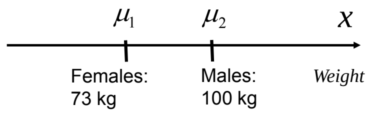
Classification

- **Simple allocation rules**
- **Extensions**
  - Not the same costs of mis-classifications
  - Prior knowledge available
  - NOT the same (co)variances
  - Multivariate feature observations
- **Other methods:**
  - K-nearest neighbour
  - PLS regression
- **Multiclass problems**

Classification – basic aim

Classification – basic aim

Simple example



Observe a weight of a new person:

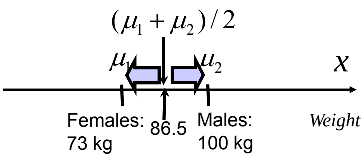
Task: predict the sex of this new person (”classify”)

Simple example

”Obvious” method:

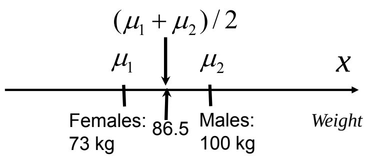
IF weigth close to female level: $c_{new} = 1(\text{female})$

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

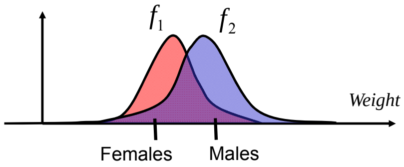


Allocation rule

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

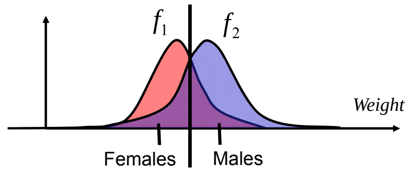


Considering population distributions:



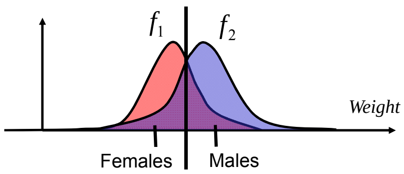
Allocation rule: The most ”probable”!

Allocation rule



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

Allocation rule



IF both distributions are normal with the same variance:

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

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

Loss function for wrong classification

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

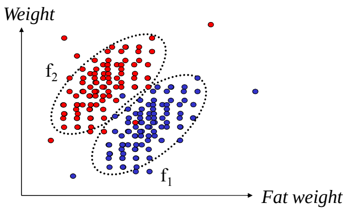
In this course: We work with equal losses: L(1|2)=L(2|1)

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:**

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

Two class bivariate example



Discriminant Analysis

- Normal distributed data

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

- To classify we must study:

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

Discriminant Analysis

- $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$:

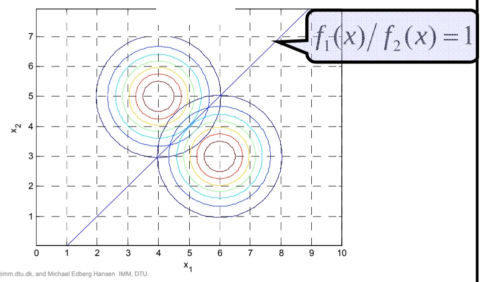
  Linear discriminant function (LDA)

- $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$

  Quadratic discriminant function (QDA)

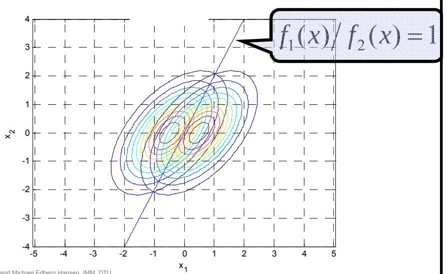
Bivariate normal with homogeneous variance

- Same x1 and x2 variance and no correlation:



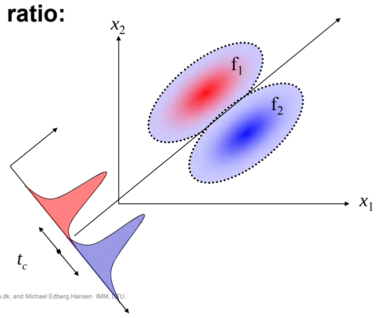
Bivariate normal with homogeneous variance

- WITH correlation:

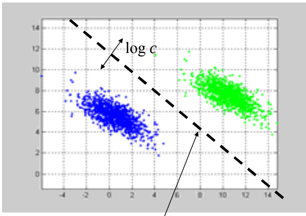


Fisher Discriminant Analysis (CVA)

- Optimal ratio:

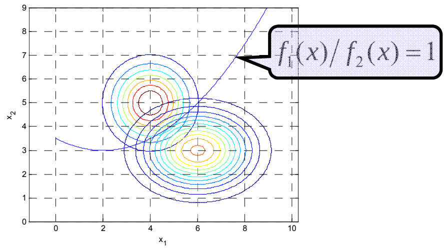


Use data to estimate distributions:

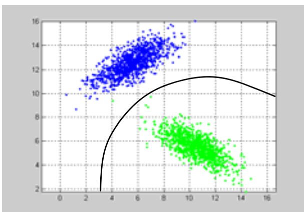


$$\underline{x}^T \underline{\underline{\Sigma}}^{-1}(\underline{\mu}_1-\underline{\mu}_2) - \frac{1}{2}\underline{\mu}_1^T \underline{\underline{\Sigma}}^{-1}\underline{\mu}_1 + \frac{1}{2}\underline{\mu}_2^T\underline{\underline{\Sigma}}^{-1}\underline{\mu}_2 - \log c = 0$$

Bivariate normal with heterogeneous covariance



Use data to estimate distributions:



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

Relations

- **Mahalanobis:**
  - Finds points of "equidistance"
- **Fisher:**
  - Finds univariate discrimination score
- **"General Bayes":**
  - Finds points of "equi-probability"

Relations

For 2 classes, normal distribution and variance homogeneity:
- All three methods are the same! (=LDA)

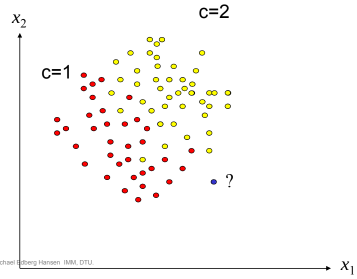
For 2 classes, normal distribution and variance heterogeneity:
- Bayes = Mahalanobis = QDA
- (Can be mimicked by an LDA with squared and cross product terms)

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".

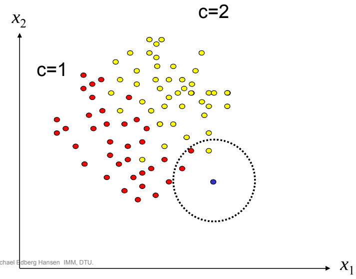
k-Nearest Neighbour method

- 3-NN



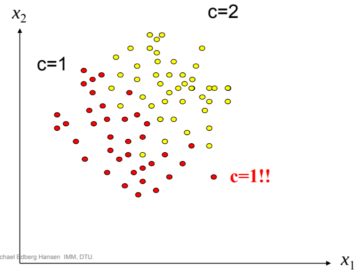
k-Nearest Neighbour method

- 3-NN



k-Nearest Neighbour method

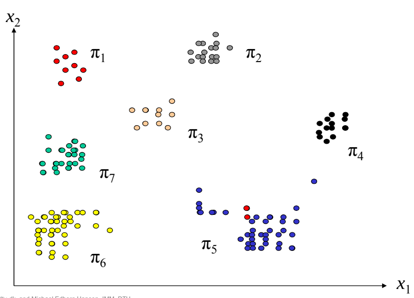
- 3-NN



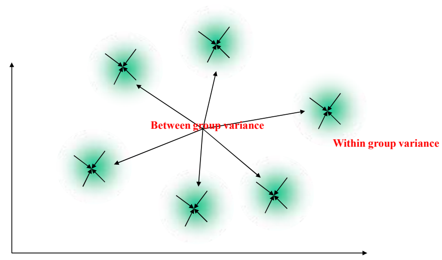
Characteristics of the NN classifier

- **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

Multiclass class problems

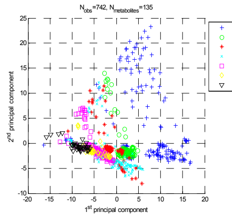


Mahalanobis/Fisher approach (LDA)

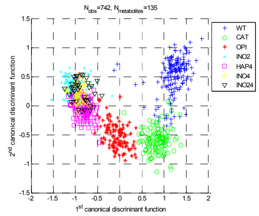


Example

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

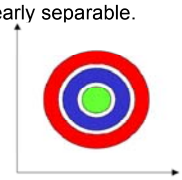


CVA:



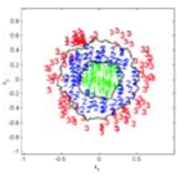
k-Nearest Neighbour method

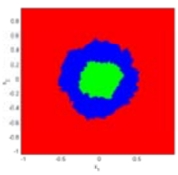
- **Data is generated for a 2-dim 3-class problem, where the likelihoods are unimodal, and are distributed in rings around a common mean.**
  - These classes are also non-linearly separable.
- **k-NN with**
  - *k* = 5
  - Metric = Euclidean distance



5-Nearest Neighbour

- **Solution**





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)

Classification in Unscrambler X

- **Linear Discriminant Analysis**
  - LDA: Multiclass, multivariate X (assuming n_i>p)
  - QDA
  - (Mahalanobis)
  - All may be combined with PCA (PCA-LDA option)
- **PLS1:** 2-group LDA, PLS-DA (a "PLS" version of PCA-LDA)
- **PLS2:** multi-group LDA, PLS-DA
- The PLS-methods can by "QDA-like" by inclusion of squared/crossed variables.

Classification

- **Simple allocation rules**
- **Extensions**
  - Not the same costs of mis-classifications
  - Prior knowledge available
  - NOT the same (co)variances
  - Multivariate feature observations
- **Other methods:**
  - K-nearest neighbour
  - PLS regression
- **Multiclass problems**