Statistical image analysis, 593B

Finn Lindgren, finn@maths.lth.se
Visiting Assistant Professor of Statistics
University of Washington, Seattle, WA, USA

Usually:
Research Associate in Mathematical Statistics
Centre for Mathematical Sciences
Lund University, Sweden

L1, 2007-03-26

LUND UNIVERSITY

Lectures and computer exercises

	Day of week	Time	Room
Lecture	Monday	10.30–12.20	PDL C301
Computer exercise	Wednesday	10.00–12.50	CMU B027
Exercises	T.B.A.		
Lecture	Friday	10.30–12.20	PDL C301

Course web-homepage: Current information and Matlab files will be available at
http://www.maths.lth.se/matstat/kurser/fms150mas228/uw07/

Literature: "Image Modelling and Estimation, A statistical approach", 2006 edition, by Finn Lindgren. This compendium, is available for download in PDF form from the course web-page. The lecture notes will also be made available from the web-page. Note that the final chapters of the compendium are being rewritten during the course and will be made available as they are done.

Projects

The examination takes the form of three small project assignments, to be solved in groups of two (or individually) students each. For all 3 projects, a written report shall be completed. The final project will also be presented to the other students in the form of a short (ca 10 min) seminar.

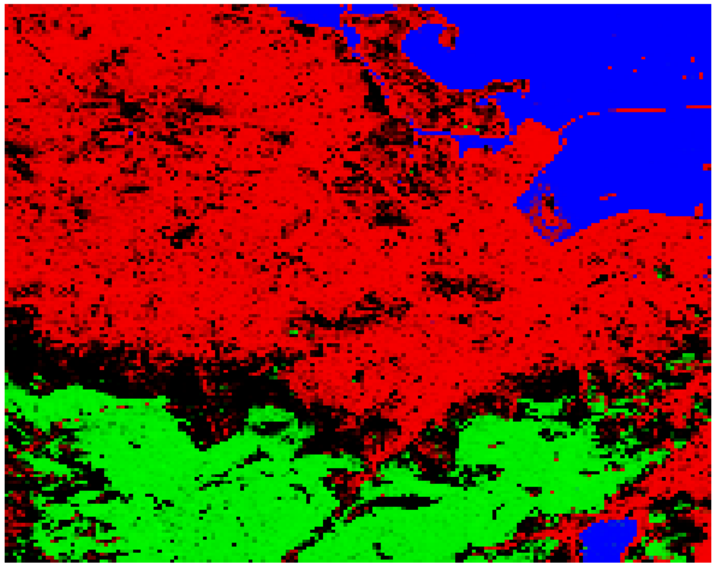
Project 1: EM and classification/segmentation.
Project 2: Markov random fields.
Project 3: Shape analysis, warping, super-resolution, parameter estimation etc.

Course contents

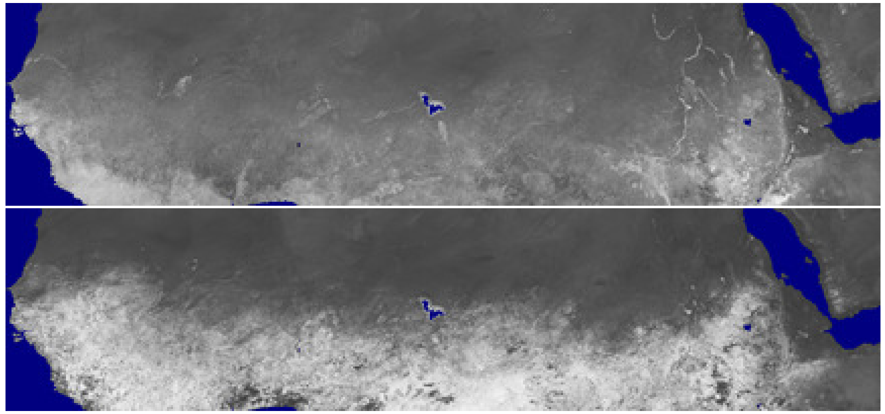
- Bayesian methods for modelling, classification, and reconstruction.
- Correlation structures, multivariate techniques, discriminant analysis.
- Markov fields, Gibbs-distributions, spatial statistics, reconstruction of fields using sparse measurements.
- Simulation methods for inference (MCMC).
- Statistical shape analysis, deformable templates.
- Warping and morphing.
- Applications such as remote sensing and environmental problems, surveillance, and general methodology.

