
NORWEGIAN UNIVERSITY OF LIFE SCIENCES
MDCCCLIX

Classification techniques

focus on Discriminant Analysis

Seminar: Potentials of advanced image analysis technology in the cereal science research

2111
2005

Ulf Indahl/IMT - 14.06.2010

Supervised Learning

- Task:
  - For a given *feature* vector $\mathbf{x} \in \Re^p$:
    - Based on an appropriate model find the correct class label $g \in C = \{1, 2, ..., L\}$ (L = number of <u>unordered</u> categories).
  - Start by building a classification rule consistent with, and based on
    - A set of <u>training data</u> (n samples) $\{(\mathbf{x}_i , g_i) \mid i = 1,...,n\}$ to "learn the relationship between feature vectors $\mathbf{x} \in \Re^p$ and the groups in C".
  - For new feature vectors $\mathbf{x}_k$:
    - Predict its class g_k with good precision (high probability).
  - As always:
    - Validation of the model building is required.

Probabilistic/statistical approach based on Bayes rile

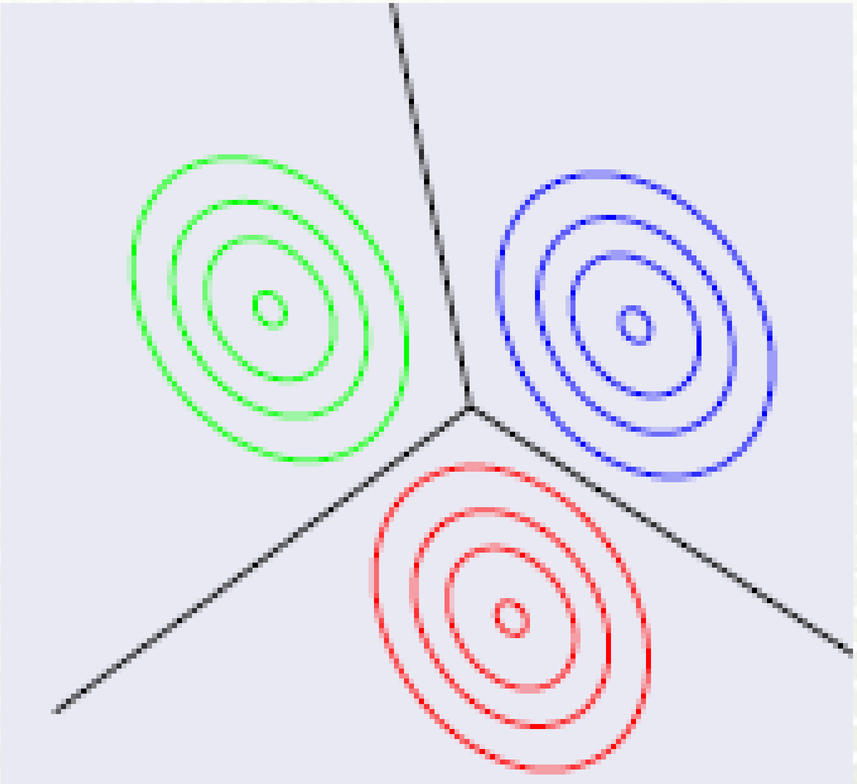
- Assume feature vectors of each class described by a probability density f(**x**|*g*).
- Aposteriori probability of each class *g*:
  - $p(g|\mathbf{x}) = \pi_g f(\mathbf{x}\,|g)/\Sigma\pi_j f(\mathbf{x}\,|j)$
- Bayes rule:
  - Assign **x** to class *k* if the aposteriori probability
    - $p(l\,|\mathbf{x}) > p(g\,|\mathbf{x})$ when $g \neq l$.
- Our focus:
  - Assume each group described by a normal distributoin

$$f(\mathbf{x}|l) = \frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}|}} exp^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_l)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_l)}$$

LDA – Linear Discriminant Analysis:

Assume all groups multinormal with identical covariance structure Σ and individual group means $\mu_1,\ldots,\mu_L$

- For LDA the groups are spearated by linear surfaces:



Case study

- Iris datasettet (Anderson/Fisher):
- 3 classes - the species of iris:
  - Setosa(1), Versicolor(2) and Virginica(3).
  - n = 150 samples (50 of each class),
- p = 4 features (in principle measurable by appropriate image analysis methodology.)
  - Measurements of Sepal- and Petal- length and width.

