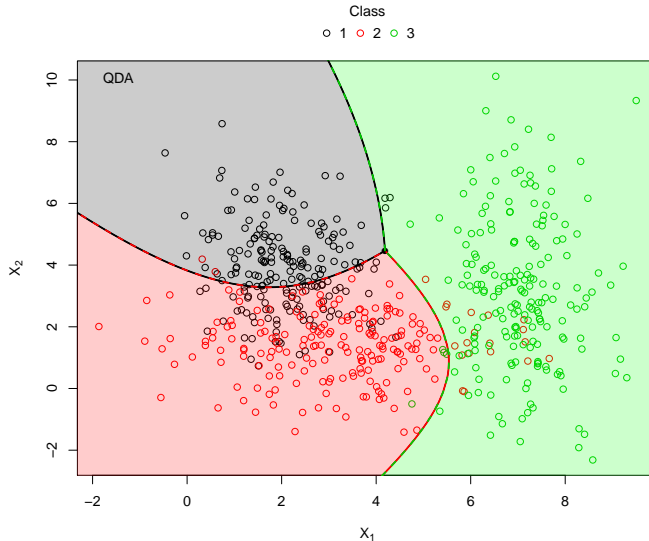
Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

Eksempel (fortsat)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

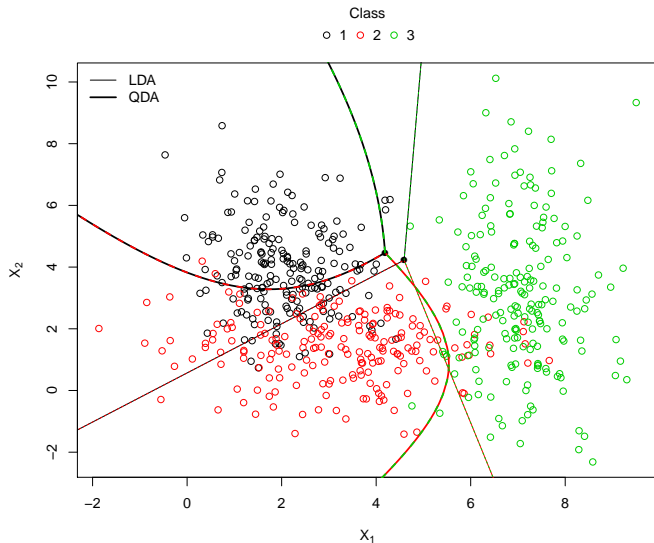
Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

Eksempel (fortsat)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

Opgaver

LDA og QDA i R



I R kan vi foretage LDA og QDA analyser vha. MASS-pakken

```
## LDA analyse af simuleret data
LDA <- lda(class~.,data=data)
LDA.pred <- predict(LDA)
data$pred.lda <- LDA.pred$class
```

```
with(data,table(class,pred.lda))
      pred.lda
class   1    2    3
      1 173   27    0
      2  31 149   20
      3   0   2 198
```

```
## Tilsvarende for QDA ##
QDA <- qda(class~.,data=data)
QDA.pred <- predict(QDA)
data$pred.qda <- QDA.pred$class
```

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)

24

Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

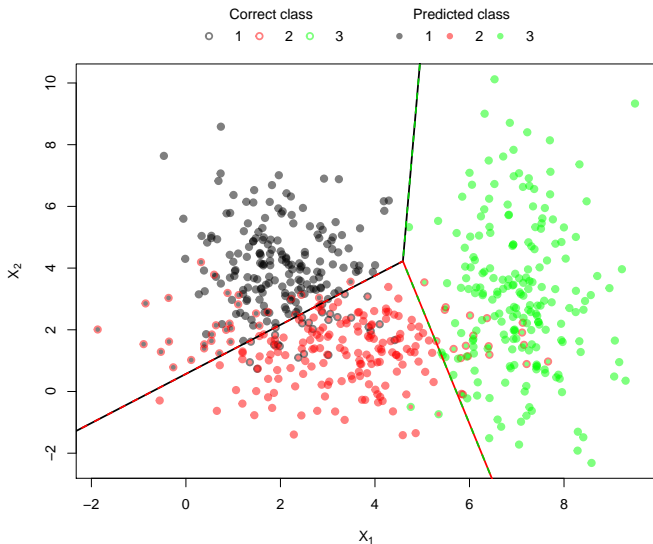
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel (fortsat)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Torben Tvedebrink
tvede@math.aau.dk



