Math Foundations of ML, Fall 2018

Homework #8

Due Friday November 2, at the beginning of class

As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

1. Using you class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

2. Suppose that we want to create a realization of Gaussian noise $\boldsymbol{e} \in \mathbb{R}^5$ with covariance matrix
$$\boldsymbol{R} = \begin{bmatrix} 1 & 1/3 & 1/9 & 1/27 & 1/81 \\ 1/3 & 1 & 1/3 & 1/9 & 1/27 \\ 1/9 & 1/3 & 1 & 1/3 & 1/9 \\ 1/27 & 1/9 & 1/3 & 1 & 1/3 \\ 1/81 & 1/27 & 1/9 & 1/3 & 1 \end{bmatrix}.$$
We have at our disposal a random number generator that creates independent and identically distributed Gaussian random variables with variance 1. We use this to generate $\boldsymbol{e}_{\text{ind}} \in \mathbb{R}^5$, and then pass the output through a matrix to give it the desired covariance structure. Find a matrix $\boldsymbol{Q}$ such that the covariance matrix of $\boldsymbol{Q}\boldsymbol{e}_{\text{ind}}$ is $\boldsymbol{R}$.

3. Let $Z[1], \ldots, Z[N]$ be a sequence of independent Gaussian random variables with mean 0 and variance 1. You observe the random vector X in $\mathbb{R}^N$ that is generated through the autoregressive process
$$X[k] = \begin{cases} Z[1], & k = 1 \\ aX[k-1] + Z[k], & k > 1. \end{cases}$$
Given $X = \boldsymbol{x}$, find the MLE for $a \in \mathbb{R}$. (Hint: Conditional independence.) (Further hint: The conditional independence structure makes this a Markov process, meaning that we can factor the distribution for $X \in \mathbb{R}^N$ as
$$f_X(\boldsymbol{x}) = f_{X[1]}(x_1) f_{X[2]}(x_2|x_1) f_{X[3]}(x_3|x_2) \cdots f_{X[N]}(x_N|x_{N-1}).$$
)

4. Let $\boldsymbol{A}$ be an $M \times N$ matrix with full column rank. Let E be a Gaussian random vector in $\mathbb{R}^M$ with mean $\boldsymbol{0}$ and covariance $\boldsymbol{R}_e$. Suppose we observe
$$Y = \boldsymbol{A}\boldsymbol{\theta}_0 + E,$$
where $\boldsymbol{\theta} \in \mathbb{R}^N$ is unknown.

(a) What is the distribution of Y and how does it depend on $\boldsymbol{\theta}_0$?

(b) Find a closed form expression for the maximum likelihood estimate of $\boldsymbol{\theta}_0$. (In this case, we are working from a single sample of a random vector.)

(c) What is the distribution of the MLE estimator $\hat{\boldsymbol{\Theta}}$? Is $\hat{\boldsymbol{\Theta}}$ unbiased?

(d) What is the MSE of the MLE, $\mathrm{E}[\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\theta}_0\|_2^2]$?

(e) Compute the Fisher information matrix $\boldsymbol{J}(\boldsymbol{\theta}_0)$ and verify that the MLE meets the Cramer-Rao lower bound.

(f) Defend the following statement: The MLE is the best unbiased estimator of $\boldsymbol{\theta}_0$.

5. A Cauchy random variable with "location parameter" ν has a density function

$$f_X(x;\nu) = \frac{1}{\pi(1+(x-\nu)^2)}, \quad x \in \mathbb{R}. \tag{1}$$

Despite its simple definition, this is a strange animal. First of all, its mean is not defined, as the integral $\int x/(1+x^2)\ \mathrm{d}x$ is not absolutely convergent. It is also easy to see that the variance is infinite. But as you can see (especially if you sketch it), then density is symmetric around ν, and ν is certainly the median.

Let $X_1, X_2, \ldots, X_N$ be iid Cauchy random variables distributed as in (1). From observed data $X_1 = x_1, \ldots, X_N, x_N$, we will compare three estimators: the sample mean

$$\hat{\nu}_{mn} = \frac{1}{N}\sum_{n=1}^{N} x_n,$$

the sample median

$$\hat{\nu}_{md} = \begin{cases} x_{((N+1)/2)}, & \text{N odd,} \\ \frac{x_{(N/2)}+x_{(N/2+1)}}{2}, & \text{N even,} \end{cases}$$

where $x_{(i)}$ is the ith largest value in $\{x_1, \ldots, x_N\}$, and the MLE

$$\hat{\nu}_{mle} = \arg\max_{\nu} L(\nu; x_1, \ldots, x_N) = \arg\max_{\nu} \sum_{n=1}^{N} \ell(\nu; x_n)$$

where $\ell(\nu; x_n) = \log f_X(x_n; \nu)$.

(a) One particular draw of data for $N = 50$ is variable `x` in the file `hw08p5a.mat`. Plot the log likelihood function, and report the MLE for ν. Your MLE will of course be approximate, but make sure yours is accurate to within 10^{-2} to the true MLE. I will give you a hint here and tell you that the true value of ν is somewhere in the interval $[0, 5]$.

(b) The file `hw08p5b.mat` contains a matrix `X`. This is an $N \times Q$ matrix, where $N = 50$ and $Q = 1000$; each entry is an independent Cauchy random variable with $\nu = 0$. Treating each column of `X` as a single draw of the data for $N = 50$, compute the sample mean, sample median, and MLE for each column. From these report the empirical mean squared error (the difference between the estimate and $\nu = 0$, squared, averaged over all Q trials) for each of the three estimators.

(c) Plot, overlayed on the same axes, the log likelihood functions $\sum_{n=1}^{50} \ell(\nu, x_n)$ as a function of $\nu \in [-1, 1]$ for each of the first 10 columns of `X` from the last problem. On top of this, plot the average log likelihood function (as a dotted line) over all $Q = 1000$ trials.