The dataset was collected by E. Anderson (1935) and analyzed by R.A. Fisher (1936)

Estimation of LDA-model from the empirical data

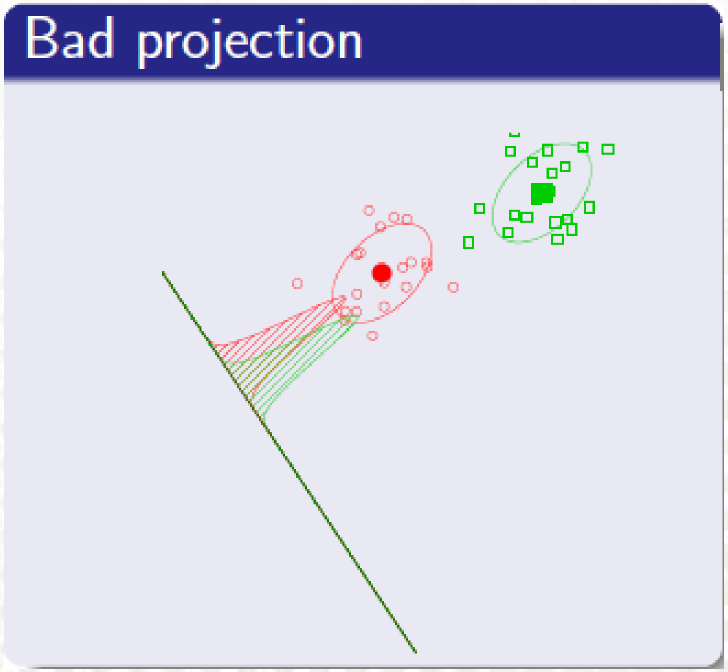
- Demo of simple MATLAB-code
  - Results (confution matrix) of leave-one-out crossvalidation:

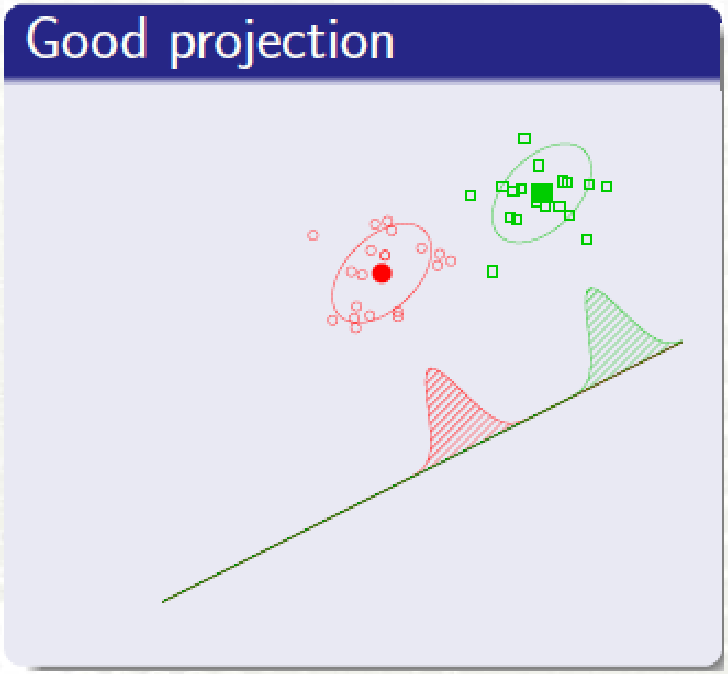
	Set	Vers	Virg
pred Set	50	0	0
pred Vers	0	48	2
pred Virg	0	1	49

- Error groups (2 & 3): 3/100
- OBS - full ”leave-one-out” crossvalidation can be estimated without explicit remodelling (a consequence of linear modeling – similar to MLR).

