Statistical image analysis, 7.5hp **Written exam**
TMS016/MSA300 **2018-05-30, 08:30 - 12:30**

Teacher and Jour: David Bolin, phone 772 53 75.

Allowed material: Chalmers allowed calculator.

Grading: Correct and well-motivated solutions (the first question only requires an answer) give the points that are written in parenthesis at each question. You can in total get 20 points on the exam, and 20 points for the project assignments. To pass the course, you must have completed all project assignments and given a project seminar. The limits for the grades are:
GU: 20 and 31 points for "G" and "VG" respectively.
CTH: 20, 28, and 34 points for 3,4, and 5 respectively.

Answers can be given in English or Swedish.

---

1. Answer the following statements with one of the choices "true", "false", or "I do not know". For each statement, a correct answer gives 1 point, an incorrect answer (e.g. answering "true" when the statement is false) gives -0.5 points, and "I do now know" gives 0 points. (6p)

   (a) If a random field is Gaussian, then it is strongly stationary if and only if it is weakly stationary.

   (b) Let $\mathbf{X} \sim \mathsf{N}(\mathbf{0}, \mathbf{\Sigma})$ be a Gaussian Markov random field with respect to a graph $\mathcal{G} = (V, E)$. Assume that $\{i, j\} \notin E$, then the element at position (i, j) in $\mathbf{\Sigma}^{-1}$ is zero.

   (c) Let A be a set of pixels in an image, and let S_{ij} be a structure element centered at pixel (i, j), then the set $\{(i, j) : S_{ij} \subseteq A\}$ is called the opening of A.

   (d) Let $N(A)$ denote the number of points in the set A for a point process. Then $E(N(A))$ is the intensity function of the point process.

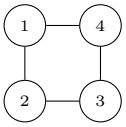
   (e) The black curve in Figure 1 shows the estimated K-function for a point pattern of locations for cancer cases in England. This estimate indicates that the point pattern is more clustered than what one would expect for a homogeneous Poisson process.

   (f) When using feedforward neural networks, the input data must have a lattice structure.

2. (a) Let $Y_1, \ldots, Y_n$ be observations taken at locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ of a Gaussian random field $X(s)$ with mean-value $\mu(\mathbf{s}) = \sum_{i=1}^{K} \beta_i B_i(\mathbf{s})$ and stationary covariance function $r(\mathbf{h}; \boldsymbol{\theta})$. Here $B_1(\mathbf{s}), \ldots B_K(\mathbf{s})$ are known covariates and $\boldsymbol{\theta}$ is a vector with parameters for the covariance function. In the course we went through two likelihood-based methods for estimating the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, the profile-likelihood approach and the restricted ML (REML) approach. Describe how the profile-likelihood approach works. (2p)

   (b) What is the advantage with using REML instead of regular ML estimation? Is there some disadvantage with the method? (1p)

   (c) Assume that we have an image with only four pixels, modelled as a discrete Markov random field $\mathbf{Z} = (Z_1, \ldots, Z_4) \sim \pi(\mathbf{Z})$, defined on the undirected graph

   What are the cliques of this graph? (1p)

   (c) To estimate the parameters of the Markov random field $\mathbf{Z}$, we used a pseudo-likelihood method. Define the pseudo-likelihood and explain why we in general had to use this instead of the usual likelihood. (2p)

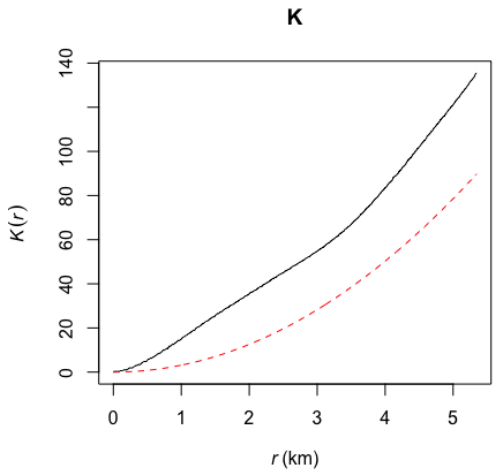


Figure 1: The estimated K function for a point pattern of spatial locations for cancer cases in England (solid line) as well as the theoretical K function for a homogeneous Poisson process (dashed line).

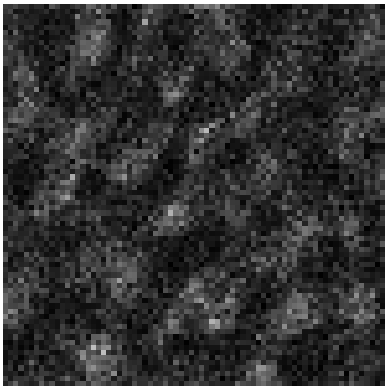


Figure 2: An image of a porous material that is used as coatings for pharmaceutical tablets.

3. Figure 2 shows a microscopy image of a part of a porous material that is used as coatings for pharmaceutical tablets. The drug release rate of the tablet depends on the pore structure of the material, and to get an estimate of the release rate it is therefore important to be able to segment the image to find the pores. In the image, dark pixels corresponds to pores, and brighter pixels correspond to the solid material. There is, however, quite a bit of noise in the image which makes it difficult to see where the pores are.

   (a) Propose a method which you think would work well for segmenting the image to find the pores. (2p)

   (b) An important characteristic for the release rate is the pore volume fraction, i.e. the relative amount of pores in the material. This quantity can be controlled in the manufacturing process, and for this particular material it is 30%. Propose a segmentation method that takes this information into account to improve the segmentation. (2p)

4. Suppose that we have a data set of 1000 labelled images of handwritten zeros and ones, as in Figure 3.

   (a) Describe some image features that can be useful for distinguishing between images showing zeros and ones. Suppose that you now want to classify a new, unlabelled, image using

Figure 3: Eight images of handwritten zeros and ones.

these features with the K nearest neighbor approach. Describe how this is done. (2p)

(b) Describe a method for how you can find a good value of K to use for the K nearest neighbor classifier based on the 1000 labelled images. (2p)

Good luck!