Supervised Classification/Discrimination

- Output y: Class (either two or more)

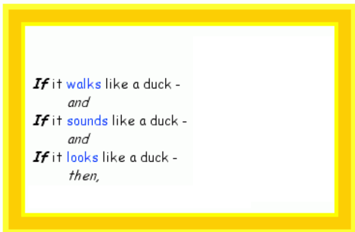
© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Supervised Classification/Discrimination

- Simple OLS/MLR based (eg. with ys=+/-1) (Logistic regression)
- LDA, QDA, (Bayes method)
- Fisher/CVA
- K-nearest neighbour
- ”X-PCA-versions” of the above
- PLS regression (=PLSDA=D-PLS)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Classification – basic aim

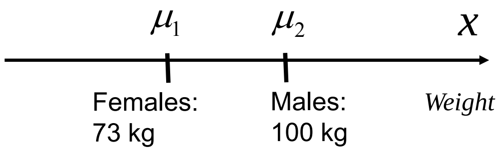


© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Classification – basic aim

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Simple example



Observe a weight of a new person:

Task: predict the sex of this new person (”classify”)

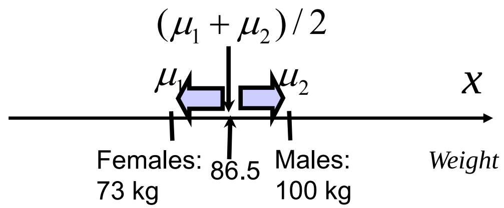
© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Simple example

”Obvious” method:

IF weigth close to female level: $c_{new} = 1(\text{female})$

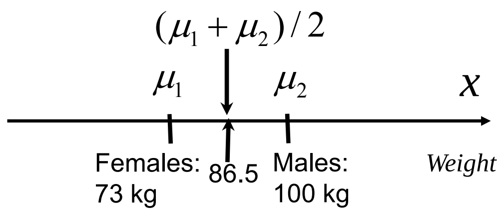
IF weigth close to male level: $c_{new} = 2(\text{male})$



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

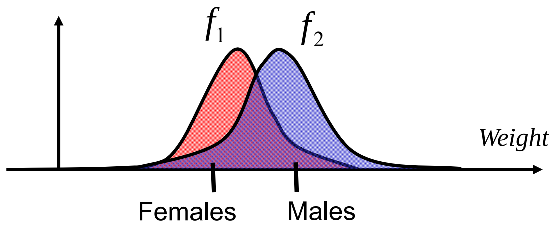
Allocation rule

$$c_{new} = \begin{cases} 1, \text{if } x_{new} < (\mu_1 + \mu_2)/2 \\ 2, \text{if } x_{new} > (\mu_1 + \mu_2)/2 \end{cases}$$



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

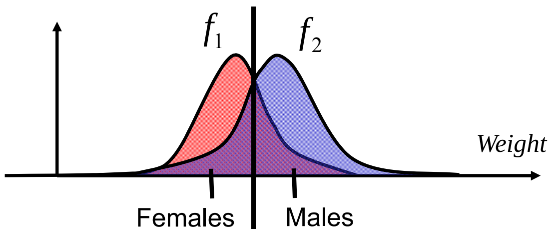
Considering population distributions:



Allocation rule: The most ”probable”!

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

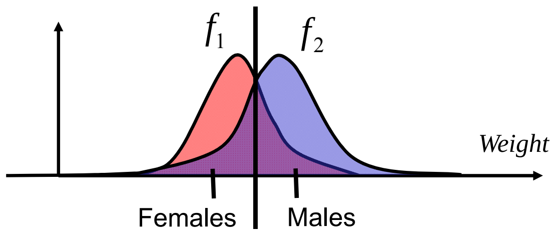
Allocation rule



$$c\,(x_{new}) = \begin{cases} 1, \text{if } f_1(x_{new})/f_2(x_{new}) > 1 \\ 2, \text{if } f_1(x_{new})/f_2(x_{new}) < 1 \end{cases}$$