Canonical variates: Visual/geometrical facet of LDA

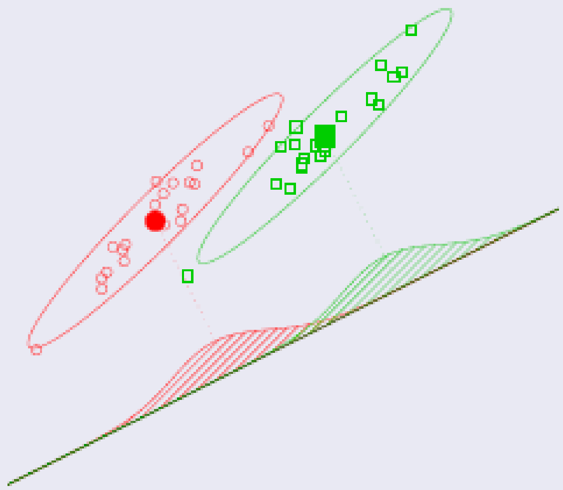
- Fisher's Idea:
  - Project data such that classes are "optimally" separated.





1st idea

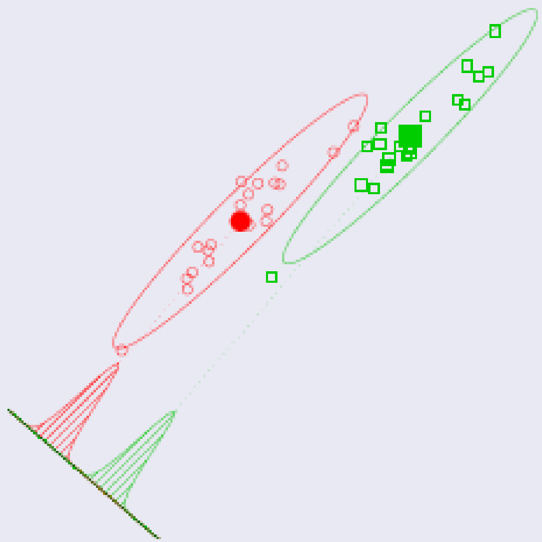
Project data such that class means are best separated



$\leadsto$ does not seem optimal

Better idea

Consider within-class covariance structure as well



NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Fisher's Formulation

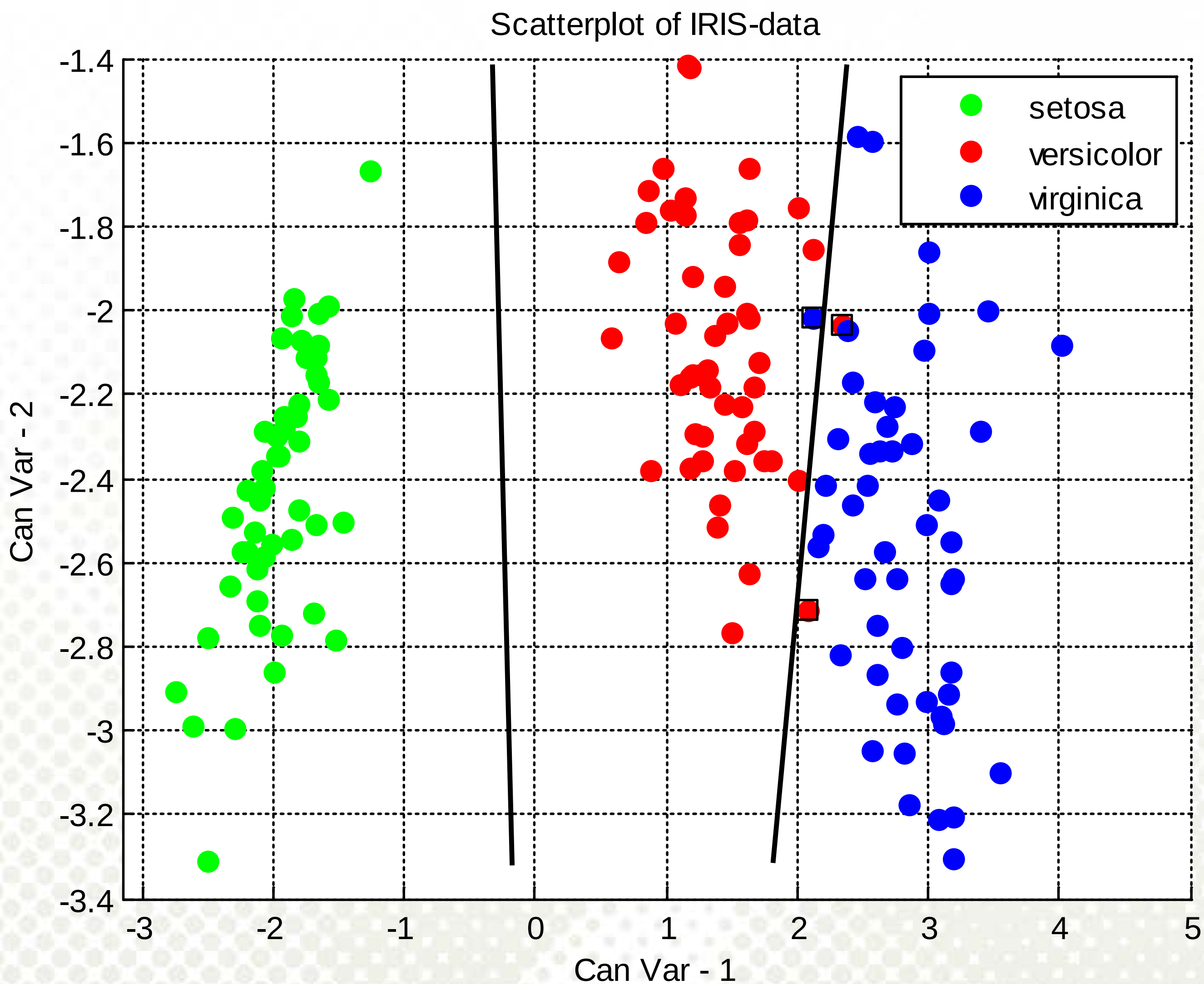
- We want to find the projection a maximizing the quatient:

$$\frac{\text{Distance between the projected class means}}{\text{Variance of the projected data}} = \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}}$$

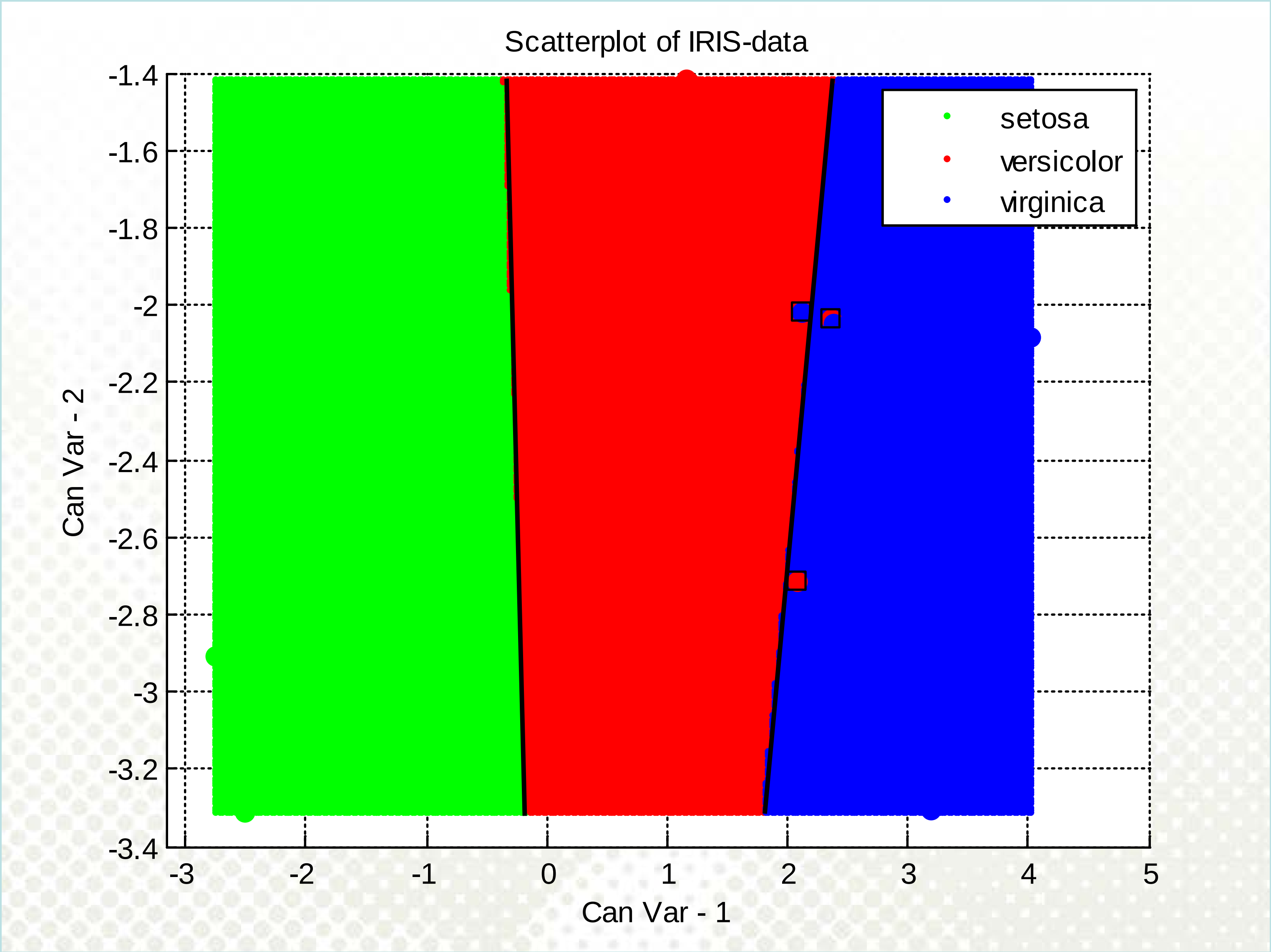
- **B** is the so-called between-groups covariance matrix measuring the dispersion of the class means:

$$\mathbf{B} := \frac{1}{L-1}\sum_{l=1}^{L}(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})'$$

- One projection is often not enough to separate data and one can use se further projections:
  - Maximize the above quotient over all **a** that are **W**-orthogonal to projection directions already found.


Scatterplot of IRIS-data
setosa
versicolor
virginica
Can Var - 2
Can Var - 1
-1.4
-1.6
-1.8
-2
-2.2
-2.4
-2.6
-2.8
-3
-3.2
-3.4
-3
-2
-1
0
1
2
3
4
5

NORWEGIAN UNIVERSITY OF LIFE SCIENCES



LDA applied to the original features = LDA applied to the data after projection onto the subspace spanned by the canonical variates

- Compare the estimated classprobabilities of LDA on original features and LDA on canonical variates.

Classification by regression

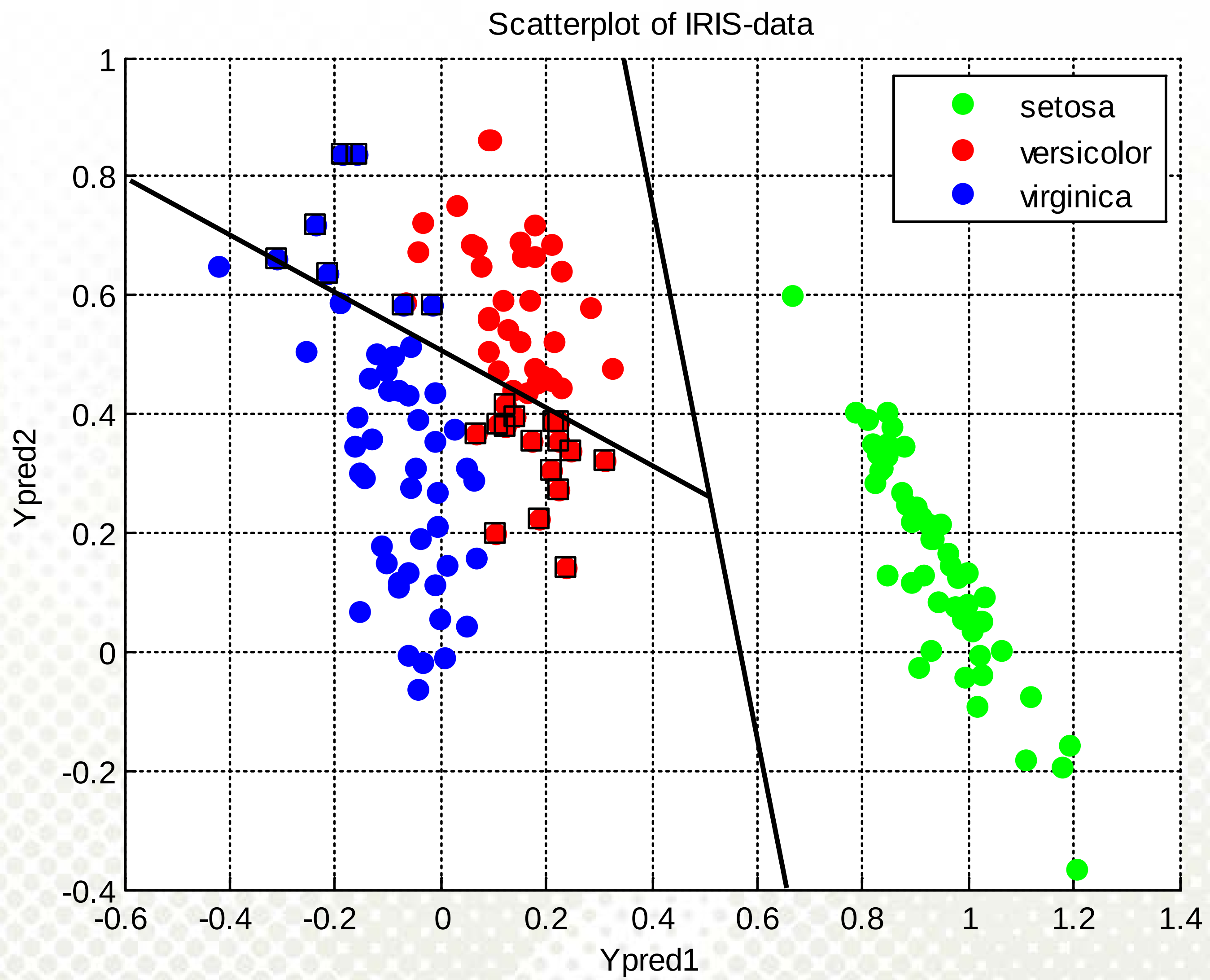
- Group memberships are dummy-coded
  - Use MLR to obtain ”fitted dummies”.
  - Classify sample **x** to the class corresponding to the largest among the L fitted values.
- This strategy is <u>not</u> optimal.
  - See MATLAB-code applied to IRIS data.
- However
  - LDA applied to the fitted MLR values = LDA applied to the original features (!)