Support Vector Machines

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

26 **Support Vector
Machines**

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

Support Vector Machines



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

27 Support Vector Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

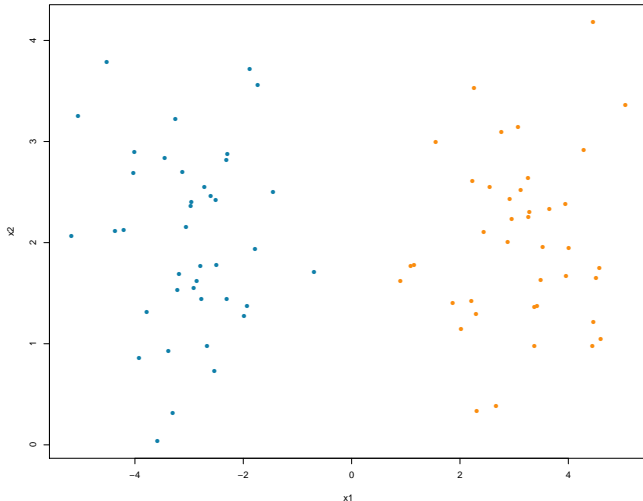
Opgaver

Support Vector Machines (SVM) er en type classifier som har vist sig meget effektiv i forskellige anvendelses områder.

Vi starter med det simpleste tilfælde..

Eksempel

Separation af to klasser



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

28

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

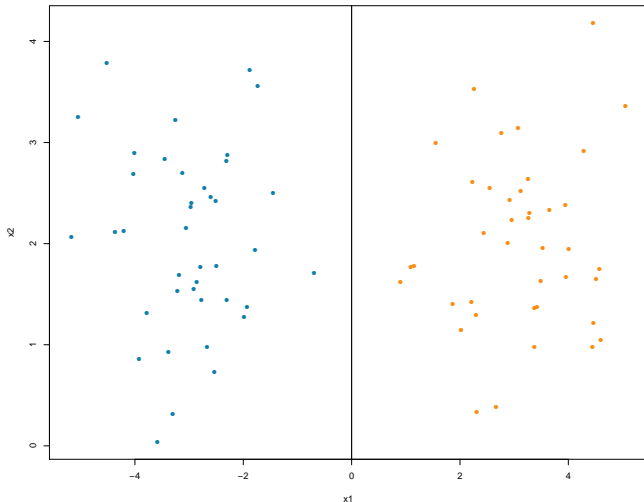
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Separation af to klasser – Et muligt valg



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

28

Den separable
situation

Slack variable
Den ikke-separable
situation
Kernel trick

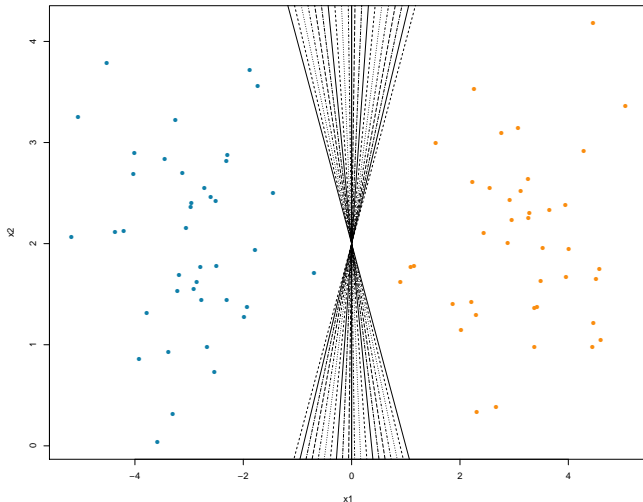
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Separation af to klasser – Uendeligt mange muligheder