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Allocation rule



IF both distributions are normal with the same variance:

$$f_1(x) = f_2(x) \Leftrightarrow x = (\mu_1 + \mu_2)/2$$

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Classification

- **Complications:**
  - Not the same cost of mis-classifying in class 1 or 2
  - Prior knowledge available about the class sizes (NOT fifty-fifty)
  - NOT the same (co)variances
- **Extensions:**
  - Multivariate feature observations
  - More than two classes

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Loss function for wrong classification

		Predicted class	
		1	**2**
True Class	**1**	**0**	L(1|2)
	2	L(2|1)	**0**

In this course: We work with equal losses: L(1|2)=L(2|1)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

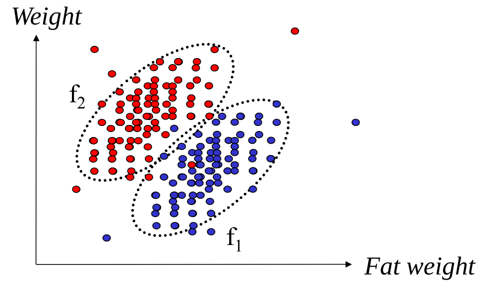
Different priors

- **Take knowlegde of class sizes into account:**
  - Prior probilities P(x|1) and P(x|2)
  - Usually: n1/(n1+n2) and n2/(n1+n2)
- **Allocation rule:**

$$c(x)=\begin{cases}1,\text{if } f_1(x)/f_2(x)>P(x|2)/P(x|1)\\ 2,\text{if } f_1(x)/f_2(x)<P(x|2)/P(x|1)\end{cases}$$

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Two class bivariate example



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Discriminant Analysis

- **Normal distributed data**

$$\mathrm{f}\left(\underline{\mathbf{x}}\right)=\frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}|}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

- **To classify we must study:**

$$f_1(x)/f_2(x)$$

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Discriminant Analysis

- **$\Sigma_1 = \Sigma_2$: the points distribution pattern co(variance) matrix**

  Linear discriminant function (LDA)

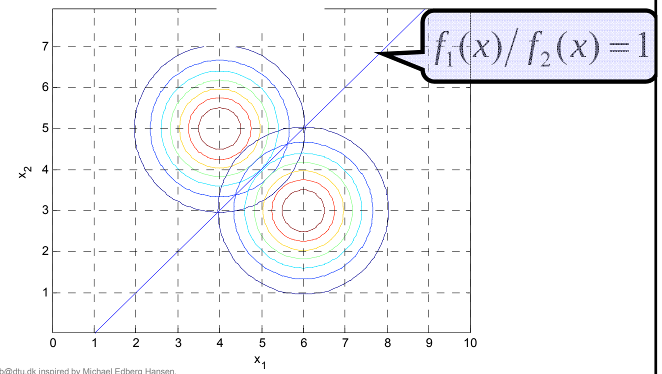
- **$\Sigma_1 \neq \Sigma_2$**

  Quadratic discriminant function (QDA)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Bivariate normal with homogeneous variance

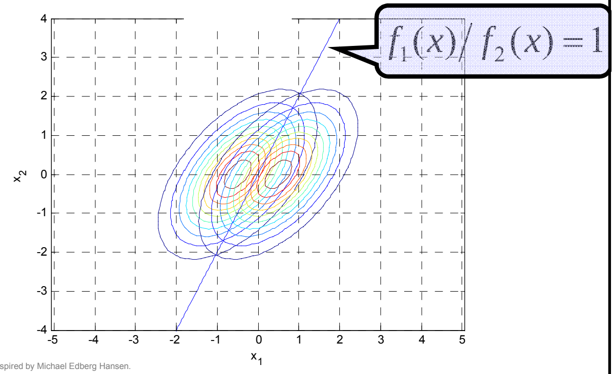
- Same x1 and x2 variance and no correlation:



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

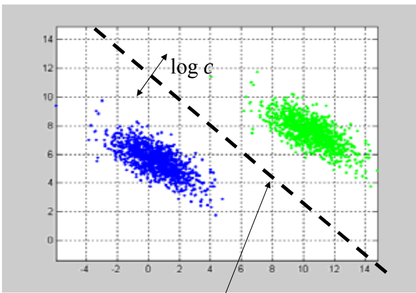
Bivariate normal with homogeneous variance

- WITH correlation:



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

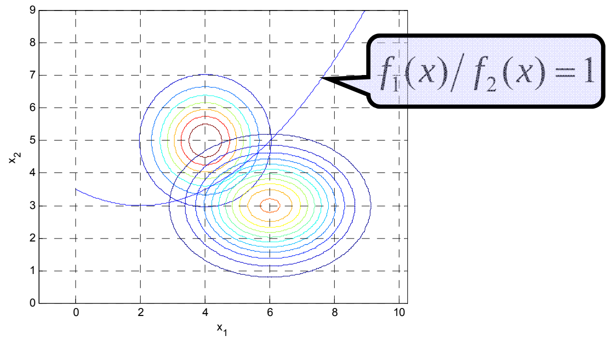
Use data to estimate distributions:



$$\underline{\mathbf{x}}^T \underline{\underline{\mathbf{\Sigma}}}^{-1}\left(\underline{\boldsymbol{\mu}}_1 - \underline{\boldsymbol{\mu}}_2\right) - \frac{1}{2}\underline{\boldsymbol{\mu}}_1^T \underline{\underline{\mathbf{\Sigma}}}^{-1}\underline{\boldsymbol{\mu}}_1 + \frac{1}{2}\underline{\boldsymbol{\mu}}_2^T \underline{\underline{\mathbf{\Sigma}}}^{-1}\underline{\boldsymbol{\mu}}_2 - \log c = 0$$

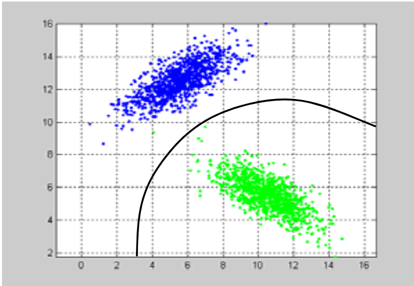
© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Bivariate normal with heterogeneous covariance



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Use data to estimate distributions:



$$\left(\underline{\mathbf{x}}-\underline{\boldsymbol{\mu}}_1\right)^T \underline{\underline{\boldsymbol{\Sigma}}}_1^{-1}\left(\underline{\mathbf{x}}-\underline{\boldsymbol{\mu}}_1\right)-\left(\underline{\mathbf{x}}-\underline{\boldsymbol{\mu}}_2\right)^T \underline{\underline{\boldsymbol{\Sigma}}}_2^{-1}\left(\underline{\mathbf{x}}-\underline{\boldsymbol{\mu}}_2\right)-2\log c=0$$

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

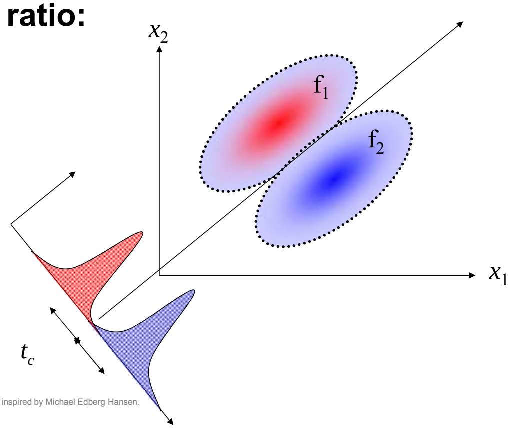
Supervised Classification/Discrimination

- Simple OLS/MLR based (eg. with ys=+/-1) (Logistic regression)
- LDA, QDA, (Bayes method)
- Fisher/CVA
- K-nearest neighbour
- ”X-PCA-versions” of the above
- PLS regression (=PLSDA=D-PLS)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

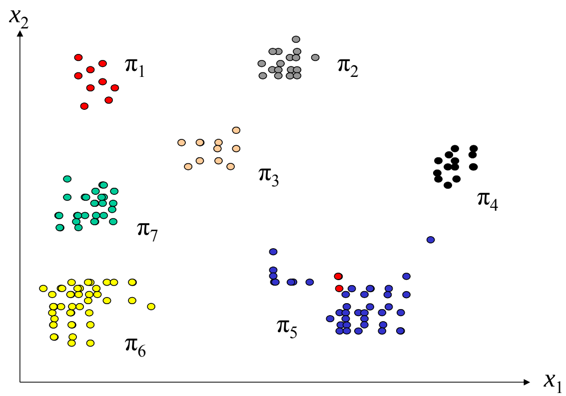
Fisher Discriminant Analysis (CVA)

- **Optimal ratio:**



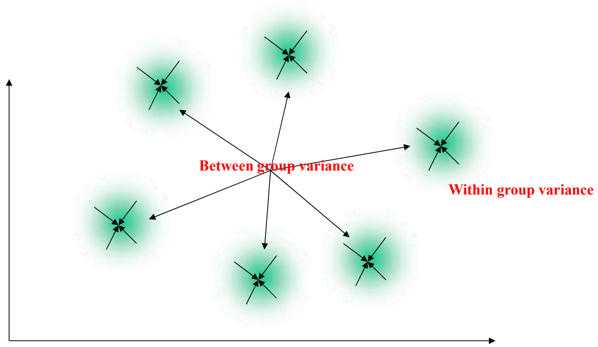
© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Multiclass class problems



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

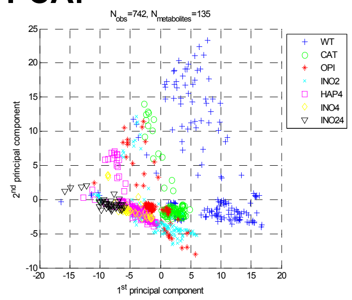
Fisher approach (LDA)



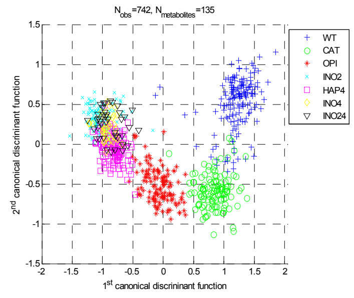
© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Example

- 742 obs: 7 mutants and 600 ions.
- PCA:



CVA:



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Relations

- **Fisher:**
  - Finds univariate discrimination score
  - Maximizes ANOVA F-ratios
- **”General Bayes”:**
  - Finds points of ”equi-probability”

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Relations

For 2 classes, normal distribution and variance homogeneity:

- All three methods are the same! (=LDA without prior)

For 2 classes, normal distribution and variance heterogeneity:

- Bayes = QDA
- (Can be mimicked by an LDA with squared and cross product terms)

- **For MULTIclasses, normal distribution and variance homogeneity:**
  - Full dimensional Fisher=LDA without prior

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

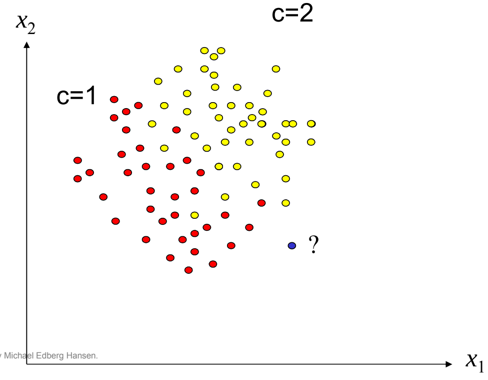
k-Nearest Neighbour method

- **The k-NN classifier is a very intuitive method**
  - Examples are classified based on their similarity with training data.
- **The k-NN only requires:**
  - An integer *k*.
  - A set of labeled examples.
  - A measure of "closeness".

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

k-Nearest Neighbour method

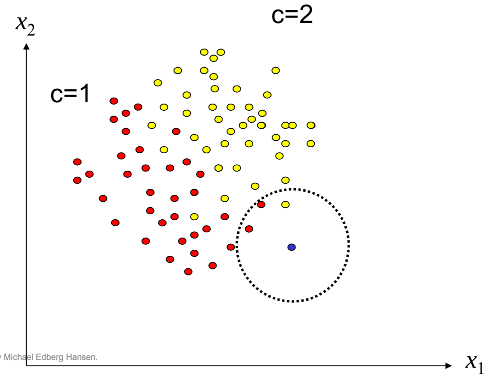
- **3-NN**



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

k-Nearest Neighbour method

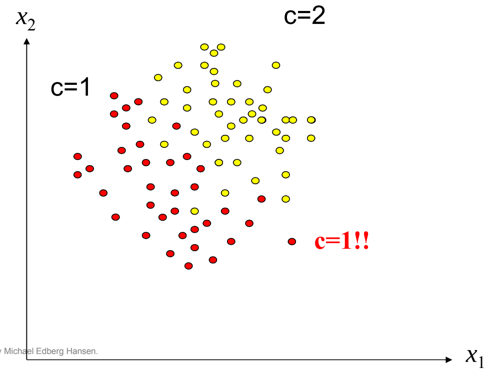
- **3-NN**



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

k-Nearest Neighbour method

- **3-NN**



© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

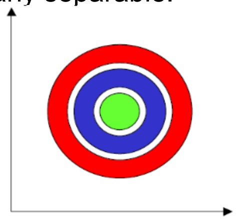
Characteristics of the NN classifier

- **Advantages**
  - Analytically tractable, simple implementation
  - Nearly optimal in the large sample limit ($N \rightarrow \infty$).
- **Disadvantages**
  - Large storage requirements
  - Computationally intensive recall
  - Highly susceptible to the curse of dimensionality
- **1-NN versus k-NN**
  - Dertermine by cross-validation

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

k-Nearest Neighbour method

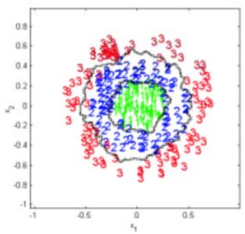
- **Data is generated for a 2-dim 3-class problem, where the likelihoods are unimodal, and are distributed in rings around a common mean.**
  - These classes are also non-linearly separable.
- **k-NN with**
  - $k = 5$
  - Metric = Euclidean distance

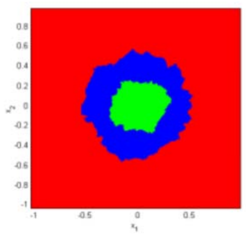


© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

5-Nearest Neighbour

- **Solution**





© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Link to regression and PLS

- **Two group LDA can be (exactly) obtained by simple regression**
- **Just use the binary y's in an MLR**
- **A version of QDA (not 100% equivalent) can be obtained by including product and squared variables in the MLR**
- **PLSR and/or PCR is also a good idea!**
- **Multi-class discrimination can be handled by PLS2**
  - Make K dummy variables (easy/inbuilt in Unscrambler)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Some prominent methods NOT covered

- Classification Trees (CART)
- Random forests
- Neural Networks (NN)
- Support Vector Machines (SVM)
- Various regularized/sparse methods
-

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.

Supervised Classification/Discrimination

- Simple OLS/MLR based (eg. with ys=+/-1) (Logistic regression)
- LDA, QDA, (Bayes method)
- Fisher/CVA
- K-nearest neighbour
- ”X-PCA-versions” of the above
- PLS regression (=PLSDA=D-PLS)

© Per Bruun Brockhoff, perbb@dtu.dk inspired by Michael Edberg Hansen.