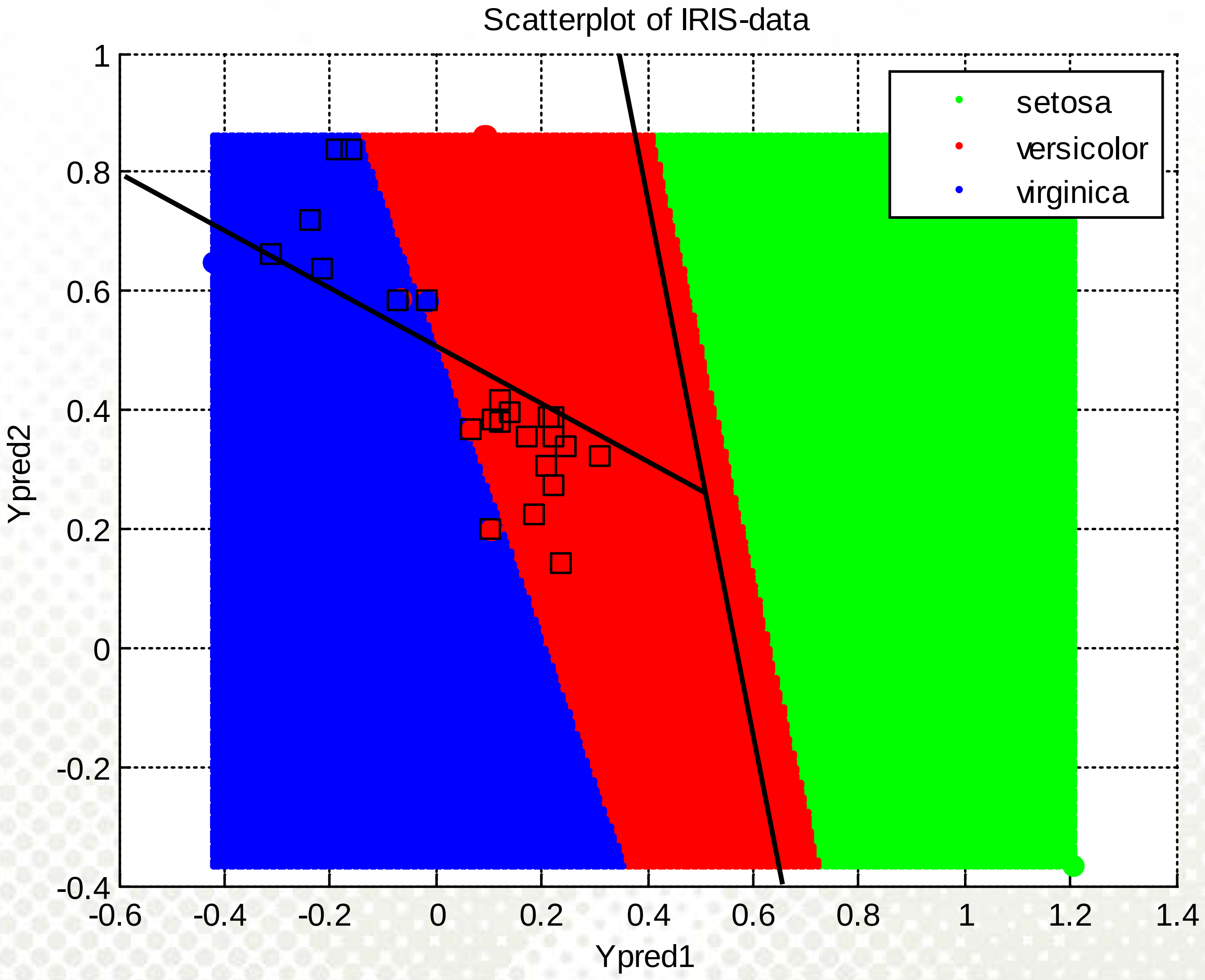
Classification by regression: The Iris data

- Demo of simple MATLAB-code
  - Results (confution matrix):

	Set	Vers	Virg
pred Set	50	0	0
pred Vers	0	34	16
pred Virg	0	7	43

- Error groups (2 & 3): 23/100


Scatterplot of IRIS-data
setosa
versicolor
virginica
Ypred2
Ypred1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4


Scatterplot of IRIS-data
setosa
versicolor
virginica
Ypred2
Ypred1
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4

Modifications

- RDA:

Regularised discriminant analysis (Friedman, 1989)

Compromise between LDA and QDA. Use covariance

$$\hat{\mathbf{\Sigma}}_l^{RDA}(\alpha) = \alpha \hat{\mathbf{\Sigma}}_l + (1-\alpha)\hat{\mathbf{\Sigma}}$$

($\alpha = 0 \rightsquigarrow$ LDA / $\alpha = 1 \rightsquigarrow$ QDA)

- PDA/FDA:

Penalised discriminant analysis (Hastie *et al.*, 1995)

When using many covariates (*e.g.* gene expression data), LDA might be too flexible and thus unstable.
$\rightsquigarrow$ Solution must be regularised: Use penalised within-class matrix (*cf.* ridge regression)

$$\mathbf{W}^{PDA} = \mathbf{W} + \lambda \mathbf{\Omega},$$

Flexible discriminant analysis (Hastie *et al.*, 1994)

Sometimes LDA (and even QDA) is too rigid.
$\rightsquigarrow$ More flexible solution possible by building discriminant analysis around a non-parametric regression technique.

Some other popular classification methods

- Nearest Neighbours
- Logistic Regression
- Classification Trees
- Support Vector Machines
- Neural Networks
- ...

Modifications cont.

- LDA combined with PLS methodology:

  – References:

    - *Ulf G. Indahl, Harald Martens and Tormod Næs:*
      – "*From dummy regression to prior probabilities in PLS-DA*", Journal of Chemometrics Volume 21 Issue 12, Pages 529 – 536.

    - Kristian Hovde Liland, Ulf Geir Indahl.
      – "*Powered partial least squares discriminant analysis*", Journal of Chemometrics, Volume 23 Issue 1, Pages 7 – 18.

NORWEGIAN UNIVERSITY OF LIFE SCIENCES

References

- Fisher, *The use of multiple measurements in taxonomic problems*, Eugenics – 1936.
- Duda, Hart & Stork, *Pattern Classification*, Wiley – 2001.
- Hastie, Tibshirani & Friedman, *The Elements of Statistical Learning*, Springer – 2001.
- Ludger Evers, *Decision theory and discriminant Analysis* (*lecture notes)* – University of Oxford – 2004.
- Friedman, *Regularized discriminant analysis*, JASA – 1989.
- Hastie, Buja & Tibshirani, *Penalized Discriminant Analysis*, Annals of Statistics – 1995.
- Hastie, Tibshirani & Buja, *Flexible Discriminant Analysis*, JASA – 1994.