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

28

Den separable
situation
Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Den optimale hyperplan



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

29

Den separable
situation

Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

Kan defineres ved at kræve:

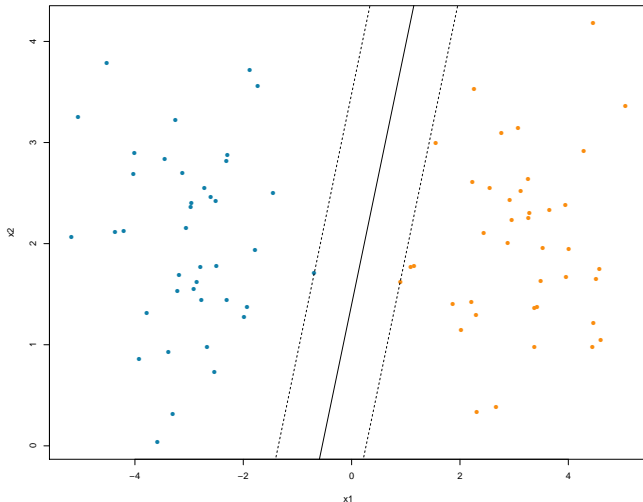
- a) Separere de to klasser i disjunkte dele
- b) Have den maximale afstand til de nærmeste punkter i hver klasse

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Den optimale hyperplan



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

30

Den separable
situation

Slack variable
Den ikke-separable
situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Optimalt for testdata?



På foregående figur så vi hvorledes SVM angav den bedste hyperplan, men at den var stærkt influeret af **den blå observation** længst til højre.

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation

31

Slack variable

Den ikke-separable situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Optimalt for testdata?

Slack variable

På foregående figur så vi hvorledes SVM angav den bedste hyperplan, men at den var stærkt influeret af **den blå observation** længst til højre.

Ved at bruge *slack variable* kan vi finde et hyperplan med andre egenskaber.

Formålet med *slack variable* er, at vi tillader at observationerne må ligge *udenfor* den rigtige klasse hvorved vi muligvis kan opnå bedre klassifikation for testdata.

“A few rotten apples in the basket”



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

31

Slack variable
Den ikke-separable
situation
Kernel trick

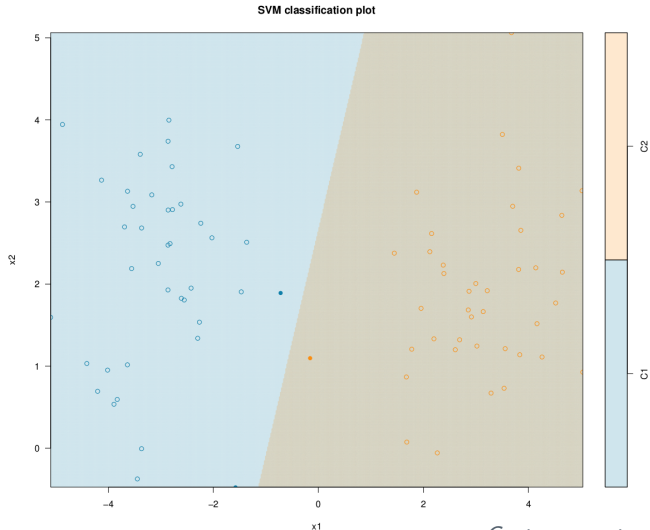
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Ingen slack variable



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

32

Slack variable

Den ikke-separable
situation
Kernel trick

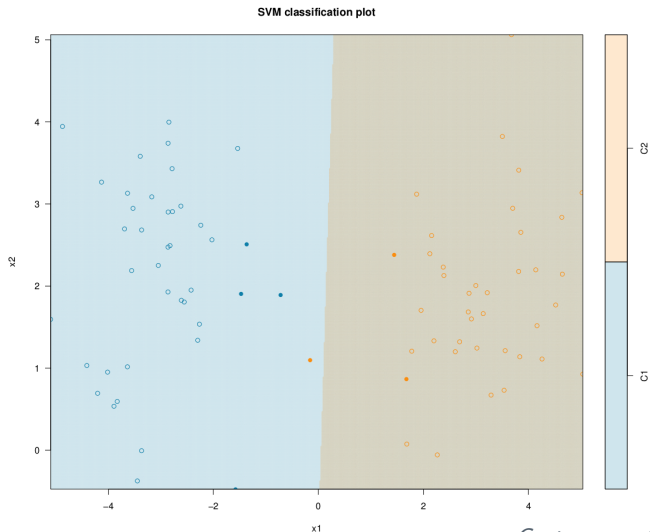
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Med *slack variable*



