Statistical image analysis, 7.5hp
TMS016/MSA301

Solutions to written exam
2019-06-05, 08:30 - 12:30

1. (a) True.

   (b) False. REML provides unbiased estimates of $\boldsymbol{\theta}$, but they typically have higher variance.

   (c) True.

   (d) False. This is the image closing.

   (e) False. The backpropagation algorithm is used to compute the gradient of the loss function.

2. (a) The function needs to be positive definite. This means that for any finite $n > 1$ and for any choice of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, the matrix $\mathbf{\Sigma}$ with elements $\Sigma_{ij} = r(\mathbf{s}_i, \mathbf{s}_j)$ is positive semidefinite: $\mathbf{c}^T\mathbf{\Sigma}\mathbf{c} \geq 0$ for any vector $\mathbf{c} \in \mathbb{R}^n$.

   (b) If the field is stationary, we have that $r(\mathbf{s}+\mathbf{h}, \mathbf{t}+\mathbf{h}) = r(\mathbf{s}, \mathbf{t})$ for any vector $\mathbf{h} \in \mathbb{R}^2$. This means that $r(\mathbf{s}, \mathbf{t})$ only depends on the difference $\mathbf{s} - \mathbf{t}$. If the field also is isotropic, $r(\mathbf{s}, \mathbf{t})$ only depends on the distance between the points, $\|\mathbf{s} - \mathbf{t}\|$.

   (c) We have to show that $r_Y(\mathbf{s}, \mathbf{t})$ is positive definite. Take $n > 1$, a set of locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, and define the matrices $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_\varepsilon$, with elements $\Sigma_{ij} = r(\mathbf{s}_i, \mathbf{s}_j)$ and $\Sigma_{\varepsilon,ij} = r_\varepsilon(\mathbf{s}_i, \mathbf{s}_j)$ respectively. We then have that $\mathbf{\Sigma}_\varepsilon = \sigma_e^2\mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Take any vector $\mathbf{c} \in \mathbb{R}^n$, we now have to show that $\mathbf{c}^T\mathbf{\Sigma}_Y\mathbf{c} \geq 0$, where $\mathbf{\Sigma}_Y = \mathbf{\Sigma} + \mathbf{\Sigma}_\varepsilon$. We have:

   $$\mathbf{c}^T\mathbf{\Sigma}_Y\mathbf{c} = \mathbf{c}^T\mathbf{\Sigma}\mathbf{c} + \mathbf{c}^T\mathbf{\Sigma}_\varepsilon\mathbf{c}.$$

   Since r is a covariance matrix, we have that $\mathbf{c}^T\mathbf{\Sigma}\mathbf{c} \geq 0$. Further

   $$\mathbf{c}^T\mathbf{\Sigma}_\varepsilon\mathbf{c} = \mathbf{c}^T\sigma_e^2\mathbf{I}_\varepsilon\mathbf{c} = \sigma_e^2\mathbf{c}^T\mathbf{c} = \sigma_e^2\sum_{i=1}^n c_i^2 \geq 0.$$

   Thus, $\mathbf{c}^T\mathbf{\Sigma}_Y\mathbf{c}$ is a sum of two non-negative terms and thus non-negative.

3. (a) In logistic regression we assume that the probabilities for the classes are

   $$P(z = k|\mathbf{y}) = \begin{cases} \frac{\exp(\beta_{0,k}+\boldsymbol{\beta}_k^T\mathbf{y})}{1+\sum_{\ell=1}^{K-1}\exp(\beta_{0,\ell}+\boldsymbol{\beta}_\ell^T\mathbf{y})} & \text{if } k < K. \\ \frac{1}{1+\sum_{\ell=1}^{K-1}\exp(\beta_{0,\ell}+\boldsymbol{\beta}_\ell^T\mathbf{y})} & \text{if } k = K, \end{cases}$$

   where $\{\beta_{0,k}, \boldsymbol{\beta}_k\}_{k=1}^K$ are parameters of the model. To estimate these parameters, we numerically maximise the conditional log-likelihood $\sum_{i=1}^n \log P(z = z_i|\mathbf{y}_i)$. This can, for example, be done using gradient-descent optimization.

   (b) The logistic regression model and the LDA model have the same forms for the probabilities $P(z = k|\mathbf{y})$. The difference is in how we estimate the models from data. For LDA we maximize the regular log-likelihood $\sum_{i=1}^n \log \pi(\mathbf{y}_i)$ to find the parameters instead of the conditional likelihood. One therefore uses the fact that the LDA model assumes that the data from a specific class is Gaussian. This improves the efficiency of the estimation if the Gaussianity assumption is satisfied. If, however, the data does not seem to be Gaussian for a given class, it is safer to use logistic regression.

4. (a) We obtain a filtered image by computing the convolution between the image pixel value $x_{i,j}$ and a filter kernel w. We could for example use a simple averaging filter with values

$w_{i,j} = 1/9$ if $-1 \leq i, j \leq 1$ and $w_{i,j} = 0$ otherwise. The filtered image is then obtained as

$$\begin{aligned}\hat{x}_{i,j} &= \sum_{k=-1}^{1} \sum_{\ell=-1}^{1} w_{k,l} x_{i-k,j-\ell} \\ &= \frac{1}{9}(x_{i-1,j-1} + x_{i-1,j} + x_{i-1,j+1} + x_{i,j-1} + x_{i,j} + x_{i+1} + x_{i+1,j-1} + x_{i+1,j} + x_{i+1,j+1}).\end{aligned}$$

Thus the value of $\hat{x}_{i,j}$ is the average over the 9 pixels values in x closest to (i, j).

(b) Let $X_{(i,j)}$ denote the value of the image a pixel (i, j). We then have that $X_{(i,j)}|X_{-(i,j)} \sim \mathsf{N}(\mu_i, \sigma^2)$, where

$$\begin{aligned}\mu_i &= \mathsf{E}(X_i|\mathbf{X}_{-(i,j)}) = \mu - \frac{1}{5}(X_{(i+1,j)} + X_{(i+1,j)} + X_{(i,j+1)} + X_{(i,j-1)} - 4\mu) \\ \sigma^2 &= \mathsf{V}(X_i|\mathbf{X}_{-i}) = \frac{1}{5}\end{aligned}$$

(c) We assume that the observed values in the pixels, $Y_{(i,j)}$ are noisy observations of the corresponding pixels in the true image, $X_{(i,j)}$, which we model using the GMRF. Thus, $Y_{(i,j)} = X_{(i,j)} + \varepsilon_{(i,j)}$, where $\varepsilon_{(i,j)}$ are iid $\mathsf{N}(0, \sigma_e^2)$ variables.

(d) We have that the distribution of $\mathbf{X}$ conditionally on the observed values $\mathbf{Y}$ is

$$\mathsf{N}(\mu\mathbf{1} + \hat{\mathbf{Q}}^{-1}(\mathbf{Y} - \mu\mathbf{1}), \hat{\mathbf{Q}}^{-1}),$$

where $\hat{\mathbf{Q}} = \mathbf{Q} + \sigma_e^{-2}\mathbf{I}$. We use the mean value of this distribution as predictor:

$$\hat{\mathbf{X}} = \mathsf{E}(\mathbf{X}|\mathbf{Y}) = \mu\mathbf{1} + \hat{\mathbf{Q}}^{-1}(\mathbf{Y} - \mu\mathbf{1}).$$