Classification

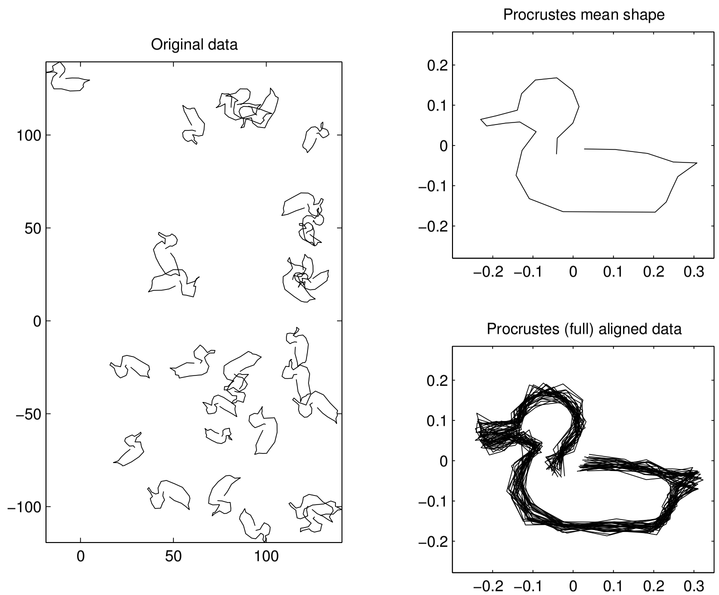
Classification



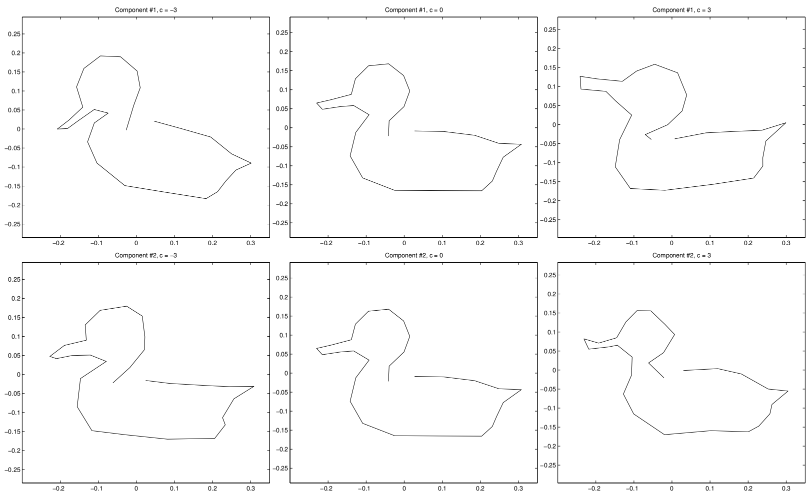
Random fields



Shape analysis



Shape analysis



Preliminary schedule

Week 1/13 (26/3–30/3)

L1: Introduction; Statistical modelling
C1: Images and probability densities in Matlab
L2: Modelling and classification I

Week 2/14 (2/4–6/4)

L3: Classification II; The EM-algorithm
The first project assignment is handed out
C2: Classification
E1: Exercises; Bayesian statistics
L4: Markov fields I; Gaussian fields

Week 3/15 (9/4–13/4)

L5: Markov fields II; Gaussian fields
C3: Gaussian Markov random fields
L6: Markov fields III; Discrete Markov fields

Week 4/16 (16/4–20/4)

L7: Markov fields IV; MCMC simulation
The second project assignment is handed out
C4: MRF-simulation, Discrete Markov Fields
E2: Exercises; Markov fields
L8: Deformable templates I; Procrustes

Week 5/17 (23/4–27/4)

L9: Deformable templates II; Shape-PCA
C5: Deformable templates
L10: Deformable templates III; Warping, Image-PCA

Week 6/18 (30/4–4/5)

L11: Deformable templates IV; Shapes from images
The third project assignment is handed out
C6: Warping and Image-PCA
E3: Exercises; Statistical shape analysis
L12: Weighted PCA, Snake-estimates

Week 7/19 (7/5–11/5)

L13: Snake estimation; Repetition; Comments
C7: Snakes
L14: Reserved

Week –/–

Seminars for project 3.

A statistical approach

- Often, we have some **prior knowledge** of the reality.
- How can natural variation be described and modelled?
- Given knowledge of the true reality, what can we predict about images and other data?
- How can we best guess "future" behaviour using past data?
- How can measurements be used to make inference about the reality?

Bayesian modelling

- We assume that there is some unknown truth, that we would like to find out about. This "reality" can be measured, usually with measurement variation. Often, only partial observations can be obtained, in the sense that the information would be incomplete *even if the observations were perfect*.
- The following components are used to combat this problem, in the framework of Bayesian modelling:
- A prior, "a priori", model for reality, $\boldsymbol{x}$, given by the probability density $\pi(\boldsymbol{x})$. This means that we view the (current) reality as a random variable, drawn from a set of possible realities, $\boldsymbol{\Omega}$.
- A conditional model for data, $\boldsymbol{y}$, given reality, with density $p(\boldsymbol{y}|\boldsymbol{x})$.

Bayes' Formula

- How should the prior and the data model be used to make statements about the reality $\boldsymbol{x}$, given observations of $\boldsymbol{y}$?
- Bayes' Formula/Rule/Theorem for continuous $\boldsymbol{x}$-distributions:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})}{\int_{\boldsymbol{x}'\in\boldsymbol{\Omega}} p(\boldsymbol{y}|\boldsymbol{x}')\pi(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'}$$

$p(\boldsymbol{x}|\boldsymbol{y})$ is called the posterior, or "a posteriori", distribution.

Bayes' Formula

- Bayes' Formula for discrete $\boldsymbol{x}$-distributions:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})}{\sum_{\boldsymbol{x}'\in\boldsymbol{\Omega}} p(\boldsymbol{y}|\boldsymbol{x}')\pi(\boldsymbol{x}')}$$

- The total distribution for data, over all possible outcomes of reality, is $p(\boldsymbol{y})$. As we will see during the course, it is not always necessary to compute this complicated distribution.
- Often, only the proportionality relation

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{x},\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})$$

is needed, when seen as a function of $\boldsymbol{x}$.

A simple example

- Three closed boxes, 1, 2, and 3, contain different amounts of very small rocks and diamonds. Box 1 contains no diamonds, box 2 50% rocks and 50% diamonds, and box 3 contains only diamonds.
- Pick a box at random with equal probability, and pick a piece from the contents. If you happened to receive a diamond, what is the probability that you have picked box 3?
- The "reality" x is the number of the box.
- The measurement y is the indicator of whether you received a diamond.
- Use Bayes' formula to compute the probability.

The prior and data probabilities

$$\pi(x = k) = 1/3, \quad k = 1, 2, 3.$$

$$p(y = i|x = 1) = \begin{cases} 1, & i = 0 \\ 0, & i = 1 \end{cases}$$

$$p(y = i|x = 2) = \begin{cases} 1/2, & i = 0 \\ 1/2, & i = 1 \end{cases}$$

$$p(y = i|x = 3) = \begin{cases} 0, & i = 0 \\ 1, & i = 1 \end{cases}$$

The posterior probabilities

The posterior probabilities for box 3 are

$$p(x = 3|y = 0) = 0$$

$$p(x = 3|y = 1) = \frac{p(y = 1|x = 3)\pi(x = 3)}{\sum_{x'=1}^{3} p(y = 1|x = x')\pi(x = x')} = \frac{1 \cdot 1/3}{0 \cdot 1/3 + 1/2 \cdot 1/3 + 1 \cdot 1/3} = 2/3$$

The total data probabilities are

$$p(y = 0) = 1 \cdot 1/3 + 1/2 \cdot 1/3 + 0 \cdot 1/3 = 1/2$$
$$p(y = 1) = 0 \cdot 1/3 + 1/2 \cdot 1/3 + 1 \cdot 1/3 = 1/2$$

Questions: Where did the prior probabilities come from? How should the prior probabilities be constructed in realistic applications?

A hierarchical model

A common use of Bayesian modelling is to apply prior distributions to the parameters in a standard model. This can be illustrated by the following example.

- The diamonds in the previous example are not all of the same size. Consider a situation where someone is offering to give you a diamond, but only if you can guess its weight simply by looking at it. You are allowed to let as many people as you like help you guess.
- A simple model: Let x be the true weight, and let y_i, $i = 1, \ldots, n$ be the guesses of each person.
- For simplicity, assume that the guesses are Gaussian and unbiased, so that $y_i \in \mathsf{N}(x, \sigma^2)$.
- The Maximum Likelihood (ML) estimate of the weight x is the mean of the guesses, $\overline{y}$.

Using prior information

- How can we use the knowledge that diamonds do not come in arbitrarily small or large sizes?
- Assume that we, from previous experience, know that diamonds in general have a weight that is $\mathsf{N}(\mu, \tau^2)$, for known values of μ and τ^2.
- Exercise 2.4 shows that

$$(x|y_1, \ldots, y_n) \in \mathsf{N}\left(\bar{y}\frac{\tau^2}{\tau^2+\sigma^2/n} + \mu\frac{\sigma^2/n}{\tau^2+\sigma^2/n}, \frac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n}\right).$$

The Maximum A Posteriori estimate of $\boldsymbol{x}$ is pulled toward μ compared to the ML estimate, so that the guessing uncertainty is moderated by the prior information that diamonds have an expected weight μ.

Estimation procedures

Given a prior and a data model, there are many possibilities for extracting information about the underlying reality.

- Maximum A Posteriori (MAP): Choose the $\boldsymbol{x}$ with largest posterior probability (density) $p(\boldsymbol{x}|\boldsymbol{y})$.
  - Solve analytically (often impossible)
  - Standard numerical optimisation methods
  - Specialised procedures, using the model structure
- Posterior Mean: Choose the expectation in the posterior distribution as an estimate of the true $\boldsymbol{x}$. Often requires simulating samples from the posterior distribution.
- Simulation: Simulate samples from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$. Estimate statistical properties from these samples. The samples can be seen as representative "possible realities", given the available data.
  - Markov chain Monte Carlo (MCMC) simulation

Exercises

- Before C1, read pages 13-17 in the book, in particular Example 2.3, that deals with treating parameters as random variables, and investigating the posterior distributions given simple prior distributions. It is also a good idea to study the "Matlab tips".
- Before E1, try to solve exercises 2.1–2.6. Hopefully 2.1 and 2.2a) are familiar from the basic course in statistics.