Cost-parameter: 1

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

32

Slack variable

Den ikke-separable
situation
Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Den ikke-separable situation



I det tilfælde hvor det ikke er muligt at finde en separerende hyperplan, er det nødvendigt at bruge slack variable – ellers findes der ingen løsning.

Slack variablene bruges til “måle” hvor langt de ikke korrekt-klassificerede observationer ligger fra hyperplanen.

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

33

Den ikke-separable
situation

Kernel trick

Opgaver

40

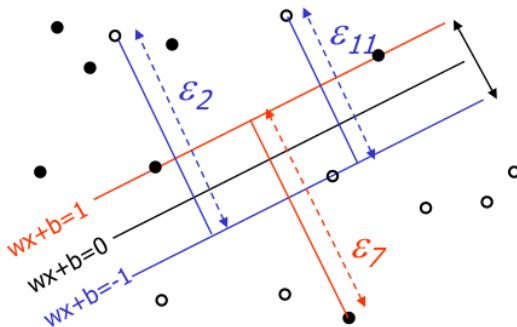
Torben Tvedebrink
tvede@math.aau.dk

Den ikke-separable situation



I det tilfælde hvor det ikke er muligt at finde en separerende hyperplan, er det nødvendigt at bruge slack variable – ellers findes der ingen løsning.

Slack variablene bruges til “måle” hvor langt de ikke korrekt-klassificerede observationer ligger fra hyperplanen.



SVM

CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

33 Den ikke-separable
situation
Kernel trick

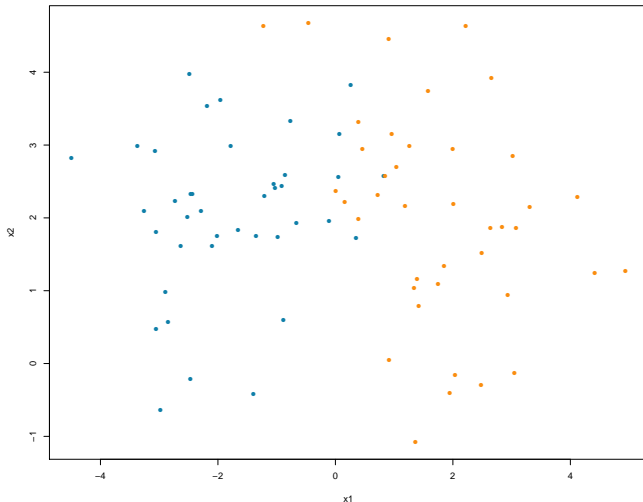
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempel

Ikke-separabel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

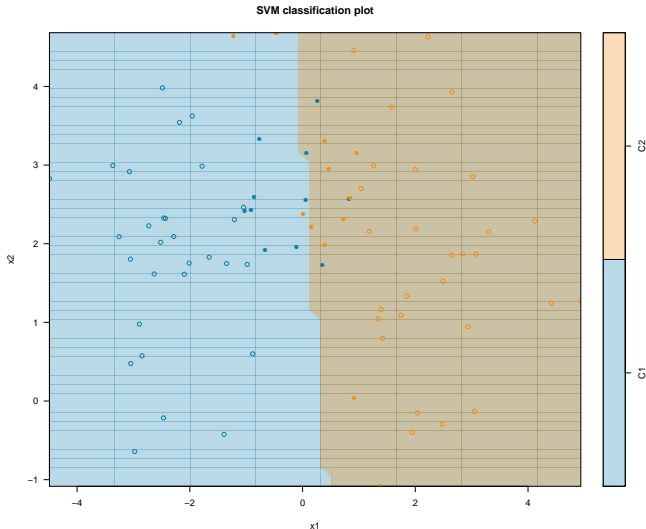
Slack variable

34 Den ikke-separable
situation
Kernel trick

Opgaver

Eksempel

Ikke-separabel



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

34

40

Torben Tvedebrink
tvede@math.aau.dk

Udvid/afbilled data til højere dimension



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

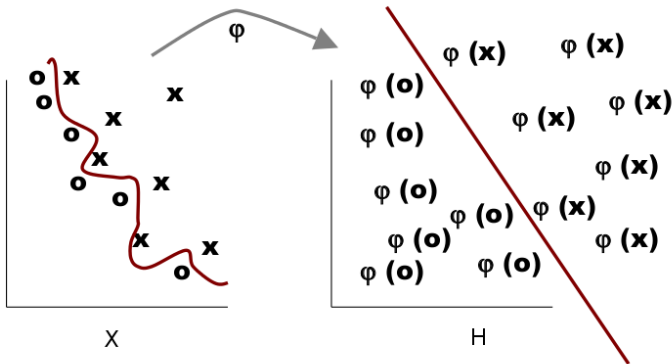
Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver



35

40

Eksempler

Afbilled ind i højere dimension



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

36 Kernel trick

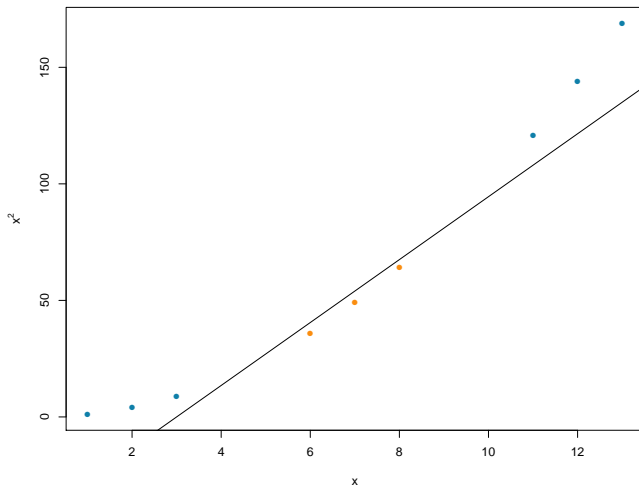
Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempler

Afbilled ind i højere dimension



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation

Kernel trick

Opgaver

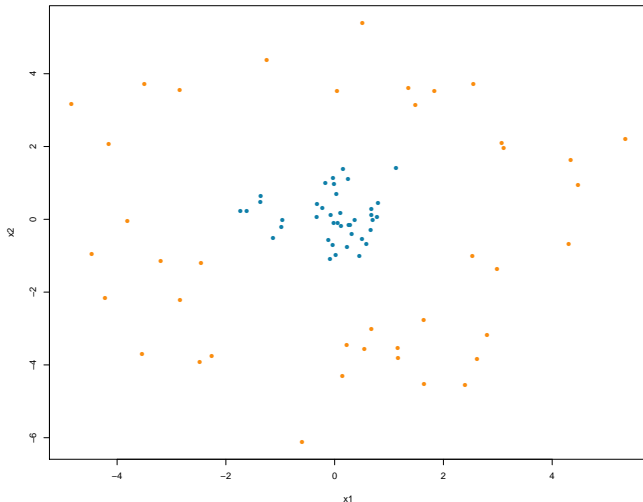
36

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempler

Afbilled ind i højere dimension



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

36

40

Torben Tvedebrink
tvede@math.aau.dk

Eksempler

Afbilled ind i højere dimension



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

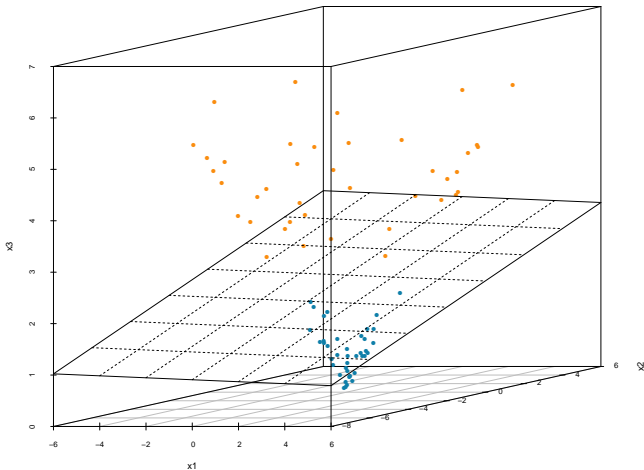
Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver



36

40

Eksempler

Afbilled ind i højere dimension



```
## RGL library giver adgang til 3D
## visualisering vha. OpenGL
library(rgl)

## Kan således rotere data i 3D:
with(circ.data, plot3d(x=x1, y=x2, z=x3))
```

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap
Krydsvalidering
Bootstrap

Diskriminant
analyse
Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation
Slack variable
Den ikke-separable
situation

36 Kernel trick

Opgaver

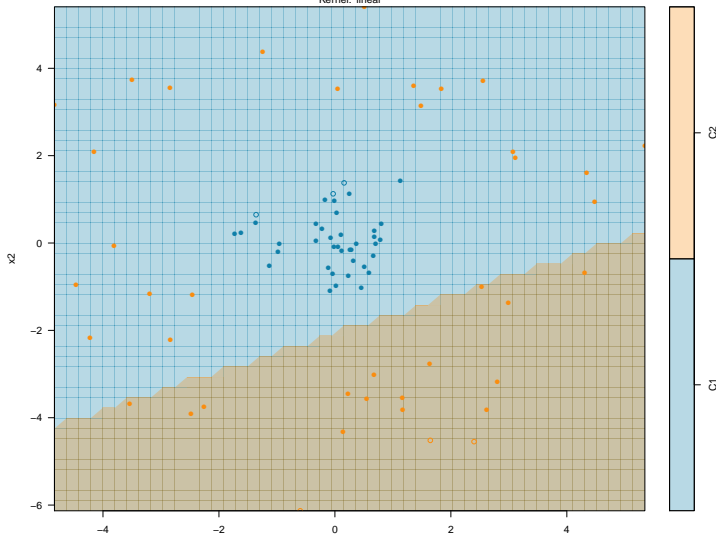
Eksempel

Fortsat - afbildet ind i højere dimension



SVM classification plot

Kernel: 'linear'



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

37

40

Torben Tvedebrink
tvede@math.aau.dk

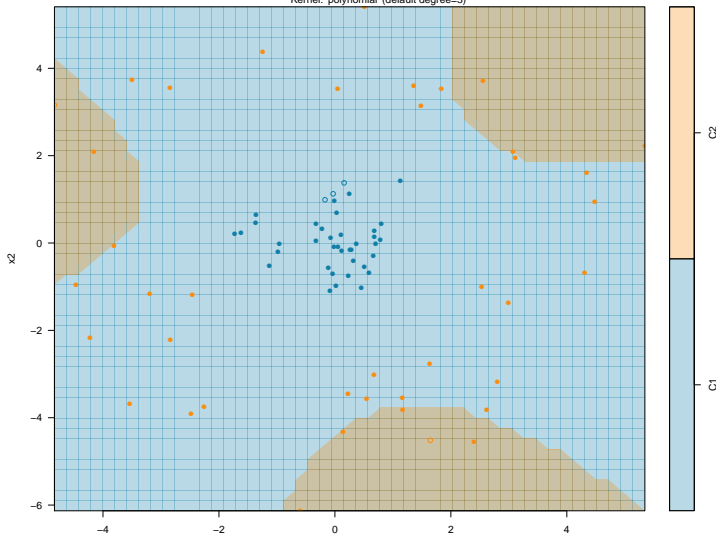
Eksempel

Fortsat - afbildet ind i højere dimension



SVM classification plot

Kernel: 'polynomial' (default degree=3)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

37

40

Torben Tvedebrink
tvede@math.aau.dk

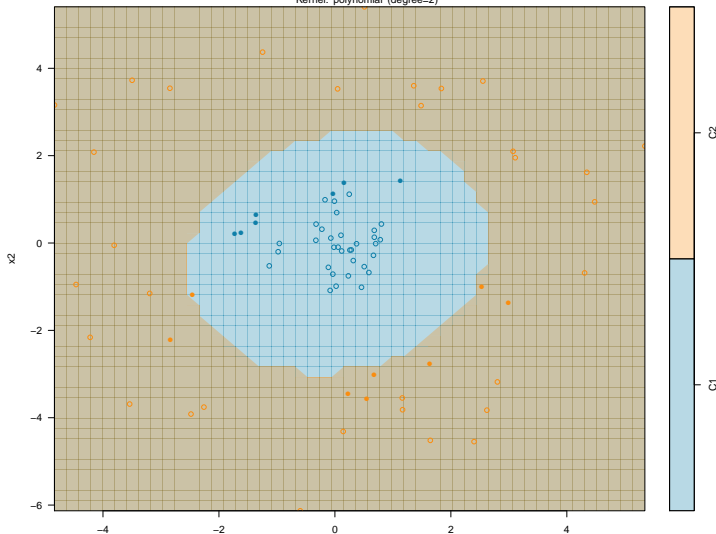
Eksempel

Fortsat - afbildet ind i højere dimension



SVM classification plot

Kernel: 'polynomial' (degree=2)



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation

Kernel trick

Opgaver

37

40

Torben Tvedebrink
tvede@math.aau.dk

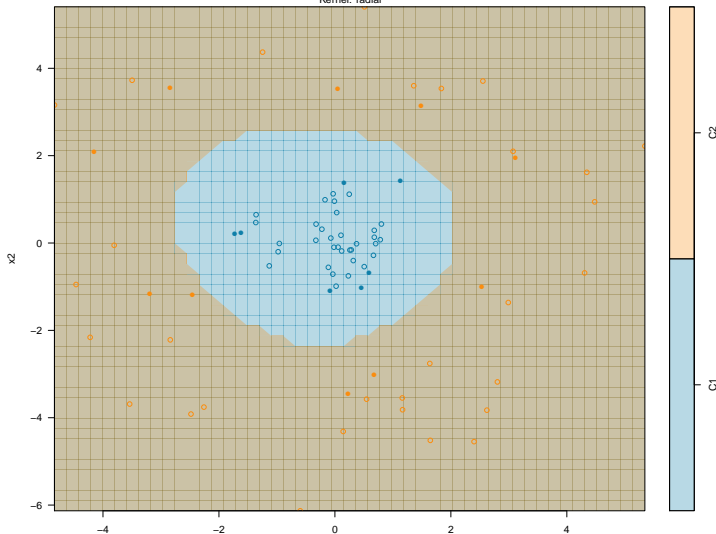
Eksempel

Fortsat - afbildet ind i højere dimension



SVM classification plot

Kernel: 'radial'



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

37

40

Torben Tvedebrink
tvede@math.aau.dk

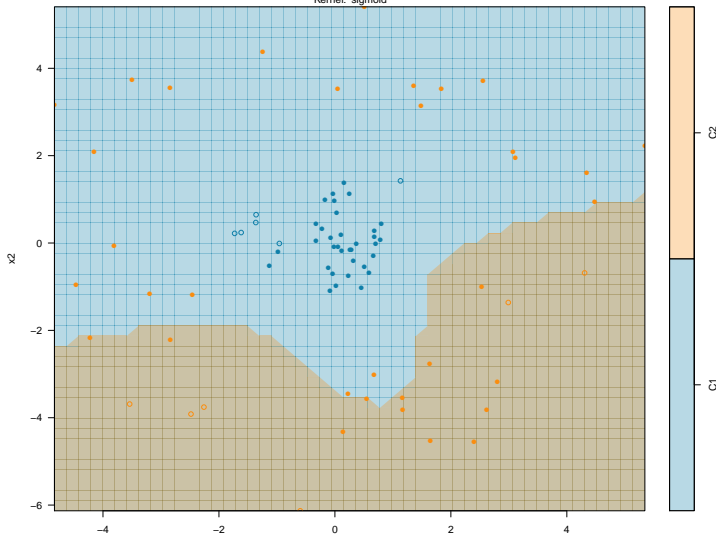
Eksempel

Fortsat - afbildet ind i højere dimension



SVM classification plot

Kernel: 'sigmoid'



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

Kernel trick

Opgaver

37

40

Torben Tvedebrink
tvede@math.aau.dk

Ofte anvendte kerner



Navn	Udtryk	R bemærkning
linear	$\langle \mathbf{x}_i, \mathbf{x}_j \rangle$	kernel="linear"
polynomial	$(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)^p$	p (degree) , γ (gamma) og r (coef0) kan tunes
radial basis	$\exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	γ (gamma) kan tunes <i>Kaldes også Gaussian pga. relation til normalfordelingen - default i R</i>
sigmoid	$\tanh(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)$	γ (gamma) og r (coef0) kan tunes

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

38

Kernel trick

Opgaver

40

Torben Tvedebrink
tvede@math.aau.dk

Ofte anvendte kerner



Navn	Udtryk	R bemærkning
linear	$\langle \mathbf{x}_i, \mathbf{x}_j \rangle$	kernel="linear"
polynomial	$(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)^p$	p (degree) , γ (gamma) og r (coef0) kan tunes
radial basis	$\exp(-\gamma \ \mathbf{x}_i - \mathbf{x}_j\ ^2)$	γ (gamma) kan tunes <i>Kaldes også Gaussian pga. relation til normalfordelingen - default i R</i>
sigmoid	$\tanh(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + r)$	γ (gamma) og r (coef0) kan tunes

Funktionen `tune.svm` kan benyttes til at 'gridde' over interessante værdier for tunings parametrene.

SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation

Slack variable

Den ikke-separable situation

Kernel trick

Opgaver

38

40

Andre R-pakker til SVM



SVM
CV, Bootstrap
LDA, QDA

Self-study opgaver

Krydsvalidering og
bootstrap

Krydsvalidering
Bootstrap

Diskriminant
analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector
Machines

Den separable
situation

Slack variable

Den ikke-separable
situation

39

Kernel trick

Opgaver

Navn	Primær funktion	Features
kernlab	ksvm()	Flere kernels og brug af egne kernel-funktioner
klaR	svmlight()	Kræver installation af svmlight
svmpath	svmpath()	Avanceret metode til parameter valg - egne kernels kan bruges

40

Torben Tvedebrink
tvede@math.aau.dk



SVM
CV, Bootstrap
LDA, QDA

1. Gennemtag analysen af Auto data fra ISLR-pakken som vist under *Krydsvalidering*.
2. Vis at sandsynligheden for at en observation er i et bootstrap samle er givet ved: $P(\text{Obs. } i \in B) \approx 0.632$.
3. Vis at det α som minimerer $\text{var}(\alpha\mathbf{x} + (1-\alpha)Y)$ er $(\sigma_Y^2 - \sigma_{XY})/(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY})$.
4. Vis at beslutningsgrænsen for LDA med $K = 2$ er som angivet på slide 19.
5. Løs opgave 1 [Wu&Kumar] (a og b) vha. logistisk regression i **R**

Self-study opgaver

Krydsvalidering og bootstrap

Krydsvalidering
Bootstrap

Diskriminant analyse

Lineær (LDA)
Kvadratisk (QDA)

Support Vector Machines

Den separable situation
Slack variable
Den ikke-separable situation
Kernel trick

40 Opgaver

6. Løs opgave 2 [Wu&Kumar] vha. `svm` i **R**
7. Løs opgave 3 [Wu&Kumar] - brug fx. `tune.svm` til at vælge/finde gode parameter værdier
8. Analyser crabs-data fra **MASS** og klassificer på art (species, `sp`).
Hint brug log-transformeret data og `kernel="